

# REVUE TUNISIENNE DE SCIENCES SOCIALES

PUBLICATION DU C.E.R.E.S.

Secrétariat d'Etat à la Recherche  
Scientifique et à la Technologie

N° **SPECIAL**: V<sup>es</sup> Journées Scientifiques du Réseau  
LTT (AUPELF-UREF)  
**LA MEMOIRE DES MOTS**  
SOUS LA DIRECTION  
DE Taïeb BACCOUCHE, André CLAS et Salah MEJRI

**Avec la participation de :**

André Roman ; Leila Messaoudi ; Mohamed Lazhar Abbès ; Hassen Hamzé ; Habiba Naffati ; Josef Dichy ; Karim Chibout et Anne Vilnat ; Douglas Skuce, Ingrid Meyer et Judy Kavanagh ; Sofia Benyahia ; Abdelhamid Camoun ; Miloud Taifi ; Ismaïl Timimi et Geneviève Lallichi-Boidon ; Christian Boitet ; Alpha Mamadou Diallo ; Ibra Diene ; Zachée Denis Bitjaa Kody ; Sylvie Porhiel ; Pierre Lerat ; Hussein Habaïli et Slim Ben Hazez ; Kamel Gaha ; Raphael Rajaspera ; Ambroise Queffélec ; Lina Sader Feghali ; Antoine Noujaim .

### **Membres du comité du Réseau**

- . André CLAS : coordonnateur du réseau, Université de Montréal, Canada
- . Anna ANASTASSIADIS- SYMÉONIDIS, Université Aristote de Thessalonique, Grèce
- . Roger GOFFIN, Université Libre de Bruxelles, Belgique
- . Antoine LIPOU, de l'Université Marien Ngouabi , Congo-Brazaville
- . Salah MEJRI, de l'Université de Tunis I, Tunisie
- . Philippe THOIRON, de l'Université Lumière Lyon -2, France

### **Rencontres Linguistiques Méditerranéennes**

Cette association a pour objectif de promouvoir la coopération entre des linguistes des différents pays méditerranéens dans le domaine de la recherche, de l'enseignement et de la formation des chercheurs .

Elle compte se placer dans la perspective du traitement automatique des langues en particulier dans la mise au point d'outil et de lexiques électroniques, et compte développer des outils pédagogiques sur différents média destinés à améliorer la connaissance des langues méditerranéennes.

Courriel: mejris @ yahoo.com

## **ATLAS LINGUISTIQUE DE TUNISIE**

### **Centres d'Études**

### **et de Recherches Économiques et Sociales**

23, rue d'Espagne-Tunisie.

Tph./Fax : 216.1.343 237

## **1ère Rencontre Linguistique Méditerranéenne**

### **RLM**

ayant pour thème

## **LE FIGEMENT LEXICAL التكلس المعجمي**

Tunis, les 17-18 et 19 septembre 1998

E-mail : mejris @ yahoo . com

# REVUE TUNISIENNE DE SCIENCES SOCIALES

**PUBLICATION DU C.E.R.E.S.**

Secrétariat d'Etat à la Recherche  
Scientifique et à la Technologie

**Directeur : Hachemi LABAIED**

**Numéro Spécial** : V<sup>es</sup> Journées Scientifiques du Réseau  
LTT (AUPELF-UREF)

Actes du colloque de linguistique:

**LA MEMOIRE DES MOTS**

Tunis 25-26 et 27 sept. 1997

Sous la direction de Taïeb BACCOUCHE, André CLAS et Salah MEJRI

avec la collaboration de :

-Nadia Khojet El Khil

-Abdelaziz Dhifi

Pour tous renseignements s'adresser au service des publications  
et de diffusion du CERES-23, Rue d'Espagne 1000 Tunis R.P.

Tél: 242 994 C.C.P. :617-37

## **Comité d'honneur**

Taieb BACCOUCHE, *Président de l'Association Tunisienne de Linguistique,*

*Professeur à l'Université de Tunis I*

Samir MARZOUKI, *Directeur de l'école normale supérieure,*

*Professeur à l'Université de Tunis I*

Abdelkader MEHIRI, *ancien Secrétaire d'état de l'Enseignement supérieur et de la recherche scientifique, ancien Président de l'Université de Tunis I,*

*Professeur à l'Université de Tunis I*

Abdessalem MESSIDI, *ancien Ministre de l'Enseignement supérieur et de la recherche scientifique, Professeur à l'Université de Tunis I,*

Malika TRABELSI-AYADI, *Directrice de l'Institut Supérieur de l'éducation et de la formation continue*

## **Comité d'organisation**

Salah MEJRI, *Président du comité, Tunis*

Rafika AMRI ABBES, *Tunis*

Abdelhamid CAMOUN, *Tunis*

André CLAS, *Montréal*

## **Comité scientifique**

Enrico ARCAINI, *Centre italien de linguistique appliquée, Rome*  
Henri BEJOINT, *CRTT, Université Lumière-Lyon2, Président d'EURALEX*

Daniel BLAMPAIN, *ISTI, Bruxelles*

André CLAS, *Coordonnateur du Réseau LTT, Université de Montréal*

Franz GUENTHNER, *CIS, Ludwig-Maximilian-Universität de Munich*

Antoine LIPOU, *Université Marien Ngouabi, Brazaville*

Bernard Quémada, *IPHE, Paris, Vice-Président du conseil de la langue française*

Philippe THOIRON, *CRTT, Université Lumière-Lyon2*

Salah MEJRI, *Université de Tunis I*

Jean-Marie ZEMB, *Collège de France, Paris*

## SOMMAIRE

Présentation : .....	9
<b>André Roman</b> : Natures et mémoires des «mots» .....	11
<b>Leila Messaoudi</b> : Technolectes bilingues (français-arabe) et modes de dénomination .....	27
<b>Mohamed Lazhar· Abbès</b> : Fréquence, polysémie, synonymie : l'exemple du verbe <i>dire</i> dans le Coran .....	47
<b>Hassan Hamzé</b> : De la racine au mot ou du mot à la racine : problématique de la création d'une nouvelle mémoire de l'emprunt en arabe .....	61
<b>Habiba Naffati</b> : Les mots français d'origine arabe. Tunisie-France : aller/retour .....	75
<b>Joseph Dichy</b> : Mémoire des racines et mémoire des mots : le lexique stratifié de l'arabe .....	93
<b>Karim Chibout et Anne Vilnat</b> : D'une figure de sens aux sens figurés : modélisation et interprétation avec les graphes conceptuels .....	109
<b>Douglas Skuce, Ingrid Meyer et Judy Kavanagh</b> : Bases textuelles et bases de connaissances sur le Web : le système IKARUS .....	129
<b>Sofia Benyahia</b> : Les mots que les femmes détestent et pourquoi .....	143
<b>Abdelhamid Camoun</b> : Mot et mémoire grammaticale .....	163
<b>Miloud Taifi</b> : La mémoire du coeur. Translation et symbolisation sémantiques .....	185
<b>Ismail Tmimi et Geneviève Boidin</b> : Cristal - GAG - 3AD95 Une coordination pour transiter de l'analyse de mots à l'analyse de phrases puis à l'analyse de discours .....	207

<b>Christian Boitet</b> : Mémoire des mots en traitement automatique des langues et «étymologie synchronique» .....	219
<b>Alpha Mamadou Diallo</b> : Tabous et euphémismes lexicaux en pular du Fouta Jalon (Guinée) .....	229
<b>Ibra Diene</b> : Signifiante et fonction culturelle des emprunts et des détournements sémantiques dans la poésie épique sénégalaise, d'expression wolof et d'alphabet arabe. L'exemple du <i>Diaza'ou Chakor</i> de Moussa Kâ .....	243
<b>Zachée Denis Bitjaa Kody</b> : Intégration des francismes dans les langues bantu .....	255
<b>Sylvie Porhiel</b> : Choix de la charge théorique et mots fonctionnels ....	271
<b>Pierre Lerat</b> : Lexique et mémoire .....	285
<b>Hussein Habaili et Slim Ben Hazez</b> : Génération automatique de mots arabes à partir de racines et de schèmes .....	293
<b>Gaha Kamel</b> : Les procédés linguistiques d'actualisation et de virtualisation et la production du mot .....	327
<b>Raphaël Rajaspera</b> : Une problématique de la dénomination technique dans deux langues à comparer : mémoire des notions et mémoire des termes .....	337
<b>Ambroise Queffélec</b> : Les emprunts : des mots qui se souviennent ?	353
<b>Lina Sader Feghali</b> : Une mémoire au service de la créativité ? .....	369
<b>Antoine Noujaim</b> : La mémoire des mots, un problème fondamental pour la traduction .....	381

## PRÉSENTATION

*Ce numéro spécial est consacré à une partie des communications des V<sup>e</sup> Journées scientifiques du Réseau Lexicologie, Terminologie, Traduction de l'AUPELF - UREF organisées à Tunis les 25, 26 et 27 septembre 1997, en collaboration avec l'Association Tunisienne de linguistique.*

*C'est grâce au soutien du Ministère de l'Enseignement Supérieur et du Secrétariat d'État à la Recherche Scientifique et à la Technologie que ces journées ont connu un très grand succès. Le nombre des participants (plus de deux cents linguistes), la qualité des communications, la diversité des centres d'intérêt, le haut niveau du débat scientifique en ont fait un moment important de l'activité scientifique.*

*Le choix du sujet «La mémoire des mots» est, de l'avis de tous les participants, très judicieux. Il a offert à tous les participants diverses possibilités de questionnements où étymologie, emprunt, figement, traitement automatique du langage, néologie, culture, morphologie, syntaxe, sémantique, synchronie, diachronie, etc. se complètent, s'interpénètrent et se prolongent l'un dans l'autre; le tout étant cerné dans le cadre du mot, grâce à la mémoire. Une mémoire, marginalisée jusque-là dans les études linguistiques, se trouve ainsi réhabilitée; une mémoire où l'acquis culturel se transforme en mots pour finalement assurer une fonction identitaire, où se forgent les outils conceptuels au moyen desquels nous pensons notre vécu et nous appréhendons l'univers dans sa complexité fluctuante; une mémoire qui se construit, se déconstruit et se défait avant de se reconstruire, se perd au gré de l'évolution et des contraintes du système pour qu'on la retrouve finalement au détour d'une forme morphologique, d'une nuance sémantique ou d'une construction syntaxique, rappelant ainsi l'universalité du mot en tant qu'instrument de dénomination sans lequel l'intelligence serait incapable de fournir des repères fiables au pensable. Aussi le mot ne devient-il mot que s'il emporte en lui-même une partie de nous-mêmes, servant d'empreinte qui assure la continuité, par le verbe, dans l'intermittence de l'expérience humaine.*

*Ce numéro spécial est aussi le fruit d'une coopération internationale dont l'initiateur est le Professeur Gaston Gross, Directeur du Laboratoire de Linguistique Informatique de Paris XIII. C'est grâce à son initiative que nous sommes entrés en contact avec le Professeur André Clas, coordonnateur du Réseau LTT, et que nous avons décidé du choix de la Tunisie comme pays d'accueil pour cette grande manifestation scientifique.*

*Nous profitons de la parution de ces Actes pour remercier toutes les institutions universitaires qui nous ont soutenus dans notre initiative et en particulier le Centre d'Études et de Recherches Économiques et Sociales qui a bien voulu assurer cette publication, et l'Institut Supérieur de l'Éducation et de la Formation Continue qui a accueilli le séminaire de formation sur le traitement automatique du langage animé par l'équipe du Professeur Gaston Gross.*

*Nous devons beaucoup à Monsieur le Ministre de l'Enseignement Supérieur qui a bien voulu inaugurer les travaux de notre colloque, à Monsieur le Ministre du Tourisme qui a dès le début soutenu nos efforts, à Monsieur le Président de l'Université de Tunis I qui a clos les Journées, à Monsieur le Maire de la ville de Tunis qui a comblé les participants de son hospitalité et à la direction de Tunis air qui nous a accordé des tarifs exceptionnels.*

*Nous remercions également Danielle Collignon (GRESLET, Université de Montréal) qui préparé tous les textes, Nadia Koujet El Khil qui a tout relu en vue de la présente édition et Abdelaziz Dhifi qui a assuré le suivi pour que ce travail soit publié.*

*Sans notre équipe d'organisation qui a montré un savoir-faire et une efficacité appréciée de tous, les V<sup>e</sup> journées n'auraient pas eu un tel succès <sup>(1)</sup>.*

*Taïeb BACCOUCHE et  
Salah MEJRI*

---

1) Un autre volume fera l'objet d'une autre publication assurée par l'AUPELF-UREF, Collection «Actualité scientifique», Serviced, Tunis, 1998.

# Natures et mémoires des «mots»

André ROMAN

*CRTT, Université Lumière-Lyon II,  
Lyon, France*

## 1. — INTRODUCTION

Toute langue humaine naturelle apparaît fondamentalement comme une systématique de signifiés et de signifiants (1) constituée contre le hasard par un usage immémorial (2).

L'Homme a su inventer  *systématiquement*  des entités et les noms de ces entités, lorsqu'il a été capable d'inventer une combinatoire, aucune «invention» n'étant possible,  *systématiquement* , que par la mise en oeuvre d'une combinatoire de composantes.

La première nomination a dû être la nomination d'entités du monde immédiatement saisissables par les sens.

---

1) Un «signifié» est une entité abstraite, un «concept», qui apparaît dans un système linguistique comme l'un de ses éléments constitutifs. Dans le sous-système de nomination, il a pour support nécessaire soit un «signifiant» sonore fourni par le sous-système phonologique, soit un signifiant «sourde» qui est établi comme «zéro» par des paradigmes du sous-système de nomination, ou, sinon, établi comme une position par des formes de ce même sous-système. Dans le sous-système de communication, il a également un «signifiant» sonore sauf s'il signifie une relation, univoque ou biunivoque, qu'il implique.

2) Si l'Homme parlait au hasard, sa parole ne serait pas prévisible linguistiquement. L'irrégularité du «hasard» empêcherait la coïncidence constante entre les locuteurs qui est évidemment nécessaire à leur communication.

L'Homme aura saisi, naïvement, certaines de ces entités comme des êtres universels, le temps, l'espace, la vie, ou comme des êtres communs aux membres de sa communauté, tel vivant ou tel non vivant (3).

Les êtres universels, ont vocation à devenir des signifiés de la langue, à la condition que celle-ci ait la capacité de leur accoupler un signifiant. Cela est possible seulement si leur saisie naïve, la naïveté des débuts, est réductible et, par là, représentable et donc intégrable dans l'organisation de la langue.

Ainsi l'espace, irréductible, multidimensionnel toujours, restera sans signifiant (4). Et les « quatre éléments », *l'air et le feu, la terre et l'eau*, ne seront que des noms, les noms propres de divinités ou, sinon, des noms communs (5).

En revanche le temps est réductible à une seule dimension, une ligne dont l'Homme peut se voir occuper tel segment, tel moment, et tout d'abord, vaguement, un présent (6).

Sans la saisie du temps, aucune expérience ne peut être verbalisée. Le temps est prégnant du langage. À partir du temps, l'Homme, comme il nommera, imaginera, dans sa langue, deux catégories de « noms » déterminés du point de vue du temps : les *res* et les *modus* : les *modus*, comme des entités *linguistiques* imaginées par lui dans un déroulement apparent du temps, dont le temps est l'une

---

3) Naïvement, une entité peut être perçue comme universelle si elle apparaît toujours présente. Sa présence constante implique sa présence partout. L'universalité ainsi entendue lie le temps et l'espace, indissociablement. Et elle n'implique avec l'espace et le temps que la vie, qui naît du temps, qui est avec le mouvement dans l'espace l'autre mesure du temps. Le ciel, l'air, la terre sont de l'espace. L'eau et le feu, qui ont pu être considérés comme des éléments primitifs, ne se trouvent pas toujours, ne se trouvent pas partout. De même que le feu, tel phénomène naturel, sur la terre, dans le ciel la chute des corps... se répète; mais il n'est présent qu'autant qu'il se répète dans un moment, dans un lieu; c'est un fait d'expérience et non pas une donnée. Or, l'organisation d'une langue n'est systématique que si elle est régulière. Idéalement, la systématique d'une langue est une organisation sans rupture et, par là, sans interruption. Les entités du monde qu'elle est à même de verbaliser comme des signifiés sont donc, semble-t-il, les entités universelles telles qu'elles ont été définies.

4) L'espace est dit, toujours, par des noms de lieu.

5) Cependant les langues « négro-africaines » ont un système de classes nominales qui « repose sur une répartition des êtres et des objets et, postérieurement sans doute, des abstractions en un certain nombre de catégories » (Meillet, A. et M. Cohen (dir.) (1952) : *Les Langues du monde*, Paris, CNRS, tome II, pp. 740-741).

6) Le temps systématique est précisé par des noms de temps.

des composantes; les *res*, comme des entités *linguistiques* imaginées par lui hors du temps, étrangères au temps, sans composante de temps.

La première combinatoire inventée par l'Homme a dû être la combinatoire binaire : elle est la première combinatoire qui s'offre et elle est puissante.

Dans l'hypothèse soutenue ici, fondée en diachronie lointaine, toute langue se serait constituée en un «système de systèmes» agençant :

- un sous-système de sons,
- un sous-système de syllabes,
- un sous-système de nomination qui, à chacune des entités inventées par l'Homme dans l'univers de ses possibles, associe un «nom», sonore, qui en est la représentation métonymique (7);
- un sous-système de communication, c'est-à-dire de relations entre ces «noms» qu'elles reliaient pour façonner des descriptions des expériences humaines, ainsi communicables à tous les Hommes d'une même langue.

Les sons proportionnés par les deux sous-systèmes qui les organisent en éléments de la langue, sons (8) et syllabes, forment la *materia prima* sonore des «signifiants».

Les «signifiés» essentiels, sonores ou non, primitifs sans doute, sont :

- les relations du sous-système de communication qui font des unités de nomination de chaque langue les constituants de ses phrases;
- les unités de son sous-système de nomination, qui dénotent :

\* *le général* : le temps général, le lieu général, par exemple, qui ont une identité intrinsèque;

\* *le banal*, dont l'identité, saisie tour à tour, hors nom, est référentielle : «je» ou «ce», par exemple;

---

7) En son principe, tout système de nomination est métonymique, autrement dit symbolique, dans le sens grec de «symbole» «objet coupé en deux (tesson) constituant un signe de reconnaissance quand les porteurs pouvaient assembler (*sumballein*) les deux morceaux» (*Le Grand Robert, s. v.*).

8) Phonèmes et intonèmes.

\* des *traits* du monde, les «modalités» de la langue (9) : l'«animéité», qui dit la vie, le «genre», qui dit le sexe, par exemple (10);

— les unités de son sous-système de nomination, qui *représentent* ou *résumant* des *res* ou des événements dont l'identité a été nommée : «il», par exemple.

Les unités de la langue produites par le système de nomination,

— ou bien nomment l'Homme; ce sont leurs noms identitaires ou **noms propres**, qui sont des «noms naturels»;

— ou bien nomment les autres entités «sensuelles» du monde ce sont les **noms communs**;

— ou bien nomment les entités gagnées plus tard sur la naïveté par la «science» et créées par elle au fur et à mesure de son développement; ce sont les **termes**, qui sont des noms «conceptuels».

## 2. — LES NOMS

### 2.1 — Le «nom propre»

Le «nom propre» est un nom généalogique. Il s'inscrit sur un arbre unaire, réduit à son tronc. Chaque homme nommé est seul à un niveau de cet arbre : né d'un couple, il est nommé métonymiquement, par le nom d'un seul de ses deux parents, celui de son père généralement; le père, dans le cas où l'enfant est nommé par son père, a été nommé semblablement. Cette nomination remonte le temps aussi loin que chaque société le croit nécessaire. Le frère ou la soeur sera sur un arbre parallèle, différent uniquement par son propre nom, qui le séparera de cet autre arbre.

---

9) A la différence des autres unités de nomination, les modalités, et comme elles les translatifs, sont des pièces du sous-système de nomination non construites sur des racines.

10) Le «translatif» est également une «pièce» importante du sous-système de nomination : il permet de passer outre ses limites étroites; en effet, il «translate», c'est-à-dire transforme, toute une phrase en une unité de nomination dont la richesse sémantique est bien plus grande que la richesse sémantique des unités de nomination systématiquement possibles. Mais la phrase traduite est éphémère. Elle n'est pas conservée par la mémoire; et elle n'entre pas dans les dictionnaires, sauf, éventuellement, comme une citation.

Le «nom propre» réalise ainsi la seule application identique du sous-système de nomination de la langue (11).

## 2.2 — Le «nom commun»

Le «nom commun» est le nom d'une entité qui, dans le cas des *naturalia* a été abstraite d'entités multiples. Cette abstraction a effacé les différences entre les différents exemplaires de cette entité. Ces différences n'ont pas été retenues par l'Homme en raison de son indifférence pour elles : elles lui sont inutiles ou inconnues. C'est par là d'abord que le «nom commun» est commun.

L'appréhension de l'entité naturelle nommée par un nom commun est pragmatique. Elle se fait par la reconnaissance naïve de couples d'entités que l'Homme distingue selon son besoin. Aussi le «nom commun» est-il, à la différence du «nom propre» et du «terme», un nom approximatif : il ne s'inscrit pas sur un arbre mais dans un champ référentiel où il trouve ses contours spécifiques (12).

Le «nom commun», alors, ne pouvait être détaché de l'espace. Il était concret. Ou, plutôt, son caractère concret, ainsi désigné, par hypallage, est apparu quand l'homme a inventé des entités abstraites, des entités détachées de l'espace et par là-même du temps, à partir donc, chaque fois, d'un *modus*. Mais ce *modus* réifié, ce nouveau nom, reconnu comme le nom d'une entité n'occupant aucun espace, ce nouveau venu, désigné comme un nom «abstrait», par la même hypallage, n'a pas été signifié comme tel par un nouveau morphème (13) : la langue, pour le discriminer n'a inventé aucun

---

11) L'autre application identique de la langue est réalisée par le sous-système de communication dans ses phrases absolues. Exemple : /ʔal:a:h-u/, «Allâh». La *res* attestée n'est qu'elle-même; elle s'autodétermine. Hors systèmes, les onomatopées, phrases ou noms, sont des applications identiques. Les «langages» des animaux ne comprendraient que cette seule relation qui peut être présentée comme une relation unaire.

12) Voir Gross, G. et A. Clas (1997) : «Synonymie, polysémie et classes d'objets», *Meta*, XLII/1, pp. 147-154.

13) En arabe, le schème /qatul/, du *modus* infinitif, où la voyelle /u/ était le signifiant de la modalité d'*agentivité nulle*. La disparition de cette modalité a transformé le schème /qatul/ en la chaîne /qutØ/, asystématique. Ainsi /ʒanuʔ/ dénotait le «fait de détester (par tempérament)»; ce *modus* infinitif avait donc pour sujet un certain vivant déclaré par le locuteur; si celui-ci voulait attribuer à son expression une valeur générale, il lui donnait un sujet général, par exemple, «les hommes»; puis /ʒanuʔ/, en devenant, /ʒunØʔ/, a signifié la «détestation», abstraitement. C'est semblablement à partir de *modus* que la langue arabe s'est donné les noms abstraits de la culture qu'elle exprime. Elle a réalisé ces noms abstraits en combinant des unités dénotant des *res*, pro-formes ou formes, et deux racines

morphème dénotant la présence ou l'absence de l'espace dans les entités nommées par ces noms.

Concrets ou abstraits, les noms des *naturalia* apparaissent comme des noms hérités.

Dans le cas des *artificialia*, le nom est créé, idéalement, normalement, avec la première entité manufacturée par l'Homme, qui la nommera. Elle est alors unique. Mais ce nom d'une entité unique, qui restera unique pendant un temps, n'est pas un nom propre.

Le nom d'une entité unique, naturelle ou artificielle, n'est pas un nom propre. Le nom du soleil, s'il est vu comme unique dans telle cosmogonie, ne deviendra un nom propre que si le soleil est imaginé comme un être vivant à l'instar de l'Homme (14).

Le nom commun, s'il est d'une *res*, dès lors même qu'il est abstrait d'une certaine collection de ces *res*, implique le nombre. Le nombre se développe à partir du nom commun et du nom commun seul. En effet, la *res* générale, de par sa généralité, et la *res* banale, qui se prête, tour à tour, à chacune de ses réalisations, sont étrangères au nombre.

Ainsi «nous» et «vous», idéalement, seront dits par des signifiants irréductibles aux signifiants de «je» et de «tu». De même l'autre *res* banale, hors «je» et «tu», saisie par l'ostension, et qui, de ce fait, apparaît comme le véritable partenaire de «je» et de «tu», comme la «non personne» que «il», qui est la «personne délocutive» (15), ne peut pas être. En arabe, /da : /, «ce», de racine √t, et son pluriel, /ʔu(:) la : (ʔi)/, de racine √l, seraient des réalisations de cette autre «personne ontique», à côté de «je» et de «tu».

---

monoconsonantiques : la première, \*√c (>/j:/), du *modus* «être», constitue le *modus* auquel est raboutée la seconde racine, √m (>/t/), de la *res* générale. Exemples : /ʔana:-ni + j:-a + t/, «égoïsme» (</ʔana:-ni + j:/, «égoïste», étymologiquement «être je», </ʔana:/, «je»); /kalØb-i + j:-a + t/. «cynisme» (</kalØb-i + j:/, «cynique», étymologiquement «être chien», </kalØb/, «chien»); /masʔu:l-i + j:-a + t/. «responsabilité» (</masʔu:l-i-j:/, «(qui peut être) questionné», </masʔu:l/, «questionné»).

14) Cf. Le verset VI/78 : /fa lam:a: raʔa: s sams-a ba:zigat-a-n qa:la ha:da: rab: i:/, «Quand il vit le soleil se levant, il dit : "Ceci est mon Seigneur".».

15) L'expression «non personne» est de É. Benveniste (1946) : «Structure des relations de personne dans le verbe», *Bulletin de la Société de Linguistique*, vol. XLIII, repris dans *Problèmes de linguistique générale* (1966-1974), Paris, Gallimard, NRF, vol. I, pp. 225-236. Joly, A., (1973) : «Sur le système de la personne», *Revue des Langues romanes*, tome LXXX/1, pp. 59-97, l'identifie comme la «personne non interlocutive».

Quant au *modus*, il n'est susceptible que de l'itération.

C'est le nom commun, abstrait, qui, en terminologie, représentera éventuellement le *modus*. En effet, comme il n'existe que par une *res*, son agent, ou dans une *res*, son lieu, le *modus*, ne saurait être un terme.

### 2.3 — Le «terme»

Le «terme» est le nom d'un concept et non plus d'un percept.

Un «concept» est une entité abstraite, étrangère au temps, qui s'inscrit sur un arbre particulier, un arbre conceptuel, binaire.

L'arbre binaire commence dans une racine imaginée par hypothèse. Il se divise, dès cette racine, en deux branches que diviseront, éventuellement, en d'autres branches, les oppositions qui pourront être greffées sur l'arbre.

Les termes sont les noms de la racine et des embranchements ainsi déterminés.

Une théorie, quand elle naît, d'une intuition ou d'une probabilité, sa racine, son hypothèse, ne pourra être symbolisée entièrement par un arbre binaire que lorsqu'elle aura abouti, ce qui ne nécessite point que toutes les branches, toutes les feuilles de l'arbre aient été reconnues simultanément. La théorie naissante est un gisement de possibles. L'«invention» des concepts qu'elle porte n'est plus que la recherche méthodique, contrainte, de ceux des concepts appariés qu'elle recèle encore, qui restent à découvrir.

La théorie, comme elle se constitue, comme elle tâtonne, l'ordre des découvertes n'est pas l'ordre idéal des raisons, court le hasard d'avoir à forger de pseudo-concepts, qu'elle nomme par des noms qui ne sont que de pseudo-termes. Les pseudo-termes sont reconnaissables car ils ne réfèrent pas à un embranchement dans l'arbre binaire qui structure la théorie mais charrient une description où des noms communs jouent aux termes, voleurs jouant aux gendarmes. Le pseudo-terme deviendra un terme si cette description, vérifiée, peut être abstraite en la définition d'un embranchement.

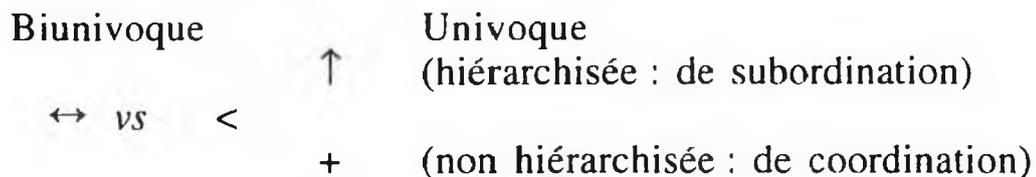
La création des termes est donc consciente, néologique.

Néanmoins le terme, dans son emploi profane, n'est plus qu'un mot plus ou moins évocateur; il devient un pseudo-terme.

Les *codes*, les *constitutions*, sont essentiellement des textes de «termes». Les *textes fondateurs* qui leur ont donné naissance sont essentiellement des textes de «noms communs».

### 3. — LES RELATIONS

Le sous-système de communication, qui est solidaire du sous-système de nomination — il est son jumeau — s'est établi, nécessairement, sur les trois relations binaires qui seules sont possibles. Ces trois relations sont donc, dans toutes les langues, les relations constitutives de la phrase :



Le système universel construit par ces trois relations est essentiellement simple du fait de leur caractère abstrait.

### 4. — LES SIGNIFIÉS

Le caractère abstrait des trois relations binaires reconnues rend possible et aussi commande le recours systématique à des morphèmes porteurs des spécifications nécessitées par la communication.

*Les morphèmes propres à la relation biunivoque.*

Ce sont d'abord :

— *la diathèse subjective* ou *objective*, qui oriente la relation biunivoque qui, dès lors qu'elle est sans orientation propre, s'offre à ces deux parcours : à partir d'un «sujet» ou vers un «objet»;

— *le mode réel* ou *potentiel* ou *irréel*, dont le *modus* est prégnant (16);

— *l'affirmation*, impliquée par la relation biunivoque, par là même qu'elle existe, et qui est une facette du réel;

— *la négation*, par opposition binaire;

— *l'interrogation* (17), *l'obligation* (18), suscitées par le mode potentiel.

16) La relation à une *res* d'un *modus* déclaré comme «passé» est «réelle», *de facto*. Au mode «réel» s'oppose, binairesment, le «non réel» qui est soit «potentiel», soit «non potentiel», c'est-à-dire «irréel».

17) Dans le cadre de l'hypothèse binaire, androgyne, chaque entité n'existe qu'avec l'entité à laquelle elle s'oppose, qu'elle annonce, que le chercheur, curieux par son commandement, doit découvrir : l'interrogation est doublement existentielle.

18) L'«impératif» comme l'envers de l'«interrogatif». Il existe d'autres modaux, secondaires, que les langues ont ou n'ont pas.

*Les morphèmes propres aux relations univoques.*

Ils se sont établis dans l'opposition générale {[non spécifié] vs [spécifié]}.

Ce sont :

— Les morphèmes définisseurs de la relation univoque, «+», qui sont les «coordonnants»;

— Les morphèmes définisseurs de la relation univoque «←», qui sont les «subordonnants»; la modalité exceptive.

Hors les subordonnants qui, généralement, lorsqu'ils sont spécifiés, miment le monde, — ils verbalisent les relations de l'Homme à son groupe, «avec», par exemple, ou à son environnement, «dans», par exemple —, les morphèmes qui viennent d'être retrouvés sont issus, immédiatement, des trois relations binaires reconnues et du temps.

*Les morphèmes propres aux res.*

Ce sont :

— la modalité d' «animéité», qui dit la vie;

— les modalités de genre, qui disent le sexe;

— la modalité, extrinsèque, de nombre, qui commence dans la combinatoire binaire et s'accomplit dans les «noms communs».

*Les morphèmes propres aux modus.*

Ce sont :

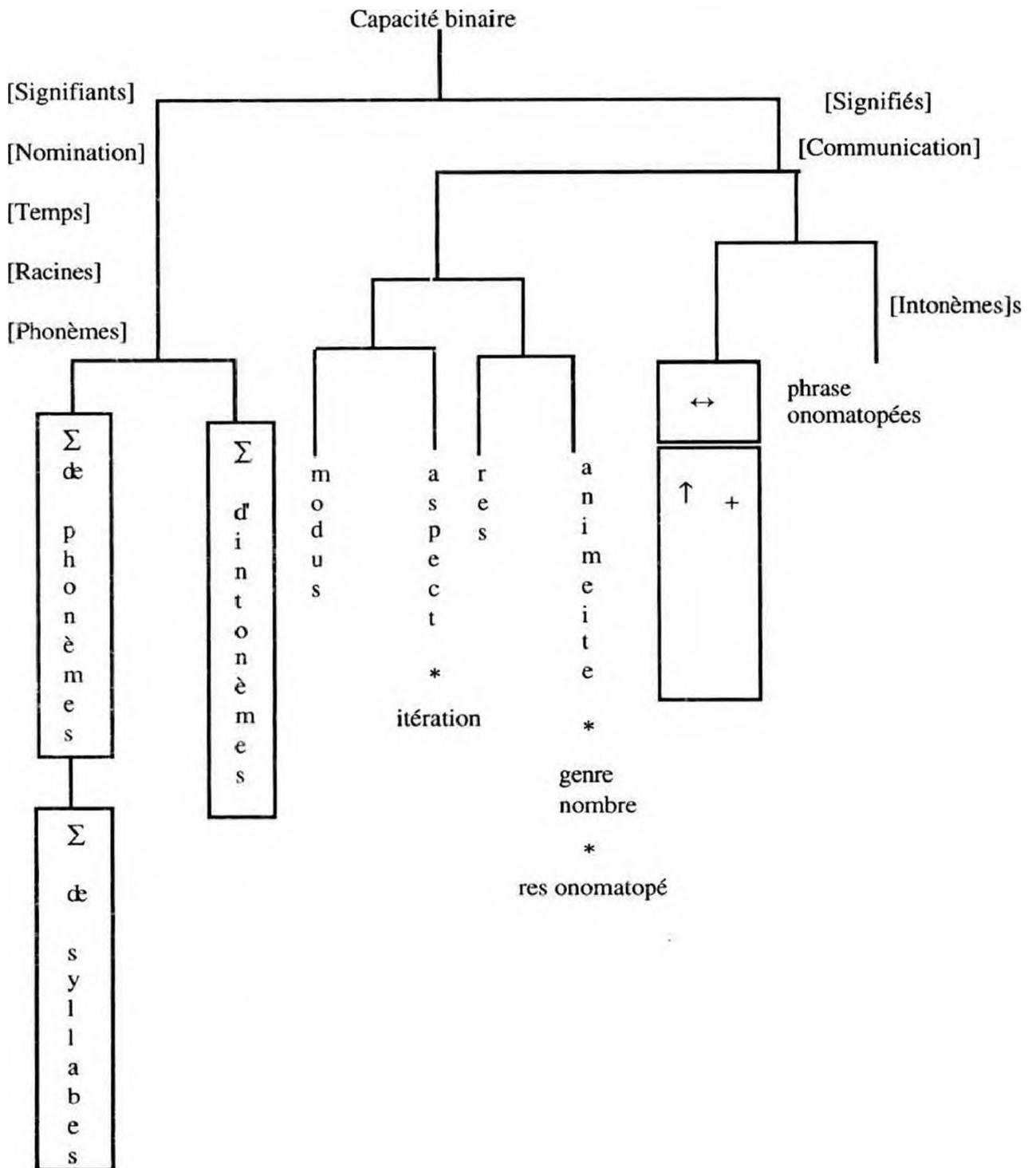
— les morphèmes de temps ou d'aspect : primitivement semble-t-il, *le passé* et *le non passé*, *présent* puis non présent, c'est-à-dire *futur*; *l'achevé*, qui est impliqué par le passé, qui termine le *modus*; le *modus* non passé, qui est *inachevé*, s'il est considéré indépendamment de son terme, qui est *achevant* ou *non achevant*, s'il est considéré par rapport à son terme, selon son progrès vers son terme (19).

---

19) Ainsi le *modus* est façonné soit par le temps, soit par la relation biunivoque sur laquelle la phrase se fonde. Le *modus* est donc seulement temporel, aspectuel, modal. Il est sans relation au monde. C'est par confusion entre la relation biunivoque, qui est linguistique, et les référents du monde, que le concept de «vrai» a été introduit, erronément, dans le fonctionnement des langues. En fait, les langues n'ont aucun morphème de vérité. Elles se sont établies, sans doute, pour verbaliser la réalité du monde mais telle qu'un locuteur donné l'exprime, dans son *discours*. Seul son discours peut être vrai ou faux, par incohérence propre ou par discordance avec ce qui est.

L'organisation générale des langues, telle qu'elle vient d'être retracée, est représentée par l'arbre ci après (20) :

### Arbre de l'organisation générale des langues



20) Cependant l'unité de nomination et aussi la phrase peuvent être structurées. L'unité de nomination, la phrase, sont alors produites comme «onomatopées», c'est-à-dire, non par le truchement des sous-systèmes de nomination et de communication de la langue, mais par une convention immédiate des locuteurs.

Chaque langue humaine, se sera faite, machinalement, par le parcours de cet arbre non point, régulièrement, à partir de sa racine première, mais à partir de telle entité saisie dans sa réalité binaire, par le truchement, peut-être, comme à l'envers, de la création possible de nouveaux signifiants signalés par l'unicité d'un son, notre phonème, qui «appellera» son partenaire dans le système se dessinant.

De fait, ce sont d'abord les sous-systèmes phonologique et syllabique qui, dans chaque langue, ont déterminé les candidats signifiants auxquels l'Homme pouvait apparier tel candidat signifié inventé par lui dans le monde de ses possibles.

Cependant un candidat signifiant peut ne pas trouver le candidat signifié sans lequel il ne peut faire partie du système de la langue. Ainsi, toutes les voyelles brèves de l'arabe, égales entre elles, avaient vocation à jouer le même rôle syntaxique : /u/ est, dans le sous-système de communication, le signifiant du «sujet»; /a/, celui du «complément»; /i/ n'a trouvé que de faux emplois dans les expansions complétives après subordonnant; exemple : /ma<sup>c</sup>a l kalØb-i/, «avec le chien», qui est une réalisation secondaire du syntagme primitif <sup>\*</sup>/ma<sup>c</sup>a l kalØb-a/; et encore, par analogie, dans l'expansion annective; exemple : /kalb u s sajk-i/, «le chien du cheikh», qui est une réalisation secondaire du syntagme primitif <sup>\*</sup>/kalb-u s sajk-a/.

## UNE HISTOIRE BRÈVE DE L'ARABE

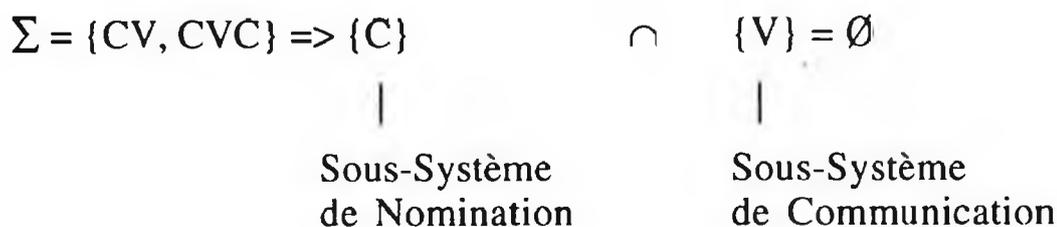
Le proto-système syllabique des langues sémitiques anciennes, {/CV/, /CVC/}, qui est resté le système syllabique propre à la langue arabe a déterminé dans leur fonctionnement une disjonction du sous-ensemble des consonnes, {C}, et du sous-ensemble des voyelles, {V} :

$$\Sigma = \{CV, CVC\} > \{C\} \cap \{V\} = \emptyset \quad (21)$$

Cette disjonction a permis l'attribution systématique de tâches différentes aux consonnes et aux voyelles. Les langues sémitiques ont construit leur sous-système de nomination sur des racines de consonnes, leur combinatoire étant plus puissante du fait de leur plus grand nombre. Et elles ont fondé leur sous-système de communication sur leurs voyelles brèves utilisées par elles comme des désinences casuelles.

21) Voir A. Roman (1981) : «De la langue arabe comme un modèle général de la formation des langues sémitiques et de leur évolution». *Arabica*, Paris, tome XXVIII/2-3, pp. 127-161.

L'organisation générale de la langue arabe peut donc être représentée par le schéma suivant (22) :



Par besoin, les «unités de nomination commune», *res* ou *modus*, ont été construites sur des racines de trois consonnes. En effet, la combinatoire de trois consonnes est la première combinatoire à même de produire en nombre suffisant les arrangements constituant les racines nécessaires de ces «noms communs».

Par contre les unités de nomination générale ou banale, et aussi les unités de représentation, elles peu nombreuses, pouvaient être construites sur des racines d'une seule consonne (23).

En arabe, l'opposition fondamentale {*res* vs *modus*} s'est construite dans les unités de nomination à racine triconsonantique, «QTL», sur le couple {«Qv<sub>1</sub>TØL» vs «Qv<sub>1</sub>Tv<sub>2</sub>L»} : «Qv<sub>1</sub>TØL» est le schème des *res*; «Qv<sub>1</sub>Tv<sub>2</sub>L», le premier schème des *modus*.

La première voyelle est, dans l'un et l'autre schème, imposée par le sous-système syllabique.

Elle a, dans les *res*, servi de signifiant à l'animéité.

Elle a, dans les *modus*, servi, parallèlement, de signifiant à la diathèse.

La deuxième voyelle des *modus* a servi de signifiant à une modalité de la diathèse subjective; cette modalité d'«agentivité» dénotait l'initiative, le degré de participation du «sujet» à la production du *modus* : elle faisait apparaître le «sujet» soit comme agissant de son propre chef, soit comme réagissant à un événement, soit comme ne pouvant rien faire que subir un *modus* qui lui échappait.

Ce premier schème des *modus*, «Qv<sub>1</sub>Tv<sub>2</sub>L», sans autre morphème aspectuel que «zéro», s'est établi, *ipso facto*, comme le

22) Cette structure est la première structure des langues sémitiques.

23) Ainsi √m est la racine de la *res* générale √n, celle du lieu général; √t, celle du temps général.

schème du *modus* infinitif (24); exemple /sariq/, «vol». Les deux schèmes des *modus* «achevant» et «non achevant» se sont construits par opposition de longueurs vocaliques; d'où «Qv<sub>1</sub> : Tv<sub>2</sub>L», exemple : /sa:riq/, «voleur dans l'acte de voler», «Qv<sub>1</sub>Tv<sub>2</sub> : L», exemple /saru : q/, «voleur par instinct» (25).

La combinatoire des racines, offerte par la combinatoire générale de la langue, son *modus operandi*, a conduit à la création de deux signifiants symétriques pour les deux autres aspects des *modus* de la langue, également symétriques, l'«achevé» et l'«inachevé», les deux arrangements possibles {√CCC + √C}, signifiant de l'«achevé» et {√C + √CCC}, signifiant de l'«inachevé». Et elle a conduit à la création des autres formes.

#### 4. — CONCLUSION

La nature des mots conditionne leur relation à leur mémoire.

Le nom propre est, par excellence, le nom de la mémoire.

Le nom commun, régulièrement, n'a, de par lui-même, de mémoire que celle, commune, de sa racine, et celle, spécifiée, de ses modalités. Exemples de la langue arabe :

— /kalØb/, «chien», dont la racine √k-l-b est le morphème chaînant toutes les unités de nomination du domaine de son application; la voyelle «Ø» signifiait l'absence de toute relation au temps; la forme était ainsi catégorisée par elle comme une *res*; /a/ était le signifiant, sans doute, de l'«animéité»; le sexe de l'animal nommé par /kalØb/ était simplement impliqué;

— /kalab/, «affection de la rage», de même racine; où la seconde voyelle faisait de la forme une forme de *modus*; la première voyelle était le signifiant de la diathèse qui ne peut exister que dans le temps; le timbre de la seconde voyelle dénotait l'agentivité du sujet du *modus*, ici sa réaction.

Ainsi, la langue arabe signifiait, globalement, par la racine √k l b, les traits particuliers à la morphologie de l'animal — mammifère, digitigrade..., les formes et les effets de la maladie «canine». Elle ne signifiait par ses modalités, c'est-à-dire, par la mémoire de son système, que bien peu de choses au regard de ce que

24) Le *masdar* de la tradition grammaticale arabe.

25) Exemple : /kalØb saru:q/, «chien voleur», où /saru:q/ est d'agentivité nulle.

signifiait la racine qui s'impose toujours comme une entité opaque (26). C'est la mémoire textuelle ou vécue du locuteur qui seule sait ce que sont le «chien» et la «rage». Les étymologies, quand elles existent, quand même elles seraient vivantes, sont enfermées dans les mêmes limites étroites des unités de nomination qui leur ont succédé. Il est remarquable que la langue arabe, qui s'est construite sur des racines de consonnes, est sans langue antérieure, sans «latin» ni «grec» (27).

Une langue est, en elle-même, constituée, forcément, sur le couple {«signifiant» - «signifié»}. Le signifié est imaginé à partir d'un sens choisi entre tous pour son universalité potentielle. Bien que le signifié soit issu du monde des sens, l'abstraction qui le constitue coupe du monde des sens le système linguistique où il s'emploie. Manifestement les systèmes de nomination sont infirmes. Les entités que l'Homme invente dans le monde sont nommées, d'abord, à travers les racines, puis par le moyen de modalités. Les modalités sont des signifiés du système. Les racines sont la porte étroite du monde. Elles calibrent les entités du monde. Elles les réduisent à la mesure de systèmes linguistiques. Les unités de nomination ne peuvent avoir d'autre mémoire que celle de leurs racines et de leurs modalités. Leurs relations, *dans la langue*, aux autres unités de nomination sont *entièrement* signifiées par le truchement des trois seules relations possibles entre elles : la relation biunivoque entre les deux éléments fondamentaux qui constituent les noyaux des phrases; les relations univoques de coordination et de subordination qui rapportent à chacun de ces deux éléments des chaînes non bornées d'extensions.

---

26) L'hypothèse d'une racine faite d'atomes de sens est irréaliste. J. Wilkins, auteur d'un *Essay toward a real character, and a philosophical language*, publié en 1688, a fait une proposition précise illustrée par l'exemple suivant : «Se (De) significa Elemento, allora (Deb) deve significare la prima differenza; la quale (secondo le Tavole) è Fuoco : e (Deb) denoterà la prima specie, che è Fiamma. (Det) sarà la quinta differenza sotto il Genere che è Meteora che Appare; (Det) la prima specie, e cioè Arcobaleno; (Deta) la seconda, e cioè Alone» (Eco, U. (1993) : *La Ricerca della lingua perfetta*, Roma Bari, Laterza, chapitre 12).

27) Une racine arabe peut avoir existé dans une autre langue sémitique, éventuellement avec quelque différence. Ainsi la racine  $\sqrt{h-l-k}$ , «périr», de l'arabe, correspond à la racine  $\sqrt{?l-k}$ , «partir», de l'akkadien. Certaines racines sont pan-sémitiques, comme la racine  $\sqrt{?-b-w}$ , qui dénote la paternité. Ruhlen, M. (1994) : *On the Origin of Languages — Studies in Linguistic Taxonomy*, Stanford, Stanford University Press, 1, aurait retracé la racine  $\sqrt{m-l-g}$ , «swallow, throat», sous différentes formes parentes, dans toutes les langues du monde; elle se trouve en arabe sous la forme  $\sqrt{m-l-z}$ , qui dénote la «sucée du sein maternel».

Dans le cas où l'unité de nomination est faite d'une base et d'une expansion, la relation entre ses deux composantes étant univoque, la cohésion de cette unité de nomination n'est généralement pas assurée par le système (28); c'est encore la seule mémoire du locuteur qui de «chaise longue» fait le «nom commun» d'un certain meuble.

En outre, les langues sont oublieuses.

L'arbre binaire originel est désormais un «arbre relique».

La langue arabe a «oublié» les modalités d'itération, d'agentivité... de ses *modus*. Elle a «oublié» les racines monoconsonantiques de ses *res* et de ses *modus*; en conséquence, en rupture avec son organisation générale, elle a laissé des analogies locales conduire à leur remplacement par des voyelles; par exemple, dans /QaTaLtu/, «j'ai tué», auparavant \*/QaTaLku/, où la voyelle /u/ est devenue le signifiant de la première personne à la place de la consonne primitive /k/ (29).

Ainsi, la nomination, en perdant plusieurs de ses signifiés, est devenue irrégulière. L'irrégularité ne peut être confiée qu'à la seule mémoire, à l'instar des racines qui nous apparaissent, presque toujours, nées du hasard.

Parce que les mots n'ont pas de mémoire, l'*Homo loquens* est tout autant un être textuel.

Si l'on prête aux mots une histoire, c'est par hypallage (30).

---

28) Elle peut cependant parfois signifier cette solidarité référentielle en combinant dans la phrase à cet effet certains de ses morphèmes; exemple repris aux travaux de Sydo : «un placard de cuisine» vs «un placard de la cuisine».

29) La consonne /k/ qui a précédé /t/ a sans doute elle-même succédé à la consonne /ʔ/ qui est le signifiant régulier de la «première personne». Quant à la consonne /t/, elle n'est rien d'autre, dans cette forme, qu'une consonne syntagmatique.

30) Le signifié de toute phrase, dès lors qu'elle est bien une phrase avec les deux constituants élémentaires de son noyau, est un signifié complet. Et tout comme celui de n'importe quel signe, il est ce à quoi la phrase «renvoie abstraitement», c'est-à-dire linguistiquement, tandis que son «sens», tout comme celui de n'importe quel signe, est ce à quoi elle renvoie concrètement, c'est-à-dire extralinguistiquement. Voir particulièrement Bureau, C. (1976) : *Linguistique fonctionnelle et stylistique objective*, Paris, PUF, p. 138. Dans ces conditions, la traduction est essentiellement une traduction des *sens* et non pas de *signifiés*.

# **Technolectes bilingues (français-arabe) et modes de dénomination**

**Leila MESSAOUDI**

*Faculté des lettres et des Sciences humaines,  
Kénitra, Maroc*

La communication que nous nous proposons de présenter s'articulera autour de la problématique de la dénomination dans les technolectes bilingues.

Dans un précédent travail, nous avons abordé les modes de désignation en nous inspirant de Ch. S. Peirce et de Cassirer. Nous comptons poursuivre cette réflexion en nous interrogeant sur les modes de dénomination, en essayant de déceler et de répertorier les procédés linguistiques — endogènes et allogènes — utilisés dans les technolectes bilingues (français-arabe) et les stratégies — d'adéquation ou de compensation — déployées dans la langue-cible (l'arabe), pour adhérer à la notion, via le terme de la langue-source (le français).

Pour ce faire, nous avons recueilli un corpus à partir de documents-sources variés tels que les dictionnaires spécialisés ou les notices pharmaceutiques (d'utilisation des médicaments). Cet exercice permettrait de mesurer le degré de mode d'appropriation des technolectes dans une langue donnée.

## **1. — QU'EST-CE QU'UN «TECHNOLECTE» ?**

Claude Hagège a été l'un des premiers à utiliser le mot «technolecte» qu'il a défini comme «un ensemble de désignations

spécifiques d'un certain domaine de l'activité humaine : sciences et leurs applications (par ex. biologie et médecine, chimie et pharmacopée, etc.), arts, vie politique et administrative, économique et sociale...». Selon lui, le technolecte est «constitué de termes» qui «doivent posséder idéalement les propriétés suivantes :

— être normalisés, c'est-à-dire admis par tous les utilisateurs du technolecte concerné;

— être monosémiques ou univoques, c'est-à-dire n'avoir, dans tous leurs emplois au sein du technolecte qu'un seul sens;

— n'être pas connotés, c'est-à-dire ne pas s'associer d'images et représentations connexes.» (1)

En revoyant les traits retenus par Claude Hagège, on pourrait se demander si «technolecte» n'est pas à rapprocher de «terminologie».

Le rapprochement est possible. Toutefois, le terme de terminologie couvre diverses acceptions dont nous citerons à la suite de Guy Rondeau (1984) les trois suivantes :

«1. L'ensemble des termes d'un domaine, comme par exemple la biochimie, ou d'une discipline, comme par exemple la linguistique. On dira alors : la terminologie de la biochimie, de la linguistique [...]

2. Les méthodes de collecte et de classement de termes, de création néologique, de normalisation des termes, de diffusion des termes : c'est ce à quoi s'emploient terminologues et terminographes.

3. Une science dont l'objet est d'ordre linguistique mais qui est essentiellement pluridisciplinaire et participe à la fois de la linguistique, de la logique, de l'ontologie, de la classologie et de l'informatique». (2)

Sans s'arrêter aux différents sens que revêt le terme de terminologie — on peut renvoyer à titre indicatif à A. Rey (1979, 1992) et à Helmut Felber (1984) — on peut retenir qu'il désigne une discipline qui veut s'ériger en science, depuis l'impulsion qui lui a été donnée par l'un de ses pionniers Eugen Wüster mais qu'il s'applique aussi aux différents ensembles de termes propres à une activité ou à un domaine donné.

Pour résumer, la terminologie désigne à la fois les produits et les procédés. Les produits sont les termes eux-mêmes, les procédés renvoient aux principes et méthodes de travail utilisés.

On pourrait ainsi opérer une première distinction en proposant de réserver le terme de terminologie à la discipline et consacrer le terme de technolecte aux seuls produits du travail terminologique que sont les termes. Proposition qui ne pourra pas être retenue car le poids de l'usage a déjà ancré des habitudes. En revanche, il serait peut-être utile d'opérer une distinction judicieuse : le technolecte ne désigne que les produits et engloberait aussi bien les unités normalisées (termes) ou non normalisées comme les unités phraséologiques. Ces dernières de type syntagmatique sont le plus souvent évitées par les terminologues; pourtant, elles ont une forte présence dans les domaines spécialisés et sont souvent l'apanage des traducteurs lorsqu'ils ne disposent pas d'un terme simple. Le technolecte ne serait pas comme la terminologie essentiellement paradigmatique mais renfermerait aussi les unités complexes que les linguistes désignent par lexies complexes (Pottier, 1974) ou synapsies (Benveniste, 1966) du type «aigle pêcheur à tête blanche».

En fait, cette réflexion menée sur les technolectes (Messaoudi, 1990) a abouti à l'idée que ce terme serait retenu utilement pour remplacer «langue de spécialité» dans un domaine scientifique ou technique et que les termes utilisés seront formés de mots uniques, multiples ou même d'unités phraséologiques. Nous constatons avec satisfaction que l'appréhension que nous en avons eue est conforme à celle avancée récemment dans le projet des «principes méthodologiques du travail terminologique» diffusé sur le réseau internet (<http://www.iula.upf.es/cpt/cptgmfr.htm>), et qui a été présenté par Louis-Jean Rousseau de l'OLF du Québec, à l'occasion de la rencontre de Barcelone (Realiter, 1995). Cette définition est la suivante : «On entend par technolecte un sous système linguistique utilisé dans un champ d'expériences particulier et caractérisé par une terminologie spécifique et par d'autres moyens spécifiques tels le style et la phraséologie. Cette notion est le plus souvent nommée langue de spécialité mais cet usage est critiqué».

Nous avons constitué un corpus de technolectes bilingues afin de procéder à une étude linguistique des unités qui les constituent et afin de réfléchir, en particulier, sur le mode de dénomination lors du passage de la langue qui sert de support notionnel (langue-source) à la langue qui sert de support formel (langue-cible).

Pour la constitution de ce corpus, nous avons choisi deux domaines : l'agriculture — plus spécialement les machines et outils agricoles — et la santé — en particulier les noms des maladies. Pour le premier, nous avons dépouillé des documents lexicographiques;

pour le second, les notices qui accompagnent généralement les médicaments. Par commodité, nous allons renvoyer au premier par AGR et au second par SAN et nous formerons des échantillons à partir desquels nous étudierons les modes de dénomination dans la langue cible (l'arabe standard = AS) et la langue source (le français = F).

### 1.1 — Échantillon AGR

Les données collectées dans cet échantillon ont été puisées dans deux dictionnaires bilingues français-arabe :

— *Dictionnaire d'agriculture Français-Arabe*, Paris, CILF-ACCT, 1985, XIX + 525 + 61 pp. Il contient 36 00 entrées environ (chiffre évalué aléatoirement).

— *Dictionnaire français-arabe des termes agricoles de Mustapha Chihabi*, Le Caire, Imprimerie misr, 1957, 694 + 28 pp. Il contient 10 000 entrées (chiffre mentionné dans la préface).

— *Dictionnaire technique : technique agricole*, M. A. Nassar, Cairo-Leipzig, 1977, 125 + 321 pp. Il contient 1370 entrées. Ces documents-sources seront symbolisés respectivement par X, Y et Z.

### 1.2 — Échantillon SAN

Pour ce second échantillon, nous avons procédé au dépouillement exhaustif de 87 notices en circulation au Maroc. Elles sont toutes bilingues (français-arabe standard) et comportent souvent les mêmes rubriques. Nous simulerons ci-après le format d'une notice au sein duquel nous insérerons les rubriques et leur(s) appellations

Rubrique 1	Nom du médicament
Rubrique 2	Présentation, forme
Rubrique 3	Composition, formule
Rubrique 4	Propriétés Indications thérapeutiques
Rubrique 5	Contre-indications Précautions d'emploi Effets secondaires Tolérance
Rubrique 6	Posologie Mode d'emploi / d'administration.

*Remarques :*

— La rubrique 1 pourrait se prêter à elle seule à une étude fructueuse sur le plan sémiolinguistique. Deux axes seraient à prendre

en considération : d'une part, l'obligation de rappeler par le nom du médicament sa composition scientifique ou son action thérapeutique principale et d'autre part, l'aspect esthétique-publicitaire et son rôle dans la dénomination. Par exemple, «sédaspir» de «sédatif» + «aspirine» mais nous ne multiplierons pas les exemples vu l'aspect publicitaire qu'ils pourraient revêtir tout à fait involontairement de notre part !

Pour ce qui est du corpus examiné, il est à noter que les noms des médicaments ne sont pas traduits en arabe mais que des efforts d'adaptation au système phonologique de la langue cible sont généralement fournis de façon que le produit soit toujours reconnaissable.

— La rubrique 2 est consacrée à la présentation du médicament (forme, emballage, conditionnement, etc.) et elle a été retenue pour constituer SAN 1. Nous y reviendrons en détails mais nous pourrions noter au passage que la place de cette rubrique est variable : elle peut se trouver comme dans la simulation ci-dessus immédiatement après la rubrique 1 ou bien tout à fait à la fin, après la rubrique 6.

— La rubrique 3 est consacrée à la composition chimique et à la formule qui la résume. Nous ne nous en occuperons pas puisque cela ne constitue pas l'objet de cette recherche qui n'est pas spécialisée en pharmacie.

— La rubrique 4 réunit les propriétés et les indications thérapeutiques qui constituent deux champs différents mais complémentaires. L'un, celui des indications thérapeutiques, est «ciblé» : il renvoie aux maux, maladies, douleurs, symptômes, etc. qui sont concernés par le produit. L'autre, celui des propriétés est axé davantage sur les processus des actions thérapeutiques des médicaments — d'où d'ailleurs au niveau discursif, profusion d'expressions telles que : «exerce une action bienfaisante, fluidifiante, sédatif, intense, rapide» ou encore «stimule les réactions de...» et «permet des efforts prolongés et exceptionnels...». En revanche, dans le cas des indications thérapeutiques, l'on note une concision au niveau discursif : les maux, maladies, douleurs, etc. sont désignés par des termes, le plus souvent les uns à la suite des autres, simplement séparés par des blancs ou par des signes de ponctuation (points, virgules ou tirets). Cette case a été exhaustivement dépouillée pour constituer le matériau de l'échantillon SAN 2.

— La rubrique 5 est un champ variable, englobant les contre-indications, les précautions d'emploi, les effets indésirables, etc. Elle n'est pas obligatoirement présente contrairement à la rubrique 4, qui

est un champ fixe. Étant donné son caractère facultatif et hétérogène, nous ne l'avons pas prise en considération ou du moins pas de façon systématique dans le dépouillement.

— La rubrique 6 qui réfère à la posologie, a fourni la matière nécessaire pour la constitution de l'échantillon SAN 3.

Après cette description sommaire des documents-sources, il nous paraît utile de signaler que les dénominations de ces rubriques en arabe sont assez disparates. Nous les avons collectées et avons évalué à l'aide de données numériques la fréquence de telle ou telle appellation dans les notices dépouillées. Pour en faciliter la lecture, nous avons pensé à les présenter dans le tableau suivant :

Rubrique	Dénomination en français	Dénomination en arabe
2	Présentation  Forme Emballage Conditionnement	taqdim (31), Cabawat (11), akl(10), aklu ddawa'(9), taqdimat(4), akaldawa'iyah(1), taCbi'ah (1), Card (1), tawdib (1), akalun saydaliyah (1) akl (5), taCbi'ah (2) Cabawat (5), taCbi'ah (2) taCbi'ah (2), Cabwah (2)
3	Composition Formule	tarkib (38) tarkib (5), siah (2)
4	Propriété (s)  Indications (thérapeutiques)	xawas (15), xasa'is (6), xasiyah (3), mîzah (1) dawâcil'isticmâl (29), irâdât (17) bayanât (8), taclimât (6), hâlâtu listiCmâl (6), 'istitbâbât(5), 'istiCmâl (4), 'iârât (4), al 'âtaru ttibiy (4), 'istitbâb (1), tawdihât (1), wasafât(1), dawâci /dalâ'ilu l'istitbâb (1), irâdât tibbiyah/ Cilajjiyyah (1), bayânât l Cilâj/ tibbiyah (1)

Rubrique	Dénomination en français	Dénomination en arabe
5	Contre indications	mawâniCu l'istiCmâl (9), hâlâtu Cadami l'istiCmâl (7), nawâhi l'istiCmâl (4), tadaddul'istibâb (4), mudaddât (3), bayânât/hâlât mudâddah (2), alhâlâtu-llatî-lâ yustacmalu ddawâ'u fihâ (1)
	Précautions d'emploi	ihtiyâtul'isticmâl(21), al'ihiyatu fi kiyfiyyati/Cinda l'isticmâl/ (1), 'ihtirâzât/tahdîrât l'istiCmâl (1)
	Effets indésirables	ta'tîrât Rayr marRûb fîha (5) hâlât Rayr marRub fîha (3), 'aCrâd jânibiyyah (3) 'âtâr jânibiyyah Rayr marRûbah (2) mafâCil Rayr marRûbah (1)
	Effets secondaires	'âtâr tânawiyyah (5) 'âtâr jânibiyyah (1)
	Tolérance	al-'ihtimâl (2), tâqatu ttahammul (2)
6	Posologie	maqâdir (48), jurCât (16), maqâyis (5), tarîqatu l'istiCmâl (3)tariqatu ttadâwi/ttanâwul (2), al kimmiyah (1)
	Mode d'emploi	tarîqatu l'istiCmâl (2), kîfiyyatu al 'istiCmâl (1)

*Remarques :*

— Les données numériques figurent entre parenthèses ( ) dans le tableau.

— La traduction mot-à-mot des équivalents arabes n'est pas donnée dans le tableau, elle sera mentionnée ultérieurement suivant les besoins de l'analyse.

Après avoir décrit de façon concise les diverses rubriques des notices pharmaceutiques et après avoir signalé la disparité des dénominations qui leur sont affectées en arabe, nous allons présenter ci-après SAN 1, SAN 2 et SAN 3.

## SAN 1 Présentation du médicament

Ampoule	'unbûbah (5), qinnînah (3), ambûlah (2), hubâbah (2), anbûlah (1), qârurah (1)
Capsule	kabsûlah (3)
Collyre	taqtîrah (3)
Comprimés	aqrâs (35), habbât (3), madRutât (3)
Comprimés effervescents	aqrâs raddâdah (2), aqrâs fâ'irah (2)
Compte-gouttes	qattârah (2)
Crème	krim (2)
Dragée	madRûtah (2), qors (1)
Étui	Rilâf (1), qinnînah (1), Culbah (1)
Flacon	Qârûrah (6), qinnînah (5), zujâjah (4)
Gel	Cajîn (1)
Gélule	mulabbasah (4), kabsûlah (1)
Gouttes	qatarât (3), nuqat (1)
Granulé buvable	mashûq liurb (2)
Lotion	dahûn (3), sâ'il (2)
Ovule	buwaydah (2)
Pommade	murham (8), duhn (2)
Poudre	mashûq (3)
Sachet	kîs (3)
Sirop	arâb (6), marûb (3)
Soluté buvable	mazîjun yurab (2), sâ'il (2)
Solution	mahlûl (2), sulisyûn (1)
Suppositoires	tah^mil (9), qawâlib (3), quwaylibât (2), labbusât (2)
Suspension buvable	muCallaq / sâ'il liurb (2)
Tube	unbûbah (6), 'unbûb (2), wiCâ (1)

### *Remarques :*

— Dans le répertoire ci-dessus, nous avons reproduit les termes tels qu'ils apparaissent dans les notices. Par exemple, s'ils y sont au pluriel, nous les présentons tels quels.

— Le contexte linguistique n'a pas été retenu pour deux raisons : premièrement, il n'est pas toujours présent et c'est généralement la sémiologie (topologique et typographique) qui introduit les termes. Ex. : «Présentation : comprimés»; «Présentation : sirop»; deuxièmement, lorsqu'il est présent, c'est généralement sous la forme récurrente : «boîte de...» ou «tube de...» référant au type de récipient.

Cette information n'est pas réellement pertinente d'autant que le terme «boîte» présente une grande fréquence et qu'il a un comportement stable aussi bien dans l'usage courant que dans le technolectal dans la mesure où il a toujours comme équivalent arabe, le terme «Culbah».

— Le terme «ovule» a pour équivalent arabe «buwaydah» qui signifie littéralement «petit oeuf». Forme morphologiquement correcte puisqu'elle a respecté le schème du diminutif mais elle n'est pas conforme aux règles phonologiques de l'AS où la forme attendue serait «buyaydah» et non «buwaydah» plutôt conforme au dialectal.

Les interférences morphophonologiques dues à la diglossie AS et arabe dialectal ne sont pas très nombreuses et c'est le seul cas rencontré dans le corpus SAN.

— Pour clore ces remarques, nous terminerons sur une observation d'ordre culturelle : le français, pour une solution buvable, mentionne les proportions en nombre de «verres à madère» tandis que l'arabe utilise les «verres thé» ! Mais il faut reconnaître que l'équivalent en unités de mesure est mis entre parenthèses, ce qui évite toute ambiguïté.

## **SAN 2**

Cet échantillon est constitué à partir de la rubrique 4.

Afin d'affiner l'analyse et par souci d'homogénéité, nous avons préféré cerner un micro-champ; celui de la dénomination des maladies dans un seul domaine : la dermatologie qui est le plus représenté dans le corpus des notices, environ 12%.

Dénomination en français	Équivalent en arabe
acné	habbu abâbi
allergie cutanée	hasâsiyatun jildiyyah
dermatite	maradu l jidi
dermatose	amrâdun jildiyyah
dermatose microbienne inflammatoire	amrâdun jildiyyatun mikrubiyyatun multahibah
dermatose surinfectée	amrâdu l jildi adîdatu ttalawwuti
eczéma	ikzîma, qawbâ'
eczémas exogènes chroniques de toute étiologie	al-'ikzîma al-muzminah xârijiyyatu l mana' mahma kâna masdaruha
érythème solaire	barqaCatu amsi
érythèmes	ihmirâru l jildi
herpès	hirbis, qawbâ'
ichtyose	dâ' u ssamak
lichen	hazâz
lupus	di'bah
lupus érythémateux chronique	al di' batu l'ihmirâriyyatu 'almuzminah
prurit	hikkah
pustulose	dâ' u l butûr
pustulose palmo-plantaire	dâ' u l butûri fî rahati l yadi wa batini l qadami
urticaire	aarâ, al artikariyâ, al hazâzât, attafah al jildiy

## 2. — FORME DES UNITÉS TECHNOLECTALES

### 2.1 — Les unités simples

Les unités simples en arabe standard se caractérisent par une exploitation exclusive de la dérivation morphologique i.e. les schèmes. L'observation de ces formes, dans le cadre de notre thèse de

doctorat consacrée aux technolèctes, a conduit à la conclusion que l'exploitation des schèmes radicaux est moins productive que celle des schèmes extra-radicaux.

Sur l'ensemble des schèmes extra-radicaux, qui sont au nombre de 175 dans AGR et 23 dans SAN et en vertu du seuil minimal retenu soit 5 %, les schèmes les plus fréquents dans AGR sont au nombre de sept : micc2ac3 (30 occurrences), clac2c2ac3ah (27 occurrences), mic1c2ac3h (22 occurrences), muclac2c2ic3ah (15 occurrences), mic1c2àc3 (13 occurrences), clàc2ic3a (10 occurrences), clàc2ùc3ah (7 occurrences). Quant aux schèmes extra-radicaux les plus productifs dans SAN, ce sont : clùc2àc3 (17 occurrences) et clac2àc3 (10 occurrences).

Ces résultats confirment ceux des études antérieures sur l'exploitation des schèmes dans la terminologie arabe.

Nous ne nous attarderons pas à ces aspects dérivationnels qui ont déjà fait l'objet de recherches approfondies et nous nous intéresserons davantage ici à la composition et à la syntagmatisation qui caractérisent les unités complexes.

## 2.2 — Les unités complexes

Ce sont des «conglomérés» (Benveniste, 1966 : 171) qui sont des constructions syntagmatiques à deux ou plusieurs éléments.

Dans le corpus lexicographique (AGR), les critères d'identification ne nous posèrent pas problème. En revanche, dans le discours des notices pharmaceutiques, il n'était pas toujours aisé de délimiter les frontières des unités complexes. Dans les cas ambigus, nous avons recours, en plus du critère formel (deux ou plusieurs éléments) à un deuxième critère : celui du degré de soudure, d'adhérence des éléments examinés en fonction de la non-séparabilité des éléments qui est vérifiée lors du test de la détermination globale par l'adjectif. Prenons un exemple tiré de SAN :

- a. 'iltihâbu l - qasabâti l - hâddu  
inflammation art.déf. bronches art.déf. aigue
- b.'iltihâbu l - qasabâti l - hâddati  
inflammation art.déf. bronches art.déf. aigues

Dans l'exemple a, l'adjectif [*hâddu*] «aigu» détermine globalement l'ensemble [*iltihâbu l qasabâti*] en appliquant la règle de l'accord en genre et en nombre avec le premier élément [*'iltihâbu*]

qui est au masculin singulier. Ce qui peut être traduit littéralement en français : «inflammation des bronches aiguë» et en technolecte médical «bronchite aiguë».

Dans l'exemple b, l'adjectif [*hâddati*] détermine le second élément [*l- qasabâti*] qui est au féminin pluriel et ne se rapporte pas à l'ensemble de la séquence. La traduction littérale de b correspondrait à : «inflammation des bronches aiguës» qui ne pourrait être «une bronchite aiguë».

Dans un cas, l'adjectif «aigu» aura le sens de «intense, violent» dans l'autre, il aura celui de «pointu».

En résumé, la méthodologie d'identification de ces unités complexes s'est appuyée sur trois critères :

1. Le critère purement formel qui consiste à relever les séquences contenant plus d'une suite graphique délimitée par des blancs. Ce travail préliminaire pouvait même être réalisé par une machine programmée en fonction de règles restrictives; par exemple, le nombre maximal des éléments constituants, la nature des fonctionnels à, de, etc. — avec le risque que l'ordinateur en balayant les unités, ne ramasse indifféremment toute suite, de nombre *n* d'unités graphiques, en ne distinguant nullement entre les suites qui présentent une adhérence, une soudure, autrement dit une cohésion suffisamment forte pour que la séquence dans son ensemble soit considérée comme un signe unique et les suites qui ne présentent pas cette caractéristique. D'où la nécessité d'un second critère.

2. Le critère syntaxique de la détermination s'est révélé très utile et permettait ainsi de vérifier le degré de cohésion d'une suite (voir exemple a et b supra).

3. Le critère sémantique qui est celui de l'appréhension précise du contenu technolactal (spécialisé) des unités complexes examinées, nous avons vu les deux interprétations possibles de «aigu» selon qu'il détermine inflammation ou bronche. Ceci présuppose le retour à la langue-source et le recours au spécialiste.

Ainsi, les questions auxquelles nous avons tenté d'apporter des éléments de réponse sont les suivantes : comment fonctionnent les unités complexes dans les technolectes en arabe standard ? Quelles sont les ressources linguistiques utilisées ? Quelles sont les stratégies déployées par cette langue pour répondre aux besoins de la communication technique et spécialisée ?

### 2.3 — Ressources linguistiques

Nous ne nous intéresserons ci-après qu'aux unités complexes et nous essaierons de présenter une typologie d'exemples types de ces constructions, rencontrés dans AGR et dans SAN. Ils sont au nombre de 16.

Exemple type 1 :      mihrât(un)      basât(un)  
                          charrue              simple  
                          «araire»

Modèle : Nom + Adjectif

Exemple type 2 :      nâfixat(un)              râfiCat(un)  
                          aspirateuse              chargeuse  
                          «aspirateur chargeur»

Modèle : Nom + Nom

Exemple type 3 :      darrâsat(un)      farrâmat(un)      harrâsat(un)  
                          batteuse              hacheuse              broyeuse  
                          «batteuse, hacheuse, broyeuse»

Modèle : Nom + Nom + Nom

Exemple type 4 :      râfiCat(un)      bi              awkat(in)  
                          élévatrice      fonctionnel      fourche  
                          «Fourche élévatrice»

Modèle : Nom + Fonctionnel + Nom

Exemple type 5 :      mihrât(u) l              kaf - (i)  
                          charrue      article      découverte      fonctionnel  
                          (génitif)      «charrue déchausseuse»

Modèle : Nom + Nom + Fonctionnel (génitif)

Exemple type 6 :      haffârat (un)      karrâyat (un)      li l xanâdiq  
                          creuseuse              cureuse              fonctionnel  
                          article fossés      «creuseuse cureuse des fossés»

Modèle : Nom + Nom + Fonctionnel + Nom

Exemple type 7 :      qattâCat(u)      l      durrat (i)      l      safrâ'(a)  
                          coupeuse              article      maïs              fonct.génitif  
                          article      jaune      «hâche-maïs»

Modèle : Nom + Nom + Fonctionnel + Adjectif

Exemple type 8 :      mut(on)      bi      'asnân (in)      marinat(in)  
                          peigne      fonctionnel      dents              flexibles  
                          «herse à dents flexibles»

Exemple type 9 : 'amrâd(un) jildiyyat(un) mikrûbiyyat(un)  
 multahibat(un) maladies dermiques  
 microbiennes inflammées  
 «dermatose microbienne inflammatoire»

Modèle : Nom + Adjectif + Adjectif + Adjectif

Exemple type 10 : mihrât(un) katîr(u) l maqâtiC(i)  
 charrue beaucoup article socs  
 «charrue polysoc»

Modèle : Nom + Adverbe + Nom

Exemple type 11 : mihrât(un) taht turb-iyy(un)  
 charrue sous sol suffixe adjectival  
 «charrue sous-soleuse»

Modèle : Nom + Adverbe + Adjectif

Exemple type 12 : 'âlat(u) qtilâC - i 'aCwçad - i lqotn - i  
 machine arrachage fonct.gén. bâtons  
 fonct.gén. coton «élévateur de coton»

Modèle : Nom + Nom + Fonctionnel + Nom + Fonctionnel + Nom

Exemple type 13 : 'âlat(u) ra - i l 'asmidat - i l  
 sâ'ilah machine action d'asperger fonct.  
 gén art engrais fonct. art(liquide)  
 «épandeur d'engrais liquide»

Modèle : Nom + Nom + Fonctionnel + Nom + Fonctionnel + Adjectif

Exemple type 14 : 'âlat(u) l hasâd - i wa l dirâs - i  
 machine art. moisson fonct.gén. et art.  
 battage fonct.gén.  
 «moissonneuse batteuse»

Modèle : Nom + Nom + Fonct. + Coordonnant + Fonctionnel + Nom

Exemple type 15 : jihâzun dawarâniy(un) li l ayy(i) bi l  
 tanqit(i) appareil rotatif pour art.  
 arrosage avec art. goutte à goutte  
 «arroseur rotatif»

Après avoir sérié les exemples-types des différentes unités complexes récurrentes dans AGR et dans SAN, nous ne saurions passer sous silence l'intervention «forte» de la syntaxe. Il est vrai que E. Benveniste avait déjà attiré l'attention sur ce phénomène concernant la composition. Il avait noté : «la composition nominale

est une micro-syntaxe» (1966 : 145). En effet, le technolecte se caractérise par des compositions qui correspondent à des schémas syntaxiques précis et aisément cernables vu leur récurrence (cf. les schémas supra).

Concernant les types dégagés, il convient de noter que les liens syntaxiques vont de la détermination simple qui s'instaure entre deux éléments selon l'ordre déterminé + déterminant à la détermination complexe à expansions. Cette malléabilité syntagmatique est l'une des possibilités créatrices des technolectes et l'une des voies ouvertes à la néologie.

Par exemple, la détermination par l'adjectif du déterminant du déterminé dans les deux formes suivantes a été fort utile lors de l'évolution de l'outillage :

- a. moton bi asnânin marinah «herse à dents flexibles»
- b. moton bi asnânin jazi'ah «herse à dents rigides»

La variation de la construction syntaxique permet aussi de distinguer entre deux notions; par exemple, entre dermatite et dermatose :

dermatite	<i>maradu l jidi</i>
dermatose	<i>amrâdun jildiyyah</i>

Par ailleurs, la syntaxe implique qu'il existe des fonctions précises au sein de l'énoncé et notamment celle hiérarchisée de la fonction prédicative. Pourrait-on parler au sujet des unités complexes de cette fonction ? Auquel cas, la distinction entre énoncé syntaxique libre et construction lexicale à schéma syntaxique complexe — appelée lexie ou synapsie — ne serait plus possible. L'analyse d'Emile Benveniste est que la fonction prédicative se trouve suspendue dans ces énoncés. Nous poserons à sa suite qu'effectivement la fonction prédicative n'est pas à repérer à l'intérieur d'une unité complexe mais rien ne s'oppose à ce qu'elle joue un rôle prédictif au sein d'un énoncé donné.

Ce n'est point là notre propos et ce qui importe, c'est la nature des rapports qui s'instaurent entre les éléments des unités complexes à travers les exemples types présentés ci-dessus. Le but est de cerner les moyens formels mis en oeuvre par la langue cible (l'arabe standard) afin de rendre des notions exprimées dans la langue-source qu'est généralement le français et parfois l'anglais.

En outre, l'on peut noter qu'il existe des unités technolectales complexes mais qui ne sont pas le résultat de combinaisons syntagmatiques d'un noyau (déterminé) et de plusieurs déterminants mais plutôt par des parties des éléments-formants. C'est un procédé qui produit les mots-valises ou les acronymes. Nous inspirant de l'analyse qu'en a faite B. Pottier, nous les appellerons des mixonymes. Ils ne sont pas très productifs dans les technolectes examinés mais il est utile d'observer qu'ils sont fréquents dans le technolecte médical. Nous en avons recueilli quelques-uns et nous avons pu constater qu'ils sont de deux types. Nous les présenterons respectivement dans ce qui suit :

### 1<sup>er</sup> type

#### Éléments-formants

taht + aljildah

taht + al fakk

qabl + al wilâdah

bayna + al asnân

#### Mixonymes

tahjildiy «sous cutané»

tahfakkiy «sous maxillaire»

qabwilâdiy «prénatal»

bay'asnâniy «interdantal»

### 2<sup>e</sup> type

#### Éléments-formants

wajaCu + al maCidah

fathu + al kiliyah

'isti'sâlu + al kiliyah

qatCu + al 'amCâ'

qatCu + al hunjarah

#### Mixonymes

wajCadah «gastralgie»

fatkalah «néphrostomie»

salkalah «néphrectomie»

qatmaCah «entérectomie»

qatjarah «laryngotomie»

Les mixonymes figurant dans le 1<sup>er</sup> type sont aisément décodables car relativement transparents. Le procédé est simple : la première syllabe appartient à un autonome (adverbe) et constitue le premier élément-formant identifiable et le second élément est un adjectif de relation (nom + suffixe -iyyun).

En revanche, les mixonymes du 2<sup>e</sup> type ne sont pas aisément décodables sachant que le premier élément n'est pas un morphème grammatical (affixe ou adverbe). Ces formations opaques ne pourront pas être décodées et d'ailleurs, elles ne sont pas fréquentes.

Pour conclure sur les mixonymes que nous incluons dans les unités technolactales complexes, nous pouvons retenir que la règle générale de formation est :

1<sup>re</sup> syllabe du premier élément-formant + Réduction du deuxième élément

Il est vrai que cette règle devrait être assortie des lois phonétiques de réduction des éléments-formants.

Pour finir, nous insisterons sur le fait que les mixonymes fonctionnent isolément à l'intérieur des technolactes et ne s'y intègrent qu'imparfaitement. Ils sont d'ailleurs souvent considérés comme hors système et ne sont point appréciés par les terminologues arabes qui y voient des formes barbares. Pourtant le procédé de la mixonymie, appelé traditionnellement «naht» par les grammairiens arabes devrait bénéficier de plus d'intérêt car c'est aussi une voie pour la néologie arabe.

Ayant essayé de décrire les procédés formels utilisés à partir des échantillons présentés dans deux domaines différents, nous pourrions souligner l'importance des procédés syntaxiques utilisés : la langue cible n'a pas toujours essayé de calquer la langue source. La notion est réinterprétée à partir d'un autre appareil formel endogène. Certes, les productions sont souvent des phrasèmes qui présentent une systématique comme cela ressort des exemples types dégagés et dans un second temps i.e. à une étape ultérieure, les phrasèmes (unités complexes) sont évités — dans la mesure du possible — par les terminologues qui présentent des unités simples et les proposent en remplacement des autres. C'est ce que nous avons constaté lors de notre travail sur le microchamp des noms de maladie. En comparant la nomenclature obtenue à partir des notices des médicaments et celle existant dans le dictionnaire de médecine unifié à l'échelle du monde arabe, nous avons effectivement pu vérifier l'hypothèse que les phrasèmes sont remplacés par des unités simples.

Ex. : *habbu abâb*, littéralement «boutons de jeunesse (juvénils)» pour «acné» est remplacé par *cuddun*; *amrâdun jildiyyah*, littéralement «maladies de la peau» pour «dermatose» par *julâd*;

*dâ'u ssamak*, littéralement «mal du poisson», pour «ichtyose» par *sumâk* et *di'bah* littéralement «louve» pour «lupus» par «*da'ab*».

C'est à l'exploitation des schèmes que va la préférence des terminologues arabes, ces exemples et surtout les documents dont ils sont puisés i.e. les lexiques unifiés à l'échelle du monde arabe l'ont

démontré. Mais les unités complexes (phrasèmes) constituent proportionnellement la plus grande partie des technolectes. Le recours aux procédés syntaxiques endogènes est important, les exemples types l'ont démontré.

L'étude formelle montre que la langue cible use de tous les moyens dont elle dispose allant de la détermination à la juxtaposition. Parfois même la mixonymie est utilisée.

Quant aux procédés allogènes qui se caractérisent le plus souvent par l'emprunt, ils ne sont exploités qu'en derniers recours et présentent une très faible productivité dans les corpus examinés. Cette conclusion va à l'encontre des idées qui circulent sur le fait que la langue-cible est passive et ne fait que recevoir et subir. Certes, le travail terminologique est de longue haleine, et il faut accorder une attention particulière aux technolectes et en particulier aux phrasèmes et aux mixonymes. Ils sont les signes de la dynamique d'une langue et des efforts déployés pour se développer. Deux stratégies sont mises en oeuvre. L'une de compensation consiste à vouloir combler les lacunes par le recours à l'emprunt de la langue source qui «compense» ainsi un vide terminologique constaté dans la langue-cible. L'autre d'adéquation puisque la langue cible utilise son propre système et en conformité (en adéquation) avec ses règles.

Comme toute langue soucieuse de se développer et d'être en mesure d'assimiler le progrès, la langue arabe use de ces deux stratégies avec une large prédilection pour la seconde. La recherche pourrait se poursuivre dans un cadre comparatiste en essayant d'évaluer le degré d'appropriation des technolectes par la langue cible en usant du double concept d'opacité et de transparence. Ceci nous conduirait à aborder en particulier les procédés sémantiques que nous n'avons pas du tout étudiés ici.

## RÉFÉRENCES

- BENVENISTE, Emile (1966) : *Problèmes de linguistique générale*, Paris, Gallimard.
- FELBER, Helmut (1984) : *Terminology manual*, Paris, UNESCO.
- HAGÈGE, Claude (1982) : «Voies et destins de l'action humaine sur les langues», *La réforme des langues*, Hambourg, Buske, pp. 11-67.
- MESSAOUDI, Leila (1986) : «L'utilisation de la base de données Lexar (de l'IERA) dans la recherche des emprunts dans la terminologie scientifique arabe», *Dialogue entre la langue arabe et la langue française*, Paris, CILF, pp. 97-109.
- MESSAOUDI, Leila (1990) : *Des technolectes*, Thèse de Doctorat d'État, Paris V.
- MESSAOUDI, Leila (1995) : «Linguistique et traduction. Le cas des technolectes», *Traduction et interprétation des textes*, Publications de la faculté des lettres, pp. 5-15.
- MESSAOUDI, Leila (à paraître) : «Un technolecte agricole en arabe dialectal marocain dans le parler des jbala».
- POTTIER, Bernard (1974) : *Linguistique générale. Théorie et description*, Paris, Hachette.
- RONDEAU, Guy (1985) : *Introduction à la terminologie*, Montréal, G. Morin éditeur.

## Fréquence, polysémie, synonymie : l'exemple du verbe *dire* dans le coran

Mohamed Lazhar ABBÈS

*Université Tunis I, Tunis, Tunisie*

Cette communication se veut une réaction par rapport à une remarque de J. Picoche (1991 : 201) sur la fréquence d'emploi des mots de la langue : «Je m'étonne qu'on n'accorde pas plus d'importance à la fréquence des mots qui est loin d'être sans relation avec leur polysémie et même la nature de cette polysémie». En effet, la consultation du dictionnaire nous montre aisément que ce sont les mots les plus courants qui sont polysémiques, que la polysémie d'un mot est en relation directe avec sa fréquence d'emploi et que cette même polysémie est la cause primordiale de la multiplication des synonymes.

En français, l'exemple nous est fourni par le verbe *dire*, l'un de ces verbes qui réfèrent à l'activité langagière et qui, dans un énoncé, peuvent réfléchir l'acte de communication lui-même. Ce verbe occupe une place privilégiée par sa fréquence d'emploi (le plus fréquent après *faire*), sa polyvalence syntaxique et sa plasticité sémantique (quatre-vingt-dix synonymes dans *Le Grand Robert*).

Depuis ses origines indo-européennes (*déic, dic* «montrer»), le verbe *dire*, avait et conserve dans un certain nombre de ses représentants un caractère solennel, religieux et juridique. La définition étymologique de *bénir, benedicere* «dire du bien», de *juger, judicare* «dire le droit» ou de *prêcher, praedicare* «dire en public»

présente *dire* comme définisseur ou archilexème d'un nombre de verbes extrêmement important.

L'emploi «abusif» que connaît le verbe *dire* est loin d'être l'apanage de la communication orale courante, ni celui de l'écriture moderne (cf. *L'amant* de Duras) ni non plus de la langue française. La langue arabe dispose de son correspondant, le verbe *qâla*. En effet, autant que d'autres civilisations de l'oralité, la civilisation arabe ancienne accorde beaucoup d'importance à la parole proférée. Plus tard, l'apparition du Coran ne fait que renforcer le prestige de la parole, puisque, on le sait, en Islam, le miracle est d'ordre verbal.

Quiconque lit ou écoute réciter le Coran constate la fréquence exceptionnelle du verbe *qâla* et de ses dérivés. Voyons donc comment se comporte ce verbe dans le texte coranique, ce texte qui, comme le souligne Massignon «se présente comme une dictée surnaturelle, enregistrée par le prophète inspiré, simple messenger, chargé de la transmission de ce dépôt...» (Grosjean, 1976 : Préface).

Notre démarche va de l'étude des structures de la langue, telles qu'elles apparaissent dans les grammaires et les dictionnaires, à la description d'usages discursifs. Le verbe *qâla* nous servira à démontrer que la description des propriétés lexicales et distributionnelles contribue à l'analyse des valeurs sémantiques.

Le dépouillement du texte coranique a permis le relevé de 1730 occurrences de la racine [KWL] sous 49 formes différentes réparties comme suit :

*qâla* (529), *qâlû* (332), *qul* (331), *yaqûlûna* (92), *yaqûlu* (68), *qîla* (52), *qâlat* (43), *qulnâ* (27), *layaqûlanna* (15), *naqûlu*, *qûlû*, *qawluhum* (12), *taqûlûna* (11), *taqûlû* (10), *yaqûlû*, *qultum*, *aqûlu* (9), *yaqûlû* (8), *taqûlû*, *qultu*, *aqul* (6), *taqûla* (4), *taqûlu*, *taqûlu*, *qâlâ*, *qûlâ*, *qîlun*, *qâilun* (3), *taqûla*, *qâlatâ*, *qulna*, *qawlukum*, *qawluhu*, *qawlî* (2), *qâlahâ*, *qultuhu*, *naqul*, *lanaqûlanna*, *yaqul*, *yaqul*, *qulna*, *qûli*, *yuqâlu*, *taqûlu*, *taqûluhu*, *qawluka*, *qawlunâ*, *qawluhâ*, *alaqawîl*, *qîlihi*, *qâiluha*, *qâilîna* (1).

*Qâla* est parmi les verbes qui forment la quasi-totalité du matériel verbal de l'arabe, que les grammairiens arabes désignent par actif pour marquer que «le sujet participe d'une manière effective et parfaite à l'accomplissement de l'action» (Blachère, 1975 : 257).

C'est un verbe transitif direct qu'on peut catégoriser, du point de vue de l'aspect lexical, comme verbe d'événement opposé à *takallama*

(parler), verbe de *processus*, en ce sens que *qâla* est intrinsèquement **perfectif** (porte en lui-même l'idée d'un terme) si bien que lorsqu'il est construit transitivement il doit être obligatoirement accompagné d'un complément alors que son correspondant **imperfectif** peut s'en passer : ainsi *takallama* (parler) et *qâlu kadha* (dire qqch). Mais en tant que verbe de dire, il est trivalent, il comporte trois actants : *Qâla kadha li foulén* (dire qqch à qqn) (Tesnière, 1982 : 255-258).

Sur le plan sémantique, *qâla* a la caractéristique de servir de terme générique dans le lexique et de substitut dans l'usage parlé spontané. Il exprime le contenu commun à beaucoup de verbes moins fréquents que lui et il appartient à la série des verbes introducteurs d'énoncés. Sa description tient à cette relation d'ordre lexical qu'il entretient avec les autres verbes selon laquelle il est hyperonyme.

Avant d'en examiner l'emploi, il convient de donner quelques indications sur la formation de ce verbe. *Qâla* est une racine concave, parmi les 40 racines trilitères avec une semi-voyelle ou semi-consonne «[w]» au milieu de deux consonnes fortes et différentes. Cette semi-voyelle, étant entre deux voyelles brèves, a disparu en formant une longue dans *qawula* > *qâla*. Son inaccompli est du type *yaSEulu* : *yaqûlu*. C'est la catégorie la plus fréquente. (Blachère, 1975).

## LE SUJET DU VERBE QÂLA

Le procès exprimé par le verbe *qâla* exige de son agent la capacité d'énoncer, la propriété *locuteur*. Dans la plupart de ses occurrences, le verbe *qâla* a pour sujet grammatical — dans le cas d'une forme verbale active — un terme dont le référent a les traits [animé] et [humain]; quand ce n'est pas le cas, le sujet — en tant qu'agent de *qâla* — reçoit dans le contexte la propriété d'être locuteur. C'est ce qui se passe dans le verset suivant où cette propriété est conférée à «leurs peaux» par extension métonymique :

«Et ils demanderont à leurs peaux : «pourquoi avez-vous témoigné contre nous ?» — [leurs peaux] répondront : «Allah nous a donné la parole, [lui] qui donna, la parole à toute chose, lui qui vous créa une première fois et vers qui vous serez amenés.»» (*Elles ont été rendues intelligibles* 21).

Dans le Coran, le dire est attribué à Dieu, aux Anges, aux êtres humains, aux Djinns, mais aussi aux animaux et aux objets inanimés.

## L'OBJET INDIRECT DE *QÂLA*

De la même façon que le sujet de *qâla* a la propriété locuteur, l'objet indirect de *qâla* si tant est qu'il en admette un (1), a la propriété allocutaire; le référent dénoté par cet objet indirect a donc en général, comme le référent du sujet, les traits [animé] et [humain]. Toutefois, des objets «inanimés» peuvent sous l'effet de personnification se trouver affectés de la faculté de recevoir des ordres comme dans le verset suivant :

«Ensuite, il se tourna vers le Ciel alors qu'il était fumé et il lui dit, ainsi qu'à la Terre : «Venez de gré ou de force !», Et le Ciel et la Terre répondirent : «Nous venons avec obéissance.»»

## LE COMPLÉMENT D'OBJET DIRECT DE *QÂLA*

Le verbe *qâla* se construit avec un groupe nominal, complément d'objet direct. Ce groupe nominal réfère à une forme ou un contenu *énonçable* (2).

«Toutefois, ne leur promettez rien secrètement sauf à [leur] dire des paroles reconnues [concevables].» (*La Génisse* 235)

Certains compléments d'objet direct ne renvoient pas à du *dit* ou du *préférable* de la même façon que «des paroles», dans l'exemple précédent; ainsi dans le verset :

«[...] ne leur dis pas : "Fi" et ne les brusque point» (*Le voyage nocturne ou les fils d'Israël* 23)

l'interjection «fi» est (3) la parole qui a été dite. Cette interjection est ici substantivée : muni de l'article zéro, elle fonctionne comme complément d'objet direct de *qâla* et non pas comme une citation au sens stricte du terme. Dans le verset suivant les mots autonymes sont des verbes à l'impératif constituant des mots-phrases :

«Ô vous qui croyez !, ne dites point : «Considère-nous (?) !», [mais] dites : «Donne-nous d'attendre (?) !»» (*La Génisse* 104)

---

1) Il faut noter qu'il n'est pas, dans le Coran, que l'interlocuteur ne soit exprimé avec le verbe *qâla* : le procès exprimé par celui-ci apparaît alors comme un acte locutif dans l'absolu, au sens où il ne met pas nécessairement en jeu un allocutaire.

2) Terme employé par M.-M. de Gaulemyn (1983 : 155).

3) *est* signifie que «fi» est ici en emploi autonymique, au sens où l'entend, à la suite de J. Rey-Debove, J. Authier (1978 : 54 ss.).

La construction <*qâla kadha*> est en général un simple renvoi à un acte énonciatif et ne représente en aucune façon le contenu des paroles produites au cours de cet acte. Les substantifs objets directs de *qâla* sont les traces d'une citation résumée.

«Qu'ils [les tueurs] soient pieux envers Allah et adressent [à leurs pupilles] un propos plein de raison.» (*Les femmes* 9)

Cette construction ne constitue donc pas une forme de rapport d'énoncé, sur laquelle nous reviendrons.

### LE VERBE *QÂLA* : VARIATION SUR «DIRE»

Dans certains des exemples cités plus haut, *qâla* a été utilisé là où la langue écrite moderne emploierait d'autres verbes exprimant un acte de parole. Selon le contexte dans lequel il se trouve, le verbe *qâla* reçoit en effet différentes interprétations.

«... Ou comme celui qui, étant passé devant une cité déserte et subversée, s'écria : «comment Allah pourrait-il faire revivre cette cité déserte et subversée, après sa mort ?»» (*La Génisse* 259)

Si *qâla* peut être traduit par «dire» en français, il a parfois le sens de «prétendre», «demander», «répondre», «commander», «conseiller», «avertir», «prévenir», «recommander».

Le sens peut provenir aussi de la construction du complément du verbe : ainsi *qâla an* «informer», *qâla ala* «prétendre».

Comme il peut être le résultat de la relation qui s'établit au niveau sémantique entre le complément d'objet indirect et son attribut. C'est une appellation que recouvre la forme *wuqâlu lahu* «s'appeler», dans le verset suivant :

«Nous avons entendu», ajoutèrent-ils, «un jeune homme nommé Abraham qui vilipendait [ces idoles].» (*Les Prophètes* 60)

Si le verbe *qâla* joue un rôle fondamental dans le rapport d'énoncé — rôle qui consiste pour l'essentiel à signaler la reproduction d'un discours — il y a un autre élément qui apparaît tout aussi important dans ce mode de rapport d'un énoncé, il s'agit de *inna*, qui, placé à la jonction de l'énoncé rapporteur et de l'énoncé rapporté, agit comme une marque de début de proposition ou de début d'énoncé. Selon Blachère (1975 : 380) «La valeur en est nettement affective : elle souligne le ton solennel du discours, accompagne le geste».

«[le Seigneur] dit : «Je vais faire de toi un guide ('imâm) pour les Hommes»» (*La Génisse* 124).

Ce qui permet d'identifier le *rapport discontinu* c'est donc la rupture entre deux situations d'énonciation (pause, changement d'intonation, procédés typographiques divers).

Si, en arabe moderne comme en français, la présence d'un verbe de locution renforce, certes, efficacement cette coupure, mais n'est pas indispensable, le contexte pouvant suffire à marquer le changement de plans énonciatifs, dans l'arabe du Coran, on ne trouve pas de citations au discours direct où les répliques sont seulement juxtaposées.

Le Coran, dans ses versions arabes, se dispense des marques typographiques (les deux points, les guillemets, les tirets) et ôte la frontière entre énoncé rapporteur et énoncé rapporté au point de rendre difficile la perception du décalage entre les deux registres. Il faut que le lecteur recoure au contexte pour identifier les diverses sources énonciatives.

À défaut des signes graphiques, le lecteur doit compter sur les signaux syntaxiques qui sont assez forts pour marquer nettement le passage du récit au dialogue, au *rapport discontinu* : passage de la troisième personne à une alternance rapide entre la première et la seconde personne, transition temporelle, emploi d'un verbe de communication.

En effet, le verbe *qâla* constitue le pivot sur lequel s'articule le rapport discontinu. C'est la séquence <*qâla* + E> que l'on considère comme schéma de base pour définir un rapport discontinu. Nous nous proposons de donner quelques exemples qui mettent en évidence la spécificité de l'énoncé E dans la séquence du type <*qâla* + E>.

### SÉQUENCE <QÂLA + E> : AUTONOMIE DE L'ÉNONCÉ

Si l'énoncé E dans la séquence <*qâla* + E> prend place et joue le rôle de complément d'objet de *qâla*, il fonctionne cependant au niveau syntaxique d'une façon tout à fait indépendante. Dans les deux exemples suivants, l'énoncé E est composé de formes verbales à l'impératif ou au prohibitif :

«Ô Adam !», **dit** [le Seigneur], «Avises-les des noms [de ces êtres] !» (*La Génisse* 33)

(«L'un d'eux dit : «Ne tuons pas Joseph, mais jetons-le dans les profondeurs de tel puits !»» (*Joseph* 12)

Mais si E peut constituer à lui seul une phrase complète le syntagme *qâla* + *sujet* reste incomplet et ne peut former une proposition à lui seul. E n'apparaît pas en fait comme un objet de même nature que *qâla* : E est un énoncé, une suite graphique ou phonique, et la seule propriété requise de E est d'être «proférable».

Le caractère proprement énonciatif apparaît dans les exemples suivants où sont rapportés des actes locutifs spécifiques :

«Celui-ci cria [alors à son frère] : «Je te tuerai !»» (*La Table servie* 27)

«[Rappelez-vous] Quand Allah demanda : «O Jésus, fils de Marie, est-ce toi qui as dit aux Hommes : «Prenez-nous, moi et ma mère comme divinité en dessous d'Allah !»» (*La Table servie* 116)

«Ils lui dirent : «N'aie pas peur ! Nous sommes deux parties en querelle l'une avec l'autre»» (*Sad* 22)

Avec le premier verset nous avons l'expression d'une menace, avec le deuxième, un reproche ou un blâme et avec le dernier l'expression d'une rassurance. Ce sont effectivement des actes énonciatifs spécifiques sur lesquels le verbe introducteur *qâla* n'apporte aucune information. Peut-être l'utilisation du discours indirect amène-t-elle l'emploi d'autres verbes substituables au verbe *qâla*. Ce qui pourrait faire apparaître des verbes comme *dénier*, *regretter*, *rassurer*. Le cas de ces verbes montre qu'il existe deux grandes catégories de verbes introducteurs selon qu'ils participent ou non à la reproduction d'une fraction de l'énoncé qu'ils introduisent, autrement dit, selon qu'ils apportent ou non une information sur son contenu.

En arabe classique, pour passer du récit au discours, le discours direct est couramment introduit par une formule composée de deux verbes de dire, le second étant *qâla* c'est-à-dire le verbe locutif, qui est sur le plan sémantique, le moins informatif. C'est là un procédé narratif largement utilisé.

Dans ce type de formule introductrice, c'est le premier composant qui donne l'essentiel de l'information sur l'acte d'énonciation rapportée. Le premier verbe locutif précise en général soit les circonstances ou le contexte dans lequel s'insère le discours (prise de parole, réponse, ...), soit les modalités énonciatives (injonction, autorisation, avertissement, suggestion, interrogation, requête, promesse, vouloir, déclaration, proclamation, etc.), et le cas

échéant les interlocuteurs de la situation d'énonciation rapportée. Le verbe *qâla* apparaît comme redondant derrière le premier verbe locutif puisqu'il ne donne aucune information supplémentaire ni sur la modalité énonciative, ni sur les interlocuteurs, ni sur les circonstances de l'acte d'énonciation rapportée.

La fonction du verbe *qâla* dans ce type de formule s'apparente à celle du gérondif français «en disant» forme que ne possède pas la conjugaison arabe.

S'il est difficile de dresser une liste exhaustive des verbes qui se rencontrent comme premier composant de la formule introductrice narrative, on peut en fournir les plus fréquents.

#### **SÉQUENCE <QÂLA + E> : LE CAS D'UN ÉNONCÉ QÂLA À L'IMPÉRATIF. «LE DISCOURS MÉDIATISÉ» (4)**

Parmi les différentes formes que peut prendre le verbe *qâla* dans la séquence <*qâla* + E>, le cas de l'impératif doit être traité séparément : la forme *qul* «dis/dites» est liée à une situation d'énonciation tout à fait spécifique. Dans les séquences introduites par le verbe associé à la première personne au masculin singulier, *qul*, qui sont de l'ordre de 331 dans le Coran, équivaut à une demande adressée par un locuteur A, en l'occurrence Allah à un interlocuteur B, Muhammed, le Prophète bien entendu, de transmettre un message à une tierce personne C; ce message est représenté par l'énoncé E. Voici quelques exemples pour illustrer «le discours médiatisé» :

«Dis : «Il est Allah, unique, Allah le seul. Il n'a pas engendré, et n'a pas été engendré. N'est égal à lui personne»» (*Le Culte*)

«Dis à ceux qui ont été impies : «Vous serez vaincus et rassemblés vers la Géhenne.»» (*Al Imrane* 12)

«[Les croyants] t'interrogent sur les boissons fermentées et [le jeu de] maysir. Réponds[-leur] : «Dans les deux cas, sont pour les hommes, un grand pêché.»» (*La Génisse* 219)

Quelle que soit la façon dont on interprète la séquence <*qâla* (impératif) + E>, c'est une situation relativement complexe qui est mise en place, lié au rôle médiateur de l'interlocuteur «présent», le messenger.

---

4) A. Ouzounian, «"Le discours médiatisé" dans lequel le locuteur demande à son interlocuteur de transmettre un message à un tiers» (1992 : 4).

En effet, le verbe *qâla* reste l'élément le plus important indiquant qu'il y a énonciation. Selon l'analyse sémantique de K. Bogacki (in Jamrozik, 1984 : 61), c'est un verbe à «structure sémantique simple. Dans le cas où il est interlocuteur du discours direct, «il est porteur d'une information non spécifiée par le contenu sémantique» [...] il a le sens ;le sujet parlant x communique quelque chose à son interlocuteur y»». Il a cette réputation de verbe neutre par excellence, une sorte de degré zéro des modalités d'évaluation assignées aux locuteurs en ce sens qu'il n'implique pas de jugements de valeur de l'énonciateur quant au caractère *bon / mauvais* de l'énoncé cité, porté par l'agent du procès comme (critiquer), ni porté par l'énonciateur comme (alléguer) où le procès est dévalué.

La traduction française de R. Blachère, par exemple, nous a aidé à relever cette tendance systématique à employer *qâla* là où le français et même l'arabe moderne utilisent des verbes plus précis et plus nuancés qui spécifient sémantiquement l'énonciation sur différents registres : *répondre* ou *répliquer*, pour la situer par rapport à une parole antérieure, *s'écrier* pour donner une information sur l'état affectif du locuteur ou *demander* quand il s'agit d'introduire une question.

«[Fils d'Israël !,] étiez-vous témoins quand, la mort se présentant à Jacob, celui-ci **demanda** à ses fils «Qu'adorerez-vous après moi ?» [Étiez-vous présents quand] ses fils **répondirent** : «Nous adorerons ta Divinité et la Divinité de tes pères Abraham, Ismaël et Isaac, Divinité Unique à laquelle nous sommes soumis»» (*La Génisse* 132)

«Ou comme celui qui, étant passé devant une cité déserte et subversée, **s'écria** : «comment Allah pourrait-il faire revivre cette cité déserte et subversée, après sa mort»» (*La Génisse* 259)

En effet, le verbe *qâla* n'est pas un «verbe modal» puisqu'il convient quelle que soit la modalité, il remplit la fonction d'«insertion», c'est-à-dire annonce et justifie l'insertion du discours direct (Bonnard, 1986).

Son statut de verbe introducteur d'énoncé de type déclaratif, interrogatif ou injonctif lui permet une assimilation aux verbes d'opinion *dhanna* (croire), de parole *ajâba* (répondre), de révélation *aqarra* (avouer), d'interrogation *saala* (demander), d'ordre *amara* (ordonner). Ces classes n'ont pas souvent de frontières bien délimitées. On peut par exemple classer dans les verbes injonctifs ou interrogatifs

sans que le trait sémantique de /communication adressée/ ni celui de /comportement verbal/ aient été pris en compte (5)

Inversement, ces traits sémantiques caractérisent des verbes qui ne partagent pas les propriétés syntaxiques de *qâla*, qui introduit toujours le discours direct avec ou sans *inna* (6). La transformation au discours indirect fait automatiquement apparaître au niveau de l'énoncé rapporteur des verbes du type *aalana* (annoncer), *akhbara* (informer), *haddatha* (rapporter), *zaama* (prétendre), *kaddama* (avancer) qui introduisent toujours l'énoncé rapporté.

Voici l'essentiel des observations que nous avons pu faire sur l'usage de *qâla* dans la présentation des répliques dans le Coran.

Les écrivains utilisent de façons très différentes les synonymes et les mots de sens équivalents. Soucieux du mot propre, ils distinguent des nuances, voire des nuances entre des mots que le langage courant a tendance à confondre. Le Coran, au contraire, se veut une langue qui «économise» en utilisant le même mot à plusieurs fins.

L'utilisation systématique du verbe *qâla* semble être caractéristique du texte coranique et contribue à le distinguer pour sa rectitude attributive.

---

5) M. Grevisse, *Le bon usage*, 8e édition, Duculot, 1964, pp. 1018 et 1024.

6) R. Blachère, *Grammaire de l'arabe classique* : «Quand elle est en attaque de phrase, elle se prononce *inna*; la valeur en est nettement affective : elle souligne le ton solennel du discours, accompagne le geste».

## RÉFÉRENCES

- ABDEL-MAJID, A. (1979) : «Madkhal fi funun al-gawl inda l'arab», *Majallat il-lugha al-arabia bi Dimachq*, LIV, fasc. 2 apr.
- ALMENBERG, D. (1958) : «La présentation des répliques chez quatre romanciers contemporains», *Studia Neophilologica*, XXX, pp. 200-213.
- ANGHELLESCU, J. (1981) : «Oser dire en Arabe», *Revue roumaine de linguistique roumaine*, vol. 26, n° 4, pp. 324-327.
- AUTHIER, J. (1978) : «Les formes du discours rapporté : remarques syntaxiques et sémantiques à partir des traitements proposés», *Documentation et recherches en linguistique allemande contemporaine-Vincennes (DRLAV)*, n° 17, pp. 1-87.
- AUTHIER, J. (1979) : «Problèmes posés par le traitement discursif dans une grammaire de phrase», *Linguisticae Investigationes*, III, 2, Amsterdam, John Benjamins, pp. 211-228.
- AUTHIER, J. (1981) : «Paroles tenues à distance», B. Conein, J.-J. Courtine, F. Gadet, F. Marandin et J.-M. Pêcheux, *Matérialités discursives*, Actes du Colloque des 24, 25, 26 avril 1980, Université Paris X-Nanterre, Lille, Presses universitaires de Lille.
- AUTHIER, J. (1982) : «Hétérogénéité montrée et hétérogénéité constitutive : Éléments pour une approche de l'autre dans le discours», *DRALV*, n° 26, pp. 91-151.
- AUTHIER, J. (1987) : «Compte rendu de *La parole d'Altri* de B. Mortara Garavelli», 1985, *Journal of Pragmatics*, vol. 11, n° 3, Amsterdam, North-Holland, pp. 427-431.
- AUTHIER, J. (1992) : *Les non-coïncidences du dire et leur représentation méta-énonciative : Étude linguistique et discursive de la modalisation linguistique*, Thèse de Doctorat d'État, Université de Paris VIII.
- AUTHIER, J. (1993) : «Le discours rapporté», *L'information grammaticale*, n° 55 et n° 56.
- BALLY, Ch. (1983) : *Traité de stylistique française*, 4<sup>e</sup> édition, vol. 2, Klincksieck.
- BLACHÈRE, R. (1957) : *Le Coran (Al Qor'ân)*, Paris, Maisonneuve.

- BLACHÈRE, R. (1959) : *Introduction au Coran*, Paris, Maisonneuve.
- BLACHÈRE, R. (1966) : *Le Coran*, coll. «Que sais-je ?», Paris, PUF.
- BLACHÈRE, R. (1975) : *Grammaire de l'arabe classique : (morphologie et syntaxe)*, Maisonneuve.
- BONNARD, H. (1986) : «Le discours», «Le lexique», «La distribution», *Grand Larousse de la langue française (GLLF)*, Paris, Larousse.
- CHAROLLES, M. (1976) : Exercices sur les verbes de communication», *Pratiques*, n° 9, pp. 83-107.
- GAULMYN, M.-M. de (1986) : «Sur les verbes de parole. Dire, demander, proposer», S. Rémi-Giraud et M. Le Guern (dir.), *Sur le verbe*, Lyon, PUL.
- GAUVENET, H. (1976) : *Pédagogie du discours rapporté*, Didier (CREDIF).
- GENETTE, G. (1972) : *Figures III*, Paris, Seuil.
- GIRY-SHNEIDER, J. (1978) : «À propos de quelques nominalisations», *Langue française*, n° 39, Paris, Larousse, pp. 30-48.
- GIRY-SHNEIDER, J. (1981) : «Les compléments nominaux du verbe dire», *Langages*, n° 63, pp. 75-97.
- GRIMSHAW, J. (1979) : «Complement Selection and the Lexicon», *Linguistic Inquiry*, vol. 10, n° 2, pp. 279-326.
- GROSJEAN, J. (1976) : *Le Coran*, Paris, Gallimard.
- GROSS, G. (1984) : «Analyse distributionnelle et étude sémantique», A. Guillet et N. La Fauci (dir.), *Lexique-grammaire des langues romanes*, Amsterdam, Philadelphia, John Benjamins, pp. 129-138.
- GROSS, M. (1992) : «Lexique et syntaxe», *Actes du colloque de lexicologie française*, Louvain-la-Neuve (21-23 mars 1991).
- JACQUINOD, B. (1988) : «Accusatif, topique et verbe 'dire'», *Logopédies, Mélanges de Philologie et de linguistique grecques offerts à Jean Taillardat*, Paris, Peeters/Selaf, pp. 103-110.
- JAMROZIK, E. (1984) : «Relations entre syntaxe et sémantique sur l'ensemble des verbes d'activité de parole», *Colloque international de linguistique romane*, Jadwisin, *Lexique et grammaire des langues romanes*, K. Bogacki (dir.), Warszawa, wydawnictwa Uniwersytetu Warszawskiego, 1987, pp. 57-72.
- LEON, J. (1988) : «Formes de discours direct dans des récits oraux», *LINX*, n° 18, pp. 107-124.

MAJMAA ALLUGHAH AL ARABIA (1981) : *Moâjam Alfadh Al Qor'an*, Dar Achurûk.

MILICKOVA, L. (1986) : «Les verbes à modalité impérative dans le discours rapporté», *Et., rom. Brno.*, t. 17, pp. 73-82.

MONTET, G. (1954) : *Le Coran*, Paris, Payot.

STRAUCH, G. (1972) : «Contribution à l'étude sémantique des verbes introducteurs du discours indirect», *Recherches Anglo-Américaines*, n° 5, Strasbourg, pp. 226-242.



# De la racine au mot ou du mot à la racine : problématique de la création d'une nouvelle mémoire de l'emprunt en arabe (1)

Hassan HAMZÉ

*CRTT, Université Lumière Lyon-2,  
Lyon, France*

## 1. — LE MOT ARABE

### 1.1 — Mots premiers et mots dérivés

Dans l'optique des grammairiens arabes, le lexique comporte deux parties :

— la première est constituée de mots considérés comme premiers, c'est-à-dire non dérivés. Cette catégorie comprend des noms communs, des noms propres et des pronoms. Az-Zajjâjî dit : « Certes, il y a dans le discours des noms premiers qui sont des noms humains et non humains qui s'appliquent d'abord pour distinguer entre les uns et les autres. Exemples : un homme, un cheval, un vêtement, un oeil, une marmite et ce qui leur ressemble. Il y a aussi des noms qui sont établis comme des [noms] propres par convention. Chacun d'eux indique la personne elle-même. Exemples : *Zayd*, *Ja<sup>c</sup>far* et ce qui leur ressemble. Il ne fallait donc pas que le nom propre fût commun [à

---

1) La transcription de certains caractères de l'arabe n'étant pas possible avec nos logiciels de traitement de texte, les phonèmes emphatiques seront en gras : **d**, **ḍ**, **h**, **s**, et **t**; les phonèmes chuintants seront soulignés : g et s.

plusieurs personnes]. Mais, avec le temps, il est devenu commun, par extension [à plusieurs personnes], et la distinction [entre les différentes personnes ayant le même nom] s'est faite au moyen des épithètes. Il y a également, des noms vagues (*mubham*) [i.e. relatifs et démonstratifs] et des noms implicites (*mudmar*) [i.e. des pronoms]» (*'Istiqâq* : 281-282).

— la deuxième comprend des mots considérés comme dérivés à partir des unités de la première partie comme les noms d'agent et de patient dérivés du verbe, les formes du duel et du pluriel tirées du singulier. Les grammairiens arabes ont identifié dans ces formes des éléments qui «chutent dans un contexte (*hâl*) et se maintiennent dans un autre. Ils ont constaté donc qu'ils dérivent de formes primaires sans augments» (*'Istiqâq* : 282-283).

## 1.2 — La dérivation

En arabe, la dérivation se fait, généralement, par le jeu de la flexion interne, parfois au moyen de la flexion externe comme c'est le cas dans le duel, le pluriel dit «sain» et le nom de relation.

Elle a été définie par Ibn Dihyat (633/1236) <sup>(2)</sup> comme une opération qui «consiste à prendre une forme (*sîgat*) à partir d'une autre avec un sens commun, une matière première commune et un ordre commun de cette matière pour indiquer, avec la seconde, le sens de la première avec un sens supplémentaire» (*al-Muzhir* : 1, 346).

Cette conception de la dérivation implique donc, comme le dit Ibn as-Sarrāj (316/928) <sup>(3)</sup>, deux conditions :

1° «Les *harf* <sup>(4)</sup> que les grammairiens arabes évaluent en F.<sup>c</sup>.L. [i.e. les consonnes radicales] de l'un des deux mots doivent se retrouver dans le même ordre, dans le deuxième. [...] La seule différence qui peut exister entre les deux se limite à une différence de

---

2) Pour les auteurs arabes anciens, la date de mort est donnée systématiquement entre parenthèses dès la première occurrence. Le premier nombre renvoie à la date de l'hégire, le second à l'ère chrétienne.

3) Voir le texte rapporté par Bohas et Guillaume (1984 : 176).

4) La voyelle longue est formée, dans l'analyse des grammairiens arabes, de deux segments : une voyelle brève suivie d'un segment consonantique d'allongement et d'adoucissement. En conséquence, le terme *harf* s'applique, dans leur terminologie, aux consonnes et aux segments d'allongement. Voir, sur la justification de cette analyse, Hadj-Salah (1979 : 2, 530, 852).

voyelles ou de *harf* augmente [...] Les deux formes sont alors différentes alors que la matière première (*'usûl*) est la même».

2° «Un sens commun aux deux mots et non le même sens. Si aucun sens commun ne les réunit, il n'y a pas de dérivation, chacun des deux mots étant étranger par rapport à l'autre. Si aucune différence ne les différencie, il n'y a pas de dérivation non plus, l'un des deux mots étant l'autre».

La définition donnée par Ibn Dihyat, ainsi que les conditions posées par Ibn as-Sarrâj, devraient remettre en cause le point de vue communément admis sur la dérivation à partir de la racine dans la tradition grammaticale arabe. En effet, dans leurs analyses, les grammairiens arabes considèrent que la dérivation consiste à créer un mot à partir d'un autre. Or, la racine n'est pas un mot; ce n'est pas une unité du lexique, mais une unité abstraite qui ne se réalise que dans le mot (5).

Dans les textes des grammairiens arabes, une forme dérivée revient toujours à un mot premier, directement ou indirectement, la dérivation pouvant être conçue comme une chaîne, c'est-à-dire comme une opération par étapes dans laquelle les formes dérivent les unes des autres. Le verbe passé, disent-ils, dérive du nom d'action (*masdar*) (6), le nom d'agent dérive du verbe passé, le passif dérive de la forme active, etc. En supposant que le nom d'action (*masdar*) qui est une forme première, vient d'une racine, non parce qu'il dérive d'une racine, mais parce qu'il est établi sur une racine, généralement triconsonantique, on peut imaginer des étapes du type suivant :

racine -----> *masdar* ----> verbe -----> nom d'agent  
 racine -----> *masdar* ----> verbe actif ----> verbe passif<sup>7</sup>

Schéma I

5) Voir H. Hamzé : Racine, dérivation et institution du langage, à paraître.

6) Dans l'analyse majoritaire, celle de Basra, le nom d'action (*masdar*) est la base de la dérivation, alors que pour les Kûfites le *masdar* est dérivé du verbe (Hamzé, 1987 : 596-605).

7) Avec une racine D.R.B.(frapper), par exemple, les grammairiens arabes donnent, selon ce que rapportent Bohas et Guillaume (1984 : 152) les étapes suivantes :  
 D.R.B. -----> DaRB -----> DaRaBa -----> yaDRiBu ---> DâRiB ---> DaRRâB  
 idée de frapper ---> le fait de frapper --> il a frappé --> il frappe --> frappeur ---->  
 frappeur avec force.

Cependant, il faut ajouter que ce schéma n'est valable que dans la mesure où on considère que la première étape, entre la racine et le nom d'action (*masdar*), ne fait pas partie de la dérivation, la racine n'étant pas considérée comme une base de dérivation.

Cependant, comme les grammairiens arabes ne parlent pas de dérivation quand il s'agit de la formation d'un nom d'action (*masdar*) ou d'un nom premier comme /KaL.B/ «chien» par exemple, il faut interpréter le schéma ainsi : le nom d'action est établi sur une racine, le verbe dérive du nom d'action, le nom d'agent dérive du verbe. Autrement dit, la dérivation n'est imaginée qu'à partir des mots construits déjà sur une racine. Le passage de la racine au mot n'est pas nommé dérivation, c'est une étape préalable à la dérivation.

Le mot arabe est analysé par les grammairiens arabes comme la réalisation d'une base consonantique qui est la matière première, suivant une forme donnée, le schème. Ils représentent les consonnes radicales du mot, généralement trois, par : F.<sup>C</sup>.L., et reproduisent le reste des segments tels qu'ils se prononcent. Avec une racine K.T.B. qui exprime la notion d'écrire, on peut représenter ainsi le premier type du schéma 1 :

K.T.B. -----> KaTB -----> KaTaBa -----> Ka : TiB (8)  
 F.<sup>C</sup>.L. -----> Fa<sup>C</sup>L -----> Fa<sup>C</sup>aLa -----> Fa : <sup>C</sup>iL

#### Schéma 2

Deux autres hypothèses sur la dérivation sont avancées :

1° l'hypothèse de Cantineau (1950 : 119-125) : elle suppose une dérivation directe, sans étapes, à partir de la racine. Le mot serait le point d'intersection de deux séries : la racine et le schème. Il appartient à chacune d'elles;

2° l'hypothèse d'A. Roman (1990 : 23-85) : elle suppose une dérivation directe à partir de la racine mais elle ne considère pas le mot comme la combinaison d'une racine et d'un schème mais de racine(s) et de modalité(s); les unités de nomination générale étant construites sur des racines monoconsonantiques et les unités de nomination particulière sur des racines triconsonantiques.

Dans l'hypothèse du croisement d'une racine et d'un schème, le mot aura une double appartenance : aux mots qui ont la même racine mais des schèmes différents, d'une part, et aux mots qui ont le même schème mais des racines différentes, d'autre part. Toutefois, on devrait, avant d'aller plus loin, distinguer deux types dans ce qu'ils appellent : «schème» :

8) Pour en faciliter le repérage, les segments radicaux seront reproduits avec des lettres majuscules.

— un schème qui est un signe linguistique puisqu'il est le signifiant d'un signifié, disons schème 1. C'est, par exemple, le cas du nom d'agent construit à partir d'un verbe triconsonantique sur le schème  $Fa : ciL, (R_1a:R_2iR_3)$ ,  $R$  étant une consonne radicale, et qui veut dire : celui qui fait quelque chose. Ainsi  $/Ka : TiB/$  serait celui qui fait l'action d'écrire, la notion d'«écrire» ayant pour signifiant le morphème discontinu, la racine triconsonantique K.T.B. et la notion de «celui qui fait» ayant pour signifiant le schème  $Fa : ciL, R_1a : R_2iR_3$ ;

— un schème qui n'est pas le signifiant d'un signifié puisque le mot n'est pas décomposable, disons schème 2. Le schème se réduit donc à une séquence de consonnes et de voyelles sans signifié. Les mots ayant cette même séquence n'ont pas forcément un sens commun. Ainsi le schème  $Fa^cL, R_1aR_2R_3$ , dans le mot *KaLB* (chien) n'est pas le signifiant d'un signifié dans la langue arabe historique. Le mot *KaLB* n'a pas un élément de sens commun avec les autres mots qui appartiennent au même schème  $Fa^cL, R_1aR_2R_3$ , comme *HaBL* (corde), *QaTL* (le fait de tuer), etc.

C'est probablement ce deuxième type de schème, schème 2, qui justifie le nombre très élevé de schèmes nominaux en arabe : Sîbawayhi (180/796) en cite 308 dans *al-Kitâb*, Ibn l-Qattâc (515/1122) en cite 1210 dans le *Kitâb l-'abniyat (al-Muzhir : 2, 4)*.

## 2. — LE MOT ARABISÉ

Aux deux parties du lexique : les mots premiers et leurs dérivés, il faut en ajouter une troisième. Elle comprend des mots «arabisés» (*mu<sup>c</sup>arrab*) appelés ainsi par opposition aux unités des deux premières, considérées par les grammairiens arabes comme des mots «arabes purs» (*carabî mahd*). C'est l'opposition entre le mot arabe pur et le mot arabisé qui nous intéresse dans la suite de ce texte.

Quelle que soit l'hypothèse retenue sur la formation du lexique et sur le processus dérivationnel en arabe, la racine demeure au centre des préoccupations puisque tout mot arabe pur doit revenir, en fin de compte, à une racine. Mais, que se passe-t-il avec les mots arabisés, c'est-à-dire lorsque l'arabe emprunte à d'autres langues ?

### 2.1 — L'adaptation phonétique

Al-Jawharî (393/1003) aurait défini ainsi l'arabisation : «l'arabisation du nom non arabe consiste en ce que les Arabes le prononcent conformément à leur manière» (*al-Muzhir : 1, 268*).

En examinant les mots considérés comme arabisés par les savants arabes on va se rendre compte que deux conditions sont requises pour que le mot soit prononcé conformément à leur manière :

a) utilisation du mot par les Arabes (*qad takallamt bi hi l-<sup>c</sup>arab*) (*al-Mu<sup>c</sup>arrab* : 176). Toutefois, il faut bien s'entendre sur la valeur des mots. Le terme «les Arabes» mentionnés ici, ainsi que dans d'autres textes de ce type, ne renvoie pas à la communauté des arabophones en général; il renvoie aux Arabes qui font autorité en matière de langue, c'est-à-dire les Arabes au langage pur (*fasîh*). Il faut donc que le mot arabisé fasse partie du corpus de référence : le Coran, la poésie arabe ancienne et les dires des Arabes jusqu'à une époque donnée, le deuxième siècle de l'hégire. Les faits de langue attestés après cette date, qu'ils soient des emprunts, des créations nouvelles ou des variations produites par les «Modernes» ne font pas autorité (Hamzé, 1987 : 128-151);

b) prononciation du mot à la manière des Arabes. C'est-à-dire une réalisation qui impose, si besoin est, une adaptation phonétique des mots étrangers pour qu'on puisse les intégrer au système phonologique de l'arabe.

## 2.2 — L'adaptation morphologique

L'adaptation au système morphologique de l'arabe ne semble donc pas une condition pour la reconnaissance du mot arabisé. En effet, dans «le chapitre de ce qui est arabisé des noms étrangers», Sîbawayhi souligne que les Arabes ne traitent pas de la même manière les mots étrangers qu'ils empruntent, «ils peuvent, dit-il, rattacher les mots arabisés à leurs schèmes comme ils peuvent ne pas les rattacher. Ainsi, ils ont rattaché *dirham* (dirham) à *hijra<sup>c</sup>* (stupide), *bahraj* (futile) à *salhab* (qui a le corps long), *dînâr* (dinar) à *dîmâs* (ce qui couvre) [...], *'Ishâq* (Isaac) à *'icsâr* (ouragan), *Yacqûb* (Jacob) à *yarbû<sup>c</sup>* (gerboise) [...], *rustâq* (village) à *qurtâs* (feuille de papier). [...] Mais ils peuvent laisser le nom tel quel lorsque ses sons ressemblent aux leurs, qu'il soit sur leurs schèmes ou non [...] Ils peuvent également changer les sons qu'ils n'ont pas tout en gardant pour le nom le schème qui est le sien en persan. Exemples : *firind* (sabre), *baqqam* (bois de campêche employé dans la teinture), *'âjurr* (brique cuite au feu) et *jurbuz* (homme fripon)» (*al-Kitâb* : 4, 303-4).

Cette explication donnée par Sîbawayhi est valable pour les savants arabes jusqu'à nos jours. On la trouve, parfois avec d'autres

formulations, dans les textes des grammairiens et des lexicographes ultérieurs (9).

Les mots donnés comme exemples de ce qui a gardé sa forme étrangère dans le *Kitâb* de Sîbawayhi sont, explicitement, appelés : «arabisés» par les auteurs arabes. Ainsi al-Jawâlîqî dit de *firind* qu'il est un nom «persan arabisé» (*al-Mu<sup>c</sup>arrab* : 473), *baqqam* est «persan arabisé» (*al-Mu<sup>c</sup>arrab* : 176). C'est la même constatation à propos de *jurbuz* : «il n'appartient pas à la langue des Arabes [...] c'est un [nom] persan arabisé» (*al-Mu<sup>c</sup>arrab* : 235).

Il semble donc que le rattachement à une forme arabe ne soit pas une condition nécessaire à l'arabisation du mot non arabe. Un savant cité par as-Suyûtî, donne l'exemple de /nawru : z/ ou /nayru:z/ arabisé du persan /nawru:z/ qui veut dire : «le nouveau jour». La première réalisation /nawru:z/, dit-il, «est plus proche de la forme phonique du mot persan, même si elle n'est pas conforme aux schèmes de l'arabe. En effet; il n'est pas obligatoire que les [mots] arabisés soient conformes aux schèmes de l'arabe. Ne vois-tu pas [les mots] 'a :jurr (brique cuite au feu), 'ibraysam (soie), 'ihlîlaj (myrobolan), 'itrîfal (sorte de médicament composé) (10) ? Si les mots arabisés sont conformes aux schèmes de l'arabe, tant mieux.» (*al-Muzhir* : 1, 291).

Sur le plan morphologique, les mots arabisés ne sont pas obligatoirement rattachés aux schèmes de l'arabe d'où deux statuts différents comme le dit Abû Hayyân (745/1344-5) :

---

9) Quelques exemples supplémentaires sont donnés dans *al-'Istidrâk* d'az-Zubaydî (379/990) qui traite des formes de l'arabe qui n'ont pas été mentionnées par Sîbawayhi. Après les chapitres consacrés aux formes du nom, il en ajoute un autre intitulé : «chapitre de ce que les Arabes ont arabisé des noms non arabes». Il justifie ainsi ce chapitre : «Nous avons seulement cité ce chapitre à la suite de [celui] des formes du nom pour qu'il soit un indicateur des mots étrangers qui ont été changés et rattachés aux formes de l'arabe, ou changés sans y être rattachés, ou qui sont restés sans changement pour que personne ne trouve un mot non arabe et croie qu'on a oublié de le mentionner dans la langue arabe» (*al-'Istidrâk* : 197).

10) Le *Lisân* ne mentionne pas 'itrîfal mais *tarfal* (médicament composé). Dans le *Supplément aux dictionnaires arabes* de Dozy on trouve 'itrîful (myrobolans, médicament composé ou électuaire, dans lequel entrent les myrobolans) (1881 : 1, 28).

a) ce que les Arabes ont rattaché aux schèmes de leur langue en fait partie; «son schème sera considéré, dans ses radicales, ses augments et l'agencement de ses segments (wazn) comme ayant le même statut que les noms arabes purs»;

b) «ce que les Arabes n'ont pas rattaché aux schèmes de leur langue n'en fait pas partie» (*al-Muzhir* : 1, 269).

Il nous semble que la dernière phrase doit être comprise de la manière suivante : ce qui n'est pas rattaché aux schèmes de l'arabe ne fait pas partie de ces schèmes, et non pas : ne fait pas partie de la langue des Arabes, ces mots étant considérés comme arabisés. Autrement dit, un mot arabisé qui ne se rattache pas à un schème arabe existant, ne formera pas un schème supplémentaire qu'on doit ajouter à ce qui existe déjà. Quant on compte les schèmes de l'arabe, il n'en fait pas partie. Il est admis, mais il reste tel quel, isolé, sans famille, à l'extérieur du système morphologique de l'arabe, sans pouvoir participer à la génération de mots nouveaux.

### 2.3 — Mot arabisé et dérivation

Qu'ils soient rattachés aux schèmes de l'arabe ou non, les mots arabisés sont étrangers à la dérivation. «[I] est absurde, dit l'un des savants arabes, que le non-arabe soit dérivé de l'arabe ou que l'arabe soit dérivé du non-arabe, les langues ne pouvant dériver les unes des autres, qu'elles soient établies par convention [humaine] ou par institution [divine]. C'est seulement dans la même langue qu'un mot peut dériver d'un autre, la dérivation étant une production et une génération. Or, il est absurde que les chamelles produisent autre chose que des chamelons et que la femme accouche d'autre chose que des enfants» (*al-Muzhir* : 1, 287).

Toutefois, on ne manque pas de constater que même des savants arabes célèbres essayaient de chercher une racine à des mots arabisés. Il nous semble qu'il s'agissait souvent, dans le rapprochement fait entre les mots arabisés et les mots arabes purs, d'une spéculation qui avait un double objectif.

a) Un objectif théorique : qui consiste à se conformer aux méthodes de raisonnement par analogie des savants arabes, d'où le recours à des pures spéculations pour parfaire leur méthode de travail et non pour répondre à une situation réelle. Il s'agit de répondre à la question suivante : si tel mot était d'origine arabe quels seraient alors sa racine et son mode de dérivation ? Dans une situation réelle, ils savent qu'un nom arabisé, dès lors qu'il est d'origine étrangère, ne

peut pas avoir un lien avec un mot «arabe pur» même s'il y a coïncidence entre leurs formes extérieures. Ainsi «'Ishâq, "Isaac", le prophète, n'est pas [tiré] du verbe *'ashaqa* qui veut dire éloigner, ni de l'un des dérivés de ce mot. Il en est de même pour le prophète Ya<sup>c</sup>qûb «Jacob» qui n'est pas [tiré] du *ya<sup>c</sup>qûb* qui est un oiseau. Il en est de même pour tout terme non arabe qui coïncide formellement avec un mot arabe» (*al-Muzhir* : 1, 292).

En admettant qu'un nom comme 'Ibrâhîm (Abraham) est non arabe, dit az-Zajjâjî, «on se dispense de le mesurer à l'un des schèmes des noms arabes; il serait même incorrect de lui donner un schème en arabe puisqu'on ne sait pas quelle est son origine» et on ne se pose pas la question de savoir si son *hamza* /' est radical ou non (*Muḳtasar z-Zâhir* : feuille 127 v.-128r).

b) Un objectif pratique. L'organisation des dictionnaires arabes par racine pose inévitablement le problème du classement des mots étrangers, qu'ils soient rattachés ou non à des schèmes arabes. Il s'agit de voir concrètement comment classer un mot censé être sans racine, dans un dictionnaire organisé suivant les racines des mots. La tâche, il faut le dire, n'était pas toujours facile.

## 2.4 — Mot arabisé et racine

Pour l'utilisateur, et parfois même pour le savant, le rattachement d'un mot arabisé à un schème, la simple coïncidence entre sa séquence sonore et celle d'un mot arabe, peuvent pousser à tisser des liens sémantiques entre les deux mots, en s'appuyant sur les liens formels apparents; du coup, à oublier l'origine étrangère du mot et à lui chercher une nouvelle généalogie.

Cette tendance à oublier est certainement accentuée par les variations qu'on peut être amené à opérer sur le mot emprunté : mettre un mot au duel, au pluriel, créer un adjectif de relation, etc. En un mot, quand on est obligé de faire appel au système morphologique de la langue sans pouvoir s'arrêter au produit fini qui a été emprunté, le mot sera alors traité comme s'il était un mot «arabe pur». Il sera considéré, dans ses radicales et ses augments et dans l'agencement de ses segments, comme le mot arabe.

En l'intégrant dans le système morphologique de l'arabe, le mot emprunté n'est plus isolé puisqu'on peut créer d'autres mots de sa famille.

En le faisant, on lui invente, inconsciemment, une racine.

### 3. — MOT ARABE ET MOT ARABISÉ

#### 3.1 — Les deux processus de dérivation

Cependant, sur le plan théorique, la création, inconsciente, d'une racine et la coïncidence avec un schème arabe ne permettent pas d'établir une équivalence entre le mot arabisé et le mot arabe pur. Demeurent deux différences essentielles entre les deux :

a) les deux mots : arabe et arabisé, ne sont pas situés sur un pied d'égalité du fait de leur appartenance à un même paradigme. Le mot arabe «pur» demeure le principe auquel se rattache le mot arabisé en tant que corollaire. C'est le mot «arabe pur» qui donne au mot «arabisé» sa légitimité;

b) le mot arabe pur provient d'une racine. Il appartient à une famille, ce qui n'est pas le cas du mot arabisé dont la racine et la famille peuvent être créées a posteriori puisqu'on invente sa racine à partir de sa forme. Autrement dit, il y a deux processus complètement différents qui distinguent le mot arabisé du mot arabe pur. Alors que le mot arabe est construit sur une racine, c'est la racine qui sera construite sur le mot arabisé. Le point de départ de l'un est le point d'arrivée de l'autre. Mais, comme la dérivation en arabe se fait toujours à partir de la racine, la constitution d'une famille pour le mot arabisé ne se fait pas directement, elle se fait par un mouvement double dans les deux sens : on passe du mot arabisé à une racine qu'on invente, puis on passe de cette racine inventée aux nouveaux mots à créer. Ces deux processus peuvent être schématisés ainsi :

Mot arabe  
racine -----> mots

Schéma 3

Mot arabisé  
mot -----> racine  
mots <-----

Schéma 4

#### 3.2 — Changement de statut du mot arabisé

Comment définir le statut des mots créés à partir du mot arabisé ? Sont-ils des mots arabes de par leur création à partir d'une racine ou des mots arabisés de par le mot emprunté qui a servi à

inventer leur racine ? En les intégrant, la langue a tendance à oublier leur origine lointaine, à effacer la première étape entre l'arabisation du mot et la création de sa racine.

Abû Bakr Ibn l-'Anbârî (328/940) ramène le mot *yahûdiyy* (juif) au verbe *hâda* (se repentir) et, par là-même, à la racine H.W.D., racine bien attestée dans les textes arabes anciens : «le juif (*yahûdiyy*), dit-il, est appelé ainsi parce qu'il s'est repenti à un moment donné. Ce nom *yahûdiyy* l'a accompagné bien qu'il ait changé son repentir et y a renoncé plus tard» (*az-Zâhir* : 2, 214). Le verset coranique et les deux vers témoins qu'il donne à l'appui de ce sens n'ont pas convaincu son élève az-Zajjâjî (337/949) qui nous dit : «*yahûdiyy* (juif) n'est pas un adjectif de relation de *hâda* (se repentir). S'il en était ainsi il serait dérivé du nom d'action (masdar) et non du verbe puisqu'on ne construit pas des adjectifs de relation sur les verbes. Il s'agit, en réalité, du fils de Jacob, d'un «nom étranger arabisé, comme disent les savants, qui a coïncidé après son arabisation avec un mot arabe» (*Sarh Mukhtasar az-Zâhir* : 146R).

Apparemment, de grands savants ont oublié l'origine étrangère du mot arabisé puisque sa racine coïncide avec une racine arabe préexistante (11).

Ne s'agit-il pas du même problème, mais cette fois-ci avec az-Zajjâjî lorsqu'il critique Ta<sup>c</sup>lab (290/903) au sujet de l'un des Attributs de Dieu : *ar-RaHMân* (Le Très Clément) ? Pour Ta<sup>c</sup>lab, c'est un mot arabisé. Pour az-Zajjâjî, c'est un mot arabe pur dont la dérivation du verbe *RaHiMa* (être clément) est bien connue. Le grand savant Ta<sup>c</sup>lab qui aurait soutenu l'origine étrangère d'*ar-RaHMân* (az-Zajjâjî, *'Istiqâq 'asmâ' llâh* : 42) ne connaissait-il pas la préexistence d'une racine R.H.M. en arabe ? Il serait trop naïf de le croire et d'envisager même cette hypothèse. Ta<sup>c</sup>lab pensait, très probablement, à deux racines : une racine arabe pur R.H.M. qui a donné l'attribut de Dieu *ar-RaHîM* (Le Très Miséricordieux), et une racine reconstituée à partir du mot arabisé (*ar-RaHMân*). Cette racine inventée ne peut être que R.H.M; elle coïncide donc avec la racine R.H.M. déjà existante dans la langue.

---

11) Cependant, ce point de vue d'az-Zajjâjî est loin de faire l'unanimité des savants arabes. Voir à titre d'exemples : az-Zajjâjî, *'Icrâb l-Qur'ân* : 1, 118; al-Qurtubî, *al-Jâmic li 'ahkâm l-Qur'ân* : 1, 294; at-Tabarsî, *Jâmic l-bayân* : 1, 125-126, qui donnent d'autres significations pour la racine H.W.D.

Il est difficile de savoir si *ar-RaHMân* est un mot arabe ou un mot arabisé. La mémoire du mot est incertaine dans la mesure où la mémoire des hommes fait défaut. Pour les arabophones, c'est, sans aucun doute, un mot arabe pur. Pour quelques savants de l'arabe, c'est un peu moins sûr.

Avec le mot *LiJâM* (bride), les savants arabes mettent le doigt sur le lien entre la constitution d'une famille et les risques de passage du mot arabisé pour un mot arabe pur. Tout comme *KiTa : B* (livre), le mot *LiJa : M* se met au pluriel : *LuJuM*, au diminutif : *LuJayyiM* et *LuJayM*, donne un verbe : *'aLJaMa* (mettre la bride) qui peut avoir toutes les conjugaisons, un nom d'action (*masdar*), un nom d'agent, un nom de patient, il peut être utilisé au sens propre et au sens figuré, etc. Cela pousse un savant, d'après ce que rapporte as-Suyûtî, à dire : «Tu serais enclin à juger ce mot *LiJâM* comme étant un mot arabe établi [à l'origine], non arabisé et non transféré, en raison de la fréquence de son emploi et de sa variation s'ils [= les savants anciens] ne l'avaient pas considéré comme arabisé de *liga : m*» (*al-Muzhir* : 1, 288).

Pour le mot *LiJâM*, on connaît l'origine étrangère, *lugâm* en persan. Pour d'autres mots on est moins sûr. Mais, qu'il soit noté ou non dans le dictionnaire, ce mot est dans la conscience des hommes, un mot arabe qui fonctionne comme tout autre mot de leur langue. Il est donc discuté entre la mémoire du lexicographe et celle de l'utilisateur.

Le mot arabisé *LiJâM* crée une vraie famille. Sa racine inventée va donc acquérir une certaine légitimité lui permettant de restaurer le processus dérivationnel dans son sens habituel, celui du mot «arabe pur». La différence entre les deux processus de formation de mots, arabe et arabisé, s'efface; le schéma 4 se remet à l'image du schéma 3; sa racine ne sera plus considérée comme le fruit du mot arabisé mais à l'origine même de ce mot. Le fils devient père non seulement de ses enfants, mais aussi de son propre père. Cela n'est possible qu'avec l'oubli.

*LiJâM* n'est plus, n'était plus au temps d'as-Suyûtî et, très probablement, avant lui, un mot arabisé <sup>(12)</sup> dans la mémoire des hommes. Il l'est toujours fixé par écrit dans la mémoire des dictionnaires. Décidément, les dictionnaires ont une autre mémoire; c'est que les langues construisent sans cesse des mémoires nouvelles.

---

12) Al-Jawâlîqî mentionne que, pour certains, *LiJâM* est un mot arabe pur (*al-Mucarrab* : 564).

## RÉFÉRENCES

### 1) Sources primaires :

ibn al-'ANBÂRÎ : *az-Zâhir*, éd. Hâtim d-Dâmin, Beyrouth, Mu'assasat r-Risâla, 1<sup>re</sup> éd., 1412/1992

#### *le CORAN*

al-JAWÂLÎQÎ : *al-Mu<sup>c</sup>arrab*, éd. F.<sup>c</sup> Abd r-Rahîm, Damas, Dâr l-Qalam, 1<sup>re</sup> éd. 1410/1990.

ibn MANDÛR : *Lisân l-<sup>c</sup>Arab*, Beyrouth, Dâr Sâdir, S.D.

al-QURTUBÎ : *al-Jâmi<sup>c</sup> li 'ahkâm l-Qur'ân*, Beyrouth, Dâr l-Kutub l-<sup>c</sup>ilmiyya, 1<sup>re</sup> éd. 1408/1988.

SÏBWAYHI : *al-Kitâb*, éd. <sup>c</sup> Abd s-Salâm Hârûn, Le Caire, al-Hay'a l-misriyya l-<sup>c</sup>amma li l-kitâb, 5 vol., 1973-1977.

as-SUYÛTÎ : *al-Muzhir*, éd. Jâd l-Mawlâ, 'Ibrâhîm et Bajjâwî, Saïda-Beyrouth, al-Maktaba l-<sup>c</sup>asriyya, 2 vol., 1986.

at-TABARSÎ : *Majma<sup>c</sup> l-bayân*, Qumm-Saïda, Matba<sup>c</sup>at l-<sup>c</sup>irfân, 5 vol., 1333/...

az-ZAJJÂJ : *Ma<sup>c</sup>âni l-Qur'ân wa 'i<sup>c</sup>râbuh*, éd. <sup>c</sup> A. Şalabî, Saïda-Beyrouth, al-Maktaba l-<sup>c</sup>asriyya, S.D.

az-ZAJJÂJÎ : *'Istiqâq 'asmâ' Llâh*, éd. <sup>c</sup> Abd l-Husayn. al-Mubâarak, Beyrouth, Mu'assasat r-Risâla, 2<sup>e</sup> éd., 1406/1986.

az-ZAJJÂJÎ : *Mu<sup>c</sup>ktasar z-Zâhir*, manuscrit conservé à Dâr l-Kutub l-misriyya, Le Caire, n<sup>c</sup> 39 S, 'adab, sous le titre : Tafsîr Risâlat Ibn Sa<sup>c</sup>îd fî 'adab l-kâtib.

az-ZUBAYDÎ : *al-'Istidrâk <sup>c</sup>alâ Sîbawayhi*, éd. Hanna Haddad, Riad, Dâr l-<sup>c</sup>ulûm, 1<sup>re</sup> éd., 1407/1987.

## 2) Sources secondaires :

BOHAS, G. et J. P. GUILLAUME (1984) : *Études des théories des grammairiens arabes, morphologie et phonologie*, Damas, Institut Français de Damas.

CANTINEAU, J. (1950) : «Racines et schèmes», *Mélanges William Marçais*, Institut d'Études Islamiques de l'Université de Paris, Paris, G. P. Maisonneuve, pp. 119-125.

DOZY, R. (1881) : *Supplément aux dictionnaires arabes*, Beyrouth, Librairie du Liban, reproduction de l'édition originale de Leyde, E. J. Brill, 1881.

HADJ-SALAH, A. (1979) : *Linguistique arabe et linguistique générale*, Thèse de Doctorat d'État ès lettres, Université de Paris-Sorbonne, Paris.

HAMZE, H. (1987) : *Les théories grammaticales d'az-Zajjâî*, Thèse de Doctorat d'État ès lettres, Université Lyon-2, Lyon.

HAMZÉ, H. : Racine, dérivation et institution du langage, à paraître.

ROMAN, A. (1990) : *Grammaire de l'arabe*, coll. «Que sais-je ?», Paris, PUF.

# Les mots français d'origine arabe

## Tunisie-France : aller/retour

**Habiba NAFFATI**

*Université de Provence, Aix-Marseille,  
France*

### INTRODUCTION

Parmi les mots français empruntés à l'arabe, notre étude s'intéressera plus particulièrement aux mots dont la forme écrite ou orale témoigne d'une origine étrangère et qui ont effectué un retour à la terre d'origine, la Tunisie, par le biais de la francophonie.

En nous aidant de dictionnaires de langue française, tel le *Petit Robert*, de dictionnaires de français dit «non-conventionnel», ainsi que d'un corpus francophone tunisien, nous observerons dans un premier temps l'évolution de ces mots en France : degré de vitalité, évolution graphique et phonétique, niveau de langue et évolution sémantique. Dans un second temps, nous observerons ces mêmes mots de retour au «pays natal» en contexte tunisien. Ainsi, nous verrons en quoi la France a pu influencer sur ces mots et comment le retour au pays a pu leur faire éventuellement retrouver la mémoire.

### 1. — EN FRANCE

#### 1.1 — Degré de vitalité

Évaluer le degré de vitalité des mots étant une tâche bien délicate, l'utilisation de certains outils d'évaluation s'avère très appréciable; en particulier celle de dictionnaires de langue française, en ce qui concerne le français dit «conventionnel»,

a) *Les dictionnaires de langue française* relèvent bon nombre d'emprunts à l'arabe. Ainsi, Mohamed Ben Smaïl (1994), dans son *Dictionnaire des Mots Français d'Origine Arabe* propose un corpus de 257 mots qu'il estime non exhaustif.

«... Il est normal et logique de penser que la moisson à laquelle nous avons abouti [...] est loin d'être définitive; il serait dérisoire et peu sérieux de prétendre arrêter là les investigations en stoppant la liste ainsi établie à ses limites actuelles.» (p. 151).

Leur intégration dans la langue française est plus ou moins ancienne (*babouche*, 1671; *raï*, 1983). La date d'intégration n'est d'ailleurs pas toujours révélatrice. En effet *youyou*, par exemple, n'est pas attesté dans le *Petit Robert* de 1990 alors qu'il l'est dans l'édition de 1993. Or, il fait partie d'une réalité assez ancienne étant déjà mentionné dans *Le français d'Afrique du Nord* d'André Lanly (1962). Ainsi, malgré une date d'intégration identique *raï* et *youyou* ne semblent pas être du même âge. Quoiqu'il en soit nous constatons que la langue française n'a pas fini d'emprunter à l'arabe, langue d'emprunt qui semble privilégiée. L'usage des dictionnaires de français non-conventionnel ainsi que l'oral nous le confirment.

b) *Français marginal, non-conventionnel et oral* : les dictionnaires d'«argot» ou, selon les termes plus justes de Cellars et Rey (1980), de «français non-conventionnel» sont des dictionnaires considérés par Cellars et Rey, comme des «compléments» ou «suppléments» officieux au *Dictionnaire de l'Académie* (p. IX). L'oral est, lui aussi, à prendre en considération dans l'évaluation, une langue vivante étant avant tout une langue parlée. Dans ce qu'on appelle «le langage des jeunes» (Girard et Kernel, 1996) ou «le français des cités» (Goudaillier, 1997), l'arabe est une des langues dominantes.

«Le français des ados [...] emprunte énormément à certaines langues étrangères, principalement l'anglo-américain, l'arabe et le gitan.» (Girard et Kernel, 1996 : 10).

Les mots empruntés sont souvent productifs, par la dérivation classique : *kif/kiffer*, *kiffant* ; ou encore, par le verlan : *fatma/mafat*. Nombre de ces mots ne sont pas cités dans le *Petit Robert* (*shitan/shatan*), mais certains le sont (*haram*, *gazelle*), malgré une différence dans les définitions qui leur sont attribuées.

C. Hagège, dans sa préface de *Comment Tu Tchatches* (Goudaillier, 1997 : 4), dit du français contemporain des cités qu'il est

«(...) une langue dont la floraison dans un cadre humain très particulier est un phénomène important dans la société française aujourd'hui.» Toutefois, l'utilisation de ces emprunts ne semble pas toujours être marginalisée. Le terme «non-conventionnel», employé par Cellard et Rey, serait plus approprié. Notons aussi l'utilisation à l'oral de termes qui ne figurent dans aucun dictionnaire, qu'il soit conventionnel ou non. C'est le cas, dans le domaine religieux de *fatwa* et *aïd* ou encore de *chorba*, *makhrouth*, *boulghour* et bien d'autres encore dans le domaine culinaire; sans oublier le *inchallah* utilisé dans tous les milieux quel que soit l'origine ou l'âge du locuteur. Ainsi, l'arabe en tant que langue d'emprunt a un très fort degré de vitalité.

## 1.2 — Évolution morphologique

Il s'agit ici de voir comment les emprunts sont accueillis dans la langue française du point de vue morphologique. La forme empruntée est-elle adaptée aux structures de la langue française, est-elle «francisée», ou bien le système de la langue d'origine est-il maintenu ? En réalité, les deux tendances sont possibles. Je signale que n'étant pas arabisante mes références seront basées sur mes connaissances plus complètes d'un arabe régional, le tunisien.

### a) Morphologie flexionnelle

— *Le genre* : en français, la répartition des mots en genre masculin et féminin est arbitraire. Ce phénomène se retrouve avec les mots empruntés à l'arabe; d'où le même risque d'erreur pour le locuteur francophone qui les utiliserait.

<i>henné</i>	n.m. / ar. n.f. ( <i>hinna</i> )
<i>baklava</i>	n.m. / ar. n.f.
<i>merguez</i>	n.f. / ar. n.m.

On remarque que le mot *henné*, au passage à la langue française, perd son *a* final, marque du féminin en arabe, pour devenir masculin. Les cas de *goule* et *felouque* sont encore plus complexes. Le *Petit Robert* les présente de la façon suivante :

**goule** : [gul] n.f. (1821; arabe *ghoûl* «démon»)

**felouque** : [f əluk] n.f. (1611; *falouque*, 1606; *pelouque*, 1595, esp. *faluca*; de l'arabe *foulk* «navire»)

*Goule* est désigné comme étant féminin; or, en arabe, *ghoûl* est masculin et son féminin est *ghoûla*. Pour ce qui est de *felouque*, si nous nous référons au tunisien, il viendrait de l'arabe *foulouka*. Dans son passage au français, il aurait donc gardé son genre féminin mais aurait perdu la marque arabe de celui-ci : le *a* final. Mais, le phénomène est bien plus complexe qu'il ne le paraît. En effet, nous pouvons aussi considérer que dans *henné* et *felouque* la marque du féminin arabe est présente, le [e] et le [ð] n'étant qu'une nuance du [a] comme nous le verrons par la suite.

— *Le nombre* : la langue française, depuis l'arrêté du 28/12/1976, autorise, dans le cas des emprunts, l'utilisation du -s. Mais la langue soignée se doit d'utiliser le pluriel d'origine. Pour les emprunts à l'arabe, le choix de la structure française semble être le cas le plus fréquent : *un djinn/ des djinns*. Cependant, la structure arabe se retrouve aussi : *un ksar/ des ksour*. Notons que dans cette tendance à l'alignement, pour le cas de mots utilisés le plus souvent au pluriel de la langue d'origine, la forme plurielle s'impose parfois comme singulier dans la langue d'accueil. À la marque du pluriel de la langue d'origine peut alors s'ajouter une seconde flexion, le *s* du pluriel français. Ainsi, on aura :

*un cleb/ des clebs* ar. sing. *kelb* / plu. *kleb*

*un fedayin/ des fedayin* ar. sing. *feddaï* / plu. *feddayin*

Remarquons que le *Petit Robert* signale pour l'entrée *fedayin* : «au sing. *un fedayin*, ou mieux *un fedaiï*». Tandis que selon Le Petit Larousse «la forme savante *fedaiï* (pl. *fedayine*) reste peu usitée».

#### *b) Morphologie dérivationnelle*

Les mots empruntés utilisent parfois les procédés de dérivation classiques du français tels que l'adjonction suffixale :

*chérif / chérif-ien* (*Petit Robert*)

*cleb / cleb-ard* (Goudaillier, 1997 : 119)

*kif / kiff-er* (Goudaillier, 1997 : 89)

Certains peuvent être tournés au verlan :

*fatma / mafat* (Goudaillier, 1997 : 154)

*bled / deblé* (Goudaillier, 1997 : 80)

On trouve aussi des cas de troncation du premier élément :

*rahat loukoum / loukoum*

Le maintien de l'article d'origine en début de lexie est aussi possible, même s'il reste un phénomène particulier et ancien :

*el oud / luth*

L'évolution morphologique de ces emprunts est une preuve de l'intérêt que porte la langue française à leur particularité morphologique et d'un effort à respecter leur structure spécifique. Au-delà des connaissances que possède le locuteur francophone de la langue originelle, l'usage se conforme au système français, ce qui entraîne parfois, comme nous l'avons vus, des phénomènes assez singuliers.

### 1.3 — Évolution graphique et phonétique

a) *Graphie et sons inconnus au système phonétique français* : le système phonétique de l'arabe possède des sons inconnus au système français. Ainsi, certaines lexies contenant des phonèmes nouveaux pour les locuteurs francophones peuvent poser problème. De plus, l'alphabet latin étant différent de l'alphabet arabe, la question de la graphie se pose, ainsi que celle de la transcription phonétique de ces sons. Selon l'alphabet phonétique du *Petit Robert*, le seul son signalé comme provenant d'un mot emprunté à l'arabe (ou à l'espagnol) est le [x] représenté graphiquement par *kh* ou *gh* pour les mots empruntés à l'arabe (*maghzen/ makhzen*). Ce n'est cependant pas le seul son inconnu au système phonétique français :

— [x] se prononce [k] avec la graphie *c* (ex. : *calif*) ou [x] avec la graphie *kh* (ex. : *khamsin*, plus récent).

— [q] se prononce [k] avec la graphie *c* (ex. : *caïd*) ou *q* (ex. : *qasida*, plus récent).

([q] : occlusive arrière vélaire sourde avec occlusion simultanée du larynx)

— [θ] se prononce [t] (= *think* anglais) avec la graphie *th* (ex. : *hadith*).

— [ð] se prononce [d] ou [t] ou [z] (= *that* anglais) avec la graphie *d* (ex. : *cadi*) ou *t* ou *oudh* ou *th* (ex. : *makrout*, *makroudh*, *makrouth*) ou *z* (ex. : *muezzin*).

— l'aspirante laryngale sourde se prononce ['] avec la graphie *h* (ex. : *ha(d)j*) ou sans marque graphique (ex. : *alfa*).

— l'aspirante laryngale (proche du *h* aspiré allemand) ne se prononce pas mais se graphie *h* (ex. : *djihad*).

— l'aspirante fricative laryngale sourde ne se prononce pas mais se manifeste par un allongement de la voyelle qui la suit avec la présence graphique du " (ex. : *aīd*) ou de l'^ ou/et du redoublement de la voyelle (ex. : *djemaā*); elle peut aussi ne pas être signalée graphiquement (ex. : *sharia*).

b) *Cas du r* : en phonétique française les sons [R] et [r] s'opposent mais correspondent à un même phonème /r/, tandis qu'en arabe ils correspondent à des phonèmes différents. Ceci pose un problème de graphie pouvant engendrer des confusions. En effet, dans le mot emprunté le [r] devient [R] avec la graphie correspondante tandis que le [R] devient [g] transcrit *gh* ou *g*.

— [r] se prononce [R] et adopte la graphie *r* (ex. : *roumi*).

— [R] se prononce [g] et se transcrit *gh* (ex. : *maghreb*) ou *g* (ex. : *goul*)

Au problème de l'adaptation et de la transcription phonétique s'ajoute donc la question de l'orthographe des sons et de l'ajustement au système alphabétique de la langue française.

c) *Nasalisation* : certaines lexies ont subi un phénomène de nasalisation. Il s'agit, dans la plupart des cas, de mots empruntés de longue date (*mesquin*; *fondouk*; *caftan*). Mais une grande partie des mots susceptibles d'être nasalisés ne le sont pas (*djinn*; *minbar*; *intifada*; *chétan*). Remarquons que *mesquin* retrouve sa prononciation originelle, c'est-à-dire sans nasalisation, et ce par le biais du «français contemporain des cités» et des jeunes maghrébins qui le parlent (Goudaillier, 1997 : 126).

d) *Voyelles et semi-voyelles (ou glides)* : selon A. Lanly (1962 : 48) les trois timbres fondamentaux de l'arabe sont [a], [u] et [i], les autres étant des nuances. Le système vocalique français étant plus riche, la variabilité d'ordre distinctif sera très faible tandis que la graphie sera plus variable, l'emprunt nécessitant une transcription adaptée à l'alphabet latin. Le choix devient souvent arbitraire. Ainsi au [w] correspondent aussi bien la graphie *ou* que *v* et *w* (*zaouia*; *baklava*; *wi(l)laya*).

Malgré les grandes possibilités graphiques et phonétiques de la langue française, il semblerait que le [y] et le [ʎ], assez rares d'ailleurs parmi les emprunts à l'arabe, soient nés du passage de ces mots dans la langue française.

*muezzin* [m edzin]

*mufti* [myfti]

*uléma, ouléma* [ylema, ulema]

L'ancienneté de l'emprunt peut être un facteur de «francisation». Aujourd'hui, alors que présentateurs de radio et de télévision essaient tant bien que mal, mais avec un plaisir certain, de prononcer les emprunts le plus fidèlement possible, on pourrait se demander si *mesquin* qui arrive avec un nouveau départ dans le français des cités, évoluera comme au XII<sup>e</sup> siècle (*meschin* > *mesquin* (nasalisé)), (Goudailler, 1997 : 126) ou si, sous l'effet d'une universalité phonétique, le [in] restera in (encore un emprunt !).

#### 1.4 — Niveau de langue

Le degré de vitalité de l'arabe en tant que langue d'emprunt se manifeste aussi dans les niveaux de langue. À l'exception du français littéraire, les emprunts à l'arabes se retrouvent à tous les niveaux de langue. La question de l'influence de ce passage d'une langue à l'autre sur le niveau de langue se pose ainsi que celle de la neutralité de l'utilisation d'un terme plutôt que de son synonyme et des connotations que cette utilisation pourrait engendrer.

Prenons comme exemple les mots des lettres A et B de notre corpus.

- |                                     |                                      |   |
|-------------------------------------|--------------------------------------|---|
| 1. <i>alfa</i> L.c.1                | 5. <i>baraka</i> fam.                | 9. <i>bey</i> hist. L.c.1   |
| 2. <i>arbi</i> pop.,<br>péj. et vx. | 6. <i>barda</i> arg.<br>mil. et fam. | 10. <i>bled</i> XIX arg. mil.<br>1. L.c. 2 fam.                     |
| 3. <i>bakchich</i> L.c.             | 7. <i>baroud</i> arg. milit.         | 11. <i>brick</i> L.c.   |
| 4. <i>baklava</i> L.c.              | 8. <i>bedouin</i> L.c.               | 12. <i>burnous</i> 1.L.c.<br>2. fam. <i>faire suer le burnous</i> . |

N.B. : L.c. = Langue courante.

On constate que, pour 50 % des exemples (6/12), le niveau de langue n'est pas signalé. On pourrait donc supposer qu'ils sont d'usage courant. 15 % environ (2/12) ont un sens courant et un sens familier. L'exemple 2 a un sens péjoratif, populaire et vieux. Les autres ont soit un sens familier, soit un sens argotique, soit les deux. Dans 30 % (4/12) des cas, on remarque un sens familier. Le terme «familier» peut avoir une connotation plus ou moins positive. Il peut définir un mot «dont [...] l'usage est ordinaire» (Cf. *Petit Robert*); mais il peut aussi faire référence à une certaine liberté ou désinvolture que connoterait l'usage de ce terme :

**bled** n.m. —(fin XIX<sup>e</sup> arg. milit.; mot ar. maghrébin, «terrain pays»). 1. En Afrique du Nord, l'intérieur des terres, la campagne. 2. (1951) Fam. Lieu, village éloigné, isolé offrant peu de ressources. => trou. *On s'ennuie dans ce bled.* => fam. *patelin*.

**burnous** n.m. — ar. *burnous*. Grand manteau de laine à capuchon et sans manches (en usage dans les pays du Maghreb). fam. *Faire suer\* le burnous*.

L'entrée ci-dessus renvoie au verbe *suer* avec lequel se trouve associé cette expression définie selon ces termes : «se disait de colons européens d'Afrique du Nord accusés d'exploiter la main-d'oeuvre indigène». Sous le mot «familier» apparaît donc une connotation dépréciative faisant référence au colonialisme. Nous constatons, si nous regroupons les mots argotiques, familiers et populaires, que 50 % des mots empruntés à l'arabe sont dépréciés dans leur sens et redéfinis jusqu'au paradoxe dans certains cas. Ainsi *sidi*, signifiant littéralement «monsieur; seigneur», et étant donc une marque de respect, devient un terme péjoratif. Par le passage au français, ces mots auraient donc perdus leur valeur originelle pour être pervertis.

Dans une autre optique, en considérant les termes dits «familiers» comme appartenant à l'usage courant sans aucune connotation négative, le total est de 75 %. De plus, les défenseurs de l'argot vous dirons :

«Nous ne croyons pas qu'elle [la notion de «non-conventionnel»] soit un produit inavouable du matérialisme haineux des classes dangereuses, enfanté dans le mystère de l'illégalité.» (Cellard et Rey, 1980 : VIII)

«C'est cela au fond qui n'a jamais été pardonné au français non-conventionnel, qu'il s'agisse de sa grammaire ou de son vocabulaire : en dépit de tous les obstacles, il existe, il vit, il oppose sans cesse à la culture estampillée une contre-culture vigoureuse, il manifeste par sa présence que rien n'est jamais joué dans la société.» (Cellard et Rey, 1980 : IX)

En effet, rien n'est jamais joué, revenons par exemple à *bled*. Comment est-il perçu en argot ?

**bled** : village, ville d'origine; par ext., pays d'origine. (Goudaillier, 1997 : 53)

Il n'apparaît aucune connotation péjorative. Ainsi, si certains mots ont subi des altérations, dans la richesse des emplois et des nuances, l'authentique demeure et résiste. Les mots empruntés à l'arabe ne sont pas, en définitive, plus mal lotis que les autres, trouvant leur place aussi bien dans le français conventionnel que dans le français non-conventionnel, deux éléments complémentaires de la langue française

## 1.5 — Évolution sémantique

a) *Conservation du sens originel* : nous constatons que pour la plupart des cas, quand le terme conserve son sens d'origine, il fait référence à sa réalité originelle, dans son contexte originel. Ainsi dans la définition, il sera précisé *en Afrique du Nord*, ou encore *dans les pays arabes*.

**casbah** : citadelle d'un souverain dans les pays arabes.

**oued** : (de l'arabe «cours d'eau»). Rivière d'Afrique du Nord. Cours d'eau temporaire dans les régions arides.

Nous constatons, de plus, que la définition de ces termes est souvent vague. Ceci peut se comprendre lorsque c'est un «petit» dictionnaire qui est consulté à la différence d'un dictionnaire encyclopédique.

**medersa** : (de l'arabe maghrébin «collège»). Etablissement religieux musulman.

Cependant, le sens des termes peut s'en trouver limité ou même dévié, ce qui mènerait à l'incompréhension, voire à l'erreur.

b) *Modification du sens originel* : l'erreur sémantique, bien que rare, n'est pas inexistante. Par exemple, selon les contes et légendes tunisiens, le *ghoul* et la *ghoula* seraient à rapprocher des ogres et ogresses occidentaux. Or, le *Petit Robert*, avec une graphie différente, propose la définition suivante :

**goule** n.f. (1821; arabe *ghoûl* «démon»). Sorte de vampire femelle des légendes orientales.

L'hypothèse d'une erreur sémantique est, dans ce cas précis, tout à fait recevable. Le sens de l'emprunt peut aussi être modifié par une restriction du sens. Ainsi, *méchoui*, qui signifie en arabe «grillé, rôti», ne concerne que le mouton dans la définition française proposée ci-dessous.

**méchoui** : Mouton rôti à la broche, sur les braises d'un feu de bois; portion de ce mouton servie au repas; le repas.

La notion de français d'outre-mer peut impliquer, elle aussi, la restriction sémantique dans le sens où le terme ne sera utilisé que lorsqu'il fera référence à une réalité d'outre-mer. Par exemple, le mot *oued* ne sera jamais utilisé pour désigner une rivière en France. Enfin

la modification peut aussi s'effectuer par glissement du sens. Ce phénomène concerne en particulier les mots dont l'emprunt est ancien. Ainsi, du sens originel de *mesquine*, «pauvre», nous avons en français standard :

1. *médiocre, sans valeur.*

Puis le terme devient polysémique :

2. *mod. personne qui s'attache à ce qui est petit, qui manque de générosité.* 3. *spéci. témoigne d'avarice, de parcimonie.*

c) *Perversion du sens* : dans la plupart des mots polysémiques, le sens premier est suivi de nouveaux sens souvent dépréciatifs, soit par leur niveau de langue (familier, argotique, populaire), soit par leur connotation péjorative qui n'est d'ailleurs pas toujours signalée. Que penser par exemple du second sens attribué au terme *ramadan/ramdam* proposé ci-dessous ?

2. *fig. et fam. (...) Ramdam [ramdam] (1896, à cause du tapage nocturne du Ramadan). Tapage, vacarme.*

Au même titre que Noël chez les chrétiens, le Ramadan est une fête religieuse. Elle célèbre une période de pénitence et ne pourrait en aucun cas avoir, dans un contexte arabo-musulman, les connotations négatives qui lui sont attribuées ici. Il est légitime de parler alors de perversion du sens. Prenons aussi le cas des mots monosémiques, tels que *fatma*. Selon le *Petit Robert*, le prénom féminin arabe désigne «une femme arabe, domestique dans les pays du Maghreb au temps de la colonisation». Cette généralisation méprisante ne pourrait en effet s'expliquer que dans le contexte du colonialisme. Avec la définition du terme *sidi*, la dépréciation par rapport au sens originel atteint son paroxysme.

**sidi** n.m. (déb. XX<sup>e</sup>; «monsieur, seigneur», 1847; mot arabe). *Péj.* Indigène Nord-africain établi en France.

La déviation est flagrante si nous rapprochons ce terme à un autre emprunt, celui-ci anglais, dont le sens originel pourrait être assimilable à celui de notre premier exemple. Sémantiquement parlant les termes *sidi*, *gentleman* et *gentilhomme* sont très proches. Ainsi, l'histoire des hommes et les réalités qui en découlent se lisent dans le vivant des mots.

d) *Émergence du sens* : C'est où on pourrait le moins s'y attendre que l'emprunt redore son blason. En effet, dans l'argot, ou

plutôt le français non-conventionnel, *fatma* signifie : «femme, fille» (Goudailler, 1997 : 126). La présence de la communauté maghrébine en France et plus particulièrement la nouvelle génération apparaît, dès lors, comme le facteur d'une évolution de l'emprunt à l'arabe vers un sens plus proche des réalités sémantiques originelles, plus vrai et par cela même plus valorisant.

## II. — EN TUNISIE

### 2.1 — Degré de vitalité

Les emprunts à l'arabe, bien que déjà bien vivants en France, de retour au pays d'origine redouble de vigueur. En effet, certains termes tels que *hammam*, *oued*, ou encore *qasida* sont considérés, en France, comme des «exotismes», c'est-à-dire des mots français faisant référence à une réalité hors de France. Pour notre étude, il s'agit d'une réalité tunisienne bien vivante. D'autres emprunts, qui évoluent en France, font référence à une réalité importée. C'est le cas de *baklava* et *tajine* par exemple qui d'ailleurs ne porte aucune marque de niveau de langue ni de marque de connotation. Ces termes restent d'usage très fréquent en Tunisie. Certains mots empruntés à l'arabe ont un équivalent en français : *sidi*, *toubib*, *barda*, *casbah*. Nous avons observé dans leur utilisation en France une certaine connotation dévalorisante. On peut alors se demander si leur retour est possible ou s'ils ne deviennent pas tabou. Bien que parfois marqués par la France, ces mots reviennent avec une fréquence plus ou moins grande. Ce retour se fait non sans action sur le niveau de langue et sur le sens, comme nous le verrons par la suite.

### 2.2 — Évolution morphologique

#### a) Morphologie flexionnelle

— *Le genre* : le locuteur tunisien francophone maîtrisant *a priori* parfaitement sa langue maternelle qu'est l'arabe, ne peut commettre d'erreur que sur la façon d'appréhender le terme emprunté dans un français standard. Le genre est un point qui pose problème. Ainsi, *merguez* se retrouve au masculin, conformément à l'arabe. Le retour au genre originel est souvent délibéré. Le terme *henné* est à l'origine un mot féminin mais dans 9 cas sur 10 il est au masculin. La seule attestation au féminin est orale et le locuteur possède une bonne connaissance du français. L'écrit serait donc plus

conforme aux normes du français standard. Cela se confirme dans un usage néologique par rapport au français standard. Là, 2/4 attestations se présentent sous la forme «francisée» (le *henné*) et ce sont des exemples écrits, alors que les deux autres sont oraux (la *henna*). Quant à *felouque*, à son retour en Tunisie, il ne réintègre pas sa marque féminine arabe *foulouka*. De plus toutes les attestations relevées sont tirées d'exemples littéraires. Le terme devient finalement aussi étranger à l'usage français qu'à l'usage tunisien. Cette volonté de préserver cette forme pourrait alors être d'ordre stylistique et on pourrait considérer que ce terme est d'usage littéraire. Enfin, il semblerait qu'aucune marque du féminin de l'arabe ne soit signalée pour les emprunts en usage dans le français standard. En Tunisie, en revanche, nous retrouvons la flexion du féminin arabe de manière systématique.

*un ghoule / une ghoula*

*un djinn / une jennia*

*un roumi /une roumia*

*un ha(d)j / une haja*

— *Le nombre* : pour le cas des emprunts en contexte francophone tunisien, la marque arabe du pluriel concurrence plus fortement celle du français. Ainsi, alors que le *Petit Robert* ne signale pas la possibilité du pluriel arabe *jnoun*, le corpus du français en Tunisie le présente comme une possibilité à part entière, au même titre que *(d)jinns* (30 % d'occurrence de *(d)jenouns*). Il en est de même pour le pluriel arabe de *roumi*, *rouamas*, qui n'est pas signalé en France mais qui apparaît dans 12 % des cas en Tunisie. De plus, on remarque que lorsque la marque arabe du pluriel est utilisée, à celle-ci s'ajoute souvent la flexion *s* du pluriel français. Ce phénomène de double marquage est très fréquent : *(d)jenouns*, *rouamas*, *ksours*. Quoiqu'il en soit, la marque arabe du pluriel semble être dominante dans ces cas là. Ainsi, dans 57 % des cas nous trouverons la forme *ksour* et non *ksars*. Enfin, l'utilisation du singulier arabe pour le cas de termes utilisés le plus souvent au pluriel est systématique en Tunisie. Ainsi nous trouvons *fidai* au singulier par opposition à *fedayines* ou *fidayines* au pluriel. Certains singuliers n'apparaissant pas en français de la métropole sont même réintégrés. Ainsi, nous avons *kelb* comme singulier de *cleb* dans 50 % des cas.

#### b) *Morphologie dérivationnelle*

La forte vitalité des emprunts en contexte originel permet un grand nombre de possibilité du point de vue de la création

linguistique. Ces termes peuvent s'approprier les phénomènes classiques de dérivation du français :

*dégourbification; gourbiville*

*après-ramadanesque; ramadanesque*

Ces dérivations peuvent être considérées comme des «tunisianismes». Toutefois, dans certains cas, il s'agit de termes réexportés et qui ont été créés en France, comme *bledard*. L'emprunt, dans son contexte originel, retrouve souvent la mémoire et réintègre alors sa forme d'origine, rétablissant les éléments tronqués : *loukoum/rahat loukhoum/ ruh halqum*; ou redonnant sa place à un article qui aurait été affixé au nom pour obtenir un seul lexème : *luth / el oud / el ud*. Il apparaît aussi des mots composés formés d'éléments français associés à des éléments arabes : *salade méchouia, oreille de kadi, mufti de la république*. Le comportement morphologique du français en Tunisie semble répondre au souci de la langue française face à l'emprunt. En effet, si cette dernière souhaite respecter pour le mieux la structure d'origine tout en intégrant le terme dans une structure française, le locuteur tunisien, en rétablissant les structures de l'arabe, donne à l'emprunt une morphologie qui s'adapte à deux systèmes en cohabitation.

### 2.3 — Évolution graphique et phonétique

a) *Graphies et sons inconnus au système phonétique français* : de retour en Tunisie, les emprunts à l'arabe contenant à l'origine des sons et une graphie inconnus au système français retrouvent la mémoire et réintègrent ces éléments que le français de la métropole avait dû adapter. Ainsi, à la graphie et aux sons qui leur sont attribués en France s'ajoutent de nouvelles possibilités de transcription.

*calif 75 % / khalif 25 %; califat 50 % / khalifat 50 %*

*caïd 67 % / gaïed, gayed 33 %; kacida 40 % / qasida 60 %*

(N.B. le [g] est une variante dialectale du [q] arabe)

*cadi, qadi 42 % / cadhi, qadhi 58 %*

*makroud, makrout 33 % / makroudh 67 %*

Certains termes, tels *muezzin*, semblent être tombés dans l'oubli (amnésie totale !). Notons, d'autres part, que deux consonnes du système français sont inconnus au système tunisien : *v* et *p*. Ainsi

*baklava*, mot d'origine turque, se présente toujours en français tunisien avec le son [w] : *baklawaw*, *baklawaw*

b) *Cas du r* : si le *gh* désigne automatiquement [R], ce n'est pas le cas de *r* qui, bien que correspondant souvent à [r], peut aussi correspondre à [R]. Ainsi, on trouve, même si cela est rare, l'occurrence *roul* comme variante de *ghoul*. Le locuteur tunisien, étant bilingue, est beaucoup moins gêné par ce genre de variantes que pourrait l'être un locuteur français unilingue.

c) *Nasalisation* : sa tendance est à l'absence de nasalisation, celle-ci étant étrangère au système arabe. La graphie peut être révélatrice de cette tendance ainsi, dans 75 % des cas, *chitane* prend un *e* final.

d) *Voyelle et semi-voyelle (ou glides)* : la variabilité graphique que nous avons constatée pour les emprunts en France se retrouve en Tunisie de façon beaucoup plus marquée. Ainsi un même mot apparaît parfois sous quatre à cinq graphies différentes :

*ulama, uléma, oulema*  
*jihād, jihad, jihed, djihad*  
*chitâne, chitane, chaïtan, chaytane, cheitane*

Cette instabilité graphique est aussi sensible au niveau du choix des consonnes : le son [s] peut aussi bien être transcrit par un *c* que par un *s* ou deux *s*.

*qacida, qasida, qassida*

## 2.4 — Niveau de langue

Nous avons constaté que le passage de l'arabe au français était souvent accompagné d'un changement du niveau de langue vers un usage familier, argotique voire populaire et péjoratif des termes empruntés. La question est de déterminer comment ces mots sont appréhendés de retour au pays.

La tendance dominante est à une neutralité du niveau de langue. Certains termes, rencontrés dans des discours sérieux voire soutenus, peuvent être jugés d'usage courant. C'est le cas de *baraka* (bénédiction divine) considéré familier en métropole et de *baroud* (lutte, combat, au sens propre et figuré) appartenant à l'argot français. D'autres termes, uniquement signalés populaires et/ou péjoratifs en France, retrouvent leur usage courant une fois de retour au pays, avec

parfois perte de toute connotation péjorative. C'est le cas de *sidi*. Le terme *arbi*, selon le contexte, pourra être jugé soit d'usage courant, soit péjoratif. Toutefois, cette utilisation du sens péjoratif dénote souvent une certaine ironie de la part du locuteur, dénonçant ainsi la perversion du niveau de langue et par conséquent du sens qui s'est opérée lors du passage à la langue française. Pour certains termes polysémiques en France, le premier sens, sans marque de niveau de langue, correspond, en général, au sens d'origine, les autres sens appartenant au registre familier. Il arrive que ce sens familier disparaisse complètement en Tunisie. Ainsi, il n'a été relevé aucune attestation de l'expression : *faire suer le burnous*. Ceci s'explique sans doute par la forte connotation péjorative qu'elle contient. Le sens familier attribué en France peut aussi être assumé, le terme emprunté devenant ainsi plus riche du point de vue sémantique. C'est le cas de *bled*.

## 2.5 — Évolution sémantique

Le retour au pays est en fait un retour à la réalité. Le mot n'est plus un exotisme qui ferait référence à un ailleurs. Le sens originel est remis en contexte ce qui rend le mot d'autant plus vivant. Il retrouve la mémoire au contact de cet environnement familier. La *ghoula* redevient l'ogresse qui hante l'imagination des enfants, et le *méchoui* est libéré de ses restrictions morphologiques et sémantiques et s'applique à nouveau à tout ce qui peut être grillé ou rôti. Le terme donne en outre naissance à un autre francophonisme tunisien : *salade méchouia*, qui désigne une salade composée de légumes (poivrons, tomates, oignons) grillés et coupés en dés. *Salamalec* devient lui aussi plus parlant autour de sa famille de mots. Ainsi est apparu un autre tunisienisme : *salam* ou *salem*, qui signifie littéralement «paix» et qui est utilisé comme formule de salut ou pour prendre congé. *Salamalec* signifie «paix sur toi» et s'utilise de la même manière que son dérivé. Le locuteur bilingue connaissant le sens exacte et la valeur des mots les traitera avec plus d'égard. Ainsi, bien que conscient du nouveau sens acquis en contexte français, le locuteur évite la déperdition du sens originel de *salamalec*. Le terme devient alors polysémique. Le sens originel dénué de connotation particulière s'ajoute au sens réducteur et souvent familier qui lui est attribué en France.

Ainsi, de retour en terre natale, où la langue maternelle est en contexte, les mots pervertis sous l'influence du colonialisme et par le poids de l'histoire, retrouvent la mémoire qui leur révèle tout leur sens et toute leur valeur. Certains sens acquis en contexte français ne sont

d'ailleurs pas imaginables de la part d'un locuteur tunisien. Ainsi, l'ambiance de tapage nocturne et l'idée de nuisance intolérable sous entendue, dans le sens donné à *ramdam*, en France, laissent la place aux joies partagées des veillées du mois de Ramadan, mois de jeûne, de recueillement, de générosité et de solidarité. *Fatma*, la Cendrillon de la métropole, laisse tomber son tablier de domestique pour redevenir une jeune femme, sans aucune péjoration, bien au contraire, *Fatima* étant le prénom de la fille unique du prophète. *Sidi* retrouve son statut de monsieur et tout le respect et la dignité d'un saint homme, ainsi que sa noblesse de saint et de seigneur.

## CONCLUSION

Nous pouvons distinguer, parmi les emprunts à l'arabe, deux manières différentes de vivre et d'évoluer en France : 1) les mots intégrés dans une réalité française; 2) les «exotismes», qui font référence à un ailleurs. Dans l'ensemble, ces mots ont tendance à s'adapter au système de la langue française, avec toutefois un souci de respect des structures d'origine. Ainsi, les «exotismes» maintiennent leur usage courant et leur sens originel. Les mots qui évoluent dans une réalité française subissent souvent une dépréciation de leur niveau de langue et de leur sens. Cette dévalorisation est explicable par un contexte historique colonial belliqueux. À ce passé de la colonisation s'ajoute celui de l'émigration, ainsi que le présent de la vie maghrébine en France d'où semble émerger une revalorisation du sens et de l'usage ainsi qu'une revitalisation du système arabe.

De retour en Tunisie, la vitalité des emprunts est décuplée. Aux acquis de leur séjour en France s'ajoute le savoir de la mémoire retrouvée. Un ajustement s'effectue alors. Dans cette francophonie qui se découvre et se construit, les mots dans leur diversité cherchent la stabilité, un équilibre entre deux langues, entre deux mondes. Revigorés à la source, un nouveau départ s'annonce, une nouvelle vie par le biais de la francophonie.

## RÉFÉRENCES

- BEN SMAIL, Mohamed (1994) : *Dictionnaire des mots français d'origine arabe*, TUNIS, STER.
- CELLARD, J. (dir.) et A. REY (1980) : *Dictionnaire du français non-conventionnel*, PARIS, Hachette.
- GIRARD, E et B. KERNEL (1996) : *Le vrai langage des jeunes expliqués aux parents*, PARIS, Albin-Michel S.A.
- GOUDAILLIER, J-P (1997) : *Comment tu tchatches ! : dictionnaire du français contemporain des cités*, PARIS, Maisonneuve et Larose.
- LANLY, A. (1962) : *Le français d'Afrique du Nord : étude linguistique*, Paris, Bordas.



# Mémoire des racines et mémoire des mots : le lexique stratifié de l'arabe

Joseph DICHY

*CRTT, Université Lumière Lyon-2, Lyon, France*

## INTRODUCTION (1)

Deux traits caractérisent le lexique arabe tel qu'il nous est parvenu :

- l'ancienneté de son vocabulaire, attestée par plus de quinze siècles de textes écrits et marquée, comme on le verra, par plusieurs formes de conservation du sens ancien des mots à côté de leur sens moderne;
- l'hypothèse, aujourd'hui si répandue qu'elle apparaît à certains comme traditionnelle, d'un système de racines (2) (au sens que prend

---

1) Ce travail se situe dans la perspective de la constitution de la base de données lexicales étendue DIINAR («Dictionnaire INformatisé de l'ARabe»). Les partenaires de ce programme, qui a abouti à la réalisation d'une première base de données (DIINAR.1), sont, à l'heure actuelle (sept. 1997) : l'Institut de Recherche en Sciences Informatiques et des Télécommunications (IRSIT, Tunis - Pr. S. Ghazeli et Dr. A. Braham), l'École Nationale Supérieure des Sciences de l'Information et des Bibliothèques (ENSSIB, Villeurbanne - Dr. M. Hassoun) et l'Université Lumière-Lyon 2 (CRTT - Dr. X. Lelubre et l'auteur); d'autres organismes doivent participer à partir de 1998 à une extension coordonnée par l'Université Lyon 2 du programme DIINAR, dans le cadre d'un projet de la Commission européenne (DG XIII). Ce travail présente mes propres propositions, et n'engage bien entendu pas la responsabilité des autres partenaires scientifiques.

2) Sauf indication contraire, *racine* sera employé ici pour «racine triconsonantique» (ou, par extension, quadriconsonantique). Il existe également des racines monoconsonantiques dont on doit la découverte dans le domaine sémitique et l'inventaire en arabe à A. Roman (1990).

ce terme dans le domaine sémitique) et de schèmes (3). Dans le cadre d'une conception largement revue de cette hypothèse (Dichy, 1997), se pose la question de l'existence d'une «mémoire des racines», à côté de celle des «mots» : il y aurait ainsi une double mémoire du lexique, inscrite dans la mémoire des textes. Le problème est, dès lors, celui de la combinaison d'un sens lexical associé à la racine et d'un sens grammatical associé au schème (dérivation «interne») mais aussi aux «morphèmes-extensions» (dérivation «externe»), auquel s'ajoute le sens lexical associé au résultat de ces «dérivations», c'est-à-dire aux *unités lexicales* elles-mêmes (4).

La première partie de cette contribution dressera un bref tableau de la perspective historique dans laquelle s'inscrit la stratification du lexique arabe (phénomènes de *néologie* ou de *polysémie stratifiées*). Cette mise en perspective est, on vient de le voir, inséparable d'un questionnement sur la «mémoire des racines» et sur celle des «mots», c'est-à-dire, sur les notions de «racine» et de «schème», dont le statut doit être révisé en fonction notamment de ce que j'ai appelé le *principe lexical de composition (PLC)*; ce sera l'objet de la dernière partie.

## ANCIENNETÉ HISTORIQUE DU LEXIQUE ARABE (5)

Antérieurement à l'avènement de l'islam, l'arabe était parlé principalement par les populations du Nord de la Péninsule arabique et de ses confins syro-iraquiens. À partir du VII<sup>e</sup> siècle, en même temps qu'il devient une langue de culture et de civilisation à vocation universelle et dotée d'un système d'écriture élaboré, l'arabe voit son aire d'extension s'élargir de manière considérable.

Or l'évolution qui s'ensuivit fut perçue par la tradition médiévale comme une décomposition ou une «corruption» (*fasâd*) par rapport à ce que l'on décrirait aujourd'hui comme la langue de la communauté linguistique au sein de laquelle a été prononcée et reçue la prédication fondatrice de l'islam.

---

3) Cette conception, due à J. Cantineau (1950 a et b), a été reprise entre autres par D. Cohen (1961/70). Elle a été critiquée notamment par P. Larcher (1995) mais aussi, dans le cadre de l'étude du système graphique de l'arabe, par l'auteur de cette contribution (Dichy, 1995/97a).

4) Voir, pour la structure de l'*unité lexicale simple* (i.e. comportant un seul noyau lexical) en arabe, Dichy (1995/97b).

5) Pour les mots arabes transcrits, voir l'annexe située à la fin du texte. Ce paragraphe reprend, avec des modifications, un passage de Dichy (1996b).

Au X<sup>e</sup> siècle, al-Xattâbî, dans son épître sur l'inimitabilité du Coran, signalait à ses contemporains qu'ils ne pouvaient faire que partiellement appel à leurs propres connaissances pour comprendre le texte coranique <sup>(6)</sup> : la langue pratiquée alors dans les centres urbains où se développait la civilisation arabo-musulmane n'était plus la même que celle des *Arabes* du Nord-Est de la Péninsule arabique (?a'râb) à l'époque du prophète. (*Arabes* est mis ici en italiques, afin d'éviter toute confusion avec l'usage actuel de ce terme.) Il semble bien d'ailleurs que l'incapacité de produire un texte similaire au Coran, qui fonde le dogme de l'inimitabilité de celui-ci (?i'jâz, mot-à-mot : «réduction à l'impuissance»), concernait tout naturellement les contemporains du prophète, plutôt que les hommes du X<sup>e</sup> siècle et des époques ultérieures.

Cette «corruption» n'affectait toutefois pas, dans l'esprit de ces savants, le *lisân al-'arab*, la «langue des Arabes» elle-même, mais ses réalisations, *lugât* <sup>(7)</sup>. Observons que l'on ne disait pas *fasâd al-lisân*, «la corruption de la langue», mais *fasâd al-luga*, «la corruption de la manière de parler» : le *lisân* est en effet la langue de la révélation coranique. Dieu, pour les musulmans, ayant fait choix des *Arabes* du VII<sup>e</sup> siècle pour la transmission du message par lequel se trouvait scellée la prophétie, les hommes de science des siècles suivants, habitant les cités de Syrie, d'Iraq, du Maghreb, de la Perse, de l'Andalus ou de l'Égypte, devenaient les dépositaires du Coran et des *dicts* du prophète (*hadit*). Mais il fallait en assurer la transmission au sein d'une nouvelle communauté linguistique arabo-musulmane, dont les membres n'avaient plus, pour l'écrasante majorité d'entre eux, le même arabe que le prophète comme langue maternelle, et qui faisaient usage, dans la vie de tous les jours, de dialectes ou de parlers bien plus éloignés de la langue «classique» que ne l'étaient les variantes dialectales des anciens Arabes — à quoi s'ajoutaient, dans bien des cas, des langues comme le persan, le syriaque, le berbère, etc. Or, dit le Coran : «Nous n'avons envoyé nul Apôtre sinon dans l'idiome de son peuple» (XIV, 4, trad. R. Blachère). La «clarté» (*bayân*) du texte sacré ne pouvait perdurer qu'au prix du maintien d'une compétence linguistique savante correspondant à la connaissance du *lisân al-'arab* <sup>(8)</sup>.

---

6) Cf. Audebert (1982 : 135-136).

7) A. Hadj-Salah (1979) et article *luga* de l'*Encyclopédie de l'Islam*, a magistralement montré la différence entre ces deux termes.

8) Rappelons que *Lisân al-'arab* est le titre du grand dictionnaire compilé par Ibn Manzûr au XIII<sup>e</sup> siècle.

On s'empresse, dès lors, à partir du VIII<sup>e</sup> siècle, de compiler toutes les informations relatives au «discours des Arabes» : poésie, lexique, relations (?*axbâr*), morceaux d'éloquence, proverbes... Parallèlement à la collecte de ce corpus, on rassemble toutes les données grammaticales et lexicales que l'on peut réunir, soit en attestant les règles ou les emplois lexicaux au moyen d'exemples tirés, le plus souvent, de la poésie, soit en interrogeant les locuteurs natifs des zones du Nord de la Péninsule non susceptibles d'avoir été «corrompues» par des contacts avec des langues étrangères. Jamais, dans l'histoire de l'humanité antérieurement à l'époque moderne, un aussi formidable travail de collecte et de description grammaticale, rhétorico-pragmatique, rhétorique, lexicale, n'avait été réalisé : le corpus et la langue de la révélation coranique se trouvaient ainsi conservés, transmis, institués, avec un souci d'exhaustivité propre aux époques médiévales (9).

## LA STRATIFICATION DU LEXIQUE, SCHÉMA DE PÉRIODISATION

Le lexicologue contemporain peut, sur cette base, mais avec les précautions nécessaires, prendre comme point de départ de ses travaux un schéma diachronique simplifié, comportant trois «grandes époques» techniquement identifiables (malgré les innombrables difficultés de la tâche (10) :

- [1] celle du *lisân al-'arab*, qui correspond au *fonds lexical ancien*. à la fois antérieur à l'islam et contemporain de sa fondation. Ce dernier comporte deux subdivisions :
  - [1.1] le *vocabulaire pré-islamique* (ou «arabe originel»), qui comprend un ensemble de notions de base (aller, venir, parler, etc.) ainsi que des notions le plus souvent exprimées par des noms, qui renvoient à la vie nomade des anciens Arabes (*dâr*, «tente»; *watan*, «lieu de vie, territoire,

---

9) «Ce souci projetait cependant une ombre : même la poésie ancienne semble avoir été enregistrée en régularisant certaines différences (Blachère, 1977); *a fortiori*, ce qui formait l'usage vivant de la grande masse des locuteurs des époques ultérieures était rejeté hors description, et, naturellement, hors norme instituée» (Dichy 1996b).

10) Citons, parmi ces difficultés, celle de la détermination, dans les œuvres des lexicographes arabes médiévaux, de ce qui relève effectivement du corpus ancien, et de ce qui ressortit à un travail de reconstruction.

quartiers», tant pour les animaux que pour les hommes (11); *ṣahâda&*, «témoignage»; *tarab*, «gaieté ou tristesse profonde», puis «émotion profonde causée par l'audition d'un poème ou d'un chant», «ivresse de l'écoute»);

- [1.2] le *vocabulaire fondateur de l'islam*, qui procède le plus souvent par néologie par rapport au lexique antérieur, dont il réinscrit les unités dans le système de significations institué par la révélation coranique et par l'enseignement du prophète. Ex. : *salâ&*, «prière»; *hudä*, «bonne voie», «Direction»; *ṣahâda&*, «prononcé de la formule rituelle par laquelle le croyant témoigne de sa foi en un Dieu unique, et en la prophétie de Muhammad» ou «témoignage», au sens que prend ce terme en droit musulman;
- [2] le *fonds lexical arabo-musulman médiéval*, qui renvoie, dans l'ensemble, aux cités de l'islam, des premiers siècles de l'Hégire à la fin du XVIII<sup>e</sup> siècle du calendrier grégorien, ainsi qu'aux prescriptions et aux doctrines religieuses, à l'activité intellectuelle, littéraire, scientifique, administrative, etc. de ces époques. Ex. : *qâdî*, «juge»; *watan*, au sens de «pays d'origine». Ici se pose le problème difficile de la délimitation des périodes — d'autant que les processus néologiques ne manquent point, ex. : *mantiq*, qui signifie originellement «paroles» ou «discours», prend plus tard, sous la plume des philosophes, les sens de «logos» puis de «logique»; *mutrib*, participe actif de sens factitif, construit sur la même racine que *tarab* ci-dessus ([1.1]), en vient à signifier «chanteur» à partir, vraisemblablement du IX<sup>e</sup> ou du X<sup>e</sup> siècle (cf. le dictionnaire réalisé par Ibn Durayd au X<sup>e</sup> siècle);
- [3] le *fonds lexical moderne*, dont le point de départ est traditionnellement projeté au début du XIX<sup>e</sup> siècle. Ex. : *sayyâra&*, «voiture»; *hâtif*, «téléphone»; mais aussi *siyâsa&*, «politique», *mujtama'*, «société», au sens *moderne* de ces termes; *watan*, «patrie», au sens nationaliste contemporain, illustré par l'expression *al-watan al-'arabî*, «la patrie arabe», citée par le «Dictionnaire de base» de l'ALECSO (1989); *ṣahâda&*, «attestation», «diplôme».

---

11) Pour l'étude du vocable *watan* dans quelques dictionnaires arabes médiévaux et contemporains, voir Dichy (1996a).

Ce schéma de périodisation doit, malgré son caractère rudimentaire, être considéré par le lexicologue comme un cadre de travail dans lequel on peut inscrire par exemple l'étude en diachronie de l'évolution sémantique des mots (cf. les exemples de *watan*, *tarab* ou *sahâda&* évoqués ci-dessus). À défaut d'un premier ensemble de points de repères diachroniques ou de conventions d'indexation, aucun travail lexicographique (réalisation de dictionnaires historiques) ou lexicomatique (constitution de bases de données indexées) ne pourra prendre en compte les phénomènes de stratification historique du vocabulaire arabe.

## LA NÉOLOGIE DE SENS CUMULATIVE

L'examen d'un grand nombre de données fait apparaître deux aspects complémentaires de cette stratification (Dichy 1996a) :

### 1. — Une fréquente conservation du sens ancien des mots à côté de leur sens moderne

On pourrait dire que les entrées du lexique arabe admettent, du fait des quinze siècles de pratique textuelle qu'elles sont amenées à refléter, une relation entre les acceptions médiévales et les acceptions contemporaines des termes, que l'on pourrait décrire comme un phénomène de *néologie de sens potentiellement cumulative*. En d'autres termes, les sens anciens ne sont pas toujours effacés par les nouvelles acceptions que prennent les unités lexicales, mais coexistent au contraire très souvent avec elles. Exemples :

- *Sayyâra&* a aujourd'hui le sens (indexé [3] ci-dessus) de «voiture, automobile»<sup>(12)</sup>, qui coexiste dans le lexique actuel avec son sens ancien ([1] ci-dessus) de «planète»<sup>(13)</sup>. En revanche, un autre sens ancien, celui de «caravane» ([1.1]), attesté par exemple dans Coran (XII, 19) ne paraît plus relever du lexique utilisé de nos jours<sup>(14)</sup>.

12) *Sayyâra&*, «automobile», a pour analogon : *tayyâra&*, «avion», attesté en arabe littéraire moderne bien que l'usage semble lui préférer *tâ?ira&*, construit sur le schème du participe actif combiné avec le suffixe +*a&* (mot-à-mot : «la chose volante»).

13) «Planète» est également exprimé au moyen du mot composé *kawkab sayyâr* (mot-à-mot : «astre qui se déplace beaucoup»).

14) Malgré son haut niveau de probabilité, cette affirmation demeure, en l'état actuel des connaissances, hypothétique : on ne dispose en effet encore ni d'études de corpus étendus permettant de confirmer ou d'infirmer de telles hypothèses, ni des outils logiciels grâce auxquels de telles analyses seraient possibles. Le programme de recherche DIINAR devrait apporter, dans ce domaine, une contribution importante.

- Les différents sens de *ṣahâda* ci-dessus, de «témoignage» ([1.1]) à «diplôme» ([3]) sont présents dans la compétence linguistique des locuteurs actuels de l'arabe littéraire.

Pour comprendre ce phénomène, le lecteur français doit imaginer que *navrer*, à côté de son sens actuel, conserve la possibilité de signifier en fonction du contexte «infliger une blessure grave». La néologie de sens consiste dans ce cas, comme on le sait, en une reprise d'un mot, souvent tombé en désuétude dans son acception originelle, en conservant certains sèmes (celui de «blessure» dans l'exemple de *navrer*), à l'exclusion de certains autres (celui de «blessure grave»), et en en modifiant d'autres (ainsi, «blessure» perd le trait «physique», et prend, par un processus métaphorique, une valeur psychologique ou affective). Dans d'autres cas, comme pour *ordinateur*, les deux acceptions, ancienne et nouvelle, n'ont plus de sème en commun. Ce processus appartient à la vie d'une langue, c'est-à-dire, à la manière dont une communauté linguistique, d'époque en époque, renomme les événements et les objets du monde. Le lexique arabe manifeste, par le caractère potentiellement cumulatif de sa néologie de sens la fidélité d'une culture aux textes qui disent sa mémoire, et dont la présence traduit son identité.

## 2. — Une interpénétration potentielle, corollairement, des sens ancien et moderne

Corollairement, on observe dans cette langue une interpénétration potentielle des sens ancien et moderne : la néologie de sens cumulative est en effet susceptible de relever de la polysémie, ex. : *ʔi'taqala*, «lier, attacher, retenir quelqu'un en captivité» (trad. Kazimirski) est réutilisé par le vocabulaire moderne au sens de «mettre en arrestation, arrêter».

La coexistence des sens ancien et nouveau maintient ainsi vivante la possibilité d'un continuum polysémie-néologie (15).

## MÉMOIRE DES RACINES OU MÉMOIRE DES MOTS ? L'ORGANISATION GÉNÉRALE DU LEXIQUE ARABE

Sur la base des observations qui précèdent, on peut poser — en résumant fortement le propos — que chaque unité du lexique arabe s'insère dans un triple réseau :

---

15) Explicitons, à l'intention du lecteur non linguiste : classiquement, la polysémie opère en synchronie, la néologie en diachronie. Les phénomènes de stratification observables dans des pans entiers du lexique de l'arabe littéraire réduisent la différence entre ces deux types de processus.

### 1. — Le réseau notionnel

Il s'agit, classiquement, du réseau des notions incluses dans des champs sémantiques, dont l'organisation varie considérablement selon les langues et les cultures.

### 2. — Le réseau des sens grammaticaux relevant de la morphologie et des relations lexique-grammaire

Le deuxième réseau est de nature morphologique. Il s'agit des systèmes de connaissances relatifs :

- à la «dérivation interne», c'est-à-dire, à ce que la grammaire traditionnelle désigne comme des schèmes ou des formes morphologiques (en arabe *wazn*). Par exemple : une forme verbale en /fa'ula/ à l'accompli a inévitablement un inaccompli en /yaf'ulu/, et correspond à un verbe de qualité ou d'état; les schèmes /tafa'ala/, /tafâ'ala/, /?ifta'ala/ comportent le «morphème écho» de racine monoconsonantique /t/ qui «réfléchit le morphème de personne à l'instar des verbes pronominaux français. Ex. : *jagtasilu*, 'il se lave'» (Roman, 1990 : 50);
- aux suffixes de la «dérivation externe», qui sont exclusivement nominaux, ex. : *lisâniyyât*, «linguistique», composé du noyau lexical *lisân* («langue») et des suffixes +*iyy* (de construction du nom-adjectif de relation) et +*ât* (pluriel du suffixe +*a&*);
- à la conjugaison et à la construction ou à la reconnaissance des déverbaux, ou «dérivés nominaux immédiats», tels que le nom de procès (*masdar*), les participes actif (?*ism al-fâ'il*) et passif (?*ism al-maf'ûl*)...;
- aux relations entre schèmes (dérivation schème1 → schème2), qui combinent dans certains cas les deux modes «interne» et «externe» de dérivation, ex. *xârij+iyy*, plur. *xawârij*, «Kharijite» (membre d'une secte des premiers temps de l'islam).

Etc. À cela ajoutent de nombreuses informations relatives aux relations entre lexique et grammaire (relations morpho- ou syntactico-lexicales) qu'il ne nous est guère loisible de présenter ici.

### 3. — Le réseau des significations au sein d'une même racine

Le troisième réseau est celui des significations observables au sein d'une même racine. Cette dernière a fait l'objet de nombreuses critiques (voir plus haut note 3). Mais celles-ci portent surtout sur l'usage, souvent abusif, qui est fait de cette notion plutôt que sur son utilité. L'un

des principaux problèmes, en ce qui concerne la conception des bases de données lexicales en arabe, réside dans le fait — qui constitue dans ce domaine un véritable obstacle épistémologique — que l'on croit pouvoir se contenter d'un dictionnaire des racines «croisé» avec un dictionnaire des schèmes.

J'ai proposé ailleurs une redéfinition de la représentation par racines et par schèmes (cf. Dichy 1997, où l'on trouvera en outre plus de références bibliographiques qu'ici). Rappelons toutefois, pour faire court, le caractère second — je dirais, l'*après-coup analogique* — de cette représentation. Le système de nomination de la langue arabe s'est constitué à l'origine à partir d'une disjonction fonctionnelle entre les voyelles et les consonnes, les racines de la langue étant triconsonantiques ou monoconsonantiques (cf. les travaux de référence d'A. Roman). Dans le lexique de cette langue, ce qu'on appelle communément «schème» jouit d'un statut :

- soit de pure représentation formelle (le «reste» — dénué de valeur sémantique — de la base lexicale une fois isolée la racine; ex. : *fi'âl* dans *himâr*, «âne»). Il s'agit en fait de *quasi-schèmes*;
- soit d'unité dotée d'un «sens grammatical», mais entrant en composition, pour la production d'un sens lexical tant avec les racines triconsonantiques qu'avec une partie des racines monoconsonantiques. C'est seulement dans ce deuxième cas que l'on est fondé à parler de *schème*. Il faut toutefois noter que la frontière entre schèmes et quasi-schèmes doit être déterminée avec précision dans le domaine des noms.

Or une majorité du lexique arabe relève de ce processus complexe de composition (*tous* les verbes et les déverbaux et une partie très importante des noms) (Dichy, 1995/97b : note 10). Les unités qui en résultent sont soumises à ce que j'ai appelé le *principe lexical de composition* (PLC) :

«Dans une langue donnée, si deux éléments <a> et <b> ayant une valeur sémantique entrent en composition lexicale, l'ensemble des traits morphosyntaxiques et sémantiques du couple <a,b> est irréductible à la sommation des traits de <a> et de <b>». (16)

La plupart des illustrations ci-dessus pourraient être citées à l'appui du PLC. Mais considérons la manière dont celui-ci éclaire la relation entre unités lexicales et racines.

---

16) Je me contenterai ici d'une brève présentation du PLC dans la perspective de la discussion de la notion de racine. Pour un développement théorique plus général, voir Dichy (1997).

Revenons pour ce faire sur l'exemple de *sayyâra&*. Le sens ancien de *sayyâr*, «qui voyage beaucoup (répétitivement)» comportait déjà des valeurs sémantiques non déductibles du sens général de la racine et de celui du schème. Le dictionnaire d'al-Zamaxsarî cite en effet l'expression *rajulun sayyâr*, qu'il glose en mentionnant des déplacements «de pays en pays» (*min baladin ilä balad*) : ce sème, qui actualise la modalité itérative de /fa'â/, est additionnel. C'est sans doute ainsi que *sayyâra&* a pu dénoter «caravane». Dans l'attestation coranique (XII, 19) déjà mentionnée, il s'agit de celle qui recueillit Joseph dans le désert, et qui se rendait *dans un autre pays* (l'Égypte) : en quelque sorte, une «caravane au long cours». Le sens de «planète» pourrait avoir conservé un sème analogue, en raison du fait que l'astre ainsi désigné semble parcourir les régions du ciel nocturne correspondant aux «maisons» du zodiaque. On observerait, dans cette hypothèse une métaphore sous-tendue par une analogie avec l'emploi de *sayyâr* pour décrire des hommes. Mais le sens de «planète» pourrait aussi, plus simplement, être affecté d'un sème «perpétuellement en mouvement», ce qui reviendrait à ne retenir que la seule valeur itérative-intensive du schème, sans l'actualisation que constitue, pour «caravane», le déplacement de pays en pays. Il en va de même pour le sens moderne de «voiture».

Or l'unité lexicale *sayyâra&* est l'objet d'un double processus de composition, qui peut être analysé ainsi (Dichy, 1995/97b) :

- 1 — Le noyau lexical *sayyâr*, «qui voyage beaucoup (répétitivement)»<sup>(17)</sup> est un adjectif construit
  - 1.1 — sur la racine triconsonantique /s-y-r/, dont le sens général, tel qu'il est glosé au X<sup>e</sup> siècle par Ibn Fâris (*Maqâyîs al-luga*, vol. 3 : 120), est «passer, se déplacer» (*mudiyyun wajarayân*), et
  - 1.2 — le schème /fa'â/, dont le sens grammatical est celui de la modalité itérative-intensive<sup>(18)</sup>.

17) Les sens de «caravane» et «planète» sont cités par le dictionnaire de Kazimirsky (1860) — mais, naturellement, pas le premier («automobile»). Pour «qui voyage beaucoup», voir également al-Zamaxsarî, *?asâs al-balâga* (XII<sup>e</sup> s.).

18) A. Roman (1990 : 68) n'indique pour le schème /fa'â/ que la modalité intensive. Des exemples tels que *sayyâr* ci-dessus, ou encore la série des noms de métier *najjâr*, «menuisier», *jazzâr*, «boucher», etc., manifestent cependant, me semble-t-il, une modalité itérative susceptible, comme dans le schème verbal de la «forme II» /fa'âla/ (Roman, 1990 : 54), d'une réinterprétation «intensive», ex. : *kattâb*, «[grand] menteur», qui mérite d'autant plus cette accusation qu'il se livre répétitivement au mensonge. C'est de manière analogue que le «[grand] voyageur» est qualifié de *sayyâr*.

2 — Le suffixe +a& correspond (avec une voyelle épenthétique /a/) au morphème de racine monoconsonantique /t/, signifiant de la «res générale» (19).

Ce double processus se réalise, conformément au principe énoncé par M. Bréal (1897/1925 : 171), selon un mode binaire d'association des éléments entrant en composition. («Cette règle n'est pas arbitraire, écrit-il : elle tient à la nature de notre esprit, qui associe les idées par couples») (20) On a donc :

- (1) noyau (*sayyâr*) = <racine triconsonantique, schème>  
 (2) unité lexicale (*sayyâra&*) = <noyau, racine monosonantique>.

Chacun de ces processus est soumis au PLC, comme on peut le constater dans le tableau suivant :

#### Analyse des sens de *sayyâra&*

	Schème /fa''âl/		Racine monoconsonantique /t/ du suffixe +a&
Sens grammatical	Modalités itérative-intensive		«Res générale», «la chose qui...»
Sens lexical 1	«qui se déplace beaucoup (21)»; «qui voyage beaucoup»	«de pays en pays»	«chose qui ***» > «caravane»
Sens lexical 2	«qui se déplace beaucoup»; «perpétuellement en mouvement»	«d'une région du ciel à l'autre»	«chose qui ***» > «planète»
Sens lexical 3	«qui se déplace beaucoup»	«d'un lieu à un autre»	«chose qui ***» > «automobile»

N.B. Le symbole \*\*\* correspond, dans la 4<sup>e</sup> colonne, à une reprise des sèmes notés dans la 2<sup>e</sup> et la 3<sup>e</sup> (et résultant de la combinaison du sens lexical associé à la racine, et des modalités associées au schème, à quoi il convient d'ajouter l'effet du PLC).

19) A. Roman (1990 : 36 et 39-41). Ex. : *darib*, «frappeur», donne, avec le suffixe +a& : *dariba&*, «impôt» (mot-à-mot : la «chose frappeuse»).

20) L'hypothèse d'une association selon un schéma binaire des éléments entrant en composition lexicale est, dans notre langage actuel, de nature cognitive. Pour M. Bréal et son époque, elle relève de «la nature de notre esprit»; rappelons que l'*Essai de sémantique* paraît pour la première fois en 1897, et la *Soirée avec Monsieur Teste* de P. Valéry, en 1896.

21) Le mot français *beaucoup*, qui comporte les deux sèmes de la «répétition fréquente» et de l'«intensité» rend bien les deux valeurs itérative et intensive du schème /fa''âl/.

Le sème «qui se déplace beaucoup» est déductible de la combinaison du sens lexical «passer, se déplacer» associé à la racine /s-y-r/ et du sens grammatical du schème /fa''âl/. Ce n'est nullement le cas en revanche pour les *actualisations lexicales* des sens grammaticaux de ce schème que l'on observe dans les sèmes «de pays en pays» et «d'une 'maison' du zodiaque à l'autre», ou pour les sens que prend tour à tour le morphème de la «*res* générale». Ces actualisations non déductibles du sens des éléments entrant en composition (ou «irréductibles à la sommation de leurs traits sémantiques») sont des manifestations du PLC, qui constitue une représentation formalisée du processus d'enrichissement du lexique au moyen de diverses formes de composition. (Ainsi, on peut écrire dans le tableau ci-dessus : «chose qui \*\*\* > caravane», mais non «=»)

Revenons maintenant à la racine /s-y-r/ : rien ne la prédestinait plus particulièrement que d'autres à accueillir un jour au sein de ses unités lexicales la voiture (*sayyâra*&) ou la circulation automobile (*sayr*)... Et pourtant, *après coup*, un lien analogique se laisse percevoir, et se prête à l'interprétation. Or la racine n'apparaît qu'en composition : elle n'a pas à proprement parler de sens en elle-même, comme le rappelle, non sans esprit, le titre de Larcher (1995) — bien que l'on puisse difficilement éluder la question d'un sens originel, ou à tout le moins d'un sens acquis d'emblée avec les premiers mots où on la rencontre. Un ensemble de traits sémantiques communs peuvent cependant être déduits des sens associés aux unités lexicales qui «relèvent» d'elle, et auxquels elle sert de «super entrée» dans les dictionnaires arabes depuis le *Kitâb al-'ayn*, au VIII<sup>e</sup> siècle. La racine est ainsi une unité seconde à double titre : parce qu'elle n'est formellement observable qu'à travers les «mots» où elle se manifeste, et parce son sens résulte lui-même de celui que l'usage accorde ces derniers.

Or cette unité seconde est le siège de processus sémantiques que M. Bréal appelait des phénomènes d'*irradiation* (1897/1925 : chap. III). L'exemple donné par lui pour expliquer ce terme est celui de la désinence latine *sco*, à laquelle le lexique a attribué une valeur «inchoative» absente de son sens originel (cf. *nosco*, «connaître», etc.), en raison de sa présence dans des verbes tels que *adolesco*, «grandir», *floresco*, «fleurir», *senesco*, «vieillir»... L'irradiation conduit les racines à accueillir des sens nouveaux : ainsi, /'-q-l/; de sens originel «attacher, lier», est également la racine de la «raison» (au sens de ce qui retient le cœur de l'homme), 'aql, puis de «l'arrestation» au sens moderne du terme (verbe *?i'taqala*). Des sous-ensembles relevant de ces acceptions se dessinent ainsi au sein de la racine; ex. *?i'tiqâl*, «arrestation»,

*mu'taqalûn*, «personnes arrêtées»; *'âqil*, «sage, raisonnable» (dit d'un humain); *ma'qûl*, «raisonnable» (dit d'une chose)...

Ces sous-ensembles relèvent des périodes historiques dont j'ai proposé un tout premier schéma. La racine est ainsi l'un des lieux de la mémoire du lexique, dont j'ai cherché à illustrer le caractère *poreux à l'histoire*. Dans cette langue en usage depuis plus de 15 siècles, les unités lexicales partageant une même racine peuvent avoir des significations remontant à des époques différentes : présentes dans la «mémoire des mots», elles sont alors «activables» en fonction des référents contextuels. La «mémoire sémantique» des racines, dès lors, évolue, mais d'une manière cumulative. D'où la notion de *lexique stratifié*, que toute base de données ou dictionnaire arabe devra intégrer à son système d'indexation.

## ANNEXE SUR LA TRANSCRIPTION DES CARACTÈRES ARABES

La transcription adoptée ici pour des raisons de «portabilité» informatique, est librement empruntée à A. Roman, notamment, 1990. Les emphatiques et la constrictive vélaire *ha:?* sont en caractères gras; le soulignement distingue, lorsqu'il y a lieu, la constrictive de l'occlusive correspondante, ou un phonème d'un phonème «voisin» (ainsi, ḡ notera le son correspondant au *ch* français : il ne s'agit pas d'une transcription phonémique au sens strict !). On a :

- 1) **VOYELLES BRÈVES** : *a*; *u*; *i*.
- 2) **VOYELLES LONGUES** : ?*alif* = â; *wâw* = û; *yâ?* = î.
- 3) **CONSONNES** (par ordre alphabétique) : *hamza* = ?; *bâ?* = *b*; *tâ?* = *t*; *tâ?* = *t*; *jîm* = *j*; *hâ?* = *h*; *xâ?* = *x*; *dâl* = *d*; *dâl* = *d*; *râ?* = *r*; *zây* = *z*; *sîn* = *s*; *sîn* = *ḡ*; *sâd* = *s*; *dâd* = *d*; *tâ?* = *t*; *dâ?* = *d*; *'ayn* = ' ; *gayn* = *g*; *fâ?* = *f*; *qâf* = *q*; *kâf* = *k*; *lâm* = *l*; *mîm* = *m*; *nûn* = *n*; *hâ?* = *h*; *wâw* = *w*; *yâ?* = *y*.
- 4) **MORPHOGRAMMES** : ?*alif maqsûra* = ä; *tâ?* *marbûta* = & (notation due à D. Cohen, 1961/70).

## RÉFÉRENCES

- ALECSO [Arab League Educational, Cultural and Scientific Organisation] (1989) : *al-Mu'jam al-'Arabî al-'asâsî* [«Dictionnaire arabe de base»], diffusion Larousse.
- AUDEBERT, C. (1982) : *Al-Hattâbî et l'inimitabilité du Coran*, traduction et introduction au *Bayân 'i'jâz al-Qur'ân*, Institut Français de Damas.
- BLACHÈRE, R. (1977) : *Introduction au Coran*, Paris, Maisonneuve et Larose.
- BREAL, M. (1897/1924) : *Essai de sémantique*, Paris, Hachette (7<sup>e</sup> éd., 1924).
- CANTINEAU, J. (1950a) : «Racines et schèmes», *Mélanges W. Marçais*, Paris, G.P. Maisonneuve et Cie, pp. 119-124.
- CANTINEAU, J. (1950b) : «La notion de racine et son altération dans diverses langues sémitiques», *Semitica*, vol. III, pp. 73-83.
- COHEN, D. (1961/70) : «Essai d'une analyse automatique de l'arabe», 1961, D. Cohen (dir), *Études de linguistique sémitique et arabe*, Paris, Mouton, 1970, pp. 49-78.
- DICHY, J. (1996a) : «La fidélité contradictoire du lexique à sa mémoire : le champ sémantique «peuple, patrie, nation» à partir de quelques dictionnaires arabes», contribution au volume en l'honneur de M. Arkoun, Tunis, T.S. Éditions, à paraître.
- DICHY, J. (1996b) : «La langue arabe dans l'histoire», *Qantara, magazine des cultures arabe et méditerranéenne*, n° 19, avril-mai-juin 1996, Paris, Institut du Monde Arabe (dossier «Langue arabe, la parole et la plume»), pp. 24-27.
- DICHY, J. (1995/97a) : «Deux grands «mythes scientifiques» relatifs au système d'écriture de l'arabe», *Savoir, images, mirages*, numéro spécial de *l'Arabisant*, Bulletin de l'Association Française des Arabisants (AFDA), Paris, 1997, sous presse.
- DICHY, J. (1995/97b) : «Pour une lexicomatique de l'arabe : l'unité lexicale simple et l'inventaire fini des spécificateurs du domaine du mot», Communication aux IV<sup>es</sup> Journées scientifiques du réseau «Lexicologie, Terminologie et Traduction» de l'AUPELF-UREF (Lyon, 28-30 sept. 1995), *Meta*, vol. 42 n° 2, juin 1997, Montréal, Presses de l'Université de Montréal, pp. 291-306.

- DICHY, J. (1997) : «Sens des mots et sens des racines en arabe», S. Rémi-Giraud et L. Panier (dir), *Journées de lexicologie et de sémiotique* organisées les 12, 13 et 14 juin 1997 par l'EA «Langue, lexique et représentation des connaissances» de l'Université Lumière Lyon-2, à paraître.
- HADJ-SALAH, A. (1979) : *Linguistique arabe et linguistique générale*, Thèse d'État, Université de Paris-Sorbonne.
- IBN DURAYD, Abû Bakr Muhammad (m. en 321/933), *Jamhara& al-luga&*, Beyrouth, Dâr Sâdir, sd., 4 vol.
- IBN FÂRIS, Abû al-Hussayn Ahmad (m. en 395/1005) : *Maqâyis al-luga*, éd. critique A. Hârûn, Le Caire, Matba'a& Mustafâ al-Bâbî al-Halabî, 1972.
- KAZIMIRSKI, A. de Biberstein (1860), *Dictionnaire arabe-français*, Paris, Maisonneuve et Cie, réimpr. Beyrouth, Librairie du Liban.
- LARCHER, P. (1995) : «Où il est montré qu'en arabe classique la racine n'a pas de sens et qu'il n'y a pas de sens à dériver d'elle», *Arabica*, vol. XLII, fasc. 3, Leiden, E.J. Brill, pp. 291-314.
- ROMAN, A. (1990) : *Grammaire de l'arabe*, Paris, Presses Universitaires de France, coll. «Que sais-je ?».
- al-XATTÂBÎ, Abû Sulaymân (m. en 388/998) : «Bayân 'i'jâz al-Qur'ân», al-Rummânî, al-Xattâbî, Abd al-Qâhir al-Jurjânî (eds), *Talât rasâ'il fi 'i'jâz al-Qur'ân*, édition crit. Muhammad Xalaf Allah 'Ahmad et Muhammad Zaglûl Sallâm, Le Caire, Dâr al-Ma'ârif, 3<sup>e</sup> éd. 1976, pp. 19-72.
- AL-ZAMAXSARÎ, Jâr Allâh (m. en 530/1143) : *ʔasâs al-balâga*, toutes réimpr.



# **D'une figure de sens aux sens figurés : modélisation et interprétation avec les graphes conceptuels**

**Karim CHIBOUT et Anne VILNAT**

*Groupe Langage et Cognition, LIMSI-CNRS,  
Orsay, France*

## **INTRODUCTION**

La métonymie <sup>(1)</sup> est, avec la métaphore, une des figures de la langue non littérale les plus usuelles. De plus, elle est la source de nombreux cas de polysémies. L'interprétation de ces figures, dont la fréquence d'occurrence est non négligeable, est une nécessité pour un certain nombre de domaines de l'informatique linguistique : compréhension de textes, interrogation de bases de données et plus généralement dialogue homme-machine en langue naturelle. Un système de traitement automatique des langues, intégrant un atelier de génie linguistique, tel que celui développé dans notre groupe (Vapillon *et al.*, 1997) ne peut se faire en ignorant ces phénomènes.

Dans la métonymie, on distingue classiquement la synecdoque de la métonymie stricte. Dans la synecdoque les relations qui lient le terme remplaçant et le terme remplacé sont des relations habituelles

---

1) La métonymie est une figure de sens particulière qui joue sur l'association d'entités se trouvant habituellement dans le même contexte. Pour cette raison, elle est appelée également figure de voisinage. Ainsi, dans la phrase *Il a bu quelques verres de trop*, le terme métonymique *verre* remplace *boisson (alcoolisée)* : le contenant est pris pour le contenu.

dans les réseaux sémantiques : liens méronymiques «partie-de» et hiérarchiques «sorte-de». L'archétype de la figure est la partie mise pour le tout :

1) *Beaucoup de personnes n'ont pas de toit* : -> *toit* remplace *maison*.

Contrairement à la figure précédente, dans la métonymie stricte les liens sont non hiérarchiques :

2) *Les bus sont en grève* : -> *bus* remplace *chauffeurs de bus*.

3) *La salle applaudit les comédiens* : -> *salle* remplace *spectateurs de la salle*.

Les travaux de linguistique comme Fontanier (publi. 1977), Dumarsais (publi. 1977) ou plus récemment Bonhomme (1987) et d'informatique linguistique (par exemple, Fass, 1991), portant sur ce trope sont fondés exclusivement sur des textes écrits, dans lesquels les métonymies rencontrées reflètent souvent un style particulier (journalistique, littéraire,...). Or, dans le langage oral les figures utilisées ne sont pas nécessairement celles employées à l'écrit et ne correspondent probablement pas aux mêmes visées communicatives. Il est donc nécessaire de disposer de corpus de dialogues en situation réelle de communication afin d'étudier les figures effectivement utilisées, les conditions de leur apparition et les motivations liées à leur utilisation. C'est dans cette optique qu'une ouverture du problème de la métonymie aux dialogues finalisés est envisagée.

## 1.— LA MÉTONYMIE EN SITUATION RÉELLE DE DIALOGUE

### 1.1 — Analyse descriptive de la métonymie dans un domaine professionnel

A partir de descriptions répétées de formes géométriques complexes et ambiguës, Krauss et Glucksberg (1977) ont mis en évidence que la formation d'un vocabulaire spécifique était liée à des simplifications progressives. Au fur et à mesure, les descriptions sont plus brèves pour se réduire dans certains cas à un mot; la simplification est essentiellement le fait de l'introduction de phénomènes d'ellipses et de métonymies. Un vocabulaire spécifique au groupe expérimental apparaît :

*Un verre à Martini renversé dans un support fer => un verre à Martini à l'envers => Le verre de Martini => Le Martini.*

*Cela ressemble à une tête de cheval => la tête de cheval => Le cheval.*

Ces résultats, obtenus dans un cadre expérimental, sont confirmés par l'analyse du vocabulaire employé par des professionnels de la bourse (Kittredge et Mel'cuk, 1983).

Pour notre étude, nous avons dans un premier temps recueilli des dialogues dans un service de radiologie conventionnelle et vasculaire d'un hôpital parisien. Ces données ont permis de mettre en évidence des raccourcis de langage et les termes spécifiques sur lesquels ils portent.

Le service compte une quinzaine de personnes :

— des radiologues, dont l'essentiel du travail est la supervision des examens et l'interprétation des clichés radiographiques. L'interprétation se fait oralement au moyen d'un dictaphone;

— des manipulateurs en radiologie, qui ont pour tâche de réaliser les examens;

— des brancardiers, qui sont principalement chargés du transport des patients (en brancard ou en chaise roulante) ou de l'accompagnement de ces derniers quand ils peuvent se déplacer. Ils fournissent également les clichés radio aux médecins demandeurs dans les autres services de l'établissement;

— des secrétaires médicales, qui ont une activité de «frappe»; elles transcrivent les comptes-rendus oraux des médecins. Elles s'occupent par ailleurs de l'accueil des patients et de la réception des demandes d'examens.

Outre les termes techniques propres à la radiologie sont apparues des expressions intéressantes employées par le personnel du service qui constituent autant de raccourcis de langage :

— des acronymes tel «A.S.P.» pour Abdomen Sans Préparation;

— des abréviations comme «*anapath*» pour anatomie pathologique ou «*artério*» pour artériographie;

— et pour ce qui nous concerne plus directement des figures elliptiques et/ou métonymiques :

4) «*Il y a encore deux poumons et un crâne qui vont descendre*».

(Patients qui vont bénéficier d'une radiographie) des poumons et du crâne.

5) «*Un transit est sur la table*».

(Un patient bénéficiant d'une radiographie du) transit (de l'intestin grêle) est sur la table (de radiographie).

6) «*Un brancard / une chaise*».

(Patient sur) brancard ou chaise roulante.

Les petites pathologies peuvent aussi servir à désigner le patient :

7) «*Amenez l'occlusion et l'épanchement pour un contrôle*».

Ces expressions sont toutefois plus rares, les radiologues étant tenus au secret médical concernant les dossiers des patients.

Par ailleurs, les exemples qui suivent mettent en évidence des phénomènes de polysémie au sein de cette figure :

8) «*Le brancardier est allé chercher l'os du deuxième*» (patient).

9) «*Affichez moi l'os*» (cliché).

10) «*Il y a un os qui vient de tomber*» (demande de radiographie).

11) «*Il faut réimprimer cet os, y a des fautes de frappe*» (compte-rendu de radio).

Ainsi, la partie anatomique objet de l'examen, mais également le type d'examen peuvent désigner l'une comme l'autre le patient, le cliché, le compte-rendu ou encore la demande d'examen.

Ces quelques exemples montrent la présence de figures régulières et réduites en nombre qui sont en grande partie à la base de la construction de ce langage professionnel. Nous proposons un certain nombre d'hypothèses portant sur les motivations des sujets liées à l'utilisation de ces figures et, au niveau cognitif, sur les représentations des connaissances du domaine.

## 1.2 — Étude par rapport au contexte d'énonciation et aux connaissances du domaine

Nous pensons que l'utilisation des métonymies répond à un principe d'économie et d'efficacité communicationnelle et que ces figures sont, au niveau cognitif, le reflet d'un mode d'organisation des connaissances particulier. Les problématiques soulevées ont principalement pour base les travaux en ergonomie du dialogue de P. Falzon (1989).

### 1.2.1 — *Économie et efficacité dans la communication*

Les langages professionnels se construisent progressivement, au cours de l'activité et sur la base d'un savoir commun. Il semblerait que ce soit essentiellement à partir de l'ellipse de langue et de la métonymie que s'élaborent ces langages spécialisés.

Les ellipses de langue, qui ne sont pas considérées classiquement comme des tropes, présentent des apparentements avec les métonymies et semblent également fréquentes dans le langage. Les liens entre les métonymies et les ellipses de langue sont encore à préciser. L'ellipse suivante montre la similitude existant entre les deux figures :

#### 12) *Le gaz est venu en ton absence (le releveur des compteurs à gaz)*

On s'est attaché ici à n'étudier que le vocabulaire apparaissant dans les dialogues entre les différentes personnes du service. En comparaison, il nous a été permis d'entendre des échanges entre les secrétaires et les patients : ce type de figures n'y apparaît pas. De même, une des secrétaires, arrivée depuis peu dans le service, avait tendance à expliciter l'ensemble des termes dans ses formulations quand elle s'adressait aux autres personnes du service.

Aussi, peut-on penser que ces deux types de figures se retrouvent dans des contextes conversationnels particuliers : elles seront utilisées lorsque les interlocuteurs en présence partagent un domaine de connaissance; le cas type étant le dialogue entre experts.

L'analyse du vocabulaire et des énoncés par rapport aux connaissances du domaine considéré a permis de mettre en évidence les termes implicites et ceux, au contraire, qui sont explicités. Il semble que le locuteur utilise les énoncés juste nécessaires et suffisants à la compréhension de son interlocuteur, étant donné le niveau d'expertise de ce dernier (ou tout au moins de l'idée que s'en fait le locuteur).

D'autre part, dans les échanges, seuls seront explicités les termes pertinents par rapport au contexte de travail. Les sujets éviteront les termes inutiles ou non pertinents; ils visent ainsi à réduire la longueur des énoncés par souci d'économie mais également d'efficacité dans la communication. De ce point de vue, un parallèle peut être fait avec les principes de pragmatique définis par Grice (1975) : maxime de quantité (suffisamment d'informations et pas plus que nécessaire), de manière (éviter les expressions ambiguës, obscures) et de brièveté. Les maximes de quantité et de brièveté sont respectées; concernant la manière, les expressions sont comprises car elles sont émises dans un milieu restreint.

### 1.2.2 — Métonymies, polysémie et structure des connaissances en mémoire

Les concepts et les relations sémantiques (i.e. les liens métonymiques) non pertinents par rapport à la situation sont éliminés. La figure suivante présente une représentation partielle du domaine de connaissance autour de la notion centrale de **radiographie**.

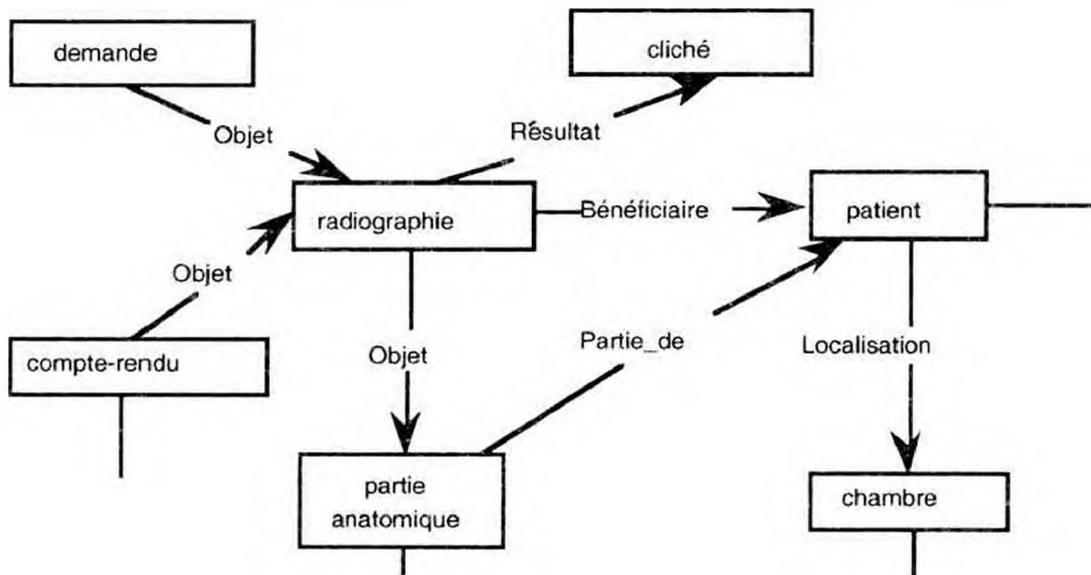
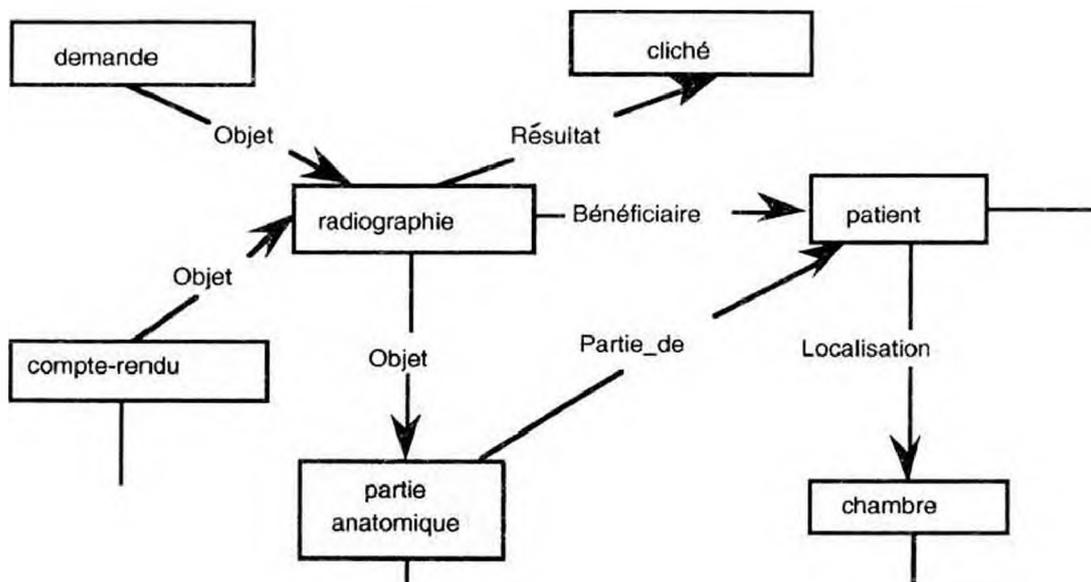
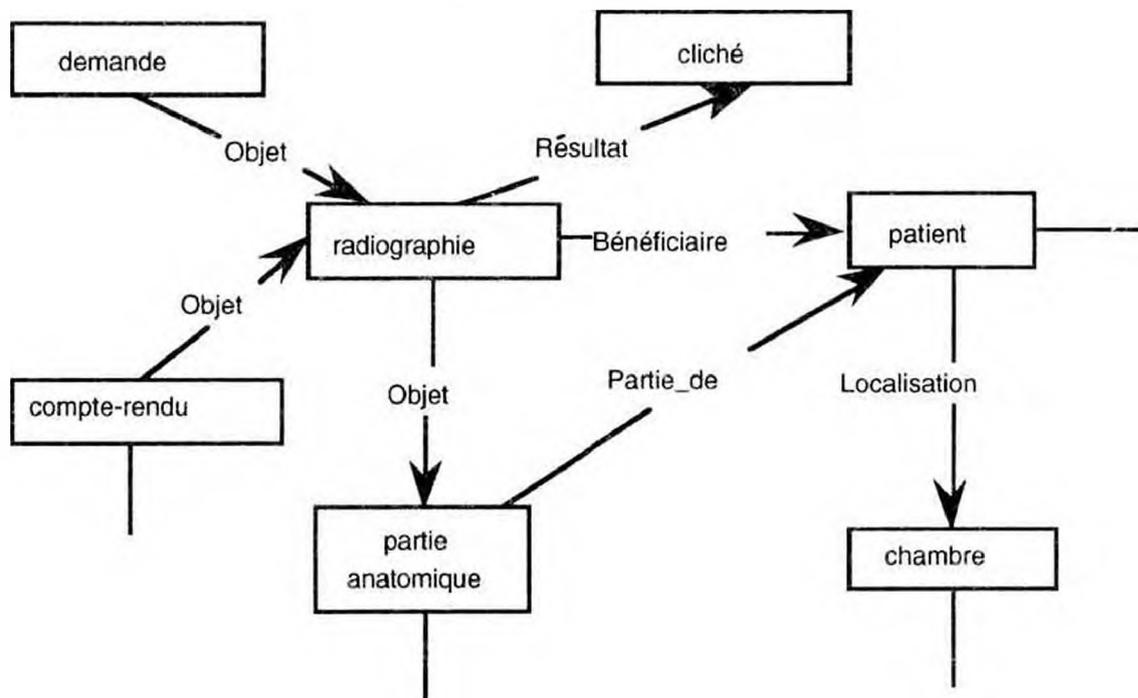


Figure 1 : Vue partielle de la représentation du domaine sous forme d'un réseau sémantique

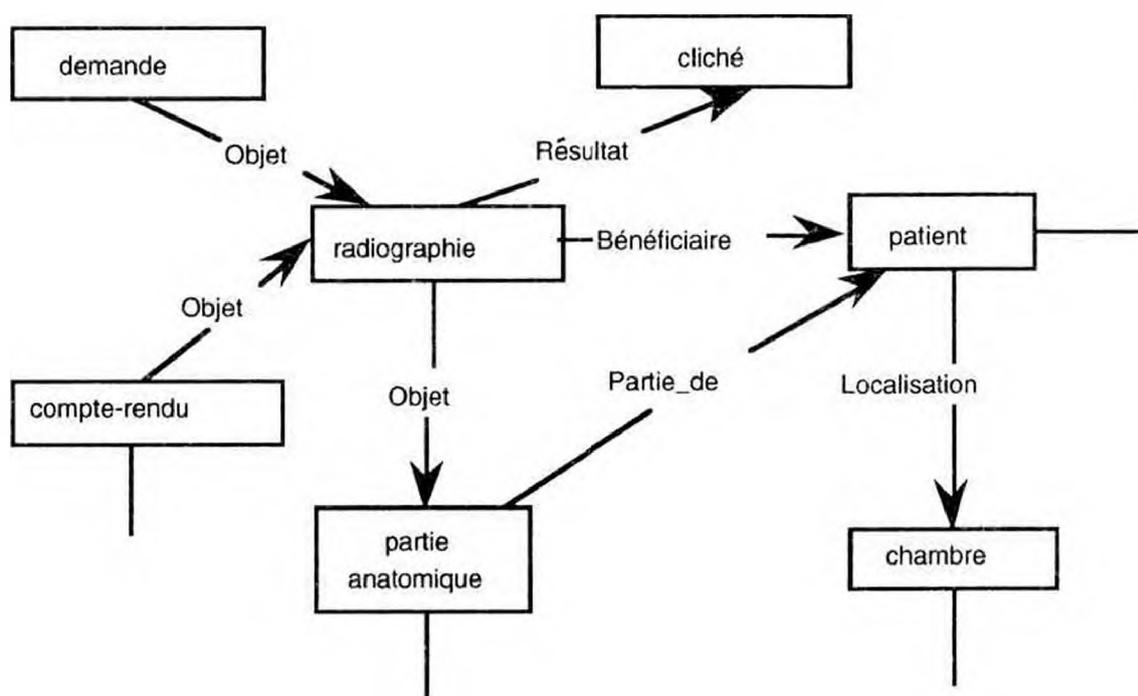
Ce schéma permet d'illustrer, à partir des exemples 8) à 11), la manière dont s'opère la construction de la figure métonymique et de la polysémie au sein de celle-ci. Les différentes significations du mot «*os*» relèvent d'une activation sélective d'une partie pertinente du schéma de connaissance (en grisé sur les figures qui suivent) en fonction du contexte énonciatif, redonnant ainsi un sens plein à l'expression employée.



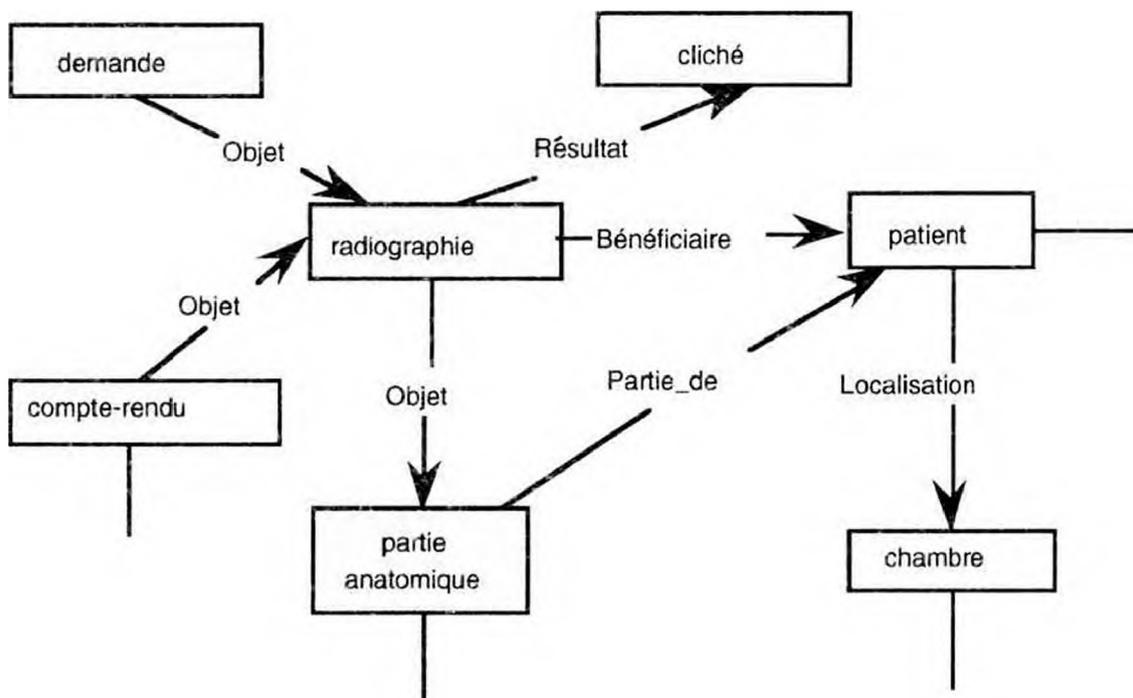
«Il faut réimprimer cet OS, il y a des fautes d'orthographe»  
 (compte-rendu d'une radiographie d'os)



«Le brancardier est allé chercher l'OS» (patient bénéficiant  
 d'une radiographie d'os)



«Affichez moi l'OS» (cliché radiographique d'os)



«Il y a un OS qui vient de tomber» (demande de radiographie d'os)

On constate que les expressions métonymiques rencontrées sont le plus souvent complexes parce qu'elles font appel à des substitutions multiples, que M. J. Reddy (1979) qualifie de **chaînes de métonymies**. Les sens des termes métonymiques utilisés sont liés aux connaissances du domaine d'une part et au contexte d'énonciation d'autre part. Le contexte situationnel et/ou le co-texte permettent de désambigüiser l'énoncé.

L'hypothèse psychologique générale est qu'il existe, dans les langages spécialisés, un lien entre la forme du discours et la structure des connaissances au niveau cognitif. Si ces raccourcis de langage sont des modes d'expression préférentiels chez les experts, c'est probablement parce qu'ils sont le reflet de l'organisation des connaissances en mémoire.

Ochanine (1978) a montré expérimentalement l'existence de ce qu'il a nommé **les images opératives**. Ce sont les représentations imagées que se font les opérateurs des objets de leur travail. Le contenu figuratif est déformé et réduit afin de ne conserver que les caractéristiques pertinentes de l'objet par rapport à la tâche en cours; c'est-à-dire finalement par rapport au contexte.

On suppose qu'il existe plus généralement des connaissances opératives simplifiées, dans lesquelles les concepts et les relations sémantiques non pertinents par rapport à la situation sont éliminés.

Les relations sémantiques en question sont précisément les liens métonymiques et elliptiques relevés. Un terme métonymique et/ou elliptique est défini comme le résultat d'un processus d'abstraction de connaissances. Dans les dialogues, les connaissances opératives se manifesteront par la fréquence d'utilisation des raccourcis de langage.

Des invariants apparaissent dans la construction des langages professionnels et par là même dans l'acquisition de connaissances expertes (i.e. structures métonymiques et elliptiques privilégiées et récurrentes). Les figures ne sont pas seulement des procédés de langue, elles correspondent à un mode d'appréhension de la réalité dans une perception globalisante (tout pour une partie) ou spécifiante (partie pour le tout).

Cette étude de terrain a mis en évidence dans les termes polysémiques des liens entre les différentes significations. Ces dernières ont pour origine une dérivation de sens à partir d'une représentation unique, en l'occurrence le schéma de connaissances du domaine de travail.

Notre étude est restreinte par la limitation de son objet; elle ne porte que sur une figure particulière (la métonymie), et un domaine de connaissances spécifique (la radiologie). Il existe des sens figurés qui ne relèvent pas de la métonymie et les termes de la langue ne portent sur aucun domaine particulier.

Ces phénomènes contextuels de sélection privilégiée d'éléments d'une représentation sémantique ont été établis expérimentalement à un niveau **micro-polysémique**. Barclay, Bransford, Franks, McCarrell et Nitsch (1974) montrent qu'un trait sémantique d'un mot peut être perçu par les sujets d'expérience comme saillant dans un contexte et pas dans un autre.

13) *«Lydia jouait délicieusement une sonate sur son piano»* .

14) *«Les déménageurs ont eu de la peine à apporter le piano»*.

Les traits sémantiques activés sont la musicalité de l'instrument en 13) et son poids et son encombrement pour la phrase 14). Cette modulation fine de la signification du mot *«piano»* est fonction du contexte phrastique dans lequel il est rencontré.

Schoen (1988) arrive à des résultats similaires en testant expérimentalement les effets du contexte phrastique sur le jugement de saillance de différentes caractéristiques sémantiques de 12 noms anglais. Denis et Le Ny (1986) observent ces effets aussi bien sur des mots que sur des représentations imagées.

Ces travaux en psycholinguistique sur la flexibilité sémantique ne sont pas sans rappeler le modèle de Ochanine présenté auparavant. L'opérativité (pour Ochanine) et la flexibilité sémantique (pour les psycholinguistes) sont selon nous un même et unique processus cognitif. Nous l'avons vu avec la polysémie dans les figures métonymiques, la flexibilité des représentations mise en évidence à un niveau très fin se retrouve également pour des acceptions plus larges. De plus, elle s'applique sur des représentations de nature différentes (mots, images). Nous envisageons l'opérativité, à l'instar de Spérandio (1982), comme un processus cognitif général qui permet d'appréhender la structure des connaissances en mémoire et leurs traitements.

Cette idée de représentation unique modifiée en contexte nous a amené à ne plus aborder la polysémie au sein d'une figure particulière, isolée du reste des faits de langue, mais dans son ensemble.

## 2. — MODÈLE D'ANALYSE DE LA POLYSÉMIE DES VERBES ET INTERPRÉTATION À L'AIDE DE GRAPHES CONCEPTUELS

Nous proposons un modèle général de la polysémie (Chibout et Masson, 1995) fondé sur la notion d'opérativité. En nous fondant sur cette hypothèse psychologique, nous travaillons à la réalisation d'un module informatique «expert» dans la détection et le traitement de sens figurés au niveau de la phrase. On s'appuiera sur le formalisme des *graphes conceptuels* pour la représentation des connaissances lexicales. En guise de préalable, nous allons voir que la métaphore, qui constitue la source principale de polysémie, présente des apparentements avec les phénomènes de flexibilité sémantique.

### 2.1 — Métaphore et polysémie

La métaphore, souvent opposée à la métonymie, est à l'origine de la plus grande partie des sens figurés. Nous pensons que l'opposition entre les deux tropes n'est pas pertinente au regard de leurs modes de construction. La métaphore lexicale se construit sur la sélection de traits sémantiques particuliers à un mot (cf. par exemple, Searle, 1979). Dans la phrase qui suit, les deux substantifs partagent le trait /sphérique/.

15) *La terre est une orange.*

Le trait caractéristique du comparant peut lui-même être utilisé dans son sens figuré; de sorte que l'expression est non littérale à double titre.

16) *Sally est un bloc de glace.*

Le trait /**froideur**/ de la glace (sens littéral) est assimilé au trait /**froideur**/ de caractère d'un individu (sens figuré).

Cette caractéristique peut d'ailleurs être réelle ou supposée (liée aux croyances) :

17) *Richard est un gorille .*

pour signifier un caractère violent, brutal (alors que le primatologue soulignera au contraire les mœurs paisibles de cet animal).

La même métaphore peut véhiculer des significations multiples en fonction du contexte, et ce par sélection de sèmes différents. En 17) *gorille* peut également signifier /**poilu**/ ou encore /**fort (physiquement)**/ qui sont autant de caractéristiques du primate.

Donc à la manière des phénomènes micropolysémiques et de la métonymie, nous avons pour la métaphore un mode de construction similaire dans la mesure où l'interprétation de la figure fait appel à la sélection d'éléments (ici un trait sémantique) d'une connaissance plus générale (ici l'ensemble des caractéristiques sémantiques qui définissent le terme métaphorique). Les significations multiples, quand elles existent, sont sémantiquement apparentées puisqu'elles relèvent d'une même représentation.

La différence entre métonymie et métaphore réside dans la nature des représentations mises en jeu; dans le cas de la métonymie, les connaissances débordent souvent du cadre strictement sémantique pour s'ouvrir à des schémas de connaissances plus larges (connaissances sur le monde). Mais, en ce qui concerne les traitements, le principe fondamental commun est la sélection d'éléments de connaissances particuliers à un terme en contexte.

## 2.2. — Le modèle proposé

La polysémie est vue comme une (re)construction en contexte à partir d'une représentation unique. Nous nous sommes intéressés plus spécifiquement à la résolution des incohérences sémantiques liées à la **polysémie des verbes**.

L'analyse des significations multiples associées aux verbes du français a permis de préciser les représentations sémantiques impliquées dans leur élaboration.

**18) *Les vagues couraient jusqu'aux rochers.***

En 18) nous sommes en présence d'une métaphore verbale interprétable en remplaçant *courir* par *se déplacer rapidement* qui sont des éléments de la description sémantique de ce verbe. Grossièrement, les traits sémantiques associés à *courir* sont :

**/se déplacer/ + /avec les pieds/ + /rapidement/...**

Nous avons systématisé ces descriptions sémantiques en précisant notamment la nature des relations liant le verbe à ses sèmes. Le mode de représentation proposé pour chaque item lexical est appelé **schéma conceptuel** et correspond à une définition dictionnaire élargie. Tout verbe est exprimé par son hyperonyme immédiat (l'événement central) et un certain nombre de cas sémantiques qui le spécifient. Outre les liens casuels habituels (agent, objet, moyen,...), quatre cas apparaissent essentiels pour la description fine d'un verbe : la **manière** dont l'événement est réalisé, la **méthode** par laquelle il se réalise, son **résultat** et le **but** intrinsèque à cet événement. À titre d'exemples, les verbes «couper» et «couvrir» sont porteurs des représentations sémantiques suivantes :

**couper** : *diviser* (hyperonyme immédiat) *objet solide* (objet) *en plusieurs parties* (résultat) *à l'aide d'un instrument tranchant* (moyen) *en traversant l'objet* (méthode).

**couvrir** : *garnir* (hyperonyme immédiat) *quelque chose* (support) *en disposant* (méthode) *une chose* (objet) *dessus pour cacher* (but) *ou protéger* (but).

Nous envisageons donc l'existence d'un sens (**proto**)typique associé à chaque verbe. Le prototype est le sens attesté le plus courant dans son emploi. Les sens dits «figurés» dérivent de ce schéma prototypique par le processus d'**opérativité** présenté ci-dessus.

Ces sens figurés sont liés à la structure sémantique du verbe proprement dite par récupération de l'hyperonyme immédiat (cas de l'énoncé 18)), mais également d'autres éléments du schéma conceptuel. Ces modes de résolution sont illustrés par les exemples qui suivent :

**19) *Le paysan coupa par le champs du voisin.***

dans lequel *couper* est remplacé par *traverser* (cas méthode dans le schéma conceptuel).

20) *La mère couvrait les fautes de son rejeton.*

21) *Cette assurance couvre le vol.*

En 20) et 21), l'interprétation de *couvrir* passe respectivement par la récupération des cas but *cacher* et *protéger* dans le schéma de ce verbe.

La simple sélection d'éléments de la représentation sémantique d'un item lexical n'est pas le seul mode de construction des significations secondes; mais elle reste cependant, comme nous allons le voir, une condition nécessaire à l'élaboration de ces dernières.

D'autres significations relèvent de métaphores conceptuelles (Lakoff et Johnson, 1985). Ainsi *abreuver quelqu'un de connaissances* s'interprète à partir de la métaphore L'ESPRIT EST UN CONTENANT. L'intérêt des métaphores conceptuelles réside dans leur généralité : elles s'appliquent non pas à un verbe particulier, mais à toute une classe de verbes. Ainsi, *gaver de connaissances, vider sa mémoire, se nourrir de lectures, dévorer un livre* seront interprétés en utilisant la métaphore L'ESPRIT EST UN CONTENANT. Ces modes d'interprétation ne remettent cependant pas en cause notre modèle de la polysémie. En effet, les structures de représentation sémantique proposées (hyperonyme + cas spécifiques) permettent de remonter jusqu'à des primitives sémantiques dont relèvent les items verbaux. Ces primitives correspondent pour la plupart à des verbes d'état : **Possesseur** (ex : *posséder*), **Localisation** (ex : *habiter*), **Contenant** (ex : *contenir*), **Support** (ex : *supporter*), **Patient** (ex : *souffrir*), **Temps** (ex : *exister*), **Expérienceur** (ex : *connaître*)...

Au plus haut niveau de la hiérarchie les verbes de processus et les verbes d'action dérivent de ces primitives d'états.

Ainsi les procès ont la forme **Devenir/Cesser + primitive**, exemples :

**Devenir contenant** (*se remplir, manger, boire*), **Devenir Temps** (*naître*), **Devenir Expérienceur** (*apprendre*)

**Cesser contenant** (*se vider*), **Cesser Temps** (*mourir*), **Cesser Expérienceur** (*oublier*)

Sur le même principe, les actions ont une structure de type **Faire Devenir/Faire Cesser + primitive**

**Faire Devenir contenant** (*remplir, abreuver*), **Faire Devenir Temps** (*créer, enfanter*), **Faire Devenir Expérimenteur** (*faire apprendre, enseigner*)

**Faire Cesser Contenant** (*vider*), **Faire Cesser Temps** (*interrompre, tuer*), **Faire Cesser Expérimenteur** (*faire oublier*)

Ces éléments de base nous ont permis de faire émerger des grandes classes sémantiques, les concepts verbaux étant finement catégorisés par l'intermédiaire de leur schéma conceptuel. Les primitives sémantiques jouent un rôle fondamental pour les sens figurés de type «métaphore conceptuelle». *Abreuver quelqu'un de connaissances* pourra s'interpréter à partir de la substitution de la primitive *Contenant* par *Expérimenteur*.

L'application de métaphores conceptuelles ne se fait pas au hasard; c'est parce qu'un item verbal est, de par sa structure sémantique, rattaché à une primitive que telle ou telle métaphore s'applique. Il n'y a donc pas de «génération spontanée» de sens; le verbe est porteur dans son schéma conceptuel de «points d'entrée» vers d'autres significations. Les exemples suivants fournissent l'illustration de l'analogie classique de l'espace au temps : 22) *couper la parole* (*INTERROMPRE*), 23) *briser une conversation* (*INTERROMPRE BRUSQUEMENT*), 24) *entrecouper ses phrases de sanglots* (*INTERROMPRE FRÉQUEMMENT*), 25) *hacher un discours* (*INTERROMPRE DE FAÇON RÉPÉTÉE*)

Les verbes *couper*, *briser*, *entrecouper* et *hacher* sont tous associés à la classe d'action **Faire Cesser Espace** (continuité spatiale). L'interprétation des métaphores passe via la substitution de cette classe sémantique par la classe **Faire Cesser Temps** (continuité temporelle). On notera par ailleurs que le cas manière porté par le schéma conceptuel de chacun des trois derniers verbes est sélectionné pour l'interprétation de l'expression, ce qui apporte une nuance fine entre ces verbes.

Tous ces modes d'interprétation représentent autant de «glissements» de sens à partir de la structure sémantique unique qui définit le verbe. Les modes de résolution s'appuient sur la simple sélection d'éléments pertinents de cette représentation ou d'une sélection de ces éléments et d'une inférence à partir d'un de ceux-ci (métaphore conceptuelle).

### 2.3 — Pourquoi les graphes conceptuels ?

En examinant les formalismes les plus couramment utilisés actuellement pour la représentation de connaissances sémantiques, on trouve essentiellement deux paradigmes : les logiques terminologiques et les graphes conceptuels. L'avantage des logiques terminologiques réside essentiellement sur l'existence d'un classifieur automatique (défini à l'origine pour KL-ONE, et repris dans la définition d'un langage tel que CLASSIC, défini par Brachman *et al.*, 1991). Ces langages servent essentiellement pour la constitution d'une base de connaissances, mais limitent (volontairement) ce qui peut être exprimé, et donc traduit, par un utilisateur.

Le formalisme des graphes conceptuels a été conçu par Sowa (1984 et 1991) dès l'origine pour servir d'intermédiaire entre le langage naturel et la logique du premier ordre, avec des extensions pour représenter des phénomènes que le calcul des prédicats interdit.

Définir une base de connaissances en graphes conceptuels consiste à définir une base canonique contenant :

- une hiérarchie des types de concepts,
- un ensemble de référents dans le monde,
- une relation de conformité entre types de concepts et référents, permettant de savoir si un référent est *compatible* avec un type de concept (si le référent *Tom* est utilisé pour un homme, il ne sera pas compatible avec le type chien, par exemple),
- un ensemble hiérarchiquement structuré de types de relations,
- un ensemble de graphes conceptuels (dits canoniques) qui indiquent pour les types de concepts leurs conditions nécessaires d'emploi.

Un graphe conceptuel est un graphe bipartite constitué de nœuds concepts et de nœuds relations, chaque concept étant constitué d'un type et d'un référent. Dans les graphes canoniques, les concepts sont génériques, et contiennent donc un référent générique (noté \*).

À partir de cette base, toute phrase exprimée en français peut être traduite par un graphe conceptuel, dont la validité est vérifiée par le fait qu'il peut être dérivée de la base de départ à l'aide d'une combinaison des quatre opérations de base définie dans le formalisme. Ces opérations sont la copie de graphe, la restriction de

types (on modifie un des concepts du graphe en remplaçant le type par un sous-type, tout en vérifiant que la conformité du référent est respectée) ou de référents (on modifie un des concepts du graphe en remplaçant un référent générique par un référent individuel, en vérifiant la conformité de ce référent), la simplification (en supprimant des éléments redondants dans un graphe) et la jointure de graphes (sur un concept commun, on joint les deux graphes de départ en un seul graphe).

L'ensemble de ces outils, y compris la possibilité de représenter des référents non encore résolus, ou des liens de coréférence entre deux concepts, rendent ce formalisme très puissant pour la représentation sémantique de la langue, ce qui nous a incité à le retenir pour le système de traitement automatique de la langue développé dans notre groupe. En particulier, les contraintes sémantiques exprimées classiquement dans les grammaires de cas, s'expriment naturellement sous forme de graphes canoniques, en ajoutant les contraintes sur les concepts pouvant remplir les cas, obligatoires ou facultatifs des prédicats.

Les graphes canoniques (ou de contrainte) ne suffisent pas pour construire la hiérarchie des types de concepts : ils n'ont qu'un rôle de *conditions nécessaires* qu'un type de concept doit remplir.

Pour préciser le sens des types de concepts, et éventuellement en introduire de nouveaux, il est possible de définir les types de concepts ou de relations, pour en donner une description plus précise, que l'ensemble des contraintes héritées de la hiérarchie des types. Cette définition peut être exploitée par les opérations d'expansion et de contraction, permettant respectivement de remplacer dans un graphe un concept par sa définition, ou de reconnaître la présence d'un type non mentionné dans un graphe, en identifiant sa définition.

Les définitions permettent de donner une description en intension des concepts. Ainsi, elles peuvent être très utiles pour aider à construire la hiérarchie des types (qu'il s'agisse des concepts ou des relations), bien que sans classifieur automatique, car les définitions ne sont pas obligatoires. En revanche un type pour lequel une définition n'a pas été donnée n'est pas obligatoirement déclaré primitif (comme c'est le cas avec CLASSIC).

Les graphes de définitions constituent la charpente de notre modèle tant au niveau de la représentation des connaissances que dans le traitement de la polysémie.

D'une part, ils implémentent les schémas conceptuels des verbes. Le schéma conceptuel de «couper» donné plus haut se réécrit comme suit :

**Type Couper (\*x) is**

**[Diviser: \*x]**

- > (Agent) -> [Anime \*y],
- > (Objet) -> [[EntiteConcrete: \*z] -> (Caract) -> [Solide]],
- > (Moyen) -> [[EntiteConcrete] -> (Caract) -> [Tranchant]],
- > (Methode) -> [[Traverser]
- > (Source) -> [Anime: \*y]],
- > (Objet) -> [[EntiteConcrete: \*z] -> (Caract) -> [Solide]]

D'autre part, la structuration de ces graphes sous la forme **hyperonyme + traits spécifiques** donne naturellement la hiérarchie des types de concepts, et ce jusqu'aux primitives.

<b>Concept</b>	<b>Type</b>	<b>Couper</b>	<b>&lt; Diviser</b>
Concept	Type	Diviser	< Separer
Concept	Type	Separer	< Faire CesserEspace
etc.			

Concernant l'interprétation de la polysémie, ces graphes de définition associés au treillis sont bien adaptés au traitement des sens multiples véhiculés par les verbes. Il s'agit, en fonction des contraintes sémantiques du contexte, de sélectionner la partie pertinente du graphe de définition (exemples 19) à 21)) ou à partir de la sélection de l'hyperonyme immédiat (par exemple en 22), pour couper, **Diviser**) de remonter jusqu'à la classe sémantique reliée à une primitive et d'appliquer une métaphore conceptuelle par substitution de primitive (**FaireCesserEspace** -> **FaireCesserTemps**).

Le graphe canonique associé au verbe permet de détecter un éventuel sens figuré :

*Canonical graph for Couvrir is [Couvrir]*

- > (Agent) -> [Humain]
- > (Support) -> [EntiteConcrete]
- > (Objet) -> [EntiteConcrete],

Par exemple dans les phrases 20) et 21), le verbe *couvrir* présente une incohérence sémantique dans la mesure où il attend comme cas support une **entité concrète** et *vol* d'une part et *fautes* d'autres part sont de type **entité abstraite**. Ces violations de contraintes sémantiques vont servir de «déclencheurs» à une interprétation figurée du verbe.

## CONCLUSION

Une étude de terrain a fait apparaître que dans un langage professionnel, les sens multiples véhiculés par un mot dépendent fortement de la représentation des connaissances du domaine. Nous avons proposé un modèle cognitif et linguistique de la polysémie en nous appuyant sur cette étude, sur des résultats expérimentaux en psycholinguistique et enfin sur l'analyse des «comportements» polysémiques des verbes du français. Les sens figurés sont contenus dans la description sémantique fine du verbe ou inférables à partir de cette description.

Enfin, nous avons montré dans la perspective d'un traitement automatique, dans quelle mesure les graphes conceptuels sont des outils adaptés à la modélisation et à l'interprétation des sens figurés.

Le traitement informatique de la polysémie est un élément central de la robustesse des systèmes informatiques; c'est-à-dire de la capacité de ceux-ci à comprendre toutes les expressions employées par les utilisateurs. Les insuffisances dans ce domaine constituent une des difficultés majeures qui freine l'utilisation du langage naturel dans la communication homme-machine. Le développement récent de l'informatique «grand public» impose une grande souplesse des systèmes. L'intérêt est ainsi évident pour l'interrogation des bases de données en langage naturel ou encore pour les interfaces en langue naturelle des systèmes d'E.I.A.O.

## RÉFÉRENCES

- BARCLAY, J.R., BRANSFORD, J.D., FRANKS, J.J., MACCARELL, N.S. & K. NITSCH (1974) : «Comprehension and semantic flexibility», *Journal of Verbal Learning & Verbal Behavior*, 13, pp. 471-481.
- BONHOMME, M. (1987) : *Linguistique de la métonymie*, coll. «Sciences pour la communication», Berne. Peter Lang.
- BRACHMAN, R., MCGUINNESS, D., PATEL-SCHNEIDER, P., BORGIDA, A. & L. RESNICK (1991). : «Living with Classic : When and How to Use a KL-ONE-Like Language», *Principles of Semantic Networks*, May 1991, Morgan Kaufman, pp. 401-456.
- CHIBOUT, K. et N. MASSON (1995) : «Un réseau lexico-sémantique de verbes construit à partir du dictionnaire pour le traitement informatique du français», *Lexicomatique et dictionnaires*, Actes des IV<sup>es</sup> Journées scientifiques du réseau LTT, Lyon, Septembre 1995, Beyrouth, FMA et AUPELF-UREF.
- DENIS, M. & J.F. LE NY (1986) : «Centering on figurative features during the comprehension of sentences describing scenes», *Psychological Research*, 48, pp. 145-152.
- DUMARSAIS (publi. 1977) : *Traité des tropes*, Éditions Le nouveau commerce, 1977.
- FALZON, P. (1989) : *Ergonomie cognitive du dialogue*, collection «sciences et technologies de la connaissance», Presses Universitaires de Grenoble, 173 p.
- FASS, D. (1991) : «MET\* : A method for discriminating metonymy from metaphor by computer», *Computational Linguistics*, vol. 17, n° 1, pp. 49-88.
- FONTANIER, P. (publi. 1977) : *Les figures du discours*, Éditions Flammarion.
- GRICE, P. (1975) : «Logic and conversation», P. Cole & J.C. Morgan (Eds), *Syntax and Semantics vol. 3 : Speech Acts*, Academic Press.
- KITTREDGE, R. et I. MEL'CUK (1983) : «Toward a computable model of Meaning-Text relations within a natural sublanguage», Actes des 8<sup>th</sup> International Joint Conference on Artificial Intelligence (I.J.C.A.I.), Karlsruhe.

- KRAUSS, R.M. & S. GLUCKSBERG (1977) : «Social and nonsocial speech», *Scientific American*, 236 (2), pp. 100-105.
- LAKOFF, G. et M. JOHNSON (1985) : *Les métaphores dans la vie quotidienne*, coll. «Propositions», les éditions de Minuit.
- OCHANINE, D. (1978) : «Le rôle des images opératives dans la régulation des activités de travail», *Psychologie et Éducation*, 3, pp. 63-79.
- REDDY, M. J. (1979) : «The conduit Metaphor- a Case of Frame Conflict in our Language about Language», Andrew Ortony (Ed.), *Metaphor and Thought*, Cambridge University Press, pp. 284-324.
- SCHOEN, L.M. (1988) : «Semantic flexibility and core meaning», *Journal of Psycholinguistic Research*, 17, pp. 113-123.
- SEARLE, J.R. (1979) : «Metaphor», Andrew Ortony (Ed.), *Metaphor and Thought.*, Cambridge University Press, pp. 284-324.
- SOWA, J. (1984) : *Conceptual Structures : processing in mind and machine*, Reading, Massachusetts, Addison-Wesley.
- SOWA, J. (1991) : «Toward the expressive power of natural language», *Principles of Semantic Networks*, May 1991, Morgan Kaufman, pp. 157-189.
- SPÉRANDIO, J.C. (1982) : «Tout traitement de l'information est opératif», *Actes du séminaire sur «l'image opérative»*, Université de Paris I, 2 juin 1981, pp. 91-98.
- VAPILLON, J., BRIFFAULT, X., SABAH, G. & K. CHIBOUT (1997) : «An object oriented linguistic engineering using LFG and CG», *ACL/EACL Workshop Computational Environments for Grammar Development and Linguistic Engineering*, Madrid, 1997.

# Bases textuelles et bases de connaissances sur le web : le système ikarus <sup>(1)</sup>

Douglas SKUCE, Ingrid MEYER  
et Judy KAVANAGH

*Université d'Ottawa, Ottawa, Canada*

## 1. — INTRODUCTION

Dans cet article, nous décrivons les derniers outils produits dans le cadre du projet *COGNITERM*, au Laboratoire d'intelligence artificielle de l'Université d'Ottawa. Le but de ce projet est de développer des méthodes et des outils qui aideront un terminologue à mieux comprendre, et à mieux représenter, les aspects notionnels des termes <sup>(2)</sup>.

L'article portera surtout sur le système *IKARUS* (*Intelligent Knowledge Acquisition and Retrieval, Universal System*), un logiciel Web qui aide l'utilisateur à construire des bases de connaissances terminologiques à partir de corpus électroniques. Par *base de connaissances terminologiques*, nous désignons une sorte d'hybride entre une banque de terminologie classique, d'une part, et une base de connaissances (comme on la connaît en intelligence artificielle),

---

1) Cette étude a été rendue possible grâce aux subventions des organismes suivants : le Conseil de recherches en sciences naturelles et en génie, le Conseil de recherches en sciences humaines, et l'entreprise Mitel, Inc.

2) Pour un aperçu plus détaillé de notre approche, voir Meyer *et al*, 1997.

d'autre part (3). Autrement dit, une base de connaissances terminologiques contiendrait non seulement toutes les données *linguistiques* qu'on trouverait dans une banque de terminologie, mais également des renseignements précis et structurés sur les *notions* désignées par les termes. IKARUS est le descendant du système *CODE* (*Conceptually Oriented Design/Description Environment*), qui a été utilisé dans plusieurs applications industrielles et universitaires, aussi bien pour la terminologie (Meyer *et al.*, 1992; Meyer et MacHaffie, 1993) que pour le génie cognitif (Skuce et Lethbridge, 1995).

CODE, programmé en langage Smalltalk, a été développé à la fin des années 80. Malgré les points forts du système (interface graphique très développée, portabilité), il était technologiquement dépassé au milieu des années 90, et nous avons donc procédé au développement d'un nouveau système, IKARUS, en 1995-96. Dans ce dernier système, nous voulions garder les points forts de CODE, tout en corrigeant quelques-unes des faiblesses de ce dernier, notamment la difficulté de le rendre accessible par le Web, l'absence d'un analyseur de corpus, l'impossibilité de communiquer avec d'autres logiciels, et une certaine complexité dans l'emploi (ressentie surtout par des non-informaticiens, comme nos utilisateurs en terminologie).

## 2. — LE SYSTÈME IKARUS

Comme c'était le cas pour CODE, IKARUS est un outil polyvalent, c'est-à-dire un outil qui pourrait aider toute personne à acquérir et à organiser des connaissances spécialisées à partir d'un corpus électronique. Le terminologue, qui s'occupe de l'acquisition et de l'organisation de connaissances notionnelles et linguistiques, est donc un utilisateur potentiel important. Compatible avec HTML, IKARUS est accessible avec un fureteur Web (4) par de nombreux utilisateurs en même temps. Le système est donc idéal pour le travail mené en collaboration aussi bien pour les efforts de standardisation entrepris à l'échelle internationale et pour la collaboration entre terminologues et experts. L'importance du Web pour la terminologie a déjà été soulignée par de nombreux chercheurs (Cf. Nuopponen, 1996; Schweighofer et Scheithauer, 1996; Jacquin et Lisouet, 1996).

---

3) Pour plus de détails sur notre «vision» d'une base de connaissances terminologiques, voir Meyer *et al.*, 1997. Pour d'autres visions très intéressantes, voir Toussaint *et al.*, 1997; Séguéla *et al.*, 1997; Condamines et Amsili, 1995; Biébow et Szulman, 1997.

4) Une version publique d'IKARUS est accessible au [http : //www. csi. uottawa. ca/~kavanagh/Ikarus.html](http://www.csi.uottawa.ca/~kavanagh/Ikarus.html). Le système est aussi décrit dans Skuce (à paraître).

Sur le plan technique, notons qu'IKARUS emmagasine ses données dans une base de données où chaque notion constitue une entrée indépendante (ou une série d'entrées). La programmation a été faite en Perl sur une plate-forme SUN, avec dbm comme gestionnaire de base de données.

IKARUS comprend deux composantes principales. D'une part, le *Text Analyzer* (Analyseur de textes) aide l'utilisateur à repérer les termes et phrasèmes possibles, et à les analyser au moyen d'un concordancier. D'autre part, le *Knowledge Base Management System* (Gestionnaire de base de connaissances) organise les concepts dans différents types de hiérarchies et explicite les relations notionnelles. Puisque IKARUS tourne sur le Web, la base de connaissances peut contenir des adresses Web, ce qui permet à l'utilisateur d'accéder directement à un site.

Dans ce qui suit, nous donnerons un aperçu général de chacune des deux composantes, ainsi que quelques exemples.

## 2.1 — L'analyseur de textes

Les analyseurs de textes sont déjà bien connus en lexicographie (voir par ex. Atkins, 1994). Et, parce que le terminologue se fie encore moins à ses propres intuitions linguistiques que le lexicographe (car, souvent, il n'est pas un vrai «expert» dans le domaine), les analyseurs de textes prennent d'autant plus d'importance. Heureusement, la plupart des fonctions que les lexicographes trouvent utiles — par ex. concordanciers, repérage de collocations, analyse de fréquence — s'appliquent également à la terminologie. Par contre, comme nous l'avons ailleurs (Meyer et Mackintosh, 1997), la terminologie, parce qu'elle diffère de la lexicographie dans sa nature même, a besoin de quelques fonctions supplémentaires. Contrairement à la lexicographie, par exemple, la terminologie implique le *repérage* d'unités lexicales, et non pas seulement leur analyse et description. En plus, en ce qui concerne l'analyse, la terminologie a besoin d'un concordancier qui montre non pas *tous* les contextes dans lesquels un terme apparaît, mais plutôt les contextes *riches en connaissances*. Ceci parce que l'acquisition des connaissances propres au domaine constitue le fondement même de la terminologie.

Pour ce qui est du repérage de termes, l'Analyseur de textes peut identifier des termes composés, ainsi que plusieurs types de collocations. Le fait que cet outil peut fonctionner avec du texte

étiqueté aide dans le repérage (par ex., pour trouver les verbes normalement employés après un substantif, etc.).

Pour ce qui est de l'analyse de termes, l'Analyseur de textes comprend une fonction appelée «*knowledge patterns*», qui filtre les résultats d'une concordance pour montrer seulement les contextes les plus riches en connaissances. Par *contextes riches en connaissances*, nous désignons les contextes qui montrent certaines relations notionnelles importantes, notamment générique-spécifique, partie-tout, fonction, synonymie, etc. Bien que cette fonction donne une quantité non négligeable de «bruit», nos recherches préliminaires dans ce sens (Davidson, 1997) indiquent qu'elle peut être très utile pour les termes apparaissant fréquemment (évidemment, un terminologue préférerait regarder, par ex., 50 contextes dont 25 % contiennent du bruit, plutôt que 500 contextes non filtrés).

La figure 1 (p. 135) montre un exemple des contextes retirés pour le terme anglais *compost*, et pour la relation notionnelle *partie-tout*. Sans filtrage, *compost* aurait produit 500 contextes. Avec filtrage, ce chiffre peut être ramené à 50 environ.

Ce type de filtrage conceptuel aide le terminologue à mieux comprendre les concepts qu'il étudie, en montrant leur place dans un réseau notionnel. Cette fonction est d'autant plus importante si l'on vise une base de connaissances terminologiques. Dans ce qui suit, nous décrivons brièvement le gestionnaire de la base de connaissances.

## 2.2 — Le gestionnaire de bases de connaissances

Traditionnellement, les terminologues ont rassemblé les résultats de leur analyse textuelle dans une base de données. IKARUS, par contre (comme son prédécesseur, CODE), se sert du modèle de la base de connaissances, modèle mis au point par des chercheurs en Intelligence artificielle, notamment en génie cognitif. En effet, les ressemblances entre terminologie et génie cognitif ont déjà fait l'objet de plusieurs études (par ex. Ahmad *et al*, 1989; Skuce et Meyer, 1991).

Dans IKARUS, le terminologue peut organiser les concepts dans divers types de hiérarchies : générique/spécifique (une voiture est un type de véhicule), partie/tout (la tête est une partie du corps), et domaine/sous-domaine (l'algèbre est un sous-domaine des mathématiques). À chaque concept, le terminographe peut associer les

caractéristiques de son choix, en utilisant une structure de «frames» semblable à ce qu'on retrouve dans les systèmes classiques comme Cyc (Lenat and Guha 1990). Dans la figure 2 (p. 136), on voit à droite des caractéristiques associées au concept de «*abstract class*», qui fait partie de notre base de connaissances sur le langage JAVA. Comme on le voit sur la Figure, si soit le nom d'une caractéristique, soit sa valeur (par ex. *superclass* : a class) sont également des concepts dans la base, des liens hypertextuels sont automatiquement établis. Sur la partie gauche de la figure, on peut voir une partie de la hiérarchie générique-spécifique contenue dans la même base de connaissances. L'héritage de caractéristiques (y compris l'héritage multiple) s'applique automatiquement à tout concept inscrit dans une hiérarchie générique-spécifique.

Les caractéristiques indiquées pour un concept sont définies par le créateur de la base. Dans l'exemple précédent (JAVA), l'accent a été mis sur des caractéristiques *conceptuelles*. Par contre, on peut également indiquer des caractéristiques *linguistiques*, comme l'a fait à Montréal notre collègue Marie-Claude l'Homme (L'Homme, 1997), pour sa hiérarchie des verbes français dans le domaine de l'informatique grand public. Évidemment, une combinaison de caractéristiques conceptuelles et linguistiques est également possible (figure 3, p. 137).

Une base de connaissances peut être visualisée de façon graphique aussi bien que textuelle. Dans la figure 4 (p. 138), on peut voir un graphe représentant une hiérarchie de concepts sur JAVA.

### 2.3 — Liens entre bases de connaissances et bases textuelles

Étant donné qu'IKARUS est compatible avec HTML, on peut incorporer des liens directs entre une base de connaissances et des sites Web. Ceci permet de lier l'analyse linguistique et conceptuelle d'un document — c'est-à-dire, une entrée dans la base — aux documents sources, si ces derniers sont disponibles sur le Web. Pour un accès plus rapide aux documents, le système permet également de les sauvegarder dans un répertoire UNIX, qui est construit automatiquement et de façon hiérarchique. La structure hiérarchique de la base textuelle peut donc correspondre à la structure hiérarchique de la base de connaissances.

Outre ces liens entre bases de connaissances et textes, des liens entre plusieurs bases de connaissances (par ex. dans des domaines liés) sont également possibles.

### 3. — DÉVELOPPEMENTS RÉCENTS : LE SYSTÈME DOCKMAN

Malgré tous ses avantages, le système IKARUS présente plusieurs faiblesses. Tout d'abord, il est assez lent, étant donné que tous les calculs sont faits sur le serveur, et que l'analyseur de textes n'utilise pas d'indexation. De plus, l'interface est limitée par les contraintes du HTML. Les possibilités d'héritage sont également plus limitées que dans CODE. Finalement, l'analyseur de textes est borné par le fait que l'analyse grammaticale se limite à un simple étiquetage.

Pour remédier à ces faiblesses, nous travaillons actuellement (depuis 1997) à un successeur pour IKARUS — le système DockMan (*Document-Based Knowledge Manager*). DockMan est presque opérationnel et sera testé dans les mois à venir dans le cadre d'un projet avec l'entreprise Nortel Inc. (télécommunications) comme aide pour répondre à des questions portant sur des textes techniques de plusieurs gigaoctets.

DockMan offrira les avantages suivants : 1) une interface améliorée, grâce à la programmation en JAVA; 2) l'utilisation d'un vrai gestionnaire de bases de données (par ex. MS Access); 3) un accès plus rapide, grâce à la programmation-client en JAVA, et aussi grâce à l'utilisation de la base de données; 4) l'utilisation d'un parseur de surface (*shallow parser*), pour l'instant le système ENGCG. La figure 5 (p. 139) montre le potentiel de ce dernier pour l'acquisition de connaissances. Dans cette illustration, le système a rangé tous les contextes dans lesquels apparaît le terme *applet*, et ceci dans l'ordre suivant : modificateur du sujet, sujet, verbe, modificateur de l'objet, et objet. Ceci permet d'avoir des renseignements conceptuels très importants sur le concept, notamment en ce qui concerne la *fonction*.

### 4. — CONCLUSION

IKARUS et son successeur essaient de fournir au terminologue la possibilité d'achever un certain nombre de tâches qui sont propres à la terminologie, et qui n'ont pas été possibles auparavant à l'aide des outils conçus pour la lexicographie. Ces tâches comprennent : l'analyse conceptuelle de corpus; la représentation du réseau notionnel qui soutient un domaine; et finalement, au moyen de la magie du Web, la collaboration entre terminologues et experts n'importe où dans le monde. Nous espérons que les outils de ce genre contribueront à une nouvelle génération de dictionnaires terminologiques qui seront non seulement créés *plus vite* que dans le passé, mais qui seront également d'une *qualité supérieure* grâce à la richesse de données conceptuelles qui complètent les données linguistiques.

Figure 1

1. Municipal solid waste ( MSW ) compost or any **compost** that **contains** sewage sludge must also meet the requirements for unrestricted grade.
2. Most **compost** also **contains** as much as 2 percent calcium .
3. If the **compost contains** animal manure , the soluble salts level can be relatively high (4.0 mmhos/cm or higher when measuring a solution of two parts water and one part compost)
4. Yard waste **compost** can be applied at a higher rate because it **contains** fewer nutrients .
5. Mixes **containing** approximately 10 percent ( by volume ) of turkey litter **compost** initially **contain** higher amounts of nutrients and particularly higher phosphate levels than currently available potting mixes .
6. Soils that **contain** lots of decomposed organic matter , like **compost** , can store more food for the plants and release it to them as they need it .
7. Nematodes: These tiny, cylindrical, often transparent microscopic worms are the most abundant of the physical decomposers - a handful of decaying **compost contains** several million .
8. None of the materials **in the compost** should be identifiable .
9. Information on the amounts of metals such as zinc, iron, lead, nickel, and cadmium **in the compost** should also be studied.
10. For container use , be cautious and use approximately 10 percent **compost** by volume **in** pine bark mixes.
11. Up to 30% of the garbage we throw out each week can go **in the compost** pile .
12. If your **compost** mix is too low **in** nitrogen , it will not heat up .
13. The nutrients stored **in compost** depend on the richness and variety of its ingredients , and on its exposure to harsh weather .
14. There is a complex food web at work **in a compost** pile , representing a pyramid with primary, secondary , and tertiary level consumers .
15. Mites : Mites are the second most common invertebrate found **in compost** .

Figure 1 : Sortie de la fonction «knowledge patterns»

## Edit Menu Knowledge Base Java User Judy

Pick the subject:	Pick an operation:	Execute	Main Menu	Help	Start Hierarchy Viewer
- Select subject -	— Selected an operation —	Start Document System		Start Text Analyzer	

<ul style="list-style-type: none"> <li><u>Object class</u></li> <li><u>System class</u></li> <li><u>Thread class</u></li> <li><u>abstract class</u></li> <li><u>public class</u></li> <li><u>subclass</u></li> <li><u>superclass</u></li> <li><u>unsorted concepts</u></li> <li><u>applet</u></li> <li><u>behaviour</u></li> <li><u>class body</u></li> <li><u>class declaration</u></li> <li><u>encapsulation</u></li> <li><u>exception</u></li> <li><u>inheritance</u></li> <li><u>interface</u></li> <li><u>message</u></li> <li><u>method declaration</u></li> <li><u>state</u></li> <li><u>variable</u></li> <li><u>member variable</u></li> <li><u>class variable</u></li> <li><u>instance variable</u></li> </ul>	<p><b>abstract class</b> Knowledge Base Java</p> <hr/> <p><b>System statements</b></p> <p>~superconcept: <u>class</u></p> <p>~subconcept:</p> <hr/> <p><b>Basic Characteristics</b></p> <p><i>instances:</i> The <u>objects</u> are called the <u>instances</u> of the <u>class</u>.</p> <p><i>member variables:</i> A Java <u>class</u> can contain two different types of <u>members</u>: <u>instance members</u> and <u>class members</u> (Java Tutorial, Line 2907)</p> <p><i>methods:</i> A <u>class</u> has both <u>class methods</u> and <u>instance methods</u></p> <p><i>parts:</i> An <u>abstract class</u> may contain <u>abstract methods</u> (methods with no implementation) (Java Tutorial, Line 3809)</p> <p><i>subclasses:</i> 0 or more subclasses</p> <p><i>superclass:</i> a <u>class</u> (or nil in the case of the <u>Object class</u>)</p> <hr/> <p><b>Linguistic Information</b></p> <p>~context1: Programmers can implement superclasses called <u>abstract classes</u> that define "generic" behaviors (Java Tutorial, Line 2621)</p> <p>~context2: some classes are written for the sole purpose of being subclassed (and are not intended to ever be instantiated). These classes are called <u>abstract classes</u> and often contain <u>abstract methods</u>. (Java Tutorial, Line 4740)</p> <p>~context3:</p> <p>~definition1: An <u>abstract class</u> is a <u>class</u> that can only be subclassed - it cannot be instantiated (Java Tutorial, Line 3524)</p> <p>~definition2:</p>
--	--

Figure 2 : Entrée de la base de connaissances pour le concept «abstract class»

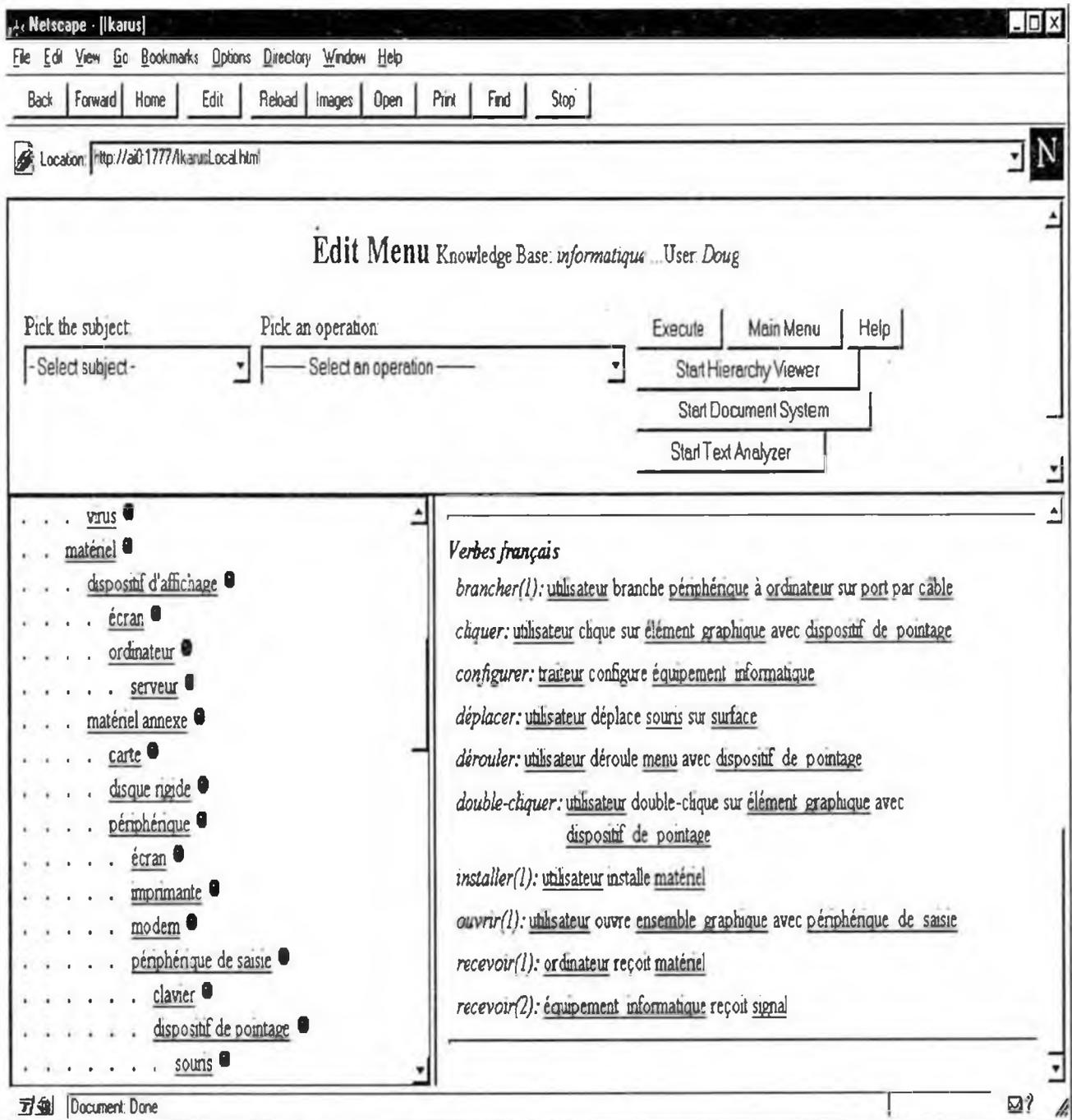


Figure 3 : Caractéristiques linguistiques du concept «souris»

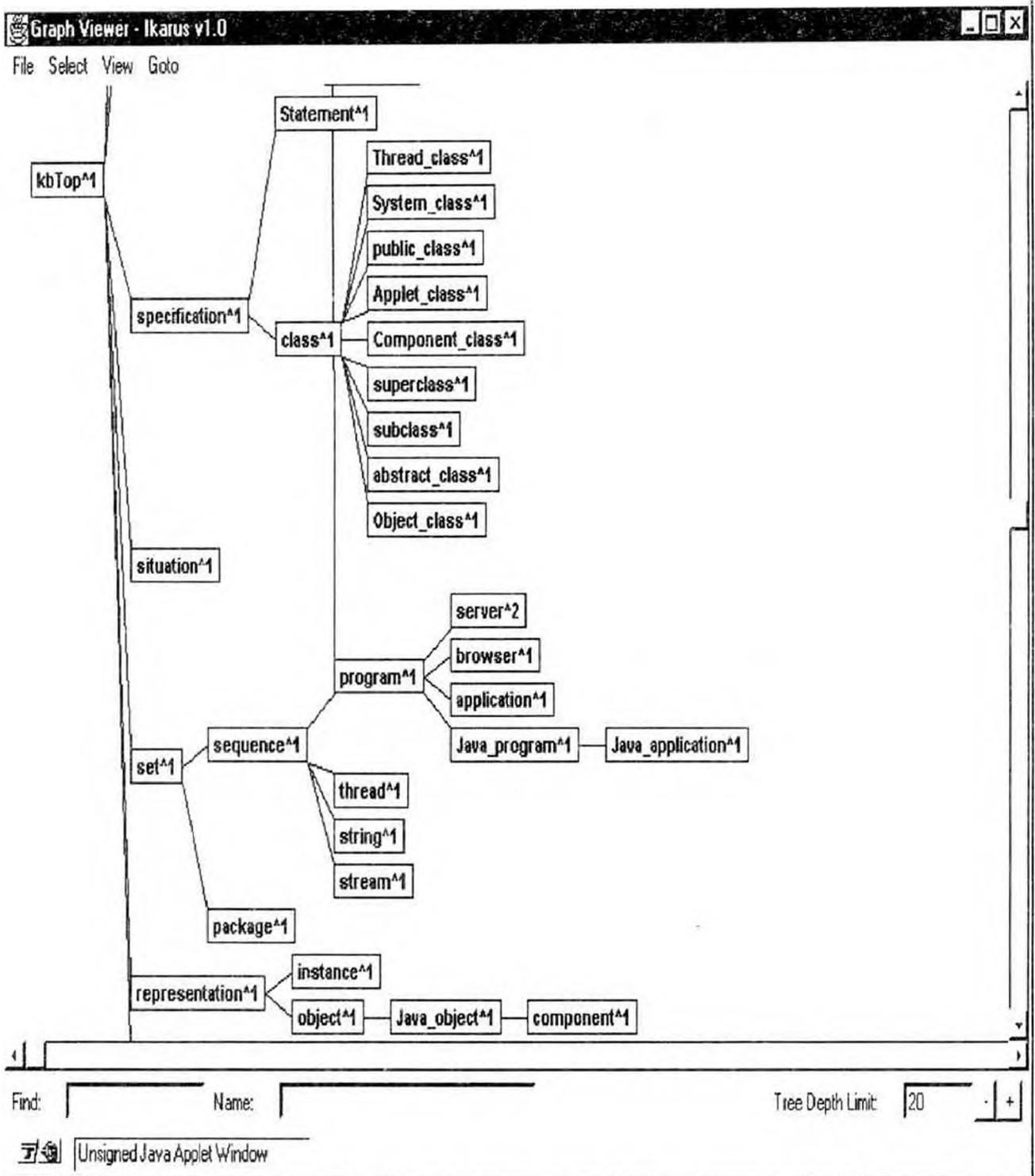


Figure 4 : Représentation graphique de la base de connaissances.

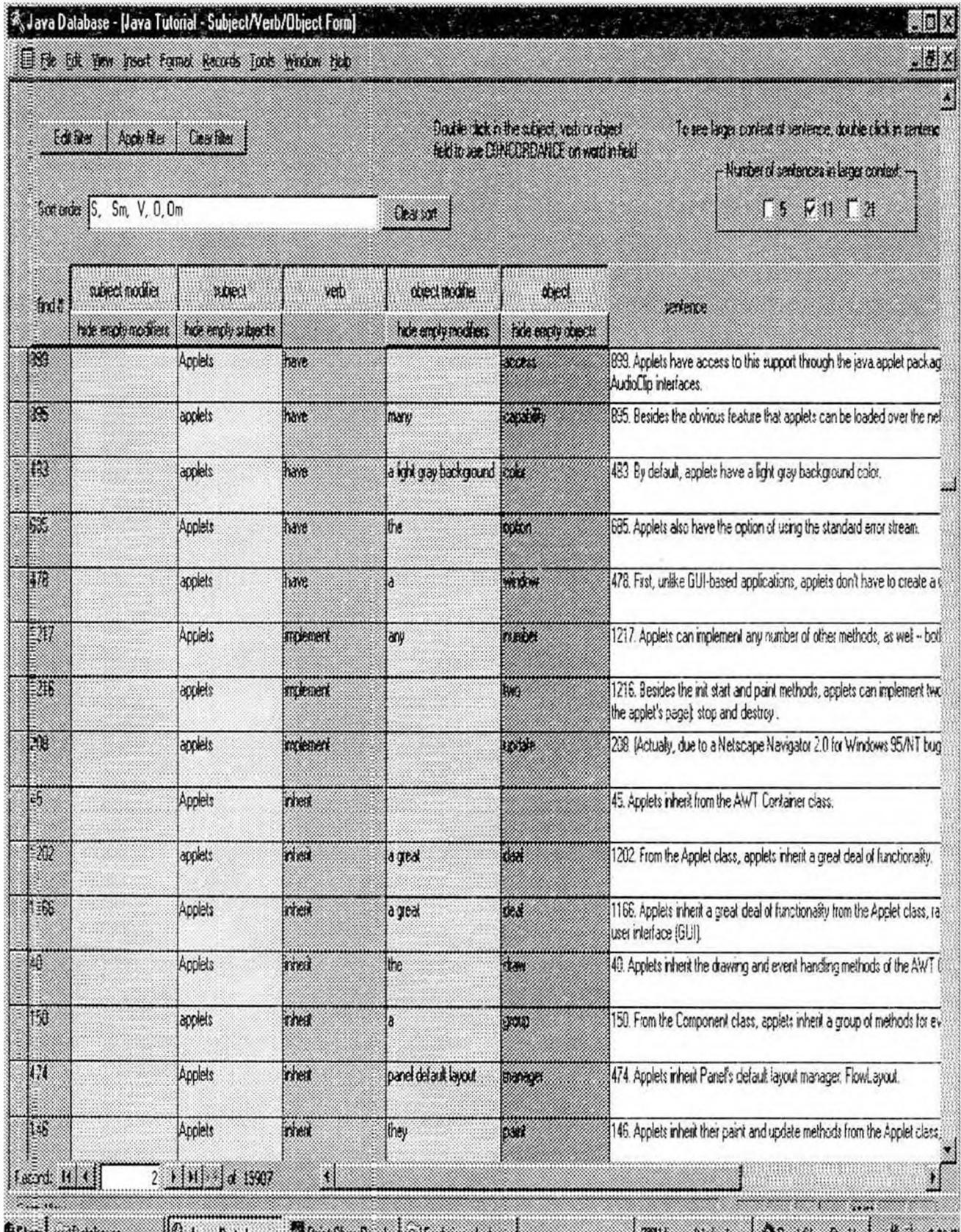


Figure 5 : Le système DockMan (fonctions d'un «applet»)

## RÉFÉRENCES

- AHMAD, K., PICHT, H., ROGERS, M. et P. THOMAS (1989) : «Terminology and Knowledge Engineering : A Symbiotic Relationship Explained», *Technical Report TR 89/1*, Guildford, University of Surrey.
- ATKINS, B.T.S. (1992) : «Tools for Computer-Aided Corpus Lexicography : the Hector Project», *Papers in Computational Lexicography (Complex 92)*, Budapest, Linguistics Institute, Hungarian Academy of Sciences.
- BIÉBOW, B. et S. SZULMAN (1997) : «Méthodologie de création d'un noyau de base de connaissances en logique terminologique à partir de textes», *Actes des deuxièmes rencontres, terminologie et intelligence artificielle (TIA-97)*, Université Toulouse-le Mirail, pp. 69-84.
- CONDAMINES, A. et P. AMSILI (1993) : «Terminologie, entre Langage et Connaissances : un Exemple de Base de Connaissances Terminologiques», *Terminology and Knowledge Engineering*, Frankfurt, INDEKS Verlag, pp. 316-323.
- DAVIDSON, L. (1997 à paraître) : *Semantically-Rich Linguistic Patterns for Terminological Knowledge Extraction* (titre provisoire), Thèse de maîtrise, École de traduction et d'interprétation, Université d'Ottawa, Canada.
- JACQUIN, C. et M. LISCOUET (1996) : «Terminology Extraction from Text Corpora : Application to Document-Keeping Via Internet», *TKE 96 : Terminology and Knowledge Engineering*, Christian Galinski and Klaus-Dirk Schmitz (Eds), Frankfurt, INDEKS Verlag, pp. 74-83.
- KAVANAGH, J. (1995) : *The Text Analyzer : A Tool for Knowledge Acquisition from Texts*, Thèse de maîtrise, Dept. of Computer Science, Université d'Ottawa, Canada.
- LENAT, D. et R. GUHA (1990) : *Building Large Knowledge-Based Systems*, Reading, MA, Addison-Wesley.
- L'HOMME, M.-C. (1997) : «Organisation des classes conceptuelles pour l'accès informatisé aux combinaisons lexicales spécialisés verbe + terme», *Actes des deuxièmes rencontres, terminologie et intelligence artificielle (TIA-97)*, Université Toulouse-le Mirail, pp. 161-174.

- MEYER, I., ECK, K. et D. SKUCE (1997) : «Systematic Representation of Concepts in a Knowledge-based System», *Handbook of Terminology Management*, S. E. Wright and G. Budin (Eds), Amsterdam/Philadelphia, John Benjamins, pp. 98-118.
- MEYER, I. et B. McHAFFIE (1994) : «De la focalisation à l'amplification : nouvelles perspectives dans une base de connaissances terminologiques», *TA-TAO : Recherches de pointe et applications immédiates*, Actes des troisièmes Journées scientifiques du réseau LTT, Beyrouth, FMA et AUPELF-UREF, pp. 425-440.
- MEYER, I., BOWKER, L. et K. ECK (1992) : «Towards a New Generation of Terminological Resources : An Experiment in Building a Terminological Knowledge Base», *Proceedings of COLING 92*, pp. 956-960.
- MEYER, I. et K. MACKINTOSH (1996 sous presse) : «The Corpus from a Terminographer's Viewpoint», *International Journal of Corpus Linguistics*, Vol. 1, No. 2.
- NUOPPONEN, A. (1996) : «Terminological Information and Activities in World Wide Web», *TKE 96 : Terminology and Knowledge Engineering*, Christian Galinski and Klaus-Dirk Schmitz (Eds), Frankfurt, INDEKS Verlag, pp. 92-99.
- SÉGUÉLA, P. et N. AUSSÉNAC (1997) : «Un modèle de base de connaissances terminologiques», *Actes des deuxièmes rencontres, terminologie et intelligence artificielle (TIA-97)*, Université Toulouse-le Mirail, pp. 47-68.
- SCHWEIGHOFER, E. et D. SCHEITHAUER (1996) : «Legal Terminology Research in an Internet/WWW Environment», *TKE 96 : Terminology and Knowledge Engineering*, Christian Galinski and Klaus-Dirk Schmitz (Eds), Frankfurt, INDEKS Verlag, pp. 92-99.
- SKUCE, D. (à paraître) : «Integrating Web-based Documents, Shared Knowledge Bases, and Local Document Searching for User Help».
- SKUCE, D. et T. LETHBRIDGE (1995) : «CODE4 : A Unified System for Managing Conceptual Knowledge», *International Journal of Human-Computer Studies*.
- SKUCE, D. et I. MEYER. (1991) : «Terminology and Knowledge Engineering : Exploring a Symbiotic Relationship», *Proceedings of the 6th International Workshop on Knowledge Acquisition for Knowledge-Based Systems*, (Banff, Oct. 1991), pp. 29-1 à 29-21.

TOUSSAINT, Y., ROYAUTÉ, J., MULLER, C., et X. POLANCO (1997) :  
«Analyse linguistique et infométrique pour l'acquisition et la structuration  
des connaissances», *Actes des deuxièmes rencontres, terminologie et  
intelligence artificielle (TIA-97)*, Université Toulouse-le Mirail, pp. 27-  
46.

# Les mots que les femmes détestent et pourquoi

Sofia BENYAHIA

*Université de Montréal, Montréal, Canada*

## 1. — INTRODUCTION

«Quels sont les mots que vous détestez, et pourquoi»? La question a été posée à cinquante-quatre Montréalaises, âgées de 30 à 75 ans, et ayant comme caractéristique commune d'être des «leaders» d'opinion. Nous avons de bonnes raisons de croire que ces personnes accepteraient avec intérêt d'examiner l'usage quotidien de la langue, au travail, à la maison, en société, bref, dans tous les milieux où elles seraient susceptibles de croiser des «irritants». Trente collégiennes, de 16 à 21 ans, se greffent à ce groupe de base. Nous les avons rencontrées dans le cadre de deux ateliers organisés au collège Rosemont, à Montréal.

*Les mots que les femmes détestent et pourquoi* n'est pas une opération d'épuration visant à nettoyer la langue de ses éléments allergènes. Ils ne sont évidemment pas les mêmes pour tous et le risque de mutisme total est trop grand. L'objectif, au contraire, en est un de débats et de discussions autour des affects relatifs aux mots avec lesquels nous cohabitons.

Le projet ne se borne pas à une analyse linguistique. L'étude renferme une dimension socio-psychologique importante. Les mots que les femmes détestent parlent de la société québécoise, et des mentalités qui la tissent. «Chacune choisira le niveau d'affectivité qu'elle veut révéler» professe Madeleine Poulin (journaliste).

Notre corpus est constitué de cinquante-quatre entrevues individuelles enregistrées, et de neuf fiches manuscrites sur lesquelles sont énumérés les exemples et les justifications des collégiennes réparties en groupes. Près de 1 600 exemples ont été recueillis. Cet article se présente comme la première classification du matériel colligé. À des fins pratiques, nous avons opté ici pour les catégories les plus fréquemment mentionnées. Ainsi, les mots que les femmes détestent concernent principalement :

- des déviations par rapport à la norme linguistique (accrocs aux règles grammaticales, distorsion du vocabulaire, influence de l'anglais, pauvreté de la langue);
- des déviations par rapport à la norme sociale (tutoiement, sacres, apostrophes);
- l'euphémisation croissante de la langue (rectitude politique ou *political correctness*, langue de bois);
- les mots à la mode;
- la féminisation des titres et des textes.

A ces catégories s'ajoute la catégorie «culturelle», dans laquelle sont cités quelques exemples d'irritants lexicaux proposés par des locutrices de deux autres langues, le pular et l'arabe tunisien.

Chacune de ces catégories est développée dans la classification suivante. Pour rendre l'aspect interactif de la recherche, les exemples sont accompagnés des commentaires justificatifs des répondantes.

## **CLASSIFICATION DES RÉPONSES LES PLUS FRÉQUEMMENT APPORTÉES À LA QUESTION-TITRE**

### **2.1 — Déviations par rapport à la norme linguistique**

La plupart des femmes interrogées ont manifesté une très grande sensibilité vis-à-vis de l'usage incorrect de la langue. Les transgressions les plus fréquemment citées concernent les règles grammaticales (confusion des genres, mauvais accords, absences pronominales), les distorsions lexicales (terminaisons amputées, mots tronqués, usages impropres), l'influence de l'anglais (anglicismes, emprunts inutiles) et la pauvreté de la langue (mots, expressions ou métaphores passe-partout).

### 2.1.1 — Accrocs aux règles grammaticales

*La confusion des genres* : **une** avion, **une** autobus, **une** argent (**elle** est bien placée), **une** ascenseur (**elle** est lente), **une** étage, **une** accident, **une** orage, **une** hélicoptère, **un** auto.

Ce phénomène arrive en tête des transgressions aux règles de fonctionnement du système linguistique. La liaison avec la voyelle initiale demeure l'explication la plus fréquemment apportée à ce phénomène. La proximité de l'anglais (*an autobus-une* autobus) ou tout simplement la paresse ont également été mentionnées.

*Les mauvais accords* : une chemise **sec**, des cheveux **sèches**, une pomme **pourrite**; les **cheval**, ils sont **spécials**.

Les mauvais accords en genre et en nombre constituent une autre source d'irritation pour les femmes. Selon certaines participantes, les règles d'accords n'ont pas été intégrées depuis l'école.

*Les absences pronominales* : **J'va** le faire, **j'va** y aller, **moi y** dire, **moi le faire**, **moi y aller**. **Moi j'aime lui** (pour celui-ci).

La paresse est encore une fois mise en cause. «On préfère les raccourcis et à un moment donné on se retrouve avec une langue amputée» prévient une journaliste.

### 2.1.2 — Les distorsions lexicales

*Les terminaisons amputées* : Un **artis...**, un **kios...**, deux **pias...**, un **dentis...**, c'est **correc...** Lorsqu'il y a des consonnes doubles, on fait chuter la dernière. On peut présumer qu'à un moment donné le **te** de dentiste tombera définitivement.

*Les mots tronqués* : **ado**, **prof**, **resto**, **scénard**, etc.

Les professeures et les enseignantes semblent davantage sensibles à ce phénomène. Même si elles usent quelques fois de ce procédé courant, caractéristique du jargon des jeunes, elles restent vigilantes et exigeantes par rapport à la clarté de la langue.

— «J'imagine qu'on en utilise tous plus ou moins. Il m'arrive de dire **prof** plutôt que **professeur**. Évidemment quand la langue est continuellement marquée par l'utilisation de mots tronqués, ça devient un jargon et c'est dangereux. Je dis à mes étudiants que la langue est un outil de communication qu'il faut préserver. La communication ça ne va pas de soi, alors il faut s'arranger pour avoir le meilleur outil possible.» (Denise Desautels, poète, professeure);

- «Le dernier en liste que j'ai entendu c'est la **littgén** pour désigner la littérature générale. Alors ça c'est très branché dans le milieu. J'enseigne la littérature alors on peut dire «Tel roman c'est dans la sf ou la littgen ?» La science fiction ou la littérature générale ? Ça m'horripile et c'est sans doute à cause de ma génération.» (Louise Dupré, professeure).

Les collégiennes interrogées sur la question apprécient l'aspect «jeune et dynamique» des abréviations. C'est un élément d'identification au groupe, un aspect de leur code comme le *look* et la recherche d'indépendance. Elles paraissent toutefois conscientes du changement de normes pour les «adultes» :

- «On utilise les abréviations quand on est en «gang», mais on parle mieux quand on change de milieu.» (collégienne).

*Usage impropre et imprécis du vocabulaire :*

Différents types d'usages impropres de vocabulaire ont été mentionnés. **Peinturer** pour **peindre**, et **s'assir** pour **s'asseoir** sont les deux exemples les plus fréquemment cités. Le manque de nuance et de précision dans le choix des mots représentent une autre source d'agacement.

- «Les mots existent; c'est amusant d'aller les chercher. Les gens utilisent **chapeau** comme mot unique pour désigner toutes sortes de couvre-chef; de la même façon, on appelle **livre** tout ce qui s'écrit que ce soit une **revue**, un **roman**, un **catalogue**, ou un **illustré**. Idem pour les couleurs; on dit **rouge** pour un **rouge tomate**, **rouge cerise**, **rouge pompier**, ou **rouge sang**, ou **vert** pour **vert bouteille**, **vert forêt**, **vert céladon**, **vert mousse**. Plus on réduit sa langue, plus on réduit son univers. C'est très heurtant pour moi.» (Béatrix Marik, directrice de publication);
- «Pourquoi parler de la **deuxième** guerre mondiale ? Il faut dire la **seconde**. Deuxième ça veut dire qu'il va y en avoir une troisième. Quand j'entends ça, ça me hérise. Aussi, je suis contre le mot **autochtone** qui veut dire au premier degré «né de la terre, du sol». Les autochtones d'ici sont venus avant nous, mais ils sont venus d'ailleurs; ils ne sont pas nés du sol. Je n'aime pas le mot autochtone.» (Louki Bersianik, romancière);
- «Il y a des mots comme **anglophones** et **francophones** qui sont utilisés à toutes les sauces. Par exemple on dit **francophone** au lieu de dire **française**. Un peu plus on va dire la langue francophone. On peut utiliser le mot **français** parfois sans qu'il s'agisse d'une

nationalité, mais on préfère francophone parce qu'on sait que dans certains cas ça s'applique donc on l'applique dans tous les cas. J'évite ça autant que possible. Je dis «la presse de langue française» au lieu de dire «la presse francophone.» (Madeleine Poulin, journaliste).

### 2.1.3 — Mots, expressions et métaphores passe-partout

Selon certaines répondantes, la sur-utilisation de ce genre de *tics de langage* masque la pauvreté de la langue, et révèle un manque de raffinement du vocabulaire. Ce sont des formules utilisées automatiquement, sans but significatif.

- «**À queque part...quelque part** c'est une expression complètement vide de sens. **Au niveau de** m'agace aussi parce qu'on le dit sans qu'il soit toujours question de niveaux ou de hiérarchie.» (Marie-Claude Fortin, journaliste);
- «Il y a eu une époque où on disait au Québec **c'est pas un cadeau**. C'est une formule qui avait été trouvée ou reconduite à la télévision. Présentement, les deux expressions à la mode qui me tannent sont **c'est pas évident** et **c'est génial**. Je prête une valise et **c'est génial** parce que voyager dans de mauvaises conditions, comme vous savez, **c'est pas évident**. L'étymologie de génie c'est «hors du commun»; ça vient de quelque chose d'infiniment supérieur. On déshabille la langue de son sens et on l'affadit à tel point que le génie devient un lieu commun. On a renversé la vapeur. Ça réduit le discours.» (Marie Laberge, romancière);
- «Le **truc**, le **machin**, la **patente**, je trouve ça triste. Qu'un enfant de six ou de dix ans trouve chose à sa portée, ça se comprend; vingt ans plus tard c'est dommage parce que on manque la chance d'utiliser un autre mot qui désigne la chose ou la patente en question. Il y a des gens qui utilisent ça en chapelet dans une phrase. «Met la patente en haut de la chose !» La patente a le dos large !» (Béatrix Marik, directrice de publication);
- «Je donne un cours de communication d'affaires au deuxième cycle et avec les étudiants on a commencé à faire une collecte de métaphores passe-partout. On est surtout attentifs, toutes les semaines, aux formules creuses avec lesquelles les politiciens se gargarisent. Alors le ministre Rochon veut **créer des synergies**, et **livrer la marchandise**. Heureusement, m'a-t-on dit aussi, que «quand le bateau coule on arrête de faire des vagues». Il y a aussi **brasser la cage**, **mettre ses culottes**, **remettre les pendules à**

**l'heure, sonner la fin de la régression.** C'est imagé, sympathique, mais l'abus fait le ridicule.» (Marie-Éva de Villers, lexicographe).

#### 2.1.4 — L'influence de l'anglais : anglicismes et emprunts

La question de l'influence de l'anglais est très sensible au Québec. Les passions s'enflamment rapidement sur ce point indissociable de l'identité québécoise. «Pour moi, tout ce qui est anglicisme est hors-circuit» proclame Marcelle Ferron (artiste-peintre). À quelques rares exceptions, toutes les femmes interrogées, semblent s'être passées le mot : attention à aux anglicismes et aux emprunts ! Les linguistes spécialisés dans le contact du français avec l'anglais rapportent que «dans le processus d'anglicisation au Québec, on constate un rejet marqué de la forme anglaise au profit de la forme française, et les procédés favorisant la substitution d'une forme française à une forme anglaise, soit le calque et l'anglicisme sémantique, ont donc la faveur des locuteurs québécois.» (Pergnier, 1988 : 71). Des exemples tirés de notre corpus d'entrevues confirment cette tendance, tout du moins en ce qui concerne l'utilisation d'anglicismes sémantiques dans certains milieux montréalais. L'anglicisme sémantique résulte de «l'attribution d'un sens anglais à un mot français (1).» (Pergnier, 1988 : 70). Les commentaires ci-dessous proposent des exemples d'anglicismes sémantiques tirés de notre corpus d'entrevues. Ils manifestent également la contrariété qu'éprouvent certaines femmes face à la synonymie inutile qui résulte de ce procédé :

— «Dans les communications scientifiques en français ou dans les conseils d'administration, tout le monde est en train d'**adresser** ou de ne pas **adresser une question**. J'ai encore vérifié dans le dictionnaire; «adresser» en français ce n'est pas du tout ça. Si je vous «adresse une question», c'est que je vous l'«envoie». Ce n'est pas dans le sens de «s'interroger à propos de tel sujet». C'est copié sur l'anglais to address an issue pour dire «aborder un problème». Les gens vont dire «Cette question n'a pas été adressée» pour dire «On ne s'est pas penché sur ce sujet». Et je parie que je ne mourrai pas sans voir ça dans le Petit Robert ! C'est en train de se répandre à une vitesse folle. Ce sont des mots français, c'est toujours là le piège. Adresser, c'est un beau mot français, une question aussi. C'est sans se méfier que les gens entrent dans ce jeu-là. Ils sont

---

1) C'est le cas de *contrôler* au sens de «diriger», ou de «régler», et aussi de *réaliser* qui s'est vu attribuer le sens de «se rendre compte» sous l'influence de l'anglais (Pergnier, 1988 : 70).

portés à penser que c'est une expression correcte qu'ils ne connaissaient pas, tout simplement.» (Francine Harel-Giasson, professeure);

- «Par mon travail, je suis sensible aux détournements de sens qui résultent de la proximité de l'anglais. Par exemple, quand on emploie le mot **agressif** au sens de «déterminé, dynamique, combatif». On dit «un vendeur **agressif**» au lieu de dire «un vendeur **motivé**». Mais ce qui est amusant, c'est que dans le Nouveau Petit Robert, la dernière édition donc, on a inséré l'exemple du «vendeur agressif» au mot **agressif**, dans le sens de «violent, qui attaque». Manifestement, on n'a pas compris l'usage qui en est fait ici. Moi, je le condamne, je le mets comme une forme fautive. Dans les anglicismes, il y a aussi l'affreux **définitivement** qui sert à tous les début de phrases au sens de «assurément, incontestablement».» (Marie-Éva de Villers, lexicographe);
- «Un mot qui m'irritait et qui est aujourd'hui accepté dans le Petit Robert c'est le mot **mature**. On n'avait peut-être pas d'adjectif pour le mot **mature** en anglais, mais on avait la **maturité** !» (Céline Hervieux-Payette, sénateure).

Dans notre corpus, nous relevons également une réaction contre les anglicismes francisés :

- «Je trouve ça triste quand j'entend des gens «**parker leur char**»; je me dis qu'on est en train de devenir des créoles d'Amérique du Nord même si je trouve que la langue créole a des expressions superbes.» (Lorraine Pintal, directrice du Théâtre du Nouveau Monde).

Ci-dessous, une autre réaction à l'anglicisation du français au Québec, relative cette fois à l'emprunt ou anglicisme intact :

- «Pourquoi **napkin** ? C'est tellement plus simple une **serviette de table**. Ça me navre. Je vois là tout une esthétique de vie qui est gommée. L'esthétique, l'art de la table, pour moi, c'est important. **Rocking-chair**, etc. Je suis d'accord pour emprunter parce que la langue est une matière vivante. On peut accepter des greffes, mais quand ça ne se justifie pas, c'est un snobisme gratuit.» (Béatrix Marik, directrice de publication).

## 2.2 — Les déviations par rapport à la norme sociale

«Dans toutes les sociétés, il existe des normes sociales [...] qui régissent les comportements, les relations entre les individus, et aussi le

langage. Il existe une bienséance, une civilité, une façon de faire, de se conformer pour maintenir l'harmonie et par conséquent les distances. Parallèlement, il y a toujours des transgressions à ces règles» peut-on lire dans l'ouvrage de Diane Vincent, *Pressions et impressions sur les sacres au Québec* (1982 : 11). Les transgressions aux normes linguistiques ont fait l'objet de la partie précédente. L'étude sur les mots que les femmes détestent fait ressortir de vives réactions contre trois pratiques sociales bien implantées au Québec : les sacres, le tutoiement spontané et les apostrophes ou interpellations.

### 2.2.1 — Les sacres

Sur ce point, les attitudes divergent. Certaines locutrices condamnent sans réserve l'usage des sacres et l'assimilent à une pauvreté de la langue ou à une volonté de se démarquer chez les jeunes :

- «Ici on est affligé de ces [sti] pour hostie, [cris] pour Christ. Tous ces sacres qui ne sont pas de véritables blasphèmes ou sacrilèges, servent (ou tentent seulement) à suppléer à l'absence de vocabulaire, ou même de donner un style apparemment délinquant ou vulgaire. Alors je ne les aime pas» (Marie-Éva de Villers, lexicographe);
- «Je viens d'un milieu où sacrer est très mal vu. Enfant, j'avais mon père qui avait un malin plaisir à en lâcher quelques-uns pour voir comment ma mère allait réagir; mais pour nous, c'était interdit.» (Suzelle Levasseur, artiste-peintre);
- «Je suis mal à l'aise quand les gens sacrent autour de moi.» (Hélène Bond, femme d'affaires).

D'autres, moins sévères sur ce point, identifient la sur-utilisation des sacres par les jeunes à une habitude passagère, que ceux-ci abandonneraient volontairement plus tard :

- «Les sacres ne m'alarment pas vraiment. Je me dis que ça doit faire partie des passages obligés. Souvent la langue sert à se démarquer. Les jeunes doivent être conscients de ça.» (Hélène Blackburn, chorégraphe).

Les collégiennes que nous avons rencontrées semblent appuyer ce point de vue, conscientes toutefois de la difficulté de s'en défaire :

- «On aimerait mieux ne pas sacrer; cette habitude dévalorise les discussions.» (collégienne);

- «On utilise toujours les mêmes sacres par habitude. Quand tu es jeune c'est interdit de sacrer donc ça nous pousse à le faire. Ça nous fait sentir supérieurs de désobéir aux interdictions. Quand on vieillit, on prend d'autres habitudes.» (collégienne).

À noter que l'étude de Diane Vincent fait ressortir que les sacres les plus courants en français québécois sont *criss*, *câlîce*, *tabarnac*, *osti*, *ciboire*, *viarge*, *sacrament*, *baptême* et *maudit* (Vincent, 1988 : 49), et que de nos jours «les jeunes filles sacrent, presque autant que les jeunes hommes de leur âge, et beaucoup plus que les hommes des générations précédentes» (Vincent 1988 : 89). Ce qui laisse entendre que les sacres étaient auparavant le domaine des hommes. Cela explique aussi en partie, la gêne de certaines femmes vis-à-vis des sacres.

### 2.2.2 — Le tutoiement spontané

Nous avons tous le souvenir d'un parent, d'un maître ou d'une maîtresse ressassant consciencieusement qu'«il faut dire «vous» aux personnes que l'on ne connaît pas et «tu» aux personnes qui nous sont plus familières». Tous ces gens-là seraient très surpris de voir à quel point leurs efforts étaient vains, et combien il semble moins «cérémonieux» aux yeux de beaucoup de tutoyer spontanément leur prochain. Tout porte à croire que le tutoiement spontané n'était pas «permis», à peine quelques années en arrière. Cela explique en partie, comme pour les sacres, la gêne, la désapprobation, et même quelquefois l'abdication de certaines femmes face à ce phénomène.

- «Ce n'est pas un mot comme tel mais c'est une façon d'être, une spontanéité qui me rend très inconfortable. C'est pour moi un sujet très délicat dans le sens où l'éducation que j'ai reçue ne permet pas de tutoyer les gens dès la première rencontre. C'est peut-être le you anglais qui est à l'origine de ce manque de distinction. C'est un espèce de nivellement; il n'y a plus d'âge, plus de génération. On est à tu et à toi avec tout le monde dans n'importe quel contexte. Je n'irai pas jusqu'au tutoiement en onde même avec des invités que je connais.» (Francine Moreau, animatrice Radio-Canada);

- «Le tutoiement m'embête mais il faut faire avec sinon on serait choqué continuellement. En fait, c'est aux enfants qu'il faut apprendre à ne pas tutoyer parce que, une fois adulte, c'est très difficile de se défaire de cette mauvaise habitude. Je vouvoyais mes grands-parents et cela ne m'empêchait pas d'être très proche d'eux. Aujourd'hui on retrouve le tutoiement dans les publicités. On

encourage cette pratique chez les jeunes.» (Greta Chambers, Chancelière de l'Université Mc Gill);

- «Je déteste le tutoiement spontané.» (Solange Chaput-Rolland, sénatrice retraitée).

Certains commentaires de notre corpus d'entrevues corroborent le fait que l'on accepte plus facilement le tutoiement lorsqu'on est jeune :

- «Je supporte beaucoup moins qu'on me tutoie dans un magasin quand j'arrive. Mais encore là ça dépend beaucoup de qui le dit et comment.» (Pascale Navarro, journaliste).

### 2.2.3 — Apostrophes et interpellations

Les apostrophes représentent un autre type de familiarités qui irritent les femmes, sans toutefois faire l'unanimité, comme le montrent les deux commentaires suivants :

- «Les apostrophes m'énervent énormément. Le **On** de **O n** magazine ?», «**On** cherche une belle valise ?». Ce **On** là qui fait que nous devenons tout à coup tout à fait copains. «**On** cherche une petite sandale d'été ?». Une fois, je regardais une paire de botte pour homme et j'ai eu droit à «**On** cherche une petite paire de bottes pour son mari ?». Alors j'ai répondu «Non. **On** n'en veut pas. **On** n'a pas de mari et **on** va laisser faire pour les bottes! ». On ne peut tout de même pas passer sa vie à se faire interpellé comme ça ! Dans cette catégorie je mets aussi l'interpellation à la troisième personne, le **A** de «**A** veut ça la madame ?». «**A** dit qu'il faut aller par là». Cette interpellation me choque beaucoup.» (Marie Laberge, romancière);
- «Quand j'entends «*La ptite mademoiselle a veux-tu queque chose ? A veux-tu que je l'aide ?*», je sais que ça dérangerait certaines femmes, mais pas moi. Ce sont parfois les chauffeurs d'autobus qui vont employer ces mots-là.» (Pascale Navarro, journaliste).

### 2.3 — Les mots à la mode

Les mots à la mode ne sont pas l'apanage de notre siècle, loin s'en faut. Dans son livre au titre précis de *Des mots à la mode, et des nouvelles façons de parler : avec des observations sur diverses manières d'agir et de s'exprimer, et un discours en vers sur les mêmes matières*, François Callières (1693) nous a laissé des observations savoureuses sur les modes linguistiques de son époque. On y apprend

par exemple que les expressions **c'est-à-dire que, il est vrai que, en effet** étaient déjà dans toutes les bouches bien articulées, et que «ceux qui sont si infatueux du rang de leurs ancêtres, et qui se servent en parlant d'eux-mêmes de ces termes vains et odieux, un homme *de ma qualité*, une femme *de ma qualité*, [...] font conclure que ce sont des gens dépourvus de toute autre sorte de mérite, réduits à se montrer par ce seul endroit qui leur soit avantageux.» (Callières, 1693 : 124-125).

Au travail comme en société, certains mots sont plus *dans le vent* que d'autres. Les propos recueillis dans le cadre de cette recherche nous enseignent que l'on se gargarise volontiers des termes provenant de la gestion et des anglicismes. Les mots qui parodent ne semblent pas éblouir pour autant tout l'auditoire.

— «Je deviens allergique très rapidement à tous les mots qui ont cette fonction-là. Je les rencontre surtout dans le vocabulaire de la gestion. Le mot **synergie** a été galvaudé dans ce sens-là. Ce n'était pas un mauvais mot au départ, idem pour **concertation**. Si bien que je ne les utilise pas. Certains mots me déplaisent aussi pour des raisons esthétiques, la **réingénierie** par exemple. Je trouve ça d'une stupidité même à l'oreille.» (Lise Bissonnette, Directrice du journal *Le Devoir*);

— «Un mot qui est devenu à la mode et qui me déplaît souverainement c'est **sociétal**. Je n'ai jamais trouvé que quand quelqu'un me parlait de «sociétal» il n'était pas en train de me parler de «social». C'est comme une forme de snobisme. C'est un mot qui m'agace.» (Francine Harel-Giasson, professeure);

— «La **mondialisation des échanges**, on entend ça constamment dans tous les média. Évidemment, en gestion ça s'emploie énormément. C'est le terme juste qui doit remplacer **globalisation** qui lui est inspiré de l'anglais, mais il devient ridicule et prétentieux en raison de l'abus qu'on en fait. Pour meubler les creux d'une conversation, on va parler de «la mondialisation des échanges et de la vive concurrence qui existe entre les entreprises, et donc qu'il faut réduire les coûts», etc. Un autre mot, la **qualité totale** que j'appellerai plutôt l'utopie totale. On commence à moins en entendre parler que l'an dernier. Il y a aussi la **réingénierie des processus**. «Il faut faire la réingénierie de nos processus». Tout découper et essayer de faire mieux finalement. Mais souvent c'est pour dire on va supprimer des emplois. On va **structurer** c'est un autre exemple. Ce n'est ni faux, ni mauvais, mais l'abus rend ridicule.» (Marie-Éva de Villers, lexicographe);

- «Je déteste le mot **must** comme dans «C'est un **must** de la saison théâtrale» par exemple. C'est un effet de mode. C'est à long terme qu'on voit ce qui s'impose, ce qui est intéressant. En fait, on essaye de forcer, de créer l'événement important. C'est artificiel.» (Marie-Claude Fortin, journaliste);
- «Moi c'est l'expression **politically correct** que je ne peux plus entendre. On l'entend tout le temps. Tous ceux qui ont des tribunes quotidiennes ou hebdomadaires, à la télé ou à la radio, n'ont que cette expression à la bouche. Et tous bien sûr se défendent d'être **politically correct**.» (Louki Bersianik, romancière).

## 2.4 — L'euphémisation croissante de la langue

Un autre aspect de l'usage quotidien de la langue qui agace beaucoup de femmes parmi les répondantes : l'euphémisation croissante et «obligatoire» prônée par les partisans de la rectitude politique. La langue de bois est l'autre type d'euphémisation dont nous tiendrons compte dans cette classification puisqu'il en a été très fréquemment question lors des entrevues. Ces deux genres de discours ont en commun la volonté de transmettre, au-delà du sens des mots, une idéologie à laquelle on souscrit.

### 2.4.1 — La rectitude politique ou *political correctness*

La *Political Correctness* ou PC a vu le jour aux États-Unis, à la suite des remous sociopolitiques des années 60. À cette époque, l'expression a désigné ironiquement qui modelait son opinion conformément aux courants politiques en vogue. À la fin des années 80, le sens se «positivise»; il est désormais considéré PC de recycler le papier, de s'opposer au port des vêtements de fourrures et d'utiliser les mot «Native American» (américain de souche) à la place d'«Indian» et «African-American» à la place de «black» (2). C'est dans les milieux académiques américains que le mouvement se radicalise pour devenir un code de comportement expiatoire vis-à-vis des minorités. Les PCers revendiquent une université plus «inclusive» dans le contenu de ses programmes et de sa population estudiantine (3). Dans les années 90, la vague PC franchit les frontières de l'université (et du continent) pour s'étendre à l'ensemble des groupes de la société, sous la forme d'une moralisation du vocabulaire. On veut nommer autrement les minorités. En France, on parle de «police du langage» et de «censure

2) THIBODEAUX, David (1992).

3) RICHER, Stéphane et Lorna WEIR (1995).

bien-pensante qui traque, condamne et bannis tout mot ou expression jugés offensants pour une communauté ou un groupe humain» (Coignards et Lanez, 1995 : 53). On l'aura compris, la «rectitude politique» ne fait pas que des adeptes. «Je déteste qu'on *abrille* comme on dirait au Québec» affirme Lise Bissonnette (directrice du journal *Le Devoir*). Comme en témoignent les commentaires ci-dessous, les femmes interrogées sur les mots qu'elles détestent n'ont pas oublié de donner leur avis sur les nouveaux mots qu'il faut utiliser pour rester «correct».

- «**Handicapé** pour moi c'est un terme brutal; je préfère **infirmes** de *infirmus* qui veut dire faible. Maintenant je suis rendue **personne à mobilité réduite**. Moi je trouve que je n'ai pas de mobilité du tout alors ! C'est vraiment aseptiser les mots.» (Anne-Marie Alonzo, poète);
- «**Malentendant, non-voyant**, je ne suis pas contre parce il y a des degrés que **sourd** et **aveugle** ne rendent pas.» (Louki Bersianik, romancière);
- «Les journalistes emploient souvent des mots pour ne pas heurter. Par exemple le fameux mot **ethnie**. Justement j'étais en train d'écrire quelque chose aujourd'hui et j'ai employé le mot **ethnie**. Je l'ai mis entre guillemets parce que ça ne peut pas faire partie de mon vocabulaire; je ne l'ai pas intégré. Dire que des gens font partie d'une ethnie, que ce sont des ethniques, c'est une aberration totale. Figurez-vous que je suis une **ethnique**, même si ça fait 25 ou 30 ans que je vis ici. Le terme **néo-québécois** convient tellement mieux. Idem quand je dis d'une fille qu'elle est **mulâtre**. Je ne me sens pas porter un jugement. Ça c'est le genre de mots avec lesquels je suis bien mal à l'aise quand il faut rédiger un texte. J'ai l'impression de jongler avec des œufs ou de marcher dessus !» (Pascale Navarro, journaliste);
- «Dans le domaine du magazine féminin, s'il y a un terme qu'on bannit, c'est le terme **grosse**. C'est tabou dans les communications. On va l'utiliser seulement dans le contexte d'un témoignage quand la personne a décidé d'elle même d'assumer son statut de grosse. On préfère le terme **ronde**. Les chaînes de vêtements spécialisées utilisent **tailles fortes**.» (Sandra Cliche, rédactrice en chef);
- «Tous les mots politiquement correct m'agacent, me choquent même. Ce sont des mots qui veulent circonscrire la liberté de pensée, autrement dit embrigader la liberté d'expression. En anglais, par exemple, on désigne quelqu'un qui est **petit** par

**vertically challenged**. La rectitude politique c'est la pensée à saveur du jour, imposée par des groupes de pression ou par les médias. Ça évolue au fil des années, tous les 10-15 ans. Le mot **contraception**, il y a 50 ans, c'était très mauvais; aujourd'hui c'est correct.» (Diane Marcelin, juge);

— «**Minorité visible** fait partie de ces mots qu'on nuance aujourd'hui. Moi je préfère qu'on me dise femme **noire** plutôt que **minorité visible**.» (Awa Bocar Ly, sociologue);

— «La vieillesse ce n'est pas un **âge d'or**; c'est un naufrage. Alors tous ces mots politically correct comme les **aînés**, le **bel-âge**, le **troisième âge**, ou l'**âge vermeil**, je ne les aime pas. Ma mère dit toujours «Je suis une vieille femme, une femme âgée, et n'essayez pas de me faire croire que c'est le troisième âge, que tout est beau, qu'on n'a plus de problème et qu'on va gambader dans la forêt!».» (Hélène Blackburn, chorégraphe) (4).

#### 2.4.2 — La langue de bois

«La première caractéristique de la langue de bois est qu'elle apparaît sous deux modalités, une modalité ouverte et une modalité cachée» affirme Françoise Thom (1987 : 15). La langue de bois partage cette caractéristique avec le langage politiquement correct. Cependant, les deux types de discours ne semblent pas circuler dans les mêmes sphères. Nous avons observé une utilisation quasi-exclusive du langage politiquement correct dans les médias (et un rejet de la part des locutrices). La langue de bois est quant à elle souvent associée, pour ne pas dire confondue, avec le discours politico-administrativo-bureaucratique :

— «Je déteste les mots qu'on utilise pour occulter certaines réalités. Par exemple, on ne parle plus de **patients** ou de **malades** dans les

---

4) Avant d'en finir avec la rectitude politique, cette observation relevée dans un article du journal *Le Point* : «Le langage des sourds ne pouvait échapper aux remaniements syntaxiques de la *political correctness*. [...] Peu à peu, les sourds corrigent les signes jugés discriminants pour les remplacer par d'autres, consensuels et égalitaires. Ainsi, le mot «japonais» ne s'exprimera plus par le geste du petit doigt bridant les yeux, mais par un demi-cercle, censé retracer le contour géographique de l'archipel nippon. Interdiction de s'écraser le nez avec le pouce pour dire «noir», ou d'agiter la main droite pour signifier «homosexuel». De nouveaux signes sont à l'étude. Quant aux sourds russes, ils ont été priés de trouver un nouveau geste pour «américain». Le signe consistant à se passer la main sur un ventre repu est qualifié de raciste par les sourds américains.» (Coignards et Lanez, 1995 : 54).

services de santé mais de **bénéficiaires**. Ça introduit l'idée que c'est une faveur qu'on leur fait de les soigner, plutôt que de placer ça sur le terrain du droit à être soigné quand on est malade, de quelque origine sociale que nous soyons. C'est un travail insidieux de conditionnement de l'opinion publique. De la même façon, on a toujours de la difficulté à parler de compression et de coupure, alors en enveloppe ça dans toutes sortes de termes : on va **rationaliser les services, les réorganiser, les adapter aux réalités contemporaines**; on va **prendre des virages, dégraisser**, etc. Les dirigeants ne veulent pas se rendre responsables des choix économiques et politiques qu'ils font.» (Lorraine Pagé, présidente de la Centrale d'Enseignement du Québec);

— «Il y a des expressions que je n'utiliserai jamais. Les **intervenants dans le milieu, les prestataires**, tout ce jargon de l'administration des petites charités mesquines qu'on fait aux gens, ça me déplaît beaucoup. C'est condescendant, ça gonflent d'importance ceux qui dispensent ces petits moyens chiches. Il me semble qu'il y a d'autres façons de parler des gens que comme prestataires.» (Madeleine Poulin, journaliste);

— «Il y a des expressions que je ne peux pas supporter comme **dégraisser, enlever le gras, couper à l'os**. Toutes ces images sont utilisées pour signifier «diminuer quelque chose». On est rendu au strict minimum, **à l'os, au squelette**. Tout ce qui touche à la chair me donne mal au cœur. Dans le contexte économique et budgétaire actuel c'est très fréquent de voir ça.» (Monique Simard, députée).

## 2.5 — Féminisation des titres et des textes

La féminisation des titres, des textes et du discours fait l'objet d'un intérêt croissant compte tenu de la richesse documentaire disponible sur le sujet (guides, actes de colloques, dictionnaires, grammaire, enquête, monographies, etc.). En dehors du circuit universitaire, on peut dire que les efforts de sensibilisation à ce phénomène ont porté leurs fruits étant donné que la plupart des femmes ont évoqué spontanément cette question. À la lumière de tous les commentaires de notre corpus d'entrevues, on peut dire que la féminisation ne fait pas l'unanimité, et que l'ambivalence persiste. L'avantage est tout de même dans le camp du oui. Quelques exemples de réactions :

— «Sur la féminisation, je suis très ambivalente. Je comprend l'idée de vouloir s'appropriier des mots, mais il y a aussi une autre approche

générique qui est de vouloir s'approprier le poste sans changer le titre qui le désigne. Gagner un rang, sans changer le terme qui décrit le rang. Je suis la première **chancelière** à l'Université McGill et tout de suite nous avons féminisé le titre. Cependant, le traducteur de l'Université, qui est français et très tatillon, a fait une vérification dans le Robert et dans le Larousse, pour découvrir que la chancelière c'était d'abord «la femme du chancelier» et ensuite, «une boîte ou un sac qui sert à garder les pieds au chaud». Donc il ne voulait pas qu'on fasse rire de nous et nous avons publié un nouveau communiqué pour annoncer que l'on optait pour **madame le chancelier**. Malgré tout, les femmes de Mc Gill ont souhaité, majoritairement, souligner l'accès, par une femme, à un haut poste de la hiérarchie de l'Université. Alors je suis redevenue la chancelière!» (Greta Chambers, chancelière de l'Université McGill);

- «La féminisation à outrance comme à l'UQAM (Université du Québec à Montréal) m'énerve au plus haut point. Par contre je suis prête à aller faire une croisade pour la promotion du substantif épïcène. Au lieu d'hésiter entre **un médecin, une femme-médecin**, quand le mot n'est pas traduit dans l'usage comme **un infirmier-une infirmière**, utilisons le substantif épïcène quand c'est possible. **Un médecin-une médecin, un dentiste-une dentiste**; on voit tout de suite qu'il s'agit d'une femme.» (Marie-France Bazzo, animatrice Radio-Canada);
- «J'aime bien qu'à l'UQAM on ait fait une convention collective en faisant voir qu'il y a des femmes là-dedans.» (Louki Bersianik, romancière);
- «Le e muet entre parenthèse je refuse ça maintenant. On n'est plus entre parenthèses. C'est toute une façon d'affirmer le sujet femme dans l'écriture. Il faut développer des automatismes même à l'oral.» (Claudine Bertrand, enseignante).

## 2.6 — Des mots et des cultures

Qu'en est-il ailleurs dans le monde ou même dans la francophonie ? «Vous poseriez les même questions à un bassin de femmes aux États-Unis ou en Europe, vous auriez des réponses très différentes» supposent quelques unes des répondantes. Voici quelques exemples de mots détestés par des locutrices d'une autre langue que le français.

- «**Femme chèvre** c'est une expression qui est utilisée dans ma langue maternelle, le pular qu'on appelle aussi quelque fois le fulami. C'est une langue parlée en Afrique de l'Ouest dans dix-neuf pays, notamment en Mauritanie, au Sénégal et au Niger. Les Peuls sont un peu partout en Afrique. C'est un peuple d'éleveurs qui s'est déplacé après chaque pression démographique. On dit en pular «debo con beo» c'est-à-dire «la femme est une chèvre». Si on se réfère un peu aux caractéristiques de l'animal, la chèvre est très capricieuse, elle saute d'une place à une autre, et elle est tout le temps en train de bêler contrairement au bœuf qui beugle une fois de temps en temps. On ne peut pas être en paix à côté d'une chèvre. Je trouve ça choquant et insultant. Ça veut dire que la femme c'est une personne bavarde, sotté, irréfléchie et bruyante, avec laquelle on ne peut pas se concentrer. On l'entend fréquemment quand les femmes sont ensemble.» (Awa Bocar Ly, sociologue);
- «En Tunisie, les mots **poisson** et **cinq** peuvent vexer car ils représentent des symboles très forts de protection contre le mauvais œil; on les fera suivre de *hachak* pour faire comprendre à l'interlocutrice(teur) qu'on ne nourrit pas de soupçons à son égard (*j'ai mangé du poisson hachak; je suis enceinte de cinq mois hackak*). Dans certains milieux, on aurait même entendu **fil de la mer** pour contourner **poisson** !» (Dalila Bejar, enseignante à la retraite).

### 3. — CONCLUSION

Pour conclure, on dira à la suite d'un chroniqueur du *Journal de Montréal*, les mots que les femmes détestent, «il y en a».

L'analyse n'en est qu'à ses débuts. Dans cet article, nous avons opté pour les catégories les plus fréquemment citées. Nous travaillons actuellement sur une classification qui reflèterait l'ensemble du matériel colligé. Nous pensons notamment à des réactions enregistrées relativement à la sonorité inélégante de certains mots (bru, marâtre, croûte, les suffixes -asse, etc.), à l'usage excessif des noms d'animaux pour insulter (une poule, un maquereau, le mouton noir de la famille, la brebis galeuse, un chien sale, etc.), à des mots relatifs à la famille (demi-frère, demi-sœur, belle-mère, concubinage, un enfant du deuxième lit, etc.), aux insultes, aux mots à consonance sexiste ou dégradante pour les femmes (hystérie, bloc honteux, grossesse, gestation). Des histoires personnelles sont également susceptibles

d'être à l'origine de la détestation de certains mots, relativement par exemple à une maladie (**halo**, **cancer**, etc.), ou à des tragédies familiales («les mots **père** et **mère** peuvent constituer des lieux de blessures» dit une poète), etc.

Certaines tendances se dessinent ou se confirment déjà même si elles ne transparaissent pas toutes dans les catégories présentées dans cet article. Les générations se divisent sur l'usage des sacres, des abréviations, ou des anglicismes. Il y a des mots que l'on aimerait pas à seize ans, d'autres à trente et d'autres plus tard : les collégiennes se sont montrées particulièrement choquées par les insultes, et les mots vulgaires auxquels elles doivent être plus confrontées que leurs aînées, les femmes de trente ans semblent allergiques au mot **madame**, les femmes de 40 ans et plus paraissant, quant à elles, beaucoup plus attachées aux normes linguistiques et sociales.

La linguistique «populaire» est une voie très agréable et très surprenante à explorer. Espérons que d'autres études de même nature s'effectueront, pourquoi pas, à travers la francophonie. N'est-ce pas un gage de bonne entente que de connaître *les mots que les femmes détestent et pourquoi ?*

## RÉFÉRENCES

- CALLIÈRES, François (1693) : *Des mots à la mode, et des nouvelles façons de parler : avec des observations sur diverses manières d'agir et de s'exprimer, et un discours en vers sur les mêmes matières*, 3<sup>e</sup> édition, Lyon, Hilaire Baritel.
- COIGNARD Sophie et Émilie LANEZ (1995) : «Le «politiquement correct» à la française», *Le Point*, n° 1186, pp. 52-61.
- PERGNIER, Maurice (1988) : *Le français en contact avec l'anglais. En hommage à Jean Darbelnet*, Paris, Didier Éruditions.
- RICHER, Stéphane et Lorna WEIR (Eds) (1995) : *Beyond Political Correctness. Toward the Inclusive University*, Toronto/Buffalo/London, University of Toronto Press.
- SCHNEIDER, Michel (1991) : «Changer les mots faute de changer les choses... Étrange et redoutable tentation.», *L'Express International*, n° 2098, p. 6.
- THIBODEAUX, David (1992) : *Political Correctness. The Cloning of the American Mind*, Lafayette (Louisiane), Huntignton Publishers.
- THOM, Françoise (1987) : *La langue de bois*, Paris, Julliard.
- VINCENT, Diane (1982) : *Pressions et impressions sur les sacres au Québec*, Québec, Éditeur officiel du Québec.



# Mot et mémoire grammaticale

**Abdelhamid CAMOUN**

*(Maître de Conférences à la Faculté  
des Sciences Humaines et Sociales de Tunis)*

**INTRODUCTION** : Nature de la communication et position du problème.

L'objet de cette communication s'inscrit en marge de ce qui nous semble être la préoccupation essentielle du Réseau : les problèmes liés au lexique, et plus spécialement ceux qui répondent à des applications d'ordre pratique (terminologie, traduction), qui intègrent, bien entendu des questions telles que la dénomination, la néologie, la formation des mots, la signification lexicale, l'évolution sémantique, due elle-même à des phénomènes de mutation inhérents à l'économie linguistique (métaphore, métonymie, ...) ou plus ou moins contingents (phénomènes sociaux proprement dits, rencontre de civilisations et de culture, bilinguisme ou multilinguisme, ...). A ce niveau, le mot est conçu comme une entrée de dictionnaire : le mot est un individu (un signe premier, isolable et définitivement isolé) «dictionnaire». Les questions soulevés par cet individu sont des questions qui préoccupent, au premier plan les lexicologues et les lexicographes, et, dans un second plan (second, selon un ordre de raison), les sociolinguistes, les théoriciens et les praticiens en matière de traduction, et enfin, dans un troisième plan et un peu obliquement, les philosophes du langage qui théorisent sur les notions, les concepts, les problèmes de dénomination,...

Il est tout à fait légitime de considérer que le mot est d'abord un individu lexical, et même plus proprement dictionnaire. Le mot, donnée première, isolée, nous sollicite d'abord, au niveau de sa signification, des différents sens qu'il peut avoir, des autres mots qu'il est susceptible de rappeler à notre conscience linguistique (mots parents morphologiquement, mots analogiques,...); bref, il est l'affaire du dictionnaire, et, empiriquement, dès que le mot est évoqué, notre imagination, nourrie par l'école, nous oriente vers la consultation du dictionnaire (alors que l'évocation de la phrase nous oriente vers la grammaire et les grammairiens).

Cependant, nul n'ignore que le mot n'est pas seulement un être lexical : outre son côté proprement «physique» (phonologique et graphique) qui assure son autonomie directement sensible, donc première en quelque sorte, le mot a une dimension grammaticale qui lui permet d'entrer nécessairement dans une classe grammaticale (*partie du discours*). Plutôt que d'évoquer ici les longues querelles qu'a suscitées la notion de mot, nous nous contentons de rappeler la définition qu'Antoine Meillet donne du mot, définition minimale mais assez juste : «Un mot est défini par l'association d'un sens donné à un ensemble donné de sons susceptible d'un *emploi grammatical donné*.» (Meillet, 1965 : T. Ier, p. 30) Cet emploi grammatical donné, par lequel A. Meillet semble clore la définition du mot, c'est ce que nous appelons *la charge grammaticale* du mot. Les deux expressions ici en italiques ont un caractère un peu vague, que nous maintenons délibérément et qui nous paraissent ainsi de nature à couvrir tous les aspects extra-lexicaux du mot (le caractère phonologique et graphique du mot étant également mis de côté). Ces traits extra-lexicaux, s'ils ne sont pas ignorés des lexicologues et des lexicographes, sont cependant considérés par eux comme seconds, et, en tout cas n'interviennent que d'une façon suffisamment globale ou épisodique pour ne pas déranger la préoccupation quasi exclusivement lexicologique qui est la leur. Ainsi, on fait état, à titre classificatoire de notions telles que *nom, verbe, adverbe,...*; si on évoque la *transitivité, l'auxiliarité, le genre,...* attachés à tel verbe, à tel nom, c'est toujours en rapport avec des problèmes de sens ou de correction d'emploi d'un mot donné. Le mot est ainsi installé dans les dictionnaires dans une sorte de quiétude qui serait dérangée par une évocation insistante de la dimension grammaticale. Certes, on semble ainsi se décharger de ces traits extra-lexicaux sur les grammairiens. Malheureusement, ceux-ci sont tellement préoccupés par la phrase que les traits grammaticaux attachés plus directement au mot sont

souvent incomplètement étudiés et implicitement confiés à l'attention des philologues, des stylisticiens ou des critiques (littéraires).

Mais, dira-t-on, c'est aux grammairiens de s'occuper de la dimension grammaticale des mots. Certes ! Mais, il faudrait que les grammairiens s'intéressent au mot en tant que mot, autrement dit à une réalité non encore phrastique quoique destinée à la phrase. Bref, nous pensons que le mot est à la fois une entrée de dictionnaire et une entrée de phrase. Que les lexicologues et les grammairiens, en spécialistes autonomes, s'intéressent, les uns quasi exclusivement au premier aspect, les autres quasi exclusivement au second, nous paraît légitime; mais, étant donné la nature complexe du mot, nous pensons qu'un champ de réflexion bi-dimensionnel (simultanément bi-dimensionnel) pour le mot aiderait à mieux en saisir la complexité et les nuances. De même que depuis quelque temps, on s'occupe de la sémantique de la phrase, il ne serait pas oiseux de s'occuper de la grammaire du mot.

La grammaire traditionnelle n'a pas négligé la charge grammaticale du mot; on lui a même reproché d'être une *grammaire du mot* plus qu'une *grammaire de la phrase*, reproche un peu immérité mais guère infamant, selon nous. Dans le mouvement linguistique moderne, qui a voulu renverser le courant en proclamant la primauté de la phrase, on peut considérer qu'un linguiste comme Gustave Guillaume s'est inscrit un peu à contre-courant en osant considérer que le mot, *être de langue*, est premier, et que la phrase, *être de discours*, est seconde; ce qui ne signifie nullement de sa part une ignorance ou une négligence du phénomène phrastique, mais plutôt que le mot, précisément, par sa charge grammaticale, *prévoit* en quelque sorte la phrase : la phrase est déjà, dans ses grandes lignes structurelles, qui sont du ressort de la langue, inscrite dans le mot qui, grâce à une espèce d'économie interne qui en fait une *partie du discours*, assure une prévisibilité de la phrase, dont le côté strictement imprévisible est, lui, du ressort du seul *discours*.

La charge grammaticale du mot est *intégrante* : elle embrasse tous les mots, autrement dit tous les signes linguistiques jouissant d'une autonomie psychique qui se traduit à l'écrit par une autonomie typographique. Le mot, en dehors de son côté proprement «physique» (phonologie, graphie), est doté d'une charge lexico-sémantique à laquelle s'associe, d'une manière obligée une charge grammaticale. Parler de la *mémoire* d'un mot, c'est, certes, d'abord, évoquer les traits lexicaux et sémantiques inscrits en lui, déposés en lui, génétiquement d'abord (tout comme les gènes dans un chromosome), mais aussi

historico-culturellement (ce qui suppose de multiples processus divers et qui sont du ressort de l'histoire, de la psychologie, de la sociologie, de l'économie du discours et de l'opération énonciative ou conversationnelle, etc.); mais c'est aussi évoquer d'autres traits, extralexicaux, ou grammaticaux, qui eux aussi ont une histoire génétique et historico-culturelle. Bref, le mot, tel qu'il se manifeste synchroniquement à nous, a, à côté de sa *mémoire lexico-sémantique*, une *mémoire grammaticale*.

C'est de cette *mémoire grammaticale* que nous nous proposons de parler dans cette communication. Le peu d'espace et de temps dont nous disposons ne nous permet qu'esquisser à grands traits ce que nous appellerons les différents *visages* de cette *mémoire grammaticale* des mots.

## 1 — LA MÉMOIRE GRAMMATICALE DES MOTS VIDES

Un premier visage de cette mémoire du mot est en rapport avec ce que nous appellerons la *grammaticité* du mot, autrement dit sa nature quasi purement grammaticale. La plupart des grammairiens distinguent les *mots pleins* et les *mots vides* (opposition due à J. Vendryes [1923 : 91 et ss. qu'il emprunte à la terminologie chinoise) : généralement, ils appellent, toujours à la suite de Vendryes, les premiers *sémantèmes*, les seconds *morphèmes*.<sup>(1)</sup> Or, nous savons que dans la terminologie grammaticale et linguistique française, on désigne par *morphème* (par opposition à *lexème* généralement), tout élément linguistique ayant une valeur purement grammaticale (alors que le *lexème* est l'élément linguistique désignant une donnée de l'expérience, autrement dit ayant un référent). Le *morphème pur*, appelé souvent *flexion* par les grammairiens, élément linguistique lexicalement vide et dont la signification est exclusivement grammaticale (expression d'une catégorie grammaticale telle que *genre, nombre, temps, mode, aspect*, ou d'un rapport syntaxique) est ce qui, dans un mot, suit le *lexème*, c'est un élément du mot, dont il est indétachable, qui n'a aucune existence autonome, ce n'est pas un mot. Le *morphème* dont il est question ici est un mot, un mot «vide» (lexicalement «vide»). Pour la plupart des grammairiens, les *mots vides* (ou *morphèmes*) sont les articles (et la plupart des déterminants), les pronoms, les prépositions, les conjonctions et les auxiliaires. Ces *mots vides* ont reçu, chez certains grammairiens, outre l'appellation de

---

1) Voir, outre Vendryes, G. Gougenheim (1969 : 47); L. Tesnière (1969 [1959] : 53); G. Le Bidois et R. Le Bidois, (1971 [1967] : I : 19); etc.

*morphèmes* (morphèmes - mots), celle de *mots - outils* (2), d'*outils grammaticaux* (3), de *mots grammaticaux* (4). Certains grammairiens ont, en outre donné à certaines sous-catégories de mot vides d'autres appellations, plus personnelles (5). Il faut noter que la plupart des grammairiens nuancent cette appellation de *mots vides*, en précisant, quelquefois, que la vacuité sémantique de ces mots n'est pas totale, mais que leur charge sémantique est néanmoins très faible, et, en tout état de cause, qu'elle est d'une nature un peu particulière. «La distinction des mots pleins et des mots vides n'est rigoureuse que dans certaines langues, en particulier le chinois. C'est d'ailleurs à la stylistique du chinois que sont empruntés les termes de mots pleins et de mots vides.», nuance L. Tesnière (6). J. Cervoni va plus loin en précisant : «Chez les grammairiens chinois, auxquels elle a été empruntée, l'opposition entre mots pleins et mots vides ne signifiait pas qu'il existât des mots vides de sens. Les mots vides étaient simplement des mots de sens moins concret que les mots pleins.» (7) La vacuité sémantique est généralement entendue comme la non référence aux données immédiates de l'expérience, et donc l'expression de rapports plus ou moins mentaux. Voici comment G. Gougenheim commente la notion de mots vides chez Vendryes. Après avoir cité celui-ci : «Le caractère du mot vide est l'abstraction; plus il s'affirme comme vide, plus il augmente sa valeur abstraite.», il précise : «Par abstraction, M. J. Vendryes entend sans doute que ces mots n'expriment pas des actions et ne correspondent pas à des

2) Voir F. Brunot, (1965 [1926] : 5); M. Grevisse (1964 [1939] : 66, § 104, n. 2), qui se réfère à A. Sauvageot : « On pourrait, avec A. Sauvageot (Franç. écrit, franç. parlé, pp. 45-52) ranger les mots en trois grandes classes : 1° les noms (substantifs et adjectifs); 2° les verbes; 3° les mots-outils (déterminatifs nominaux : articles, démonstratifs, possessifs, etc.; - déterminatifs verbaux (pronoms; - particules rectives : prépositions; - éléments articulatoires : conjonctions).», etc.

3) Voir Tesnière (1969 [1959] : 53); Le Bidois et Le Bidois (1971 [1967] : I : 20)

4) M. Grevisse et A. Goosse (1980 : 49)

5) Dans la terminologie de J. Damourette et E. Pichon, les morphèmes reçoivent le nom de taxioèmes [par opposition aux sémièmes, c.à.d. les mots pleins]. Les taxioèmes comprennent trois espèces : les flexions, les struments [= les mots vides, moins les auxiliaires] et les auxiliaires. On lit dans *des mots à la pensée* (1968 I : 80) : «Nous désignons par cette appellation [les struments] les taxioèmes formant vocables indépendants. (...) L'auxiliaire est un sémième ayant une pleine existence, mais qui se trouve temporairement dégradé, parce que, vu, sous un certain aspect, son sémième [= signifié du sémième] devient taxième [= signifié du taxioème]».

6) L. Tesnière (1969 [1959] : 55-56)

7) Cervoni, J. (1991 : 128 )

notions concrètes. D'après le contexte, nous supposons qu'il a eu en vue essentiellement des notions grammaticales (opposés aux concepts lexicaux); D'autre part, il admet des degrés dans le caractère «vide» des mots : un mot peut être plus ou moins vide, sans doute selon que la notion grammaticale est plus ou moins précise ou plus ou moins concrète.» (8)

Les *mots vides* posent beaucoup de problèmes, quant à leur identification, à leur signification précise, à la nature de leur vacuité sémantique, à leur place dans le système des signes linguistiques, à leur rôle précis dans le discours, aux rapports qu'ils entretiennent, et dans la langue et dans le discours avec les mots pleins, ... Ce qui nous intéresse ici, c'est de montrer qu'ils ont une mémoire grammaticale très forte. Cette mémoire est d'abord dans le port de cette charge grammaticale dont sont dépourvus les *mots pleins*. La conscience linguistique du locuteur a mémorisé cette opposition entre *sémantèmes* et *morphèmes*, confiant ainsi à chacune des deux catégories des champs conceptuels différents, opposés. Nous aurons, plus loin, le loisir, de préciser un peu plus cette opposition, qui, prise au plan le plus général, assure aux mots vides une place à part aboutissant, lors de leur évocation, un déclenchement mémoriel particulier.

L'opposition, au plan le plus général, entre *mots pleins* et *mots vides* nous fournit ainsi le premier visage de ce que nous avons appelé la *mémoire des mots*. Nous sommes ainsi en présence, au plan du système des signes linguistiques, et plus précisément de celui des *parties du discours*, de deux aires opposées : l'aire lexicale et l'aire grammaticale, que notre mémoire linguistique nous aide à différencier. A ce niveau, la mémoire linguistique opère un peu brutalement, et dans un premier temps (celui de la pure distinction des aires), sans nuance, ou presque. Les nuances viennent par la suite, quand il s'agit d'entrer dans le détail des propriétés des différentes sous - espèces de mots pleins et de mots vides, et des différents mots

---

8) G. Gougenheim (1970 : 19-20). Ce sont surtout certaines prépositions (à, de) que beaucoup de grammairiens considèrent comme des mots véritablement vides. Ebbe Spang-Hanssen a écrit un ouvrage intitulé *Les prépositions incolores du français moderne* (1963). L'expression préposition incolore, empruntée à W. von Wartburg et Paul Zumthor [*Précis de syntaxe du français contemporain*, Berne, 1958], est à peu près synonyme de celle de préposition vide (préposition très abstraite). Ebbe Spang-Hanssen a réduit le nombre des prépositions incolores à trois (de, à, en), plus nombreuses chez Wartburg et Zumthor. Dans son article, cité ci-dessus (1970 : 39). G. Gougenheim, conclut que des deux prépositions à et de, appelées vides par certains grammairiens, «de est la seule préposition «vide» de la langue française.

(pleins ou vides). Mais d'ores et déjà, la conscience attribue à chacune des deux grandes espèces, une mémoire d'un type particulier : une *mémoire lexicale* (ou lexico-sémantique) aux *mots pleins*, une mémoire grammaticale aux *mots vides*. Que les mots vides ne soient pas entièrement vides, ceci n'est qu'un trait, un peu particulier comme nous le verrons, de cette mémoire grammaticale.

## 2 — LA MÉMOIRE GRAMMATICALE DES MOTS COMPOSITES

Cependant, nous savons qu'un mot, très souvent, comprend à la fois un *lexème* et un *morphème*, autrement dit un élément inférieur au mot (lexème) de nature lexicale, généralement en première position, et un autre élément inférieur au mot (*morphème*), de nature grammaticale, généralement second et attaché graphiquement au mot (*flexion*). Les mots à deux *monèmes* (selon la terminologie d'A. Martinet), sont les noms (substantifs et adjectifs) et les verbes, et vraisemblablement certains adverbes. Le *monème grammatical* (*morphème*, autrement appelé *flexion*), permet d'asseoir le concept lexical dans des catégories grammaticales. Ainsi le *genre* et le *nombre*, attachés au nom, des catégories telles que le *mode*, le *temps*, l'*aspect*,... attachés au verbe, sont exprimés par les morphèmes. A vrai dire, l'identification et le rôle du *morphème* faisant corps avec le mot, ont été plus rapidement et plus sûrement étudiés par les grammairiens. Il n'empêche que les *morphèmes intégrés* (flexions) et les *morphèmes isolés* (mots vides), méritent, les uns et les autres l'appellation commune de morphèmes, dans la mesure où les uns et les autres ont une valeur extra-lexicale, autrement dit grammaticale. Cependant, il y a, entre les deux espèces de morphèmes, au moins ces deux différences :

1 — Alors qu'on est à peu près certain que dans les morphèmes intégrés (flexions), tout est extra-lexical, on a vu que le problème de la vacuité sémantique des mots vides appelle des nuances;

2 — Les mots vides (ou du moins certains d'entre eux) sont plus chargés d'histoire et sont plus souvent polyvalents que les flexions, généralement moins fluctuantes.

L. Tesnière après avoir distingué mots pleins (ou sémantèmes) et mots vides (ou morphèmes), introduit la notion de mots composites : «Beaucoup de langues, et notamment nos langues d'Europe (...) réunissent souvent dans le même mot des éléments

pleins et des éléments vides (...). Nous qualifierons les mots de ce type de composites.» (9)

Ces mots composites, si nous adoptons la terminologie de L. Tesnière, posent de nombreux problèmes. Le premier, et non des moindres, a rapport à l'identification même du mot. Etant donné plusieurs formes à lexème commun et appartenant à une même partie du discours (nom ou verbe, généralement), dira-t-on qu'on a affaire à un seul mot à plusieurs formes ou à autant de mots que de formes ? Y a-t-il un seul mot dans *chien, chiens*, ou deux mots ? De même dans *chanter, chanterai, chantions, chantasses, ...*? Autrement dit, une commutation de morphème de même paradigme crée-t-elle un autre mot, ou maintient-elle la forme nouvelle dans le même mot ? Sur ce point, les grammairiens sont divisés. Pour L. Bloomfield, E. Buysens, G. Guillaume, L. Tesnière, chaque forme compte pour un mot. *Chanter, chanterai, chantions, chantasses*, sont quatre mots différents (les mots composites de Tesnière). Buysens, parlant des formes verbales, s'exprime ainsi : «On définit la conjugaison d'un verbe comme l'ensemble des formes que peut prendre ce verbe; en réalité, il ne s'agit pas d'un mot qui change de forme, mais d'un ensemble de mots qui ont en commun une partie de leur forme et une partie de leur signifié (...)» (10). Pour J. Lyons, au contraire, ces différentes formes constituent un seul mot : «Pour Bloomfield, par exemple, la tradition scolaire a tort de dire que des unités comme *livre, livres*, ou *aimer, aimons, aimâtes*, etc... sont des «formes différentes du même mot». Mais c'est Bloomfield qui a tort. C'est à nous de décider quel sens nous voulons donner au terme mot. (...) C'est bien entendu au sens abstrait que le terme mot est employé en grammaire classique.» Et de regretter un peu plus loin : «Il est malheureux que la linguistique moderne n'ait pas gardé la définition traditionnelle du mot comme l'unité plus abstraite.» (11) Lyons, suivant en cela la grammaire traditionnelle, comme il le dit explicitement, adopte cette position pour des raisons de théorie linguistique certes, mais, ce faisant, il semble identifier le mot comme une entrée de dictionnaire, avant tout. Le mot, pour lui, comme pour la tradition, est d'abord un *individu dictionnaire*. J. Damourette et E. Pichon, pour leur part,

---

9) Tesnière (1969 [1959] : 54) Cependant, il ajoute quelques lignes plus loin : «Il va sans dire que les mots composites doivent être considérés syntaxiquement comme des mots pleins, puisqu'ils sont chargés, de par leur élément plein, de la fonction sémantique inhérente aux mots pleins.»

10) Buysens, E. (1975 : 49)

11) Lyons, J. (1970 : 152)

distinguent *mot* et *vocable* : «Les mots contenant une flexion sont de ce fait, porteurs et d'un sémième [= signifié lexical] qui leur est propre et d'un taxième [= signifié du morphème] attaché à leur flexion. Aussi existe-t-il différents mots contenant le même sémième, et se différenciant seulement par des flexions différentes, c'est-à-dire au point de vue sémantique par des taxièmes différents. Tous ces mots ayant un sémième commun sont dits appartenir au même vocable, et leur ensemble ne donne lieu qu'à un seul article de dictionnaire. C'est ainsi que *cheval* et *chevaux* sont deux mots du vocable *cheval*; *aimer*, *aimons*, *aimeriez*, *aimant*..., sont des mots du vocable *aimer*; *vais*, *allons*, *irai*..., sont des mots du vocable *aller*.» (12) Pour notre part, nous privilégierons, pour des raisons que nous expliquerons plus loin, le point de vue de la pluralité des mots (un mot par forme).

Quoi qu'il en soit, qu'on parle de formes différentes d'un même mot, ou de mots différents, à ce niveau, la mémoire grammaticale du mot est patente. Entre ces mots à même lexèmes (ou ces formes différentes d'un même mot abstrait) s'instituent des rapports morphologiques structurant des paradigmes et donc formant des espèces de micro-systèmes, morphologiques d'abord, et à signification grammaticale ensuite. Au plan morphologique, la mémoire, au sens le plus strict du mot, opère pour associer et distinguer les formes appartenant au même micro-système et également pour associer et distinguer ces micro-systèmes dont l'assemblage constitue un système plus grand. Il va sans dire qu'à ce niveau, la mémoire grammaticale (qui permet de retenir et de distinguer les formes d'un micro-système ou d'un système), opère sur deux plans, simultanément : celui des morphèmes constitutifs du paradigme (ou de l'archi-paradigme), et celui des mots *composites* (les *vocables* de Damourette et Pichon) constituant micro-système ou système. On aura ainsi, simultanément, en mémoire, d'un côté, le paradigme purement morphématique constitutif de l'imparfait de l'indicatif : *ais - ais - ait - ions - iez - aient*; mais aussi, parallèlement et par association mécanique du paradigme à tel lexème : *chantais - chantais - chantait - chantait - chantions - chantaient*. Plus concrètement, il y a même un troisième plan mémoriel qui associe, aussi mécaniquement, ces formes à des particules préverbaux (*je chantais - tu chantais - il chantait - nous chantions - vous chantiez - ils chantaient*). Cette mémoire, de type purement morphologique, est première et constitue une sorte d'assise pour une mémoire de type plus mental et plus complexe. Elle opère spontanément et

---

12) Damourette et Pichon (1968 : I : 82)

empiriquement pour les natifs de la langue, l'école intervenant après coup d'une manière énergique et autoritaire pour l'installation et la mémorisation de ces paradigmes; pour les non natifs, l'école est l'instance quasi exclusive où puisse opérer cette mémoire (aidée certes par les mécanismes cognitifs de l'apprenant).

A ce niveau, celui de ce qu'on pourrait appeler la *mémoire paradigmatique*, des complications peuvent intervenir. Nous citerons celles occasionnées par les *allomorphes lexématiques* et par les *allophones*. Ainsi, par exemple, dans les paires lexicales *coq / poule* ou *porc / truie*, on a affaire à des paradigmes apparemment purement lexicaux, mais dont la distribution est grammaticale (la catégorie du *genre*), laquelle est sous l'effet d'une mémoire grammaticale. Plus apparents sont les allomorphes de lexèmes verbaux qui viennent intensifier la mémoire grammaticale en créant une espèce de *surcharge paradigmatique*. Ainsi, les allomorphes du verbe aller : *vais, allai, irai, aille*. La conscience du locuteur mémorise, d'une part le paradigme allomorphique, et d'autre part le paradigme flexionnel. Les allophones, eux, aussi, augment la mémoire paradigmatique : le pluriel des oiseaux fait intervenir à la fois le morphème X (paradigme du nombre : Zéro/ X [X étant lui-même un allomorphe morphématique]) et le morphème Z (allophone à l'origine).

A un niveau supérieur, la mémoire grammaticale du mot composite intervient pour assurer l'emploi juste du mot en fonction de sa valeur et de sa fonction syntaxique (fonction au sens large). La valeur et l'emploi du mot résultent de ses potentialités grammaticales, qui sont en quelque sorte la résultante de son appartenance à telle partie du discours et des signifiés grammaticaux de son morphème, souvent complexe (tel cas, tel genre, ou tel mode, tel aspect, tel temps ...). Une opération mentale à la fois analytique et synthétique, tantôt pré-discursive, tantôt discursive, est, outre l'intention proprement énonciative, derrière l'entrée en phrase du mot composite. C'est à ce niveau, et pour assurer cette opération qu'intervient cet autre visage de la mémoire grammaticale du mot, précédé nécessairement par la mémoire plus mécanique dont il a été question plus haut. Ce nouveau visage mémoriel intervient dans le choix de telle forme compte tenu des possibilités syntaxiques, des nuances, des niveaux de langue, des choix stylistiques, ...

### 3 — MÉMOIRE GRAMMATICALE ET PARTIE DU DISCOURS.

Une manifestation importante de la mémoire grammaticale du mot est, sans aucun de toute dans son catalogage dans les parties du

discours. Quoi qu'en pensent certains, la distribution des mots en classes, en parties du discours, pour reprendre l'expression consacrée, n'est ni simplement formelle, ni due à une espèce d'empirisme primaire qui serait dû lui-même au hasard. La théorisation en matière de classification catégorielle des mots est loin d'être oiseuse; elle suppose toute une philosophie du langage et permet même, aux yeux de certains linguistes, de classer les langues en plusieurs types. Un linguiste qui est allé très loin, dans ce sens (trop loin au gré de certains), est Gustave Guillaume. Celui-ci a essayé de «typologiser» les langues, selon les types de mots qui les composent. Une partie du discours (appelée partie de langue dans la terminologie guillaumienne) a une genèse qui explique son économie interne; et c'est précisément cette économie interne qui permet, d'une part, d'expliquer, dans une langue donnée, le système entier des parties du discours, et par là même de tous les mots qu'elles intègrent; d'autre part, de comprendre la place qu'occupe la partie du discours (et par là même le mot) dans le fonctionnement discursif de la langue, autrement dit le niveau d'intervention du mot dans la phrase, et, par voie de conséquence de montrer que les différentes langues ont des «logiques» de fonctionnement qui varient selon la configuration et la place du mot (son niveau d'intervention), qui trouvent elles-mêmes leur explication dans la genèse des différentes parties du discours d'une langue donnée, intégrant à l'égard des mots constitutifs de cette langue. Il va sans dire que les problèmes sont complexes, dont le moindre n'est pas la notion de mot, dont la conception, la définition, risquent d'amener le linguiste à une certaine subjectivité due à un certain «européo-centrisme». Ce danger a été prévu par notre linguiste. Quoi qu'il en soit, pour lui, il y a, aux deux extrêmes de la typologie, deux types de langues : d'une part, les *langues holophrastiques* ou langues à *mots-phrases* (historiquement, c'est le type le plus primitif, type plutôt théorique, mais dont se rapprochent des langues agglutinantes comme le basque et le hongrois), langues où la distinction entre la phrase et le mot est encore indistincte, bref où le mot est phrase et la phrase mot; d'autre part, les *langues à caractères* (tel le chinois), où les éléments formant phrase n'ont pas acquis la dimension formalisante qui en ferait des parties du discours (13). Des langues comme le français, l'anglais, ... (d'une façon générale les langues modernes d'origine indo-européenne) sont des langues où le mot est entièrement formé en langue (avec quelques rares exceptions). Dans ces langues, le mot est un être de langue : la

---

13) Voir de G. Guillaume, principalement les *Leçons de linguistiques* (1948-1949, Série B, pp. 33-40). Voir aussi A. Camoun (1992 : 197)

langue est formé de mots entièrement institués et prêts à entrer tels quels en phrase. La phrase, elle est en principe du ressort du discours cependant, elle est, dans son économie structurelle, prévisible : il y a un moment pré-phrastique, suspensif en quelque sorte, et où se dessinent les deux dimensions de la phrase : la dimension prévisible et la dimension imprévisible. C'est cette dernière dimension qui est véritablement du seul ressort du discours; la première est prévisible dans la mesure où elle est l'effet immédiat de ce qui ressortit à la langue, ce qui veut dire qu'elle est inscrite dans l'économie du mot en tant qu'appartenant à telle ou telle partie du discours. Le mot, de par son appartenance à telle partie du discours, inscrit en lui-même sa future syntaxe. En d'autres termes, le mot français (ou anglais, ...) est à la fois un être lexical et un être grammatical, au sens le plus fort du mot, c.à.d. qu'il est porteur, non seulement de valeurs grammaticales perceptibles dans son morphème, mais aussi de sa syntaxe (fonction dans la phrase). Et, en effet, dans la psychosystématique (linguistique guillaumienne), le mot en tant que partie de langue (= partie du discours) a une genèse (qui n'est ni historique, ni étymologique, mais synchronique et mentalement structurelle) dont le mécanisme général est ce que G. Guillaume appelle le tenseur binaire : deux mouvements mentaux successifs, le premier, rétrécissant (qui va du large à l'étroit, ou de l'universel au singulier) est appelé le discernement; le deuxième, généralisant (qui va de l'étroit au large, ou du singulier à l'universel) est appelé l'entendement. Le discernement est matérialisant : il produit la matière du mot (son côté lexical : le concept d'un être ou d'un procès, ...), appelée sémantème; l'entendement, lui, est formalisant : il produit le moule grammatical (autrement dit la partie de langue) dans lequel sera versé le sémantème. Ce moule grammatical (la partie de langue), est le résultat de l'entendement porteur de catégories grammaticales particulières mais en affinité avec le type de sémantème choisi, et doit être suffisamment général pour pouvoir recevoir n'importe quel sémantème qui soit en affinité avec la partie du discours; ce qui veut dire que se dessine une sorte de corrélation entre le sémantème (signifié du lexème) et la partie du discours concernée. Outre les catégories grammaticales, ce qui formalise le mot (ce qui lui donne sa classe grammaticale), c'est le mécanisme d'incidence (apport d'une matière notionnelle à un support formel), producteur des rapports syntaxiques entre les parties de langue : ainsi, par exemple, le substantif a une incidence interne en langue, qui devient externe en discours, il est incident à l'article; l'adjectif a une incidence externe : il est incident au substantif; le verbe, lui, a une ou deux incidences externes : il est incident au substantif sujet, et peut être incident à un

objet, etc. C'est donc dans l'entendement, d'abord, que se situe la mémoire grammaticale du mot, puisque l'entendement est porteur des catégories qui formalisent le mot et, le situant dans la partie du discours auquel il a droit, lui assigne une couleur grammaticale. Cette mémoire grammaticale est, comme on le voit, d'essence génétique. Un mot, de par son appartenance à une partie de langue (partie du discours) emporte avec lui une mémoire : la conscience du locuteur a mémorisé des catégories grammaticales qui se superposent ou s'interpénètrent et aussi une prévision de sa fonction syntaxique dans la phrase, qui reste bien entendu puissancielle au niveau de la langue et qui se résout dans le discours en telle fonction précise, inscrite dans le mécanisme d'incidence mais dépendant aussi du contexte.

Cependant, le discernement joue, lui aussi, dans certains cas, un rôle grammatical, et se trouve alors, lui aussi, porteur d'une mémoire grammaticale. Ceci se produit dans les fameux mots vides des grammairiens. Nous surprenons ici l'économie tout entière du système des parties du discours. Pour G. Guillaume, il y a deux grandes espèces de parties de langue : les parties de langue prédicatives et les parties de langue non prédicatives. Les parties de langue prédicatives construisent leur sémantèse dans le cadre d'une matière grave (ou lourde) : concept tiré de notre expérience de l'univers (le hors-moi); les parties de langue non prédicatives construisent leur sémantèse dans le cadre d'une matière subtile : la pensée pensante. Leur sémantèse est une forme en position de matière : il s'agit en quelque sorte de l'expression d'un rapport que la pensée se construit à partir d'elle-même, de son propre fonctionnement. En d'autres termes, les parties de langue non prédicatives, dont le rôle est essentiellement grammatical, ne sont pas pour autant dénués de signification; simplement, celle-ci appartient à une autre sphère mentale que celle des parties de langue prédicatives. Leur signifié (de puissance, au plan de la langue, qui se résout en signifiés d'effet au plan du discours), est de type abstrait, subtil : c'est une forme de pensée, ce qui signifie, d'une part que ce signifié lui-même est de type grammatical, d'abord; mais d'autre part que ce signifié garde quelque rapport avec la sémantèse des parties de langue prédicative, et c'est en cela que c'est une forme, mais en position de matière. C'est ainsi que certaines parties de langue non prédicatives sont conçues comme obtenues par transcendance de parties prédicatives, par évidemment notionnel, ou plus exactement transformation d'une sémantèse lourde en sémantèse subtile, l'incidence (prévision syntaxique) restant la même : ces parties sont dites transprédicatives (ainsi, par exemple, l'article et le pronom sont

obtenus par la transcendance du substantif). Il va sans dire que, dans la genèse des parties non prédicatives, le tenseur binaire opérant, l'entendement sous-tendu par des catégories donne lieu, en fin de parcours, à des catégories de mots dont le signifié est en définitive d'abord d'ordre grammatical et dont le rôle est d'abord un rôle grammatical. C'est ce qui fait la complexité de ces parties non prédicatives, de ces mots vides (14).

Pour revenir à la typologie des langues selon G. Guillaume, typologie dont la base est le statut du mot dans la langue, le linguiste considère que, à la différence des langues d'origine indo-européenne où le mot est entièrement formé en langue, est un être de langue, les langues sémitiques sont des langues qui appartiennent à un autre type, à une autre espèce de langue, dans la mesure où le mot ne s'y présente pas entièrement formé au plan de la langue (par opposition au discours) : dans ces langues, en effet, le mot résulte de deux saisies lexicales : la première, exophrastique, appartient au plan de la langue; elle livre la matière du mot traduite sémiologiquement par la racine tri-consonantique (ex. : k-t-b-); la deuxième, endophrastique, appartient au plan du discours et formalise le mot, ce qui se traduit sémiologiquement par l'adjonction de voyelles. Dans ces langues, le mot commence en langue et s'achève (ou se parachève) en discours. Dans certaines de nos recherches, nous nous sommes intéressé à la théorie du mot arabe, en essayant de lui appliquer la thèse guillaumienne; nous avons montré, en nuancant un peu la théorie, la justesse des vues de G. Guillaume touchant le mot arabe (15). Ce qui nous intéresse ici, c'est un point qui nous semble d'une grande portée : si dans certaines langues (le français, l'anglais, ...), le mot, entièrement formé en langue, mais qui déjà pose quelques problèmes (mots à lexème commun = un seul mot ou plusieurs mots ?), dans d'autres langues, comme l'arabe, où le mot n'est pas toujours entièrement formé en langue, l'appartenance du mot au double plan de la langue et du discours est de nature à le doter d'une mémoire grammaticale un peu plus complexe; cette mémoire est, curieusement, à la fois plus légère et plus lourde que celle du mot français : plus légère et quasi nulle, si l'on envisage le plan de la racine consonantique, plan de la langue où tout est matière, mais bien plus lourde côté discours, puisqu'elle concerne, dans un premier temps,

---

14) Pour plus de détails sur le système des parties de langue dans la linguistique guillaumienne, voir G. Guillaume (1973 [1939a] : 86-98); et G. Moignet (1981 : 12-26)

15) Voir G. Guillaume (1973 [1939b] : 107-119; et A. Camoun (1992 : 177-185; 199-215; 230-236).

l'achèvement formel du mot, et dans un deuxième temps, sa véritable entrée en phrase. Dans une langue comme l'arabe, à côté d'une racine dépouillée, réduite au strict minimum lexical, se dresse une accumulation de données grammaticales qui se bousculent pour donner au mot, d'abord son appartenance à telle partie du discours (kataba : verbe - kitab : substantif, ...) ensuite, son sens final (avec tout ce qu'il suppose de polysémie, de connotations, etc.) et le doter aussi des catégories grammaticales et de sa fonction phrastique. Dans le mot arabe, la valeur lexicale proprement dite (où la matière diffluyente de la racine est dépassée) est, dans une mesure beaucoup plus grande que dans le mot français, liée à sa forme. La partie du discours est plus marquée grammaticalement.

#### 4 — MÉMOIRE GRAMMATICALE ET MODIFICATION CATÉGORIELLE

Un autre visage de la mémoire grammaticale des mots est en rapport avec un phénomène, à manifestations multiples et auquel on pourrait peut-être donner le nom de modification catégorielle (faute d'un autre terme plus approprié). Par ce terme, nous entendons toute sorte de changement intervenant au niveau du mot, en dehors du phénomène de glissement de sens, changement supposant une intervention extra-lexicale, donc grammaticale (au sens très général du mot). Appartiennent à ce phénomène général, des opérations aussi diverses que la composition, le figement, la production de locutions (pleines ou vides), le changement catégoriel, lequel peut intervenir, soit au niveau de la langue (lever => le lever), soit au niveau du discours (des pommes frites => des frites), et peut sous-tendre la composition (peut être : 2 verbes => peut-être : adverbe), la translation (selon la terminologie de Tesnière), dont nous retiendrons surtout la translation du second degré consistant, essentiellement, à considérer les propositions subordonnées d'une phrase complexe comme les équivalentes de mots simples ayant la même fonction dans la phrase <sup>(16)</sup>, les noms de discours (= les phrases devenues subordonnées et jouant le rôle d'un substantif, d'un adjectif ou d'un adverbe, chez Guillaume <sup>(17)</sup>, la subduction (notion guillaumienne, qu'on peut définir très schématiquement comme une espèce d'évidement matériel [= lexical] que la psychosystématique explique par une intervention précoce de l'entendement à la suite d'un discernement interrompu, ou écourté, et qui est au départ du

---

16) Voir Tesnière, (1969 [1959] : 543-617)

17) Voir Moignet (1981 : 247)

phénomène de l'auxiliarité, entre autres), et d'autres opérations où intervient un changement (extra-lexical) affectant le mot, à quelque niveau que ce soit (langue ou discours) et de quelque nature qu'il soit, pourvu que l'esprit, à l'occasion de son fonctionnement, évoque le mot. Nous n'avons malheureusement pas le loisir d'entrer dans le détail de chacun des aspects de cette modification affectant le mot, mais nous soulignons avec force que, quel que soit le cas, la grammaire intervient, et chaque fois, la conscience du locuteur est chargée d'une mémoire grammaticale propre au mot. Ainsi, pour ne prendre qu'un exemple, celui de la subduction affectant le verbe être, l'évident matériel comportant des degrés est de nature à charger ce verbe d'une lourde mémoire grammaticale propre à expliquer ses emplois très divers (verbe plein : *Dieu est* - verbe attributif : *Paris est la capitale de la France* - verbe copule : *Pierre est malade* - auxiliaire de la voix passive : *Pierre est puni par l'instituteur* - auxiliaire de l'aspect : *Pierre est tombé; Pierre s'est trompé*) (18). Dans ce phénomène de la modification, les mots vides jouent souvent un rôle de premier plan, lequel permet souvent de les définir : ainsi, les conjonctions de subordination, propre à transformer la phrase en nom de discours, sont considérés comme des nominalisateurs. Dans le système total des parties du discours (dans une perspective guillaumienne) on voit ainsi le rôle très subtil dévolu à cette classe de mots : ramener un fait de discours (la phrase) à un phénomène de langue (en l'occurrence une partie du discours : le nom [au sens général : mot plein non verbe]). De même, le phénomène de translation montre comment telle préposition ou telle conjonction de subordination est propre à ramener un syntagme prépositionnel ou une subordonnée à une partie du discours (substantif, adjectif, adverbe). Ainsi, grâce à ces mots tenus, vides, l'économie du discours se retrouve dans la catégorisation ressortissant à la langue, bref dans la logique première de la langue, pré-discursive, pré-syntaxique. C'est dire le grand rôle joué par ces mots vides, et plus précisément par les plus vides d'entre eux (les autres conservant une bonne charge lexicale, quoique d'une nature un peu particulière).

## 5 — MÉMOIRE GRAMMATICALE ET HISTOIRE

Il faut souligner, d'une façon très générale le rôle du facteur historique (ou diachronique) dans le façonnement de la mémoire grammaticale du mot. L. Tesnière souligne une tendance historique

---

18) Voir G. Moignet, auquel nous avons emprunté également les exemples (1981 : 265 et ss.) Sur la théorie de la subduction, voir G. Guillaume (1973 [1938] : 7 3-86)

générale allant dans le sens de la grammaticalisation des mots : «Au cours du développement historique des langues, les mots pleins ont tendance à se vider de leur contenu sémantique pour devenir les outils grammaticaux que sont les mots vides. Aussi nombre de mots vides s'expliquent-ils comme d'anciens mots pleins, tandis que l'inverse ne s'observe jamais (...)» (op. cit., p. 55, chap. 28, § 13) Cette affirmation doit être considérablement nuancée certes, mais elle est juste dans une très large mesure, et ainsi présentée, elle explique en grande partie la formation d'une grande quantité de locutions conjonctives, prépositives, adverbiales, etc. L'histoire est un facteur essentiel dans les phénomènes de composition, de figement, à orientation de plus en plus abstraite, donc grammaticale. Ce qui le prouve, c'est que souvent on assiste à une formation en devenir de locutions grammaticales. Ainsi, si des locutions telles que *parce que*, *bien que*, *tant que*, sont complètement grammaticalisées, on ne peut en dire autant de locutions comme *en attendant que*, *de manière que* (unités en voie de formation, non encore tout à fait synthétisées et où l'analyse est encore concevable : *en attendant* [verbe au gérondif] + *que*; *de manière* [substantif] + *que*). La part véritable de l'histoire se laisse percevoir surtout dans ces situations intermédiaires où l'on voit se dessiner clairement cette tendance à l'évidement lexical qui préside à la grammaticalisation des mots. Mais même dans les cas des locutions entièrement grammaticalisées, le mot ainsi formé est vu comme soumis à une mémoire grammaticale à base historique, laquelle permet d'analyser les éléments constitutifs du mot synthétisé.

## 6 — MÉMOIRE GRAMMATICALE ET COMPLÉMENTARITÉ STRUCTURELLE

Un autre visage de la mémoire grammaticale du mot est décelable dans ce que nous appellerons la complémentarité structurelle. Cette complémentarité se manifeste, d'abord dans les structures analytiques, ou structures à deux temps. Nous voulons parler d'une certaine tendance d'une langue comme le français à s'exprimer, au plan syntaxique, d'une manière progressive, ce qui suppose une distribution sémantico-logique à base syntaxique. Examinons des structures comme les structures attributives, les structures passives, ou encore les syntagmes prépositionnels compléments circonstanciels ou la plupart des subordonnées circonstancielle. Dans les phrases : *Paul est triste* - *Paul est puni* - *Les enfants jouent dans la cour* - *Paul s'est absenté parce qu'il est malade*; les segments : *est malade*, *est puni*, *dans la cour*, *parce qu'il est malade*; sont des structures à deux temps. Chacun des segments :

*est malade, est puni*, exprime dans sa totalité un état; mais cet état est analysable ainsi : *est* présente l'état dans sa généralité; *malade, puni*, expriment, chacun, un état particulier qui vient s'inscrire dans cet état général qui lui sert d'assise logique. De même, chacun des segments : *dans la cour, parce qu'il est malade*, exprime une circonstance; mais cette circonstance est analysable ainsi : *dans, parce que*, présentent, chacun telle circonstance dans sa généralité (*lieu intérieur* général, *cause* dans sa généralité), dans le cadre de laquelle vient s'installer la circonstance correspondante dans sa particularité, circonstance particulière exprimée respectivement par le SN *la cour* et la phrase *il est malade* (lieu intérieur précis, cause précise). Il y a, entre les deux éléments de la totalité fonctionnelle, une complémentarité dont la logique interne est : généralité formelle + particularité matérielle. La première composante est plutôt grammaticale (et peu sémantique); la deuxième est plutôt sémantique et peu grammaticale. C'est cette complémentarité qui crée entre les deux mots (mots de langues ou de discours) une mémoire grammaticale, inscrite en langue (dans l'économie du système linguistique pré-phrastique) et qui doit se manifester et opérer en discours.

Cette complémentarité structurelle existant entre deux mots se manifeste également et d'une façon encore plus étroite dans des ensembles synthétiques constituant mots institués en langue; nous voulons parler essentiellement des formes verbales du français constituées de plus d'un élément verbal. Appelons-les formes pluriverbales, ou mieux, formes verbales périphrastiques : ce sont les formes composées et les formes surcomposées de la voix active (ex.: *j'ai aimé; j'ai eu aimé*); les formes composées de la voix pronominale (ex. : *je me suis lavé*); la totalité des formes de la voix passive : formes minimales et formes maximales (ex. : *je suis puni; j'ai été puni*) et les périphrases verbales (ex. : *je vais chanter; j'allais chanter*) (19). Dans les formes verbales périphrastiques, le premier élément est un auxiliaire ou un semi-auxiliaire, verbe subduit, évidé de sa matière (tout en conservant un résidu sémantique de nature abstraite), le deuxième élément est une forme quasi-nominale (participe passé, infinitif) dont l'office est d'exprimer l'unique procès de la forme verbale. On peut représenter cette complémentarité ainsi : Forme totale = Auxiliaire ou semi-auxiliaire, élément faible sémantiquement (sens voisin de zéro), fort syntaxiquement + Forme

19) Nous distinguons deux types de périphrases verbales : les périphrases qui ressortissent à la langue (par exemple les tempo-aspectuelles, qui se construisent avec aller comme semi-auxiliaire et les factitives) et celles ressortissant au discours (par exemples les modales).

quasi-nominale, élément faible syntaxiquement, fort sémantiquement. Le premier élément donne à la forme sa syntaxe et sa valeur (qui découle de son résidu sémantique), le deuxième lui donne sa matière (expression du procès). Ces formes périphrastiques sont des mots synthétisés mais structurellement analysables, et c'est précisément leur mémoire grammaticale, une mémoire très forte qui permet de les construire en langue.

## 7 — MÉMOIRE GRAMMATICALE ET AVENTURE STYLISTIQUE

Nous terminons cette revue des visages de la mémoire grammaticale des mots (revue non exhaustive) par l'évocation rapide d'un visage un peu particulier, et qui est en rapport avec ce que nous appellerons pittoresquement l'aventure stylistique du mot, laquelle aventure est d'origine diachronique et en rapport avec les niveaux de langue. Elle se manifeste par exemple dans l'emploi de certaines formes verbales, comme l'imparfait et le plus-que-parfait du subjonctif, lequel emploi est senti comme plus ou moins naturel dans un niveau littéraire ou académique mais comme manifestement déplacé dans le niveau standard. En réalité, les choses sont plus complexes, les écrivains modernes (et également les journalistes, les universitaires, etc., bref les intellectuels dans leurs écrits), n'adoptant plus, en matière de niveau de langue une attitude commune (comme c'était à peu près le cas dans les siècles passés); et c'est dans cette mesure (l'éclatement relatif des niveaux) qu'on est en droit de parler de choix, donc de style. Il y a, en matière d'emploi de certaines formes verbales, en gros deux attitudes stylistiques : appelons l'une, l'attitude classique, un peu archaïsante (celle des écrivains employant les formes verbales évoquées) et l'autre, l'attitude modernisante, ou éclatée (celle des écrivains qui évitent ces formes). Il se trouve aussi des écrivains qui emploient sporadiquement ces formes (voir certains romans de Marguerite Duras). Mais, même au niveau littéraire, il y a, à côté de ce qu'on peut considérer comme une norme (l'emploi, selon une certaine concordance, des formes en question, emploi naturel dans ce niveau, mais senti comme peu naturel, archaïque, voire incorrect dans le niveau standard), des archaïsmes à proprement parler : ainsi, par exemple l'emploi du plus-que-parfait du subjonctif à la place d'un conditionnel passé ou d'un plus-que-parfait de l'indicatif modaux (*on eût dit; s'il fût venu, il eût appris...*). Quoiqu'il en soit, dans ces emplois et dans d'autres cas similaires, le choix de certains mots (de certaines formes) est sous l'effet d'une certaine

mémoire, mémoire grammaticale fonctionnant d'une certaine façon : l'histoire en est l'origine et la motivation en est d'ordre esthétique.

## CONCLUSION

Nous avons pleinement conscience du caractère très relativement exhaustif de cette recherche, et plus exactement de cette revue, un peu rapide, des différents visages de ce que nous avons appelé la mémoire grammaticale du mot; et également de sa structure qui n'est pas toujours très rigoureuse. Il faudrait vraisemblablement synthétiser davantage, autrement dit mieux construire ce discours, de manière à mieux serrer les rapports très complexes existant entre ces différents visages. Nous nous excusons donc du désordre relatif qu'on peut déceler dans notre propos. Notre intention était, avant tout de sensibiliser l'auditeur (ou le lecteur) à un aspect du mot, dont certes tout le monde est conscient, mais qui nous a paru quelque peu négligé, et quelquefois, consciemment ou inconsciemment occulté : à côté de la mémoire proprement lexicale du mot, mémoire immédiatement perceptible mais souvent complexe et en tout cas d'une très grande richesse et d'une énorme diversité, le mot est aussi lourdement chargé d'une mémoire grammaticale. Cette mémoire est, elle aussi, souvent d'une grande complexité, précisément parce qu'elle est moins directement perceptible que la mémoire lexicale. En réalité, la perceptibilité (ou l'imperceptibilité) de cette mémoire grammaticale connaît des degrés. Ainsi, par exemple, cette mémoire est plus perceptible dans le caractère des mots grammaticaux en tant que tels, ou dans les flexions véhiculant des catégories grammaticales que dans l'appartenance du mot à telle partie du discours. Or, curieusement, c'est probablement (mais c'est une simple hypothèse) là où la manifestation grammaticale est le moins perceptible que la mémoire grammaticale du mot se révèle d'une plus grande densité.

## RÉFÉRENCES BIBLIOGRAPHIQUES

- BRUNOT, F. (1965 [1926]) : *La pensée et la langue*, Paris, Masson et Cie, XXXVI - 982 p.
- BUYSENS, E. (1975) : *Les catégories grammaticales du français*, Editions de l'Université de Bruxelles, 94 p.
- CAMOUN, A. (1992) : *Etudes de psycho-systématique française et arabe*, Tunis, Publications de la Faculté des Lettres de la Manouba, 347 p.
- CERVONI, J. (1991) : *La préposition*, Paris, Duculot, 309 p.
- DAMOURETTE, J. et PICHON, E. (1968) : *Des mots à la pensée. Essai de grammaire de la langue française*. (Tome premier), Paris, D'Artrey, 678 p.
- GOUGENHEIM, G. (1969) : *Système grammatical de la langue française*, Paris, D'Artrey, 377 p.
- GOUGENHEIM, G. (1970) : « Y a-t-il des prépositions vides en français ? », in *Etudes de Grammaire et de Vocabulaire français*, Paris, A. et J. Picard, XVI - 432 p.
- GREVISSE, M. (1964 [1939]) : *Le bon usage*, Paris / Gembloux, Hatier / Duculot, 1194 p.
- GREVISSE, M. et GOOSSE, A. (1980) : *Nouvelle grammaire française*, Paris / Gembloux, Duculot, 352 p.
- LE BIDOIS, G. et LE BIDOIS, R. (1971 [1967]) : *Syntaxe du français moderne* (2 vol.), Paris, A. et J. Picard, XX - 560 p. + XI - 794 p.
- GUILLAUME, G. (1973 [1938]) : «Théorie des auxiliaires et examen de faits connexes», in *Langage et Science du langage*, Paris / Québec, Nizet / Presses de l'Université Laval, pp. 73-86
- GUILLAUME, G. (1973 [1939 a]) : « Discernement et entendement dans les langues. Mot et partie du discours », in *Langage et Science du langage*, Paris / Québec, Nizet / Presses de l'Université Laval, pp. 87-98
- GUILLAUME, G. (1973 [1939b]) : «Comment se fait un système grammatical», in *Langage et Science du langage*, Paris / Québec, Nizet / Presses de l'Université Laval, pp. 108-119

- GUILLAUME, G. (1971[1948-1949]) : *Leçons de linguistique*, Série B., Québec, Les Presses de l'Université Laval, 222 p.
- LYONS, J. (1970) : *Linguistique générale. Introduction à la linguistique théorique*, traduction de F. Dubois-Charlier, Paris, Larousse, 384 p.
- MEILLET, A. (1965) : *Linguistique historique et linguistique générale*, Paris, Champion, 1965, 335 p.
- MOIGNET, G. (1981) : *Systématique de la langue française*, Paris, Klincksieck, 346 p.
- SPANG-HANSEN, E. (1963) : *Les prépositions incolores en français moderne*, Copenhague, Forlag, 256 p.
- TESNIERE, L. (1969 [1959]) : *Eléments de syntaxe structurale*, Paris, Klincksieck, XXVI - 670 p.
- VENDRYES, J. (1968 [1923]) : *Le langage*, Paris, Albin Michel, 444 p.

# La mémoire du coeur. Translation et symbolisation sémantiques

Miloud TAIFI

*Université de Fès, Fès, Maroc*

*«Y'en a qui ont le coeur si large qu'on y entre  
sans frapper. Y'en a qui ont le coeur si large  
qu'on n'en voit que la moitié»*

Jacques BREL

## INTRODUCTION

Le corps humain constitue par excellence un repère et un réceptacle de cognition et de symbolisation. C'est par rapport à son corps que l'homme élabore des conceptions et construit des systèmes symboliques, se fondant sur un isomorphisme entre l'ego corporel et la réalité environnante. Un tel isomorphisme est double et transitif : il concerne d'abord la dénomination des choses en référence au corps, mais aussi, par le retour des choses, l'investissement symbolique des organes du corps, auxquels on assigne des qualités et des propriétés. Les langues naturelles préservent et fossilisent, dans leur lexique et leurs systèmes sémantiques, de tels transferts sémantiques et symboliques. C'est ainsi, par exemple, qu'un *fleuve* — par translation dénomminative — a un *bras*, et qu'un *artichaut* a un *coeur*; mais aussi — par symbolisation — que le *bras* signifie la *force* et la *protection* et que le *coeur* peut être *noir*, *blanc*, *gros*, etc. selon les caractères et les états d'âme.

Nous proposons, dans cette communication, d'examiner comment les langues fossilisent le sens et comment les mots du corps ont une mémoire que l'analyse — dans une perspective d'anthropologie lexicale — peut interroger et réactiver. L'organe dont nous interrogerons la mémoire est le **coeur**. Ce viscère musculaire qui palpite en chacun de nous a certainement des choses à raconter. Pour le faire parler, nous examinerons, en les comparant, les champs sémantiques du terme «coeur» en arabe marocain (**qelb**), en berbère (**ul**) et en français, considérant notamment les transferts de sens et les différentes propriétés attribuées au coeur dans les trois langues. L'objectif étant de dégager les articulations de signification dominantes et de cerner la symbolique qui sous-tend l'apparition des sens métonymiques et métaphoriques à travers les locutions et les expressions figées. Les données qui nous serviront de matériaux sont tirées des dictionnaires (voir références).

Le *coeur* peut être *aveugle, froid, mort, blanc*, etc.; autant de propriétés physiques ou morales, plus abstraites, sont attribuées au coeur dans des expressions imagées où les transferts de sens fonctionnent par figuration dans un va et vient entre le concret et l'abstrait, entre le sens littéral et le sens figuré. Ainsi, avoir le coeur mort en arabe (dorénavant A) (A : *qelb-u meyyet*) et en berbère (dorénavant B) (B : *immut-as wul*) c'est être apathique, manquer d'ambition. En berbère, l'expression signifie aussi, sans doute par euphémisme, l'impuissance sexuelle. De même, l'honnêteté, la sincérité et la franchise, en tant que qualités morales, sont signifiées par la blancheur du coeur (A : *lqelb lbyed* B : *ul umlil*)

La richesse sémantique des expressions et des locutions dans lesquelles le terme *coeur* est l'objet d'un transfert de sens, est déroutante, non seulement parce que l'origine de la translation sémantique n'est plus saisissable en synchronie — le lien entre le sens littéral de départ et le sens figuré d'arrivée s'étant historiquement relâché — mais aussi, et surtout, parce que l'organe désigné est fortement investi d'une symbolique complexe dont les langues gardent des traces, sortes de vestiges et de fossiles sémantiques qui témoignent de l'évolution culturelle.

Malgré la multiplicité de ses acceptions constitutives, le champ sémantique du terme *coeur*, aussi bien en arabe, en berbère qu'en français, n'est pas cependant anarchique et désordonné. Il y a lieu de distinguer tout d'abord les significations dues aux transferts de sens par extension, analogie ou métonymie, et ensuite celles qui relèvent de la figuration ou de la métaphorisation.

## 1. — ANALOGIE, EXTENSION ET MÉTONYMIE

### 1.1 — Analogie

Le champ sémantique du terme *coeur* se déploie, dans les trois langues, par analogie et extension. Le premier transfert de sens permet de désigner par le vocable *coeur* tout objet ayant la forme de l'organe coeur. C'est donc sur une similitude physique et morphologique que se fonde l'analogie : un bijou, un bibelot, un fromage (*coeur à la crème*) sont des coeurs s'ils ont ou évoquent la forme et l'aspect du coeur. Le même sens analogique se retrouve aussi en arabe et en berbère : (A : *qelb men n-neqra* «un coeur en argent»; B : *ul n ury* «un coeur en or»). L'analogie sémantique concerne aussi les représentations scripturales : un dessin, une figure qui représentent le coeur : *as de coeur*.

### 1.2 — Extension

L'extension du sens confère au coeur les sèmes de centre et de milieu. C'est parce que le coeur est considéré, de par sa fonction, comme la partie centrale de l'organisme dans laquelle réside le principe de la vie, que l'extension sémantique est possible. On parlera ainsi du *coeur d'une ville* (A : *qelb lemdina*) pour en désigner le centre, la partie la plus animée par rapport à la périphérie. Plusieurs autres objets comportent aussi un coeur qui est leur partie centrale ou intérieure : en arabe, toute objet qui présente une polarité intérieur/extérieur a un coeur qui est son centre intérieur (A : *qelb d-dar* «coeur de la maison»), *qelb t-tawa* «coeur de la marmite», *qelb lyar* «coeur du trou», etc.). La polarité dimensionnelle intérieur/extérieur est rendue en français pour des légumes et des fruits : *coeur d'une laitue, de l'artichaut, du palmier, du bois, d'un fruit...*

L'extension sémantique va au delà du concret, puisque le centre est investi de sèmes abstraits : c'est ainsi que l'hiver et l'été ont un coeur qui en est le point culminant relativement, respectivement, aux rigueurs du froid et aux chaleurs torrides. Le centre est aussi l'étape essentielle ou le point capital d'un processus, d'une action : *le coeur du sujet, de la question, ou d'un débat* en est la substance, le point nodal. Si en français, on rentre dans le vif du sujet, en arabe, c'est dans son coeur qu'on rentre (A : *dxel f qelb lmewdu ?*), pour en aborder les aspects essentiels.

Le sens extensionnel prend d'autres détours pour s'accrocher cette fois-ci sur des manifestations physiologiques dont le coeur est supposé être le siège, déculpabilisant ainsi un autre organe, en

l'occurrence l'estomac, qui est en réalité l'acteur principal dans tous les maux dus aux ripailles excessives. Que l'on en juge : quand on a *un dîner sur le coeur*, on a forcément *mal au coeur*; si le *coeur est barbouillé*, on peut l'avoir (*le coeur*) *sur le bord des lèvres* ou *avoir des haut le coeur*. Les méfaits de l'estomac sont ainsi transposés sur le coeur. Ce dernier joue la doublure jusqu'au bout; car une odeur fétide, nauséabonde, ainsi qu'une chose répugnante peuvent *soulever le coeur* et provoquer des nausées ou l'envie de vomir, mais c'est en fait l'estomac qui se contracte pour expulser la vomissure libératrice.

Cette translation sémantique d'un organe sur un autre a aussi donné lieu en arabe et en berbère à quelques expressions qui entérinent une telle extension du sens; ainsi si le coeur se lève (A : *nad l-u lqelb*; B : *ikker-as wul*) ou monte (A : *tle ʕ l-u lqelb*) c'est qu'on a envie de vomir; si les nausées sont violentes, c'est que le coeur veut sortir (A : *bʕa ixrez l-u lqelb*) ou s'écarter (B : *yuñf wul-nes*). Même l'aigreur et l'acidité stomacales sont, en berbère, localisées dans le coeur : (B : *isemmum-as wul* «son coeur est aigre, acide»). De même, les spasmes ou les crampes de l'estomac malade ou affamé sont transposées sur le coeur : (B : *la ittuyabbay wul-nes* «son coeur se coupe de faim») : l'expression a deux sens : «il a faim», ou «il a mal au ventre». A l'instar du français, le coeur arabe et le coeur berbère peuvent être soulevés et dégoûtés (A : *had ji ka ynewwed lqelb*; B : *ayad la yskkar ul*, «cela soulève le coeur»).

### 1.3 — Métonymie

L'expression *serrer* (quelqu'un ou quelque chose) *contre son coeur* confère au terme *coeur* un nouveau sens obtenu, non par analogie ou par extension, mais plutôt par transfert de sens métonymique : le coeur ici signifie, de par la relation de la partie au tout, la poitrine ou le torse. La métonymie fonctionne aussi et surtout dans les expressions où le terme *coeur* désigne la personne. Ainsi, dans *avoir le coeur sensible*, la propriété de sensibilité n'est pas attribuée à l'organe coeur, mais à la personne qui en est le possesseur. Le sens initial de coeur «organe...» s'opacifie ainsi pour désigner la personne. La métonymie ne fonctionne évidemment que dans des cas où les attributs sont incompatibles avec le terme *coeur* dans son sens propre : la preuve c'est qu'on peut avoir le coeur malade, faible ou volumineux; ces différents états sont physiologiques et relèvent de la déficience physique.

C'est à partir du sens métonymique du terme *coeur* que naît la métaphorisation; puisque l'homme est désigné par son coeur, les

qualités attribuées au premier sont localisées dans le second. La figuration est cependant complexe dans la mesure où elle est fondée, comme nous verrons, sur la symbolique, dépassant ainsi la construction du signifié par simple translation sémantique pour s'articuler sur des faits culturels et anthropologiques. Indépendamment de quelques significations idiosyncrasiques, est relativement au corpus des expressions que nous avons collectées, le champ sémantique du terme *coeur* est organisé en quatre domaines sémémiques dominants.

## 2. — LES ARTICULATIONS SIGNIFICATIVES DOMINANTES

### 2.1 — Conscience, pensée

Le *coeur* est le lieu de l'intellectualité et de la spiritualité. Les pensées viennent du *coeur* qui est le siège et le centre de la connaissance dans le sens large; la culture y localise l'intelligence et la sagesse. La réflexion intime qui précède et motive une décision ou un comportement se produit dans le *coeur*; c'est lui qui voit et distingue le bien du mal. Les expressions (A : *b lqelb*; B : *s wul* «avec *coeur*») indiquent dans leurs emplois que le *coeur* est conscience qui décide. Ainsi les énoncés (A : *ka yhder b qelb-u*; B : *la ysawal s wul-nes* «il parle avec son *coeur*») signifient que celui qui parle le fait avec sincérité, qu'il dit ce qu'il pense vraiment, ce qu'il croît intimement être vrai; en bref, ce que sa conscience lui dicte. Parler franchement sans trahir sa conscience, c'est, en français, *avoir le coeur sur les lèvres*; la parole qui *vient du coeur* est pure et limpide, cet élan de sincérité provoque de l'effusion et c'est alors à *coeur ouvert* que la logorrhée verbale fuse et se déploie. En arabe, la formule (A : *hadffi ba/heddet-ni qelb-i* «c'est de cela que mon *coeur* m'a entretenu») est souvent employée comme épilogue, pour clore l'exposé d'un avis ou d'une opinion à propos d'une question ou d'une affaire soumises au débat et à l'appréciation de tous.

En berbère, on dit de quelqu'un qui délire ou qui déraisonne qu'il parle sans son *coeur* (B : *la ysawal bla ul-nes*). Par contre, faire quelque chose avec son *coeur* (A : *ka yxdem b qelb-u*; B : *la yxeddem s ul-nes*), c'est le faire consciencieusement en ayant des raisons suffisantes pour le faire et en y mettant toute l'ardeur nécessaire. En français, c'est *donner du coeur à l'ouvrage*. En kabyle, réfléchir c'est *rester avec son coeur* (B : *yet\$imi d wul-is*), le *coeur* signifiant évidemment la conscience; dans le même dialecte, agir contre sa conscience, malgré soi, c'est *piétiner son coeur* (B : *ye\$fes xef wul-is*).

Parallèlement à ce premier sens, le coeur est aussi le siège de la croyance et de la spiritualité. Les croyances religieuses, qu'il ne faut pas confondre avec l'intuition, relèvent du coeur. En Islam, le coeur du croyant est considéré comme le trône de Dieu; les égarés sont ceux dont le coeur est aveugle. La vision spirituelle émane du coeur : les expressions (A : *ema qelb-u*; B : *iema y as wul*) sont employées pour parler de quelqu'un qui agit contre les principes moraux de la religion. Avoir ou ne pas avoir la foi sont donc des états du coeur, car c'est en lui que réside tout sentiment de religiosité. Ainsi, en arabe, le coeur peut être croyant (A : *qelb mumen*), impie (A : *qelb kafer*) ou franchement athée, dans ce cas le coeur est littéralement illicite (A : *qelb mehrum*).

La spiritualité suppose la méditation et la contemplation. C'est dans et par le coeur qu'on s'adonne à ces deux activités de l'esprit. La religion étant aussi une axiologie, des propriétés — qualités ou défauts selon les principes religieux — sont attribuées au coeur. Nous verrons ci-après les plus importantes.

## 2.2 — Intuition et pressentiment

La deuxième articulation de signification dominante du terme *coeur* concerne les notions d'intuition et de pressentiment. Ainsi, la connaissance indépendante de la réflexion et du raisonnement se situe dans le coeur. C'est cette connaissance qui permet de prévoir les événements et de présager de leurs conséquences. En arabe et en berbère, notre coeur nous parle pour nous avertir, nous conseiller, nous prévenir à propos de choses ou de faits qui échappent à notre conscience, à notre connaissance intellectuelle : (A : *xber-ni qelbi* ou *lqelb xbir* «mon coeur m'a informé, m'a averti») dit-on en arabe pour signifier l'intuition et le pressentiment qu'on a d'un fait ou d'un incident qui ont eu lieu ou qui auront lieu dans l'avenir. Le même sens est rendu en berbère par des verbes dire ou trouver : (B : *inna y it wul-inw* ou *yufa-t wul-inw* «mon coeur me l'a dit» ou «mon coeur l'a trouvé»). La connaissance intuitive, qui nous permet d'appréhender des situations ou des événements indépendants de notre savoir, provient du coeur; il est le siège de la prémonition et de la prescience. En français, on *écoute son coeur* qui, en principe, ne nous trompe jamais; et lorsque on agit tout en étant d'accord avec sa conscience, on le fait *de bon coeur, de grand coeur, de tout coeur*, ou encore mieux *de gaieté de coeur*; car il n'y a pas plus grand plaisir que de dire ou faire quelque chose en accord avec sa conscience. Qu'on se souvienne de l'atroce douleur du remords. Mais les *cris du coeur* se

font rares de nos jours, la spontanéité a cédé la place au calcul mesquin dû à l'amour excessif du gain.

L'appréciation des choses relève aussi du coeur. Ainsi, en berbère, si le coeur refuse et dicte une appréciation négative, on agit en conséquence sans chercher à savoir s'il a ses raisons, estimant que l'intuition est plus forte que la connaissance : (B : *ul-inw a wr irin*, «c'est mon coeur qui ne veut pas»), point c'est tout. Peut-on d'ailleurs agir contre l'intuition et le pressentiment dictés par son coeur ? La même signification se retrouve en arabe : l'expression (A : *qelb-i ma mertah* / «mon coeur n'est pas en repos») est employée pour dire qu'on n'est pas favorable intuitivement, sans savoir réellement pourquoi, à un fait, à une situation ou un individu. La vie nous confronte souvent à des situations qui nous obligent à agir de telle ou de telle manière sans avoir de raisons suffisantes pour faire le bon choix. La solution, encore une fois, est tout simplement d'*écouter son coeur*, car la sagesse lui octroie un pouvoir de discernement plus fort, plus fiable que l'esprit qui peut s'égarer dans des raisonnements tendancieux.

### 2.3 — Amour-propre et for intérieur

La dignité de l'homme, son amour-propre, qui font de lui un homme honorable et respectable, se situent dans le coeur. Le sens de l'honneur exige qu'on préserve son amour-propre de toute offense, outrage ou injure. Perdre son amour propre, c'est avoir le coeur mort (A : *qelb-u meyyet*; B : *immut-as wul*) ou n'avoir pas de coeur du tout (A : *ma εend-u qelb*; B : *war ul* «sans coeur», *ur dig-s ul*, *ur yesε i ara bbwul* «il ne possède pas de coeur»). La mort du coeur signifie l'arrêt de la vie et métaphoriquement la déchéance et le déshonneur qui peuvent pousser au suicide, car c'est l'amour-propre qui fonde même la vie sociale et tient lieu de référence aux rapports humains.

Le coeur peut aussi manquer de l'ardeur et de la vigueur nécessaires pour défendre son amour-propre et son honneur. Il est alors froid (A : *qelb-u bared*). Dans une autre expression arabe, plus terrifiante, parce que relative à l'anthropophagie, le coeur (humain bien entendu) est comestible : n'avoir plus d'amour-propre, ni le sens de l'honneur et de la dignité, c'est *manger son coeur* (A : *kla qelb-u* «il a mangé son (propre) coeur»). En berbère, on ne mange pas son coeur mais on l'ensevelit, on l'enterre. On dira ainsi de quelqu'un qui ne défend pas son amour-propre, qui se résigne au déshonneur et à la déchéance, qu'il a enseveli son coeur : (B : *yuder i wul-nes*). On n'enterre évidemment qu'un coeur qui meurt et nous avons déjà noté

qu'en berbère ainsi qu'en arabe, n'avoir plus d'amour-propre, c'est avoir un coeur mort. En français, si l'on *a du coeur*, l'on a l'honneur et la fierté dont on peut s'enorgueillir. Mais si le *coeur nous manque* pour préserver notre dignité, c'est la déchéance qui nous guette.

Le coeur est aussi le siège de l'intimité, du secret que chacun porte profondément en soi, en son for intérieur. L'expansivité ou la discrétion sont des états du coeur; il peut dans un double mouvement s'ouvrir ou se fermer : une personne expansive a *un coeur ouvert* (A : *qelb-u meftuh*); par contre, le discret, celui qui garde ses secrets par retenue et décence a *un coeur replié* (A : *qelb-u metwi*), ou recroquevillé (A : *qelb-u mkemmes*). Il faut signaler aussi que le repli du coeur caractérise celui qui nourrit des arrière-pensées néfastes ou qui fomente en secret des intrigues pour nuire aux autres. L'ouverture du coeur indique la jovialité et la gaieté du caractère; le coeur fermé, replié, recroquevillé est toujours maussade et morose.

#### 2.4 — Affectivité et sentiments

C'est dans ce domaine notionnel que le terme coeur est investi fortement de sens métaphoriques expressifs. Le coeur est en effet le siège de l'affectivité presque dans toutes les cultures traditionnelles. Les différents sentiments de l'homme y sont localisés : l'amour, la haine, la joie, la tristesse, etc. Il faut remarquer cependant qu'en arabe et en berbère, le foie (A : *kebda*; B : *tasa*) constitue aussi le siège de certains sentiments qui normalement reviennent au coeur : la tendresse, l'affection, la pitié, ainsi que les sentiments relatifs à l'amour filial et maternel. Le partage locatif des sentiments entre le coeur et le foie est un problème sémantique complexe relevant de l'anthropologie culturelle.

Pour notre propos, c'est le coeur qui indique, par ses différents états, des qualités ou des travers qui se manifestent dans les comportements de l'homme. Pour exprimer ces diverses significations, l'arabe et le berbère combinent l'item lexical *coeur* et divers verbes ou adjectifs qui désignent les propriétés attribuées au coeur et qui, par métaphore, rendent le sentiment, la qualité ou l'état de la personne. Nous présentons, dans ce qui suit, les différentes expressions relevées en proposant un classement sémantique.

Mais signalons tout d'abord que le coeur est le siège de l'amour dans toutes ses dimensions et manifestations. Qui n'a pas gravé sur un arbre ou sur un mur, un coeur transpercé de la flèche lancée par Cupidon. C'est en effet le coeur qui aime qui chérit, qui

adore, qui idolâtre. Les *affaires de coeur* sont diverses et nombreuses : le *coeur* peut être ainsi *enflammé, embrasé, ardent* d'amour; on peut aussi *offrir* ou *refuser son coeur* ou simplement le *jeter à la gribouillette* (= aimer à l'aventure); mais attention ! les *bourreaux des coeurs* guettent la moindre défaillance pour s'en emparer. Comme on peut donner son coeur, on peut, par le retour des choses, tenter de *conquérir* ou de *gagner le coeur* d'autrui; bien plus le *coeur* peut être *ravi* dans un excès de passion. L'amour fait aussi des dégâts en arabe et en berbère. Un coeur amoureux est un coeur qui bouillonne (A : *qelb-u ka yḡli*), ou qui se penche (A : *qelb-u ka ymil*). En berbère, un coeur épris descend, choit et s'affale (B : *ider wul-nes*), ou bien encore s'enfuit, se libérant ainsi de tout contrôle (B : *irwel-as wul*).

Le coeur ne renferme pas seulement cet amour qui nous porte instinctivement vers l'autre sexe, il regorge aussi d'autres sentiments à l'égard du genre prochain ou des choses. L'affection, l'amitié proviennent du coeur, mais aussi l'inimitié, l'aversion; la philanthropie et la misanthropie dépendent des élans du coeur, de ses humeurs; la sympathie et l'antipathie en tant que penchants instinctifs émanent aussi du coeur. Ainsi, ne pas *porter* quelqu'un *dans son coeur*, c'est le détester, ne pas le supporter pour des raisons qu'on n'arrive pas souvent à formuler. En arabe, exécrer quelqu'un c'est *le sortir, le chasser du coeur* (A : *xerre3t-u men qelb-i* «je l'ai sorti de mon coeur»); en berbère, c'est lui *fermer le coeur*, de telle sorte qu'il ne puisse plus y accéder (B : *qqenx ḡif-s uli-nw*, litt. : «j'ai fermé sur lui mon coeur»).

### 3. — LES PROPRIÉTÉS DU COEUR

Des attributs physiques assignés au coeur par métaphore symbolique servent à signifier des qualités ou des états qui caractérisent les personnes dont le coeur est ainsi décrit.

#### 3.1 — La couleur du coeur : le blanc, le noir et le vert

La blancheur du coeur symbolise l'honnêteté et la droiture (A : *lqelb lbyed*; B : *ul umlil*) et aussi la candeur et la naïveté. Ainsi, le coeur d'un enfant est toujours blanc parce qu'il est innocent. C'est en grandissant et en devenant adulte que son coeur peut virer au noir. La blancheur a, du point de vue axiologique, une valeur positive, en opposition à la couleur noire. *Avoir un coeur noir* (A : *lqelb lkhel*; B : *ul abexxan/aberjan*) c'est en effet être haineux, méchant, envieux, défauts et travers qui vont à l'encontre de toute morale. L'expression

arabe (A : *qelb-u mkehhel b leɛma*, litt. «son coeur est noirci d'aveuglement») rend compte du sentiment de haine poussé jusqu'au paroxysme. L'expression référerait au khôl, l'antimoine que les femmes s'appliquent sur les cils en guise de fard. On aura saisi dans cet exemple l'opposition significative entre le sens de *mkehhel* (fardé de khôl) qui renvoie à la beauté et à l'esthétique et le sens négatif de *leɛma* «cécité» qui comporte le sème d'obscurité et, par métonymie de la cause à l'effet, la notion de noir.

Un coeur vert indique la volupté et caractérise le bon vivant, celui qui continue, malgré son âge avancé, à s'adonner aux jouissances de la vie. C'est seulement en arabe que cette couleur est attribuée au coeur (A : *qelb-u xder*). La notion de verdure est à prendre dans ses deux acceptions, limitrophes d'ailleurs. Le vert renvoie en effet, au delà de la couleur, à la vie de la végétation en opposition à sa sécheresse. La sève coule dans un arbre encore vert et lui donne vie; mais lorsqu'il se dessèche, la sève, qui donne vigueur, le quitte et l'arbre meurt : *avoir un coeur vert*, c'est donc avoir encore, à l'instar de l'arbre, la force et l'ardeur suffisantes pour les plaisirs voluptueux.

### 3.2 — Dimension, volume et mouvement

#### a) Grandeur et petitesse

Le coeur peut être grand ou petit. La première propriété signifie la générosité et la noblesse du caractère (A : *lqelb lkbir*; B : *ul ameqran/axatar*). On dit en berbère : *ul axatar g bu tsart ay llan* «un grand coeur se trouve chez qui en est digne». La grandeur du coeur est une qualité positive qui prévaut chez celui qui reste toujours au dessus des mesquineries et des bassesses imposées souvent par les vicissitudes de la vie. La petitesse, son antonyme, indique la vilénie et la lâcheté (A : *lqelb ssyir*; B : *ul amezian*). La personne dont le coeur est petit se laisse facilement acheter, corrompre, renonçant ainsi à son honneur pour quelques faveurs ou avantages. En berbère, on dit aussi de la personne caractérisée ainsi que son coeur est insuffisant (B : *idrur g-g<sup>w</sup>ul-nes*, litt. «il est insuffisant dans son coeur»). La suffisance cardiaque, en tant que métaphore, fonde même, dans la culture berbère, la dignité de l'homme; car il n'y a pire défaut que la vilénie. En kabyle cependant, la petitesse du coeur signifie l'impatience qui peut conduire au manque de discernement (B : *ul-is meeztuh* «son coeur est petit»).

### b) L'ample et l'étroit

Le coeur peut être ample, large ou étroit. Cette opposition concerne les traits de caractère. La largeur du coeur (A : *lqelb lwaseε; ul iwesεen*) équivaut à la souplesse du caractère, à la bienveillance et à la clémence. Celui dont le coeur est large pardonne facilement, accorde son indulgence et oublie bien vite les vexations, les humiliations et les affronts dont il est victime. L'étroitesse du coeur signifie par contre, la susceptibilité, la rancune et la rancœur (A : *qelb-u deyyeq*; B : *ul anaqmar : ul aneqmar ur ittasy midden*, «un coeur étroit ne supporte pas les gens»); il faut entendre par ce dicton berbère l'impossibilité de supporter les sarcasmes, les railleries et les vexations subis si l'on est doté d'un caractère susceptible, irascible et belliqueux et que l'on a une tendance à garder de la rancune à ceux qui nous offensent.

C'est dans le coeur étroit que naît le sentiment de vengeance, car il ignore le pardon et l'indulgence. Ainsi les expressions (A : *ka yrfed f qelb-u*, litt. «il prend dans son coeur»; B : *la ykemma g-g<sup>w</sup>ul-nes* «il noue dans son coeur», ou bien B : *la yttamez g-g<sup>w</sup>ul-nes* «il retient dans son coeur») rendent compte de l'état sentimental vindicatif de ceux qui sont par nature rancuniers.

La bonté et la compassion sont liées à la minceur du coeur (A : *lqelb rrrhif, lqelb rrriq* «coeur mince»). La minceur suggère la transparence, celle des sentiments : avoir un coeur mince (transparent) c'est nourrir de bons sentiments à l'égard des autres. On dit aussi en arabe qu'un coeur bon est un coeur de soie (A : *qelb-u men lel'rir*).

### c) Le lourd et le léger

Les expressions françaises *avoir le coeur lourd* ou *en avoir lourd sur le coeur* : traduisent ce sentiment pénible et accablant qui nous envahit et nous assaille. Les afflictions, les peines profondes alourdissent le coeur. *Un coeur léger*, par contre, appartient à celui qui est insouciant et qui prend grand plaisir à vivre dans un bonheur permanent.

### d) Le dur et le mou

Cette opposition de propriétés est rendue en arabe, d'une part par le terme *qasl* qui recouvre plusieurs sens apparentés : «dur, solide, compact, ferme raide» et de l'autre par les termes *h'if* «tendre, fondant, friable» et *rteb* «tendre, souple, mou moelleux».

*lqelb lqash* «coeur solide» # *lqelb lh'if, lqelb rteb* «coeur mou»

En berbère, la même opposition est signifiée par le terme *qseh* avec les mêmes acceptions qu'en arabe et les termes *hrurey* «être visqueux, délayé, liquide» et *lwiγ* «être tendre, mou, souple».

*ul iqeshen* «coeur dur» # *ul ihrurin, ul algg<sup>w</sup>ay* «coeur mou»

Transposée au coeur, l'opposition entre le dur et le mou indique des traits de caractère différents : la dureté du coeur qualifie celui qui est insensible, impitoyable, cruel. Nous verrons ci-après que la dureté du coeur est comparée à celle de la pierre ou du fer. En revanche, avoir un caractère doux, sensible, compatissant, c'est avoir un coeur mou. La mollesse du coeur étant une qualité positive, celle-là même qui fonde notre humanité : la pitié et la commisération qui nous obligent à ne pas rester indifférents aux malheurs des autres.

### e) *Le mouvement descendant*

Le coeur peut connaître un mouvement descendant. Aller vers le bas, tomber, choir sont des états du coeur de celui qui, par ses comportements, est vil, ignoble, abject sur le plan moral : (A : *qelb-u saqt, qelb-u tayh*; B : *ittuttey-as ul, ider-as wul*). La notion de chute est symboliquement négative, elle est associée à la bassesse, à l'abaissement et à l'ignominie.

Nous avons déjà noté qu'en berbère, la chute du coeur indique aussi un amour impossible, malheureux parce qu'il n'est pas réciproque; ainsi (B : *ider-as wul xf...*, litt. «il a le coeur qui descend sur...»), c'est s'éprendre d'amour fou pour...; on retrouve ce sens dans le chant suivant : (B : *ur as samehx i titt, mer idd i wraεa, ur iy iddar wul xf unna wr diy-i iseksiwn* «Je ne pardonnerai pas à ma prunelle, sans la vue, mon coeur ne serait pas épris de celui qui ne faisait aucun cas de moi»). Laisser ainsi choir son coeur par passion folle est aussi une sorte de bassesse et d'ignominie.

### 3.3 — La température du coeur

Nous avons déjà signalé que la froideur du coeur équivaut au manque d'amour-propre et du sens de l'honneur, lorsqu'on ne réagit pas pour défendre et sauvegarder sa dignité. Le coeur froid (A : *qelb-u bared*) est lié, dans le même ordre d'idées, à l'apathie, à la paresse, à l'indolence et à la nonchalance. Ce dernier sens n'a pas été relevé en berbère, c'est plutôt la mort du coeur, comme on a vu, qui rend la même signification. La froideur du coeur indique aussi la patience et surtout l'impassibilité et le sang-froid. Cet état d'âme nous permet d'appréhender les choses avec calme et sérénité. En français, avoir *un coeur froid* c'est être insensible, impitoyable.

Paradoxalement, du moins dans notre corpus d'expressions, la chaleur n'est pas attribuée au coeur pour dire les états contraires aux précédents : l'ardeur, le dynamisme, l'activité. Le coeur peut cependant être cautérisé ou brûlé (A : *qelb-u mekwi, mehruq*; B : *iqqed wul-nes, iyus wul-nes*) dans des situations où l'on souffre moralement à cause d'un malheur, d'un souci, ou d'un incident fâcheux. Les peines morales cautérisent, brûlent le coeur et le consomment. En français, ce sont plutôt, comme il a été signalé, l'amour et la passion qui enflamment *le coeur* qui devient alors *embrasé, ardent*, palpitant de mille feux.

### 3.4 — La propreté du coeur

La franchise, la loyauté et l'honnêteté ainsi que des qualités similaires, sont dues à la propreté du coeur, à sa pureté : (A : *qelb-u safi*; B : *izeddiy wul-nes*). En kabyle, un coeur pur (B : *zedig wul-is*) caractérise celui qui ne tient pas rancune. La souillure du coeur, son impureté ne font ni en arabe ni en berbère l'objet de métaphorisation; l'hypocrisie, la déloyauté et la malhonnêteté sont signifiées par d'autres procédés lexicaux.

### 3.5 — Le plein et le vide

*Avoir le coeur gros* en français, c'est *avoir le coeur plein* en arabe et en berbère, mais pas seulement de chagrin : (A : *qelb -u εamer*; B : *iεemmer wul-nes*). Le coeur supporte en effet plus que cela : des colères refoulées, des griefs non exprimés, des vexations ou des injures auxquelles on n'a pas répondu stagnent dans le coeur, en plus des peines, des inquiétudes et des soucis quotidiens. Si c'est l'angoisse et l'anxiété qui envahissent le coeur, celui-ci est alors «(r)attrapé au vol» (A : *qelb-u mzelqef*), il faut comprendre l'expression dans le sens de «coeur ravi, enlevé» n'appartenant plus à son propriétaire. En berbère, le coeur qui est dans le même état est suspendu (B : *yugel wul-nes*), accroché quelque part. Le coeur suspendu, accroché (A : *qelb-u mεelleq*) signifie en arabe, l'inquiétude et le tracas lors de l'attente d'une nouvelle ou des résultats d'une entreprise ou d'une action.

Un coeur plein peut éclater si l'on ne l'épanche pas. Ainsi se mettre en colère violente, c'est avoir, en arabe, le coeur fracassé (A : *qelb-u mfergeε*); en français, c'est plutôt la rage qui envahit le coeur : *avoir la rage au coeur*. On peut néanmoins soulager son coeur en le vidant seulement au moyen de la parole thérapeutique : la confidence ou la confession (A : *xwa qelb-u*, ou *ferrey qelb-u* «il a vidé son

coeur») dit-on en arabe pour parler de quelqu'un qui dit ce qu'il a sur le coeur, qui ouvre son coeur. En berbère, épancher, décharger son coeur c'est le dénouer ou plus exactement desserrer le noeud qui l'entrave (B : *unna wr issilwin i wul-nes, d ad ittiqs*, litt. «celui qui ne dessert pas le noeud de son coeur risque de le voir éclater»).

### 3.6 — Le cuit, le moulu et le coupé

Trois propriétés supplémentaires sont attribuées au coeur : *un coeur cuit* (A : *qelb-u tayb*; B : *inwa wul-nes*) est d'abord un coeur plein de chagrin et de tristesse. La propriété de cuisson indique aussi le caractère de celui qui est nerveux, irritable et irascible, qui a les nerfs en pelote, comme on dit familièrement en français. Le coeur peut même être cuit à l'excès (A : *qelb -u hari*) et devient insensible, ne pouvant plus avoir un quelconque élan d'humanité.

Le *coeur moulu* ou *pulvérisé* (A : *qelb-u methun*) signifie une douleur, une torture morale; souffrir moralement c'est aussi avoir le *coeur coupé* (A : *qelb-u meqtu?*; B : *la ittuyabbay ul-nes*), autrement dit, c'est éprouver non seulement de la peine et de l'affliction, mais aussi de la souffrance due aux remords et aux regrets.

### 3.7 — La bête, la pierre, la citrouille et les autres

Nos amies les bêtes représentent souvent nos qualités et nos défauts, selon le bon vouloir de la culture, mais aussi selon leur mode de vie. C'est en effet en observant les comportements des animaux que nous avons tendance à les affubler d'attributs qui nous caractérisent nous-mêmes. L'appréciation est cependant toute humaine; car le comportement de l'animal n'est, en soi, ni bon ni mauvais; les animaux sont mus par l'instinct et vivent en dehors de toute axiologie. Ainsi, l'âne n'est ni plus bête, ni plus intelligent, qu'un tout autre congénère; pourtant le pauvre bougre est donné, dans plusieurs cultures, comme représentant typique de la bêtise... humaine naturellement. Nous héritons ainsi des coeurs de plusieurs animaux :

*Avoir un coeur de vipère*, c'est être perfide, fourbe, traître; la vipère étant donnée comme exemple typique de ces différentes propriétés qui caractérisent le comportement des hommes. La dureté du caractère, la férocité et l'impitoyabilité ressortent plutôt du tigre : *avoir un coeur de tigre*. Mais si le tigre est féroce, le lion est courageux : *avoir un coeur de lion*, c'est montrer un grand courage allant parfois jusqu'à la témérité.

Contrairement au lion, la pauvre poule — un volatile familier, grâce à auquel nous dégustons des omelettes aux fines herbes — est une poltronne; car *avoir un coeur de poule* (mouillée parfois), signifie en effet la couardise, cette crainte malade qui vous écrase et vous oblige souvent à vous faire petit pour passer inaperçu de peur d'avoir à vous défendre. La même propriété est attribuée à la poule arabe et berbère (A : *qelb d3a3a*; B : *ul n tfellust*); bien plus, les croyances populaires déconseillent aux enfants la consommation du coeur de la volatile de peur qu'ils ne deviennent des peureux comme des poules.

Le renard est rusé, on le sait. Mais avoir un coeur de renard en berbère (B : *ul ise eb*) c'est être plus que rusé, c'est la ruse au service de la duperie, du leurre, de la duplicité en vue d'un gain, d'un avantage, d'un privilège. La ruse nécessite la versatilité et des changements opportuns selon les circonstances et le but visé. L'instabilité comportementale et le changement de veste sont représentés en arabe et en berbère par le caméléon (A : *qelb tata*; B : *ul n tata*, coeur de caméléon).

Les minerais fournissent eux aussi des propriétés attribuées au coeur humain. *Un coeur de pierre, de rocher ou de roc* (A : *qelb-u men leh3er*; B : *ul-nes zg iselli*) signifie l'insensibilité, la dureté de caractère et l'impitoyabilité. Un coeur insensible reste indifférent aux malheurs d'autrui, il ne s'apitoie jamais, gardant l'impassible froideur du marbre : *avoir un coeur de marbre*. Ces deux attributs peuvent être poussés jusqu'au paroxysme, et c'est alors la cruauté, l'atrocité, la sauvagerie qui s'installent, réduisant l'homme à la dureté du fer, d'acier ou de bronze : *avoir un coeur de fer / de bronze* (A : *qelb d lehdid, qelb d lhend*; B : *ul n wuzzal*). La cruauté suppose la méchanceté et, par conséquent *un coeur noir*; cette couleur étant liée, comme nous l'avons signalé, à la malveillance et à la haine : le coeur est alors de charbon (A : *qelb-u fehma*; B : *ul-nes d tirginn*) ou de poudre à canon (A : *qelb-u barud*). Le charbon et la poudre signifiant par métonymie la couleur noire, cette dernière symbolise la méchanceté et la haine, le transfert de sens est ainsi opéré.

Du côté du végétal, le coeur acquiert d'autres propriétés symboliques représentées par des légumes, des arbustes ou des fruits. Nous avons noté que certains légumes ou fruits ont, par extension du sens, un coeur, signifiant l'intérieur. Il ne s'agit plus ici de la localisation d'une partie du légume ou du fruit, mais d'une qualité qu'on prédique du végétal et qu'on attribue par symbolisation au coeur humain dans, naturellement, son sens métonymique. Ainsi, *avoir un coeur de citrouille*, c'est d'abord être mou; il s'agit, bien

entendu, de mollesse de caractère, cet état flegmatique et nonchalant qui annihile les réactions à chaud et des ripostes rapides. L'expression signifie aussi la lâcheté et la pusillanimité. Voilà pour la citrouille, qu'en est-il de l'artichaut ? *Avoir un coeur d'artichaut*, c'est être volage : l'inconstance en amour caractérisant la jeunesse en mal d'aimer, mais aussi les époux infidèles.

Un arbuste et deux fruits sont associés aux coeur en arabe et en berbère. Le premier est le laurier qui symbolise l'amertume, l'aigreur, la rudesse du caractère, états dus aux espoirs déçus ou au sentiment corrosif de l'envie et de la jalousie. *Avoir un coeur du laurier* (B : *ul ulili*), c'est se monter âpre, aigri, méchant dans ses critiques, ses reproches. Les remarques et les commentaires d'un jaloux, d'un envieux sont toujours amères comme les feuilles du laurier. L'amande est mieux appréciée : *avoir un coeur d'amande* (A : *qelb luza*) c'est, en effet, être bon, compatissant, à l'écoute des malheurs des autres. La grenadine, quant à elle, de par sa forme hermétique : (A : *qelb d remmana lmeymda*) représente le caractère taciturne, renfermé et morne de celui qui ne laisse rien apparaître de ses sentiments et de ses pensées.

### 3.8 — Le chrétien et le juif

La culture marocaine de langue arabe ou berbère réfèrent au chrétien et au juif pour symboliser deux propriétés qui sont attribuées au coeur de celui qui s'en distingue (A : *qelb nnesrani*; B : *ul urumey*, «coeur de chrétien»). La comparaison est utilisée pour exprimer trois idées qui présentent d'ailleurs des affinités : le chrétien symbolise d'abord l'insensibilité, la dureté de caractère allant jusqu'à la cruauté; mais avoir un coeur de chrétien c'est aussi être d'un tempérament actif, dynamique et entreprenant; la troisième idée est afférente à l'inflexibilité du caractère, à la rigueur dans le comportement et à la rectitude dans le jugement.

Le juif représente plusieurs attributs relatifs à la notion de peur (A : *qelb lihudi*; B : *ul n wuday* «coeur de juif»), allant du simple sentiment craintif dû à la timidité et au manque de confiance en soi, jusqu'à l'appréhension et la pleutrerie par manque de courage et aussi par lâcheté et veulerie.

### 3.9 — Avoir le coeur quelque part

L'emplacement du coeur est fixe, il est situé à droite entre les deux poumons. Mais le transfert de sens par métaphore peut le déplacer pour des besoins expressifs. *Le coeur* est d'abord *bien* ou

mal *placé* : si l'emplacement du coeur est bon, c'est que celui qui le porte appartient à la classe noble, à la classe supérieure, car, dans ce cas, le coeur n'est pas seulement bien placé, mais c'est un *coeur haut placé*.

On peut avoir *le coeur à gauche* ou à *droite* selon ses opinions politiques et ses options idéologiques. En arabe, le coeur à droite (A : *qelb-u ŧel limen*) est celui du croyant, du pratiquant; le coeur à gauche (A : *qelb-u ŧel liser*) est porté par l'impie, le mécréant, l'athée.

Le coeur s'accroche par ailleurs à d'autres organes du corps : *avoir le coeur sur la main* signifie la bonté, la générosité et la bienfaisance. En arabe et en berbère, le coeur peut être placé sous les pieds (A : *ŧend-u lqelb teht r-re3in*; B : *illa y as wul ddaw idarn*, litt. «il a le coeur sous les pieds»); la personne qui piétine ainsi son coeur est insensible, n'ayant aucun sentiment de pitié et de compassion. Il a été déjà noté qu'en kabyle, piétiner son coeur c'est agir contre sa conscience.

En français, on a *le coeur sur les lèvres*, si l'on dit toute sa pensée, si l'on parle avec franchise, sans cacher ni farder la vérité. En arabe et en berbère, c'est plutôt la langue qui supporte, sans ce cas, le coeur franc et sincère (A : *qelb-u Ela lsan-u*, B : *ul-nes xf yils* «son coeur est sur la langue»).

#### 4. — A L'ASSAUT DU COEUR

Nous avons présenté dans ce qui précède les différentes propriétés qui sont attribuées au coeur; propriétés qui expriment par métaphore fondée sur la symbolique, divers états d'âme, des qualités variées et des traits de caractère différents qui sont des attributs de l'homme selon ses comportements, ses réactions, ses agissements.

Le coeur peut par ailleurs subir plusieurs et diverses actions qui peuvent le remplir de joie et de gaieté ou bien le peiner et l'affliger. Le coeur est ainsi pris d'assaut et assailli, malmené, subissant des coups et des attaques parfois corrosifs. Nous avons déjà noté que le coeur peut être soulevé lorsque l'estomac fait des siennes. Mais ce n'est pas de ce genre d'actions qu'il s'agit. Celles qui nous intéressent ici concernent les sens métaphoriques.

La peur et la crainte *glacent le coeur*, l'effroi le *transit*. L'engourdissement du coeur par le froid montre que nos réactions face au danger émanent du coeur. Le refroidissement du coeur

signifie autre chose en arabe : (A : *berred qelb-u*, «il a refroidi son coeur»), il s'agit ici de la satisfaction tirée d'une action effectuée jusqu'à satiété pour se libérer, se dédommager ou se venger : par exemple, manger à satiété, c'est refroidir son coeur, de même proférer des injures et des insultes à l'égard de quelqu'un qui vous a outragé, c'est aussi refroidir son coeur. En berbère, le même sens est rendu par l'extinction du coeur (B : *isehsey ul-nes*, «il a éteint son coeur»), le coeur qui recèle un désir inassouvi ou une humiliation et une avanie brûle d'un feu ardent qu'il faut éteindre en assouvissant son appétence ou en se vengeant pour laver l'affront.

Un coeur plein de chagrin, de rancœur ou de ressentiment demande à être dégonflé et désenflé (A : *fess qelb-u*, «il a dégonflé son coeur»); raconter aux autres nos petits malheurs dégonfle le coeur, mais le remède le plus efficace, sans doute, est de céder aux chaudes larmes consolatrices et réconfortantes.

Le coeur étant le siège de l'affectivité et des sentiments, c'est lui qui supporte les peines, les chagrins, les tourments, les émotions..., toutes les souffrances morales que nous infligent les aléas de la vie. En arabe et en berbère, le chagrin et l'affliction creusent dans le coeur (A : *ka tehfer f lqelb*; B : *la teqqaz g-g<sup>w</sup>ul*); une parole blessante ou une remarque désobligeante broient, pulvérisent ou moulent le coeur (A : *nesheq l-u lqelb*; *then l-u lqelb*; B : *la y i yberrey g-g<sup>w</sup>ul*, *la y i y33ad g-g<sup>w</sup>ul*, litt. «il me broie/moud dans le coeur»); un sentiment violent de répulsion ou de vengeance arrache le coeur ou plus exactement le déracine (A : *qelle? l-u lqelb*); enfin, un spectacle piteux, émouvant ou tragique coupe le coeur (A : *ka yqte? lqelb*; B : *la ytebbey ul*).

En français, on peut *fendre le coeur* en lui faisant éprouver un vif sentiment de chagrin et de pitié; on peut *agiter le coeur* en y répandant le trouble et les tourments; on peut *arracher le coeur* en l'affligeant vivement, on peut *briser le coeur* par des peines profondes ou des émotions pénétrantes; l'adversité, le malheur et divers déboires *percent le coeur* et le font souffrir; l'angoisse, l'anxiété et l'inquiétude *serrent le coeur*; enfin chaque jour apporte son lot de petites peines : déception, désenchantement, désillusion, découragement, etc. qui nous *crèvent le coeur*; mais heureusement que toute médaille a un avers. En effet, il arrive que le coeur connaisse un sort plus agréable.

La joie *dilate le coeur*, et c'est d'ailleurs à *coeur joie* que nous nous adonnons aux plaisirs de la vie; l'espoir *gonfle le coeur* et lui

donne l'ardeur suffisante pour aller de l'avant; on *réchauffe les coeurs* déçus, découragés, désabusés, en les consolant, en les reconfortant, en leur insufflant de nouvelles espérances.

## CONCLUSION

Au terme de cette étude, somme toute limitée, ayant laissé dans l'ombre plusieurs aspects de l'attribution des propriétés au coeur, nous pouvons résumer notre description en présentant les différentes étapes du transfert de sens dont le terme *coeur* fait l'objet, aussi bien en arabe, en berbère qu'en français.

1. *La métonymie* : Le terme *coeur* désignant d'abord l'organe est employé pour désigner la personne porteuse de l'organe. Le coeur devient ainsi le substitut de la personne. Il s'agit tout simplement d'un transfert de sens par synecdoque particularisante du type partie pour le tout, le coeur étant la partie et la personne le tout.

2. *La métaphore* : La métonymie ne suffit pas cependant pour expliquer la blancheur, la grandeur ou la mollesse du coeur (signifiant la personne) telles qu'elles sont employées dans les expressions étudiées. Les propriétés sont en effet comprises dans leur sens métaphorique et non pas dans leur sens littéral, concret. Par exemple, un *coeur blanc* ne veut pas dire une personne de couleur blanche. Il faut comprendre la blancheur du coeur dans son sens figuré obtenu par métaphore fondée sur l'analogie entre le concret et l'abstrait : le coeur est dit blanc parce que la personne qui l'a ainsi présente des qualités de franchise, d'honnêteté et de droiture et qu'elle n'est pas souillée par quelques travers et défauts. Une telle personne ressemble donc à une chose de couleur blanche et immaculée.

3. *La symbolique* : La métaphore par analogie n'est cependant possible que si la culture dont la langue est composante et véhicule, fait de la blancheur ou de toute autre propriété attribuée au coeur, un symbole. Il faut entendre par culture la vision et la conception du monde qui se manifestent à travers des interprétations diverses dont la symbolique constitue la forme centrale autour de laquelle s'articulent et s'organisent les structurations sémantiques. Ainsi la grandeur symbolise la générosité; la cuisson, l'irascibilité; la largeur, la clémence; etc. Ce sont ces représentations symboliques qui confèrent aux mots d'autres sens que la métaphore analogique prend en charge.

Comme notre comparaison l'a montré, c'est au niveau de la symbolique que l'arabe, le berbère et le français présentent quelques

ressemblances remarquables. La mémoire du coeur serait-elle commune ? Nous ne pouvons rien en savoir tant que le coeur n'a pas tout avoué. L'interrogatoire que nous avons mené nous a permis de déceler partiellement ses secrets. Mais la mémoire du coeur est intarissable. Si *le coeur vous en dit*, interrogez-le vous-mêmes, et vous verrez qu'il vous révélera des réminiscences insoupçonnées; vous en aurez ainsi *le coeur net*.

## REFERENCES

### Dictionnaires

#### Arabe

MERCIER, Henry (1951) : *Dictionnaire Arabe-Français*, Rabat, les Éditions La Porte.

#### Berbères

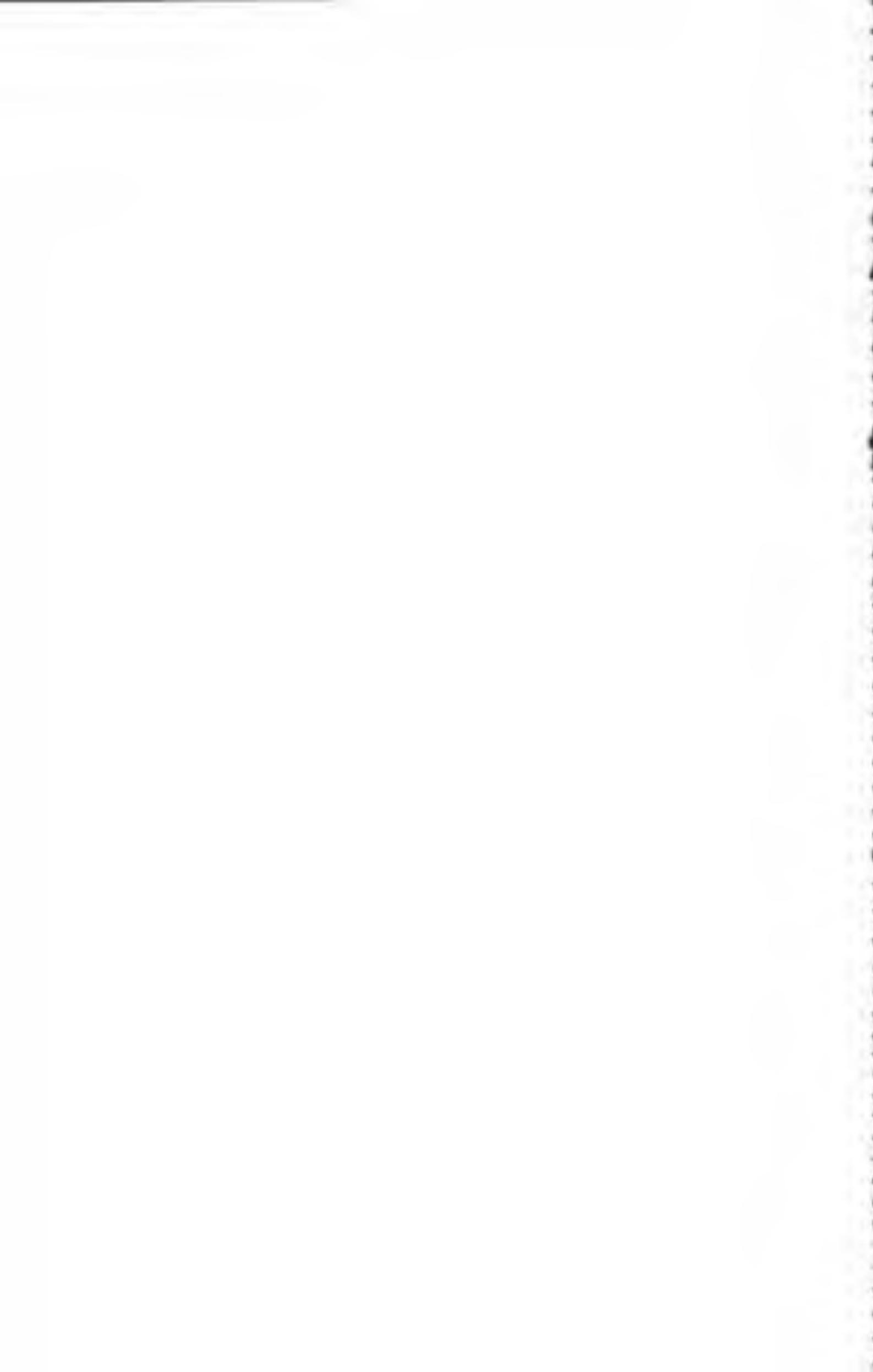
DALLET, Jean-Marie (1982) : *Dictionnaire Kabyle-Français (Parler des At Mangellat, Algérie)*, Paris, SELAF.

TAIFI, Miloud (1991) : *Dictionnaire Tamazight-Français (Parlers du Maroc central)*, Paris, L'Harmattan-Awal.

#### Français

*Le Petit Robert, Dictionnaire de la langue française* (1992) : Paris, édition Le Robert.

*DFC : Dictionnaire du Français Contemporain* (1971) : Paris, Librairie Larousse.



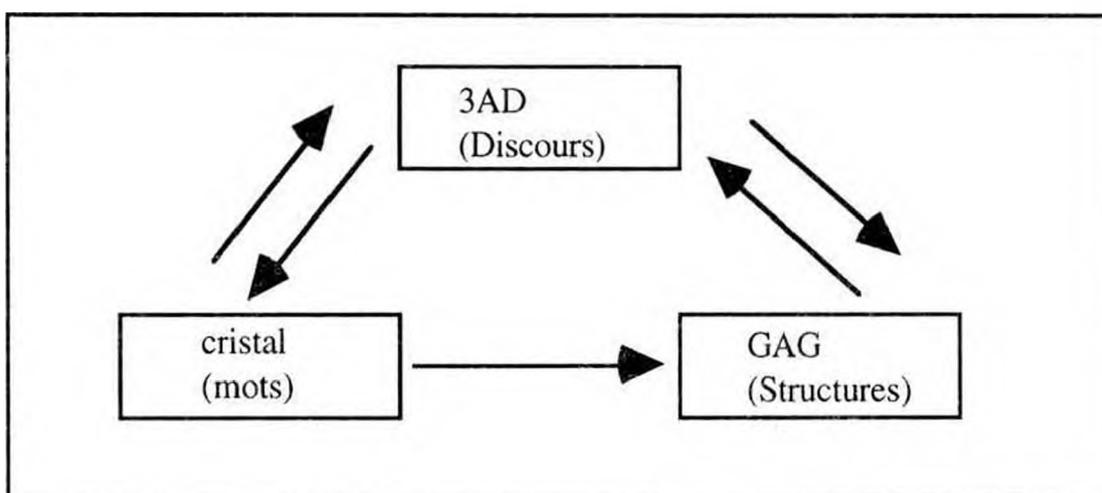
# **Cristal - GAG - 3AD95**

## **Une coordination pour transiter de l'analyse de mots à l'analyse de phrases puis à l'analyse de discours**

**Ismail TIMIMI**  
**Geneviève LALLICH-BOIDIN**

### **1 — Préambule :**

#### *1.a — Les partenaires de la coordination :*



Dans le cadre des travaux de recherche de l'équipe CRISTAL (Connaissances, Recherche d'information, Interfaces et Systèmes du Traitement Automatique des Langues), des programmes ont été développés indépendamment les uns des autres pour des buts bien précis. Récemment, trois logiciels se sont trouvés en parfaite

collaboration pour transiter d'une analyse de mots, à une analyse de structures de phrases, puis à une analyse de discours axée sur la paraphrase.

D'abord, un logiciel dit Cristal a été élaboré pour une analyse morpho-syntaxique, donnant en sortie une analyse morphologique des mots d'un corpus. Ensuite, un autre logiciel dit GAG a été conçu pour la recherche d'information dans les énoncés d'un corpus, il consiste à en extraire les énoncés répondant à une séquence syntaxique précise demandée par l'utilisateur. Enfin, un troisième logiciel, dit 3AD95 (Approximation de l'Analyse Automatique du Discours, version 95) et qui fera l'objet principal de ce papier, a été élaboré pour extraire des corpus les énoncés en relation de paraphrase. D'autres fonctionnalités en matière d'analyse textuelle peuvent être éventuellement déduites des fichiers créés lors des phases préliminaires de ces trois outils informatiques.

### *1.b — L'objet de la coordination :*

Partant d'un corpus formé de discours produits dans des conditions socio-linguistiques homogènes (une série de questionnaires, des articles sur un même événement, ou des résumés d'un même document, etc.), la méthode 3AD s'est démarquée des méthodes classiques de l'analyse de contenu, en se présentant comme un outil de dépouillement de corpus qui consiste à en extraire des classes d'énoncés, supposés en relation de paraphrase. Dans le sens que deux énoncés se trouvant dans une même classe paraphrastique partagent une information en commun, dit invariant en commun, pivot paraphrastique ou Noyau Informatif (NI). Ce qui se rapproche du concept des segments répétés dans la théorie des statistiques textuelles.

Notons au passage que l'homogénéité des conditions de production des discours assure, en principe, une répétitivité dans le lexique du corpus pour garantir par conséquent la présence des noyaux informatifs des classes. Les corpus produits par des enfants répondent parfaitement à cette particularité vu que le vocabulaire employé est très restreint et les structures syntaxiques sont presque identiques.

En dehors de la paraphrase qui fait son originalité, la méthode 3AD, avec ses classes d'énoncés «proches», peut être utilisée également dans l'indexation et la recherche d'information, et dans le dépouillement d'enquêtes (traitement des questions ouvertes).

Enfin, pour confronter des énoncés et en extraire ceux qui sont «proches», la méthode 3AD conjugue plusieurs disciplines, particulièrement des algorithmes mathématiques (algébriques et métriques) pour calculer les distances entre énoncés et des formalismes linguistiques (morphologie flexionnelle et dérivationnelle, syntaxe, sémantique) pour focaliser les points de parenté entre les constituants des énoncés.

## 2 — OUTILS MATHÉMATIQUES :

### 2.a — Généralités :

Soit A un énoncé séquentiel formé d'éléments c. B désigne également un énoncé.

*Opérateur de Transition élémentaire* : Un opérateur de transition élémentaire, noté OTE, appliqué sur un énoncé A est une des fonctions suivantes :

- Effacement  $E(c)$  : effacer de l'énoncé une occurrence quelconque de l'élément c;
- Insertion  $I(c)$  : insérer dans l'énoncé une occurrence quelconque de l'élément c;
- Substitution  $S(c,c')$  : substituer dans l'énoncé c'à une occurrence quelconque de c.

*Coût d'un OTE* : à chaque opérateur OTE, une fonction coût  $g$  associe un nombre réel positif ou nul qui vérifie :  $\gamma(E(c)) = \gamma(I(c))$  et  $\gamma(S(c,c')) = \gamma(S(c',c))$ .

*Suite de Transition* : Soient A et B deux énoncés, il existe au moins une suite d'OTE qui permettent de transiter de A vers B. Par exemple, on peut transiter par l'effacement de tous les éléments de A puis l'insertion de tous les éléments de B. Notons  $\Omega(A,B)$  l'ensemble infini des suites d'OTE permettant la transition de A vers B.

*Coût d'une Transition* : Soit s une suite appartenant à  $\Omega(A,B)$  et permettant la transition de A vers B. Le *coût de transition* de s, noté  $\bar{\gamma}(s)$  est la somme des coûts des OTE éléments de s.

*Distance de Transition* : Modulo un facteur de cadrage près, dépendant de la longueur des énoncés, la *distance de transition* de A

vers B, noté  $\gamma^*(A,B)$ , est le coût de transition minimal dans les suites de transition de A vers B :  $\gamma^*(A,B) = \min \bar{\gamma}(s)$  où  $s \in \Omega(A,B)$ .

*Relation Directe* : deux énoncés A et B sont dits en *relation directe* si leur distance de transition est inférieure à une borne  $\alpha$  fixée préalablement par l'utilisateur. Cette borne est dite seuil de distorsion ou limite de tolérance de déformation (Fuchs, 1994). A et B sont en relation directe ssi  $\gamma^*(A,B) \leq \alpha$ .

*Classe Paraphrastique* : Une classe paraphrastique d'un énoncé est formée des énoncés du corpus qui sont en relation directe ou liés par la fermeture transitive à cet énoncé.

Classe (A) =  $\{B / \exists \text{ une suite d'énoncés } C_{i=1\dots n} \text{ du corpus : } \gamma^*(C_i, C_{i+1}) \leq \alpha \text{ avec } A=C_1 \text{ et } B=C_n\}$

## 2.b — Application :

*Remarque* : Les exemples d'énoncés cités ci-dessous, sont extraits d'un corpus en cours d'expérimentation (Zlotowicz, 1978) formé de récits de rêves d'enfants.

Dans les normes de 3AD, le corpus est formaté de sorte que les éléments c d'un énoncé A soient des couples (b/t) où b est la base lexicale lemmatisée du mot et t son trait morphologique. Par exemple, l'énoncé : «la nuit, j'avais des rêves mystérieux» est analysé et présenté ainsi : «(le/D) (nuit/F-nom) (je/Y) (avoir/V) (des/D) (rêver/F-nom) (mystère/F-adj)».

Comme l'effectif de traits morphologiques dans un corpus est plus restreint, et par suite plus maîtrisable que celui de bases lexicales, les calculs ci-dessus ne seront effectués qu'en fonction de deux paramètres :

- Les traits morphologiques et
- Le fait si les bases lexicales en jeu sont égales ou différentes.

Ainsi, la fonction coût devient entièrement déterminée par le tableau CIE, et les deux matrices symétriques CSE, CSD (deux autres matrices sont en option, cf. infra) :

• Pour l'insertion et l'effacement, le coût est : CIE (t) =  $\gamma$   
 (E(b/t)) =  $\gamma$  (I(b/t))

• Pour la Substitution à bases Egales : CSE (t<sub>1</sub>, t<sub>2</sub>) =  $\gamma$   
 (S(b/t<sub>1</sub>, b/t<sub>2</sub>)) =  $\gamma$  (S(b/t<sub>2</sub>, b/t<sub>1</sub>))

• Pour la Substitution à bases Différentes :  $CSD(t_1, t_2) = \gamma(S(b_1/t_1, b_2/t_2)) = \gamma(S(b_2/t_2, b_1/t_1))$

Par exemple, soient A : (le/D) (garçon/F-nom) (attraper/V) (le/D) (sorcier/F-nan) (/T);

et B : (le/D) (garçon/F-nom) (croiser/V) (le/D) (sorcier/F-nan) (/T);

Pour transiter de A vers B, il existe plusieurs suites de transition, notamment :

la substitution à bases différentes de (attraper/V) par (croiser/V) dont le coût est  $CSD(V, V)$ ;

ou l'effacement de (attraper/V) puis l'insertion de (croiser/V) dont le coût est 2 CIE (V).

• En outre, la condition imposée  $CSE(t_1, t_2) \leq CSD(t_1, t_2)$  favorise les regroupements lorsque les bases lexicales sont égales. Par exemple : substituer (rêver/V) par (rêver/F-nom) doit coûter moins cher que substituer (rêver/V) par (crime/F-nom).

En attribuant à chaque trait un poids numérique qui correspond au coût de substitution entre deux couples (b/t) partageant ce même trait, et qui reflète l'importance de son apport informatif dans l'énoncé, les matrices suivantes répondent parfaitement à l'objectif du système :  $CIE(t_i) = 2p_i$        $CSD(t_i, t_j) = \max(p_i, p_j) + |p_i - p_j|$   
 $CSE(t_i, t_j) = 2 |p_i - p_j|$ .

Elles assurent, entre autres, que deux énoncés ayant un même verbe, par exemple, seront plus proches que ceux ayant seulement une préposition en commun.

### 3 — MODELE LINGUISTIQUE :

#### 3.a — Interface CRISTAL-3AD95 (dimension morphologique)

Pour pouvoir confronter les énoncés entre eux et en extraire ceux en paraphrase, le 3AD95 a besoin dans son étape de prétraitement, d'une couverture lexico-morphologique du corpus qui permet d'en reconnaître aussi bien les bases lexicales b que les traits morphologiques t correspondants. Ensuite, tout le calcul est basé sur le genre et le nombre d'OTE dans chaque transition. Sans cette procédure, des formes telles que rêveur, rêve, rêver, rêvées, ... dépourvues des liens morphologiques entre elles, seraient considérées des items complètement différents. On ne pourrait alors rendre

compte que ces formes peuvent contribuer éventuellement à la détermination d'un noyau informatif (NI); ce qui entraîne explicitement des résultats incomplets et imparfaits. Par conséquent, cette normalisation des énoncés en langage naturel permet de rapprocher les formes de surface, augmenter leur nombre d'occurrences, et extraire par suite les énoncés axés sur les NI.

Effectivement, cette couverture est du ressort de l'analyseur Cristal. Modulo quelques modifications d'adaptation, une interface de liaison est instaurée pour établir une coopération entre les deux systèmes, ce qui a rendu le prétraitement de 3AD quasi automatique. L'analyse (Lallich-Boidin, 1990) associe les deux dimensions de la morphologie flexionnelle et dérivationnelle pour permettre de scruter le maximum de noyaux informatifs entre les énoncés. Elle donne en sortie la base lexicale (de la forme de surface) subordonnée par le trait morphologique correspondant qui contient la catégorie grammaticale éventuellement précisée par des variables telles que le genre, le nombre, le temps, etc. Par ailleurs, pour régulariser les ambiguïtés possibles lors de l'analyse (homographes et autres ...), Cristal utilise un filtre statistique fondé sur la théorie des chaînes de Markov (Bouchaffra, 1994) de sorte qu'à chaque forme du corpus ne corresponde qu'un seul couple (b/t). Les traits principaux utilisés dans 3AD sont principalement ceux adoptés par Cristal, un ajout d'un nouveau trait reste envisageable pour une étude particulière; par exemple, pour avoir une classe formée de tous les énoncés pivotés par un invariant à très haute occurrence, il suffit de doter ce dernier d'un trait spécial et de le pondérer fortement. Les traits utilisés régulièrement par 3AD sont réparties en deux séries :

- Série LEX (pondération forte) : V verbe; F(nom et/ou adj); W adverbe.
- Série GRAM (pondération faible) : Q subordonnant; C coordonnant; P préposition;

Y pronom préverbal; D déterminant; T ponctuation.

F comporte trois situations : F-nom pour les formes toujours nom (valise); F-adj pour celles toujours adjectif (joli); F-nan pour celles tantôt nom, tantôt adjectif (méchant).

Enfin, dans le cas des corpus caractérisés par un nombre élevé de liens dérivationnels, et nécessitant une analyse rigoureuse, on introduit en option une matrice, dite Coût de Substitution à bases Lexicalement égales (CSL). Elle prend en compte le nombre

d'opérations suffixales qui séparent deux dérivés et évalue plus correctement le degré de proximité entre eux (Timimi, 1997). Ainsi entre les deux couples (nationalisation/F-nom) et (nation/F-nom), c'est le premier qui sera plus proche de (nationaliser/V). Autrement dit, s'il existe un lien dérivationnel entre deux formes, certes, elles ne seront pas à bases lexicales différentes, mais elles ne seront pas, non plus, à bases lexicales égales forcément. Cette matrice CSL réduit le nombre de classes de paraphrase mais elle permet de les générer très rigoureusement.

Remarque : pour ne pas alourdir les exemples cités ci-après, nous ne ferons pas figurer l'analyse morphologique devant les formes. On se contentera des formes de surface.

### 3.b — Interface GAG-3AD95 (dimension syntaxique) :

Une fois que le corpus est couvert par l'analyse morphologique, le système 3AD confronte les énoncés pour en extraire, sous forme de classes, ceux en paraphrase. Ceci se fait à partir du calcul du coût minimal parmi les suites possibles de transition.

Soit A un énoncé de longueur n, notons  $A_{(i)}$  le i-ème élément de A, et  $A_i$  la séquence formée de  $A_{(1)}A_{(2)}...A_{(i)}$ . Il en résulte que  $A_n = A$  et  $A_0$  est l'énoncé vide. Soit B un énoncé de longueur m. Sellers, Wagner et Fisher proposent [Sellers, 1974] le schéma récurrent suivant qui permet de calculer  $\gamma^*(A,B)$ . Sachant que :

$$\gamma^*(A_0, B_j) = \sum_{k=1}^j \gamma(I(B_{(k)})) \text{ et } \gamma^*(A_i, B_0) = \sum_{k=1}^i \gamma(E(A_{(k)}))$$

$$\gamma^*(A_{i-1}, B_{j-1}) + \gamma(S(A_{(i)}, B_{(j)}))$$

$$\gamma^*(A_i, B_j) = \text{Min } \gamma^*(A_i, B_{j-1}) + \gamma(I(B_{(j)}))$$

$$\gamma^*(A_{i-1}, B_j) + \gamma(E(A_{(i)}))$$

Cet algorithme implique que les arcs représentant les différentes transitions ne se croisent jamais, c'est-à-dire qu'il prend en compte la position de l'élément dans l'énoncé. Alors, l'avantage réside dans le fait que la distance entre les deux énoncés «le sorcier a pu attraper l'enfant» et «l'enfant a pu attraper le sorcier» est bien non nulle. Mais l'inconvénient réside dans le cas des allotaxies; la distance entre «j'ai fait un mauvais rêve» et «j'ai fait un rêve mauvais» est malheureusement non nulle alors que les deux énoncés sont en parfaite relation de paraphrase.

Pour pallier ce défaut, on intègre un nouvel opérateur de transition élémentaire, dit «transposition», et dont le coût est nul. Le principe consiste à localiser les énoncés qui nécessitent une modification dans la position de leurs éléments et demander à l'utilisateur d'approuver ou de désapprouver une éventuelle transposition au sein de l'énoncé. Ceci permet d'avoir une uniformité relative dans la syntaxe des énoncés du corpus. Pour ce faire, on se sert du système GAG, un programme réécrit en langage C par I. Timimi et qui génère tous les énoncés conformes à la structure demandée par l'utilisateur (Antoniadis, 1984).

Par exemple, l'utilisateur fixe un cadre à structure partiellement libre (ou partiellement déterminée) du style  $X^* F(\text{nom}) F(\text{adj}) X^*$ ; et le système extrait tous les générateurs maximaux (ici les énoncés) formés d'un nom précédant un adjectif; puis l'utilisateur décide ou non l'application de l'opérateur «transposition». Cette adhésion de GAG a donné au système 3AD, un nouvel élan associant à l'automatisation globale une pertinence dans le dépouillement.

### *3.c — Et la dimension sémantique ?*

La paraphrase soutenue par 3AD ne fait appel à aucun dictionnaire externe, elle se fonde uniquement sur les invariants en commun qui peuvent exister dans les énoncés à l'issue de l'analyse morphologique. Ainsi, les formes «rêver, rêve, rêveur, rêvées, ...» deviennent en parenté grâce au procédé morphologique, et contribuent par suite à l'élaboration de classes de paraphrase dans 3AD. Néanmoins, la forme «songer» qui est normalement proche de l'ensemble, dans le sens de paraphrase, se trouvera malheureusement à la même distance que la forme «se laver» qui n'a aucun rapport avec l'ensemble. D'où, des imperfections surgissent et nécessitent des aménagements pour une génération pertinente de classes de paraphrase.

Exemple 1:

- |                              |                            |
|------------------------------|----------------------------|
| a) j'ai croisé un fantôme    | b) j'ai croisé un clochard |
| c) j'ai rencontré un sorcier | d) j'ai acheté un cartable |

Dans cet exemple, les énoncés (a) et (b) peuvent faire l'objet d'une même classe de paraphrase pivotée par le noyau «j'ai croisé», alors que (c) et (d) seront à la même distance de la classe bien que (c) soit plus proche de la classe que (d).

### Exemple 2 :

- a) je préfère la campagne
- b) je choisis la campagne
- c) j'aime la campagne
- d) je déteste la campagne

Dans cet exemple, extrait d'un autre corpus, les quatre énoncés sont équidistants. Si l'un se trouve dans une classe, nécessairement les trois autres y seront également, bien que (d) présente un élément parasite dans une éventuelle classe.

### Exemple 3 :

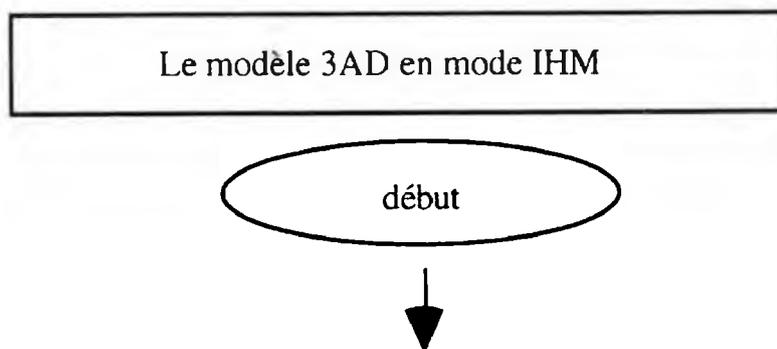
- a) Les enfants redoutent les sorciers et les voleurs
- b) les gamins craignent les magiciens et les cambrioleurs

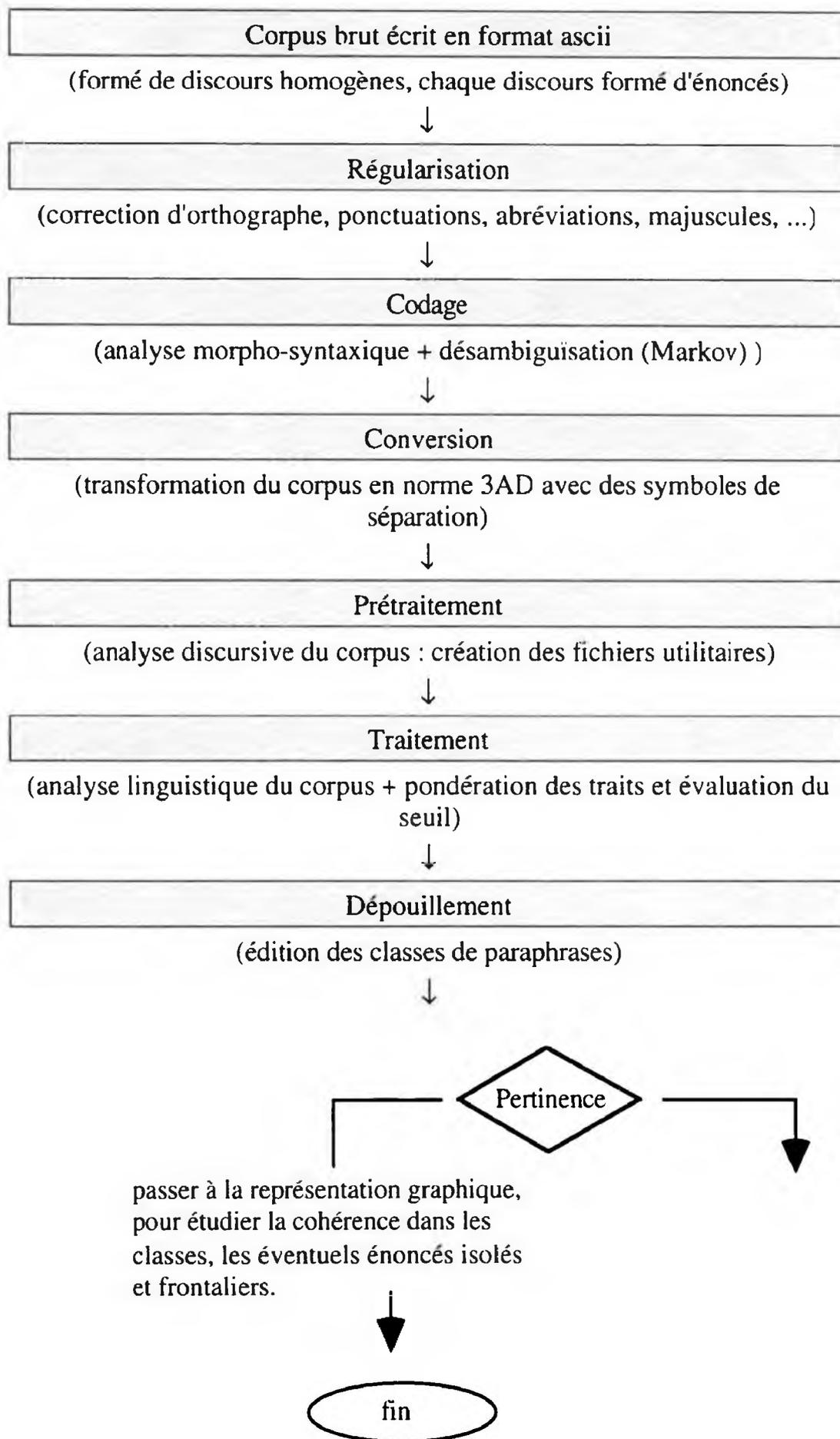
Ces deux énoncés ne se trouveront jamais dans une même classe bien qu'ils soient en parfaite relation de paraphrase car ils n'ont pas d'invariant en commun selon la théorie 3AD.

Pour remédier à ces imperfections, il s'est avéré judicieux d'intégrer un dictionnaire externe de mots synonymes. Ainsi les formes «gamins» et «garçons» ne seront scrutées ni à bases différentes, ni à bases égales, mais plutôt à *bases synonymes*. D'où l'introduction d'une nouvelle matrice de coût, dite Coût de Substitution à bases Synonymes (CSS) dont les valeurs sont comprises entre celles de CSD et de CSE.

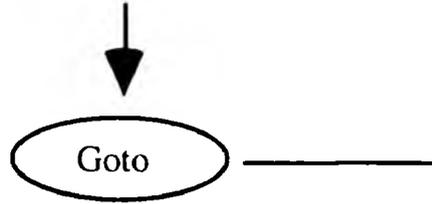
## CONCLUSION ET PERSPECTIVES :

D'autre part, un verbe modifieur et un verbe central ne devraient pas avoir le même poids car ils n'ont pas le même apport d'information, idem pour les fonctions grammaticales des noms etc. Donc il faut transformer le calcul de distances entre chaînes séquentielles de couples (f/t) (énoncés du corpus) au calcul de distances entre arbres (syntagmes du corpus) pour prendre en compte aussi bien la position que la fonction des formes (Tanaka, 1988).





Il faut modifier la pondération des traits et le choix du seuil. Si l'imperfection persiste, on peut être obligé de repartir à zéro et d'adopter soit une nouvelle segmentation via GAG et/ou une nouvelle catégorisation via Cristal, ou carrément retravailler le texte en restituant les anaphores et rendre les énoncés plus uniformes en syntaxe.



Annexe : Menu de 3AD95 (Timimi, 1995)

Fichier	Saisie	Traitement	Fonction	Calcul	Analyse	Aide
Nouveau	Corpus	Morphologie	Indexation	Matrices	Graphe	Doc
Liste	Marqueurs	Syntaxe	Numérisation	Seuil	Invariants	Commandes
Impression	Pondération	Sémantique	Traits	Classes	Cohérence	Démo
Quitter	Eloignement	Conversion				Scripts

## BIBLIOGRAPHIE

- ANTONIADIS, G. (1984) : ...laboration d'un Système d'Analyse morpho-syntaxique d'une langue naturelle : application en Informatique Documentaire, coll. «Cahiers du CRISS», Grenoble II, vol 5.
- BOUCHAFFRA, D. & J. ROUAULT (1994) : «Capturing observations in a nonstationary hidden Markov model» in P. Cheeseman and R. W. Oldford (eds.) - Springer - Verlag.
- DEL VIGNA, C. (1977) : Paraphrase discursive - essai d'évaluation quantitative - le système informatique AAD75, coll. «T. A. Informations», vol 1, pp. 45-58.
- FUCHS, C. (1994) : «Paraphrase et énonciation», Paris, Ophrys, pp. 58-67.
- LALLICH-BOIDIN, G, G. HENNERON et R. PALERMITI (1990) : Analyse du français : achèvement et implantation de l'analyseur morpho-syntaxique, coll. «Cahiers du CRISS», Grenoble II, vol 16.
- SELLERS, P. H. (1974) : An algorithm for the distance between two finite sequences, «Journal of Combinatorial Theory (JCT)», Series A, vol. 16, n° 2.
- TANAKA, E. et K. TANAKA (1988) : The tree-to-tree Editing Problem, «International Journal of Pattern Recognition and Artificial intelligence», vol. 2, n °2, pp. 221-240.
- TIMIMI, I. (1995) : «Manuel d'utilisation de 3AD95, version 1.0», document interne de GRESEC-CRISTAL - Université Grenoble 3.
- TIMIMI, I. et J. Rouault (1997) : «La paraphrase comme relation d'équivalence dans l'analyse automatique du discours», Grenoble, 4ème Conférence annuelle sur le Traitement Automatique du Langage Naturel (Taln '97).
- ZLOTOWICZ, M. (1978) : «Les cauchemars de l'enfant», PUF.

# Mémoire des mots en traitement automatique des langues et «étymologie synchronique»

**Christian BOITET**

*GETA, CLIPS, Institut IMAG, (UJF & CNRS),  
Grenoble, France*

## INTRODUCTION

A priori, la «mémoire des mots», au sens où «les mots se souviennent», n'a qu'un rapport ténu avec le traitement automatique des langues, puisque, dans la mesure où on parle de souvenir, on pense naturellement qu'il s'agit d'étymologie historique, ou diachronique. Une longue pratique de l'informatique linguistique, et plus particulièrement du TAL et de la TAO, nous montre cependant qu'il n'en est rien.

Il y a en effet un aspect de l'étymologie, que nous appellerons «synchronique» par opposition au précédent, qui peut être très important et utile, non seulement pour diverses applications en TALN, dont la TAO, mais aussi pour la création terminologique «naturelle» dans des langues peu soutenues financièrement.

Cette notion est directement inspirée des études de C. Gruaz (1987, 1990, 1995) sur la «structure des mots». Cependant, elle ne concerne pas que la structuration de familles de mots par des principes linguistiquement motivés, mais aussi la traduction systématique d'une structure de mots d'une langue dans une autre, et des factorisations accidentelles, mais économiquement utiles, comme par exemple la fusion de parties de paradigmes de paires comme parer/parier.

L'étymologie synchronique intervient aux trois niveaux du stockage, de la reconnaissance et de la création de mots ou plus généralement de termes.

## **I. — STOCKAGE**

### **I.1 — Mémoire «brutale»**

Au premier niveau, la mémoire des mots en TALN consiste tout simplement à les stocker, pour ainsi dire brutalement. Dans le cas le plus simple, par exemple pour la détection d'erreurs orthographiques, on dispose d'une très grande liste de formes fléchies, qu'on cherche à comprimer. L'approche la plus répandue consiste à représenter cet ensemble de chaînes de caractères comme un automate d'états finis, qu'on minimise par des algorithmes généraux très performants.

### **I.2 — Mémoire de la structure morphologique et «compression paradigmatique»**

B. Meddeb-Hamrouni (1996) a montré dans sa thèse qu'on peut considérablement améliorer le taux de compression en utilisant une connaissance assez grossière de la structure des mots, réduite à la liste des paradigmes telle qu'on la trouve dans tous les manuels et dictionnaires. L'idée consiste à changer de problème en ajoutant aux caractères du système d'écriture de la langue considérée autant de nouveaux caractères que de paradigmes. Sur l'arabe, l'amélioration obtenue est de l'ordre de 30 %, et permet de comprimer 40Mo de formes en moins de 1,4Mo (une disquette). Il faut dire que, dans ce cas, on dispose des paradigmes flexionnels et dérivationnels.

Une étude plus fine de la structuration des mots d'une langue donnée peut conduire à des compressions encore plus importantes. S'il s'agit par exemple de stocker non seulement des chaînes, mais les propriétés associées, et si on associe aux formes et lemmes une structuration arborescente, on pourra beaucoup gagner en ne stockant pas toutes les propriétés, mais en représentant les règles d'héritage, et en ne stockant que ce qu'elles ne permettent pas de calculer.

### **I.3 — Mémoire de la structure lexico-sémantique locale et globale**

Enfin, cette idée peut se généraliser à toute l'organisation d'une base lexicale multilingue avec acceptions interlingues, dans laquelle les éléments monolingues sont éventuellement structurés (arbres de

composition, fonctions lexico-sémantiques), et sont reliés entre eux et aux acceptions par des liens de différents types qui structurent l'ensemble.

Dans une telle base, les acceptions interlingues sont dans une sous-base centrale, à laquelle sont reliées les sous-bases monolingues. Au niveau de chaque base monolingue, on trouve une structuration par les fonctions lexico-sémantiques retenues. Au niveau des acceptions interlingues, on trouvera l'analogie, c'est-à-dire le passage d'une entité à la qualité associée, au processus dont elle est le résultat; d'un prédicat binaire (à agent et patient) à la propriété associée (où le patient a disparu, comme dans «boire» — «x boit y» dénote l'action et «x boit» la propriété d'être buveur — mais nous nous plaçons ici au niveau interlingue).

Un exemple simple de la factorisation qu'on peut obtenir est celui des relations purement sémantiques, comme la synonymie, l'antonymie, l'hyponymie, etc. Dans une telle organisation, on peut ne stocker ces relations qu'au niveau interlingue, et les déduire pour chaque base monolingue, dans laquelle chaque «sens de mot» est relié à une acception interlingue.

Bien plus, une telle factorisation pourra permettre d'éviter ou au moins de réduire sensiblement les incohérences : si deux mots (ou plutôt sens de mots A1 et B1 de la langue L1 ont des traductions A2 et B2 en L2 obtenues en passant par des acceptions antonymes A et B, A1 est antonyme de B1 en L1 et de même pour A2 et B2 en L2, par nécessité. Au contraire, dans une organisation classique, rien n'empêche que A1 soit antonyme de B1 dans le dictionnaire monolingue de L1 alors que A2 n'est pas antonyme de B2 dans celui de L2.

Il ne semble malheureusement pas possible de se passer des fonctions lexico-sémantiques dans les dictionnaires monolingues, en les déduisant des relations sémantiques analogues entre acceptions interlingues. C'est en effet qu'elles ont un caractère fortement lexical.

Par exemple, on ne voit pas comment définir une fonction sémantique «Magn» sur les acceptions interlingues, s'appliquant à des entités, et donnant des ensembles de propriétés. En effet, si #fever(body) dénote l'acception correspondant au sens fièvre.1 (symptôme de maladie), Magn(#fever(body)) devrait contenir les acceptions correspondant aux adjectifs ou expressions adjectivales apparaissant dans les différentes langues en collocation avec un terme correspondant à #fever(body) pour exprimer la force de ladite fièvre.

Mais alors, on obtiendrait des fonctions lexico-sémantiques surgénérantes dans les différentes langues : «haute fièvre» car «high fever», etc.

Au niveau interlingue, on pourrait simplement avoir une entité représentant la fonction (universelle) Magn, par exemple #[Magn]. Au niveau monolingue, on stockerait pour chaque lemme les valeurs de Magn, par exemple Magn(fièvre.l) = {forte, élevée, de cheval}. En traduction automatique, si on est capable en analyse de transformer «fièvre élevée» en quelque chose comme

fièvre.l—qual—>élevée[Magn], on pourra ensuite passer dans une représentation interlingue à :

#fever(body)—mod—>#[Magn], puis revenir en anglais à

fever.l—qual—>high[Magn], donnant finalement une traduction correcte.

## II. — RECONNAISSANCE

### II.1 — Analyse flexionnelle et dérivationnelle

Il est presque trivial de dire que la connaissance de la structure des mots permet de les reconnaître sans devoir tous les stocker, ou en n'en stockant que des morceaux. Mais ce n'est pas si sûr. Déjà, l'analyse morphologique flexionnelle pose un certain nombre de problèmes intéressants dans les langues à morphologie concaténative riche (langues romanes, langues slaves) ou très riche (langues finno-ougriennes). Quant aux langues à morphologie interne comme l'arabe, il ne semble pas exister de modèle abstrait satisfaisant et efficace, comme peuvent l'être les automates d'états finis ou les grammaires hors-contexte attribuées pour les cas précédents.

### II.2 — Analyse compositionnelle

Dès qu'une langue présente une morphologie compositionnelle riche, c'est-à-dire si elle peut former de façon libre des mots composés complexes, il est absolument impossible de procéder par stockage exhaustif. On cite toujours le cas de l'allemand, mais ce phénomène se présente dans toutes les langues. La différence est qu'il est confiné à des classes très particulières de termes dans la plupart des langues (par exemple, termes de chimie en français), alors que dans d'autres, comme l'allemand, il fonctionne de façon générale. Quoi qu'il en soit,

la reconnaissance de mots composés demande au minimum de reconnaître un tel mot comme une (ou plusieurs) suite de mots plus simples. Mais cela ne suffit en général pas pour calculer les propriétés du mot complet. Pour cela, il faut s'appuyer sur une structuration interne réursive.

### **II.3 — Analyse des mots inconnus ou incomplètement connus**

Nous venons de parler de la reconnaissance de mots explicitement ou implicitement connus (car on connaît leurs morceaux et on dispose de règles compositionnelles pour calculer leurs propriétés globales). Dans de nombreuses applications de TALN, il faut aussi reconnaître des «mots inconnus». Si on dispose d'une bonne connaissance de la structure des mots en général, et encore mieux d'une bonne connaissance de la structure des mots non décrits par les données (bases, affixes, paradigmes, règles) du système, ce sont exactement les mots dont une ou plusieurs bases sont absentes du système. Ils seraient donc analysables normalement si on complétait le dictionnaire des bases.

Dans le cas de systèmes restreints, il est aussi possible que l'analyse de tels mots nécessite l'introduction de nouveaux paradigmes, ou la relaxation de certaines contraintes dans les règles. En tout état de cause, la reconnaissance de «mots inconnus» dépend crucialement d'une bonne connaissance de l'étymologie synchronique, au moins en ce qui concerne les mots des corpus qu'on a à traiter, et de sa relation précise avec le système d'analyse implémenté. Dans Guilbaud et Boitet (1997), nous avons ainsi récemment montré comment on pouvait réutiliser les règles, les affixes et les paradigmes d'un analyseur morphologique très couvrant, mais bien sûr non exhaustif, pour construire très rapidement un dictionnaire des affixes de mots inconnus et une «grammaire des mots inconnus» très réduite, dont les articles et les règles sont des adaptations des affixes et des règles utilisés pour les mots complètement décrits, arrivant ainsi à un système reconnaissant très finement les mots inconnus.

### **III. — CRÉATION**

L'étymologie synchronique nous semble aussi importante pour la création terminologique, qu'on parle de TALN ou de création néologique directe (dans une langue) ou indirecte (par traduction ou transduction).

### III.1 — Traductions par analogie en TAO

Reconnaître la structure de mots inconnus permet de proposer des traductions «par analogie». En TAO, cela est bien sûr très utile, voire indispensable quand il y a une classe ouverte non bornée en pratique. C'est le cas des noms de lieu dans le système METEO (Chandioux 1988), car il n'est pas envisageable de stocker les noms de toutes les villes, rivières, montagnes, etc., alors qu'on peut atteindre une qualité quasi-parfaite avec un dictionnaire de moins de 3 000 entrées. C'est ou ce sera également très utile dans les systèmes d'interrogation multilingue de bases de données documentaires.

Un cas extrêmement simple est celui de la traduction des noms propres du russe en français (Boitet et Nédobejkine 1981). Si on a reconnu (grâce à la majuscule initiale, ou au suffixe) qu'il s'agit d'un nom propre, il suffit de passer au lemme et de le transcrire dans la translittération choisie. Le groupe de Prague [Hajicova, Kirschner] a également fait un travail célèbre de traduction par analogie dans le cadre d'un prototype de traduction anglais-tchèque. Il s'agit là d'une transduction morphématique mise en œuvre dans certains cas bien répertoriés (mots en -logy, -ism, -ist...).

Il s'agit donc ici à la fois de reconnaître un mot inconnu, qui peut être aussi bien un néologisme qu'un mot attesté mais d'une classe partiellement couverte ou ignorée par le système, et d'en proposer une traduction dans la langue cible. Il y a évidemment des risques d'inexactitude, qui sont au fond la contrepartie du compromis retenu entre ce qui est stocké et ce qui est calculé.

### III.2 — Création néologique «directe»

Nous voudrions enfin parler du rôle de l'étymologie synchronique dans la création néologique usuelle. Dans notre laboratoire, nous avons en effet fait un peu de «néologie directe», et créé, par analogie avec les paires «logiciel» <—> «software» et «collecticiel» <—> «groupware», quelques néologismes comme la paire «linguiciel» <—> «lingware», et «courriel» pour «e-mail [message]», «adressielle» pour «e-mail [address]», «partagiciel» pour «shareware», «gratuiciel» pour «freeware», dont certains ont «pris» et d'autres pas.

Pour que de tels néologismes paraissent heureux, l'analogie de surface est indispensable (terminaisons en -[c]iel <—> -ware, par exemple), mais ne suffit pas. Il semble que ce qui fonctionne suppose des analogies plus profondes, tant dans la langue visée (ici le français) que dans la *lingua franca* de référence (ici l'anglais).

Par exemple, le -iel ou le -ciel semble être maintenant perçu en français comme un lexème analogue à -ware, quoique le ciel ne soit évidemment pas soudainement devenu une quelconque marchandise ! Il faut aussi des schémas de valence et des schémas dérivationnels parallèles. Par exemple, «naviguer» ne peut pas être un bon candidat pour «to browse», pour la simple raison (il y en a d'autres) qu'il n'est pas transitif.

À l'inverse, la récente proposition de «mél» pour «messagerie électronique» n'a quasiment aucune chance d'être acceptée par l'usage, car elle est contraire à la phonétique française et repose sur une fausse analogie. En effet, aucun mot français ne semble pouvoir se terminer par «-él», pour des raisons phonétiques. Par contre, «-el» est fréquent : «bel, ciel, bridel, gel, caramel, tunnel, appel, sel, hôtel...». C'est sans doute pour cela que «tél-» de «téléphone, télégraphe...» s'est transformé en «-tel» dans «minitel, télétel, supratel...».

Quant à la fausse analogie, Henri Landroit, spécialiste de la question, a récemment souligné sur le... courriel ses ressorts cachés et inavouables dans un texte savoureux. Citons pour le plaisir :

«Mél.», que plus haut je saluais hypocritement comme une abréviation graphique, n'en est pas une, sinon elle s'écrirait «M. él.» [...] En effet, l'abréviation graphique française exige un point abrégatif après chaque mot abrégé par apocope. (Il suffit de le savoir. Non ? Si ? Quelle est la position de la commission générale ? En tout cas, le moindre correcteur débutant le sait, ainsi que la plupart des maquettistes et un nombre considérable de secrétaires.)

Ainsi, par exemple, «par exemple» s'abrège en «p. ex.» et non en «pex.» [...] Ce «pex.» vous semble grotesque ? Pas plus que votre «Mél.» [...] En omettant le point abrégatif après le «M» et en le maintenant après le «L», nos terminologues commissionnés mél.angent tout et viennent de créer un monstre, à la fois par troncation ou abrègement (osé mais audible : él) et siglaison (M), donc siglaison «acronymique» (inavouée mais évidente : M + él), jusqu'ici ça va encore, c'est classique, ET abréviation graphique (prétendue telle : Mél.). À l'arrivée, un machin informe, inutilisable (sauf sur ordre[...]), que l'on ne sait comment prononcer. Une abréviation, comme «Tél.», lu «téléphone» ? Une troncation, comme «Péritel», lu comme ça s'écrit ou, plus précisément, écrit comme ça se prononce [...] Un acronyme honteux ? Un mot-valise affublé d'un point abrégatif bouffon ?

Parlons-en de ce petit point noir [...] Avec lui, nos experts laissent entendre qu'il s'agit de l'abréviation graphique d'un mot

commençant par «Mél-», par exemple «mélasse». S'ils éliminent leur seul point abrégatif, ils obtiennent un acronyme monosyllabique, ridicule et abscons, sauf pour ceux qui le considéreront comme une copie servile de «mail». Bien fait [...] Ils se retrouveront dès lors avec l'accent aigu, incongru dans cette position [...] Il faudra aussi supprimer l'accent : «mel» [...] Tant qu'on y est, vu ce qui reste, autant tout supprimer [...] Je suppose qu'ils ont été satisfaits de l'heureuse ressemblance entre «Tél.» et «Mél.», elle aurait pourtant dû leur paraître très suspecte, que dis-je ?, nécessairement coupable [...] S'ils comptent sur des «loclaud.» et des «oplaud.» (selon la nouvelle mode), c'est mal barré [...] Vont-ils bientôt nous proposer l'acronyme «géomanal» pour remplacer l'abréviation «géom. anal.», que nous lisons stupidement, ringards que nous sommes, «géométrie analytique» ?

### III.3 — Création néologique «indirecte»

Enfin, il semble que la connaissance de l'étymologie synchronique de deux langues permette peut-être de contribuer de façon efficace à la création terminologique dans des langues qui ne disposent pas d'organismes s'occupant de la création terminologique. Il s'agit de la même idée que celle du groupe de Prague. Mais l'application nous semble devoir être un peu différente. En effet, il ne s'agit pas de proposer pour ainsi dire à la volée une traduction éphémère mais compréhensible. Il s'agit de proposer des termes destinés à durer, de façon que la langue considérée reste ou devienne une langue dans laquelle on peut exprimer tout ce qui concerne le monde moderne, technique en particulier, mais aussi économique et culturel (au moins pour les aspects universels).

Un scénario possible est le suivant. On part d'une langue lexicalement proche dans laquelle ces termes existent et respectent le système de la langue (ce ne sont pas des emprunts ressentis comme tels). On applique des règles de transduction entre les deux langues, fondées sur les correspondances entre les structures internes des mots et les morphèmes mis en jeu. On obtient un ou plusieurs candidats en langue cible. On cherche alors à les évaluer, en déterminant surtout s'il y a des collisions ou des incompatibilités avec des mots déjà attestés.

Il y a de nombreuses paires intéressantes, comme le malais-indonésien en source, et le tetun, le cham ou le malgache en cible, ou encore le français en source et le corse, l'occitan ou le provençal en cible. Prenons quelques exemples avec l'anglais en source et le français en cible.

Si «brauser», «braser» et «brouser» étaient proposés à partir de «to browse», il faudrait à coup sûr éliminer le second. S'il ne restait aucun bon candidat, on pourrait alors tenter d'en produire de nouveaux, en utilisant les règles d'analogie interne de la langue cible (existence de phonesthèmes ou paires phonético-sémantiques parallèles comme lavage <—> lavement, qui ont des prononciations et des sens voisins), ou en passant à des familles sémantiquement voisines. La première idée pourrait par exemple conduire à «bauser» (par inversion de type «formage»/«fromage»), puis peut-être, encore mieux, à «bauser» ou «borauser», euphoniques et isolés, qu'on pourrait préférer ou non à «butiner», déjà chargé d'une nuance frivole.

Pour illustrer la seconde idée, imaginons encore que, à partir de «dirting», hypothétique mais possible, on produise dans un premier temps «salissure», «salition», «salissage», «salissement», et «saliment», par analogie avec «vomissure», «finition», «blanchissage», «blanchissement» et «blanchiment». Il y a une collision («salissure»), et deux voisins très proches («solution», «aliment»), ce qui laisse «salissage» et «salissement». Si aucun des deux n'est accepté, par exemple pour des raisons d'euphonie, on pourrait passer à une autre famille lexicale reliée, celle de «crasse», et obtenir «crassement» (collision avec l'adverbe), et enfin «crassage», peut-être le meilleur de tous, auquel l'intuition immédiate n'aurait sans doute pas manqué.

Nous croyions ces idées assez spéculatives. Mais, peu avant la rédaction de la version finale de cette communication, nous avons cru comprendre d'une brève conversation avec le Pr. Yahya Hlal (EMI, Rabat) que son équipe travaille sur la création de terminologie moderne en arabe à partir de méthodes analogues, au moins en ce qui concerne la génération de propositions à partir d'éléments de la langue elle-même (ici, les racines et les schèmes dérivationnels). C'est donc un encouragement à approfondir et implémenter cette approche.

## CONCLUSION

Il nous semble donc que, si nous entendons par «mémoire des mots» l'étymologie synchronique, c'est-à-dire la connaissance fine des règles de construction des mots attestés et des mécanismes linguistiques de la néologie, tant directe qu'indirecte, nous disposons d'un concept qui sous-tend des activités très modernes : stockage, reconnaissance et création dynamique de termes dans diverses applications de TALN, et techniques automatisables de proposition de termes nouveaux pour des langues «cibles» à partir de ceux déjà créés dans des langues «sources».

## RÉFÉRENCES

- BOITET, C. et N. NÉDOBEJKINE (1981) : «Recent Developments in Russian-French Machine Translation at Grenoble», *Linguistics, Linguistics*, 19, pp. 199-271.
- CHANDIOUX, J. (1988) : «10 ans de METEO (MD)», *Traduction Assistée par Ordinateur. Actes du séminaire international sur la TAO et dossiers complémentaires*, A. Abbou (dir.), Paris, Observatoire Francophone des Industries de la Langue (OFIL), mars 1988, pp. 169-173.
- GRUAZ, C. (1987) : *Le mot français, cet inconnu*, Rouen, Presses Universitaires de Rouen, 250 p.
- GRUAZ, C. (1990) : *Du signe au sens*, Rouen, Presses Universitaires de Rouen, 250 p.
- GRUAZ, C. (1995) : «Les variations sémantique et formelle du mot français et d'autres langues romanes : les familles synchroniques de mots», *Proc. XXI<sup>e</sup> Congrès International de Linguistique et Philologie Romanes*, Palerme.
- GUILBAUD, J.-P. et C. BOITET (1997) : «Comment rendre une morphologie robuste du français encore plus robuste en traitant finement les mots inconnus avec les données disponibles», *Proc. TALN-97*, Grenoble, 12-13 juin 1997, 11 p. (à paraître).
- MEDDEB-HAMROUNI, B. (1996) : *Méthodes et algorithmes de représentation et de compression de dictionnaires de formes*, Nouvelle thèse, Université Joseph Fourier, Grenoble.
- POGNAN, P. (1996) : «Approches grammaticale et textuelle pour l'élaboration de systèmes d'analyse automatique et d'indexation terminologique», *Proc. Séminaire Realiter*, Nice, 1-2 juillet 1996, Université de Nice-Sophia Antipolis, pp. 93-109.

## **Tabous et euphémismes lexicaux en Pular du Fouta Jalon (Guinée)**

**Alpha Mamadou DIALLO**

*Université de Conakry, Conakry, Guinée*

En français, on continue à dire : «le soleil se lève, le soleil se couche» alors que depuis plus de quatre siècles Copernic puis Galilée ont démontré que cela se passe autrement. Toutes les langues du monde véhiculent ainsi, plus ou moins explicitement, des croyances venues du fond des âges et qui passent de génération en génération depuis les temps les plus reculés.

Et s'il en est ainsi pour les langues à longue tradition d'écriture et d'études scientifiques, qu'en sera-t-il des langues essentiellement orales et non encore tout à fait ouvertes aux réalités, à la logique et à la rationalité du monde moderne ? Et on peut affirmer que plus les locuteurs d'une langue sont démunis face aux aléas de la nature, plus des croyances primitives de ce genre persistent et se perpétuent dans la vie quotidienne reflétée par le langage.

En pular du Fouta Jalon, de telles croyances anciennes existent et nous voulons évoquer ici celles qui s'attachent à certaines unités lexicales frappées d'interdit total ou partiel. Nous nous proposons également de montrer comment la langue donne les moyens de contourner l'interdit par l'utilisation des euphémismes qui remplacent ces tabous lorsque la situation exige l'évocation de leurs signifiés ou de leurs référents dans la communication courante.

Mais avant d'entrer dans le vif du sujet, nous aimerions planter le décor en disant quelques mots sur le Fouta Jalon, ce qui permettra certainement de mieux saisir l'origine du phénomène abordé dans cette communication.

## 1. — PRÉSENTATION SOMMAIRE DU CADRE SOCIO-HISTORIQUE

Le Fouta Jalon est l'une des quatre régions naturelles de la Guinée caractérisée géographiquement par son relief montagneux, son climat relativement doux et le fait que la plupart des grands cours d'eau d'Afrique Occidentale y prennent leurs sources. C'est une région située dans la zone tropicale sèche où alternent deux saisons de durée à peu près égale. C'est ainsi qu'à la végétation luxuriante de la saison des pluies succède une saison sèche au cours de laquelle la plupart des cours d'eau finissent par tarir.

L'élevage extensif y est pratiqué par des moyens rudimentaires en s'adaptant aux changements climatiques du milieu par le phénomène de la transhumance qui consiste, pour les éleveurs, à descendre les troupeaux dans les bas-fonds pendant la saison sèche à la recherche de l'herbe verte et des points d'eau, et à remonter vers les hauteurs pendant la saison des pluies.

Du point de vue historique, plusieurs vagues migratoires semblent avoir déferlé sur la région. Avant l'arrivée des Peuls éleveurs, les Landouma, les tанда, les Baga et les Dialonkés auraient déjà occupé le terrain, successivement ou au même moment.

Quant aux Peuls, ils se seraient établis sur les lieux par vagues successives à partir du XVI<sup>e</sup> siècle en cohabitant avec les Jalonkés. Au début du XVIII<sup>e</sup> siècle, ils fondèrent sur le massif un empire théocratique plus ou moins centralisé reposant sur la pratique de l'islam.

C'est ainsi qu'il se constitua une aristocratie peule appelée parfois «aristocratie du livre et de la lance» à cause du fait qu'elle n'avait pour principale occupation que l'instruction (doublée de la pratique religieuse) et la guerre sainte qui permettait non seulement de répandre la religion, mais aussi d'acquérir des esclaves. Et la société du Fouta Jalon de cette époque se caractérisait ainsi par une hiérarchisation bien nette.

Au sommet de la hiérarchie trônaient les familles et les lignages des neuf marabouts initiateurs de la guerre sainte (*jihad*). A Timbo, la capitale, siégeait l'Almami, commandeur des croyants, chef spirituel de tout l'empire théocratique, mais dont le pouvoir n'était effectif que pour la province du même nom. A la tête de chacune des huit autres provinces, le chef était choisi dans la même famille initiatrice de la guerre sainte. Toujours au sommet de la hiérarchie, s'ajoutent à ceux-

là les grands notables et marabouts influents membres des différents conseils des anciens qui assistaient les almamis et les différents chefs de provinces.

Immédiatement après cette couche d'aristocrates, venaient les «hommes libres», Peuls éleveurs et/ou cultivateurs appartenant aux premières vagues d'immigrants animistes récemment convertis et n'ayant pas participé directement à la jihad. Ils pouvaient se livrer à toutes les activités économiques, acquérir des esclaves, s'instruire ou s'enrichir et, par ce biais, accéder à de hautes fonctions dans la hiérarchie sociale.

Après les hommes libres venaient les «hommes de caste», familles d'artisans exerçant les différents métiers nécessaires au bon fonctionnement de la société : forgerons, bijoutiers, potiers, artisans du bois, cordonniers, griots. Les tisserands et teinturiers n'étaient pas toujours castés. Les hommes de caste étaient à la fois méprisés (à cause de la place qu'ils occupaient dans la hiérarchie sociale), et redoutés (certainement à cause de leur savoir-faire souvent doublé d'un certain pouvoir occulte). Une stricte endogamie accompagnée d'une certaine mystification caractérisait ces castes.

Enfin au bas de l'échelle sociale se trouvaient les esclaves classés en deux catégories : les esclaves de maison (souvent descendants des premiers occupants réduits en esclavage pour refus d'embrasser l'islam) qui étaient mieux traités que les esclaves des champs et des pâturages qui, eux, avaient été soit capturés à la guerre, soit tout simplement achetés.

L'avènement de la théocratie s'est accompagné d'un triomphe de la langue et de la culture des Peuls dans un environnement sociolinguistique où dominaient les langues du groupe mandé, principalement le jalonké (très proche de la langue soso) et le maninka. La toponymie du Fouta Jalon reflète très largement cette situation d'antériorité des langues mandé. Le pular, langue à classes du groupe ouest-atlantique a, de son côté, subi beaucoup d'influences au contact de ces mêmes langues mandé.

Le décor est donc clair : nous avons à faire à une société fortement islamisée, très hiérarchisée, ayant comme activité économique principale l'élevage avec des moyens rudimentaires, et dont la langue entre en contact avec un groupe linguistique radicalement différent (le groupe mandé).

Il est vrai que ce monde est définitivement révolu, qu'après la colonisation et l'indépendance, une autre langue (le français) est entré

en contact avec le pular et beaucoup d'autres transformations socio-économiques sont intervenues, mais :

— certaines traces de croyances anciennes persistent toujours puisque les langues ne changent pas du jour au lendemain;

— tout les facteurs de ce décor sont importants pour la compréhension du phénomène des tabous et euphémismes lexicaux que nous avons relevés en pular du Fouta Jalon.

## 2. — TABOUS ET EUPHÉMISMES PULAR

Les mots tabous pular au sens large sont des mots que les locuteurs s'abstiennent de prononcer dans certaines circonstances de leur vie :

— par peur, quand il s'agit d'un mot désignant une réalité dangereuse pour l'homme, son entourage ou ses biens (faquves et autres animaux dangereux pour l'élevage, l'agriculture ou la vie de l'homme; maladies, etc.);

— par pudeur, quand il s'agit d'un mot en rapport avec certains besoins naturels de l'homme (uriner, évacuer les excréments, etc.);

— par respect, quand il s'agit de mots réservés à certaines catégories sociales, à certains biens socialement valorisés ou à certaines pratiques religieuses.

Ces deux derniers cas sont surtout intéressants par les euphémismes qu'ils produisent, car au lieu de mots tabous au sens strict, on pourrait plutôt parler de mots indécents ou «déplacés».

### 2.1. — Les tabous inspirés par la peur et les euphémismes correspondants

Par rapport à l'élevage, les mots qui désignent des fauves sont tabous. C'est le cas de l'hyène «fowr» (1), de la panthère «buutoori», du lion «ngayuuri», tous les trois très redoutables pour les troupeaux.

---

1) Pour les sons spécifiques au pular (par rapport au français), nous avons utilisé les signes suivants :

— Pour les glottalisées : β(bilabiale), d(dentale) et y(palatale)

— Pour les affriquées, c(sourde) et j(sonore)

— Pour les nasales, ñ(palatale) et μ(prépalatale)

Ces mots ne doivent jamais être prononcés la nuit (quel que soit le lieu) ou en dehors du village (quel que soit le moment), car prononcer leur nom, équivaudrait à les appeler, à les attirer vers le troupeau.

En restant dans le même domaine, le lait fait l'objet d'un vocabulaire particulier. Il est interdit d'utiliser les mots ordinaires «seedā» (un peu de...), «lannugol» (finir), «siugol» (suinter) ou «maafugol» (utiliser le lait comme sauce). Violer cet interdit en utilisant des mots ordinaires, peut entraîner la disparition d'abord des vaches laitières puis du troupeau tout entier.

D'autres noms d'animaux sont également frappés d'interdit pendant la nuit, mais pour d'autres raisons. C'est le cas de «saḡalde» (porc-épic), «nīwa» (éléphant) et «girayngi» (sanglier), certainement en raison de leur action nuisible sur les cultures. Pour ce dernier animal, la raison est également religieuse, puisqu'il est sacrilège de prononcer son vrai nom, surtout pendant certains mois lunaires de l'année (le ramadan et les deux mois qui le précèdent). Le nom du serpent «mboddi» s'inscrit dans la même veine, de même que les fourmis maḡnans «cula» pendant la saison des pluies.

Le phénomène des noms tabous existe également dans le domaine des maladies. C'est le cas de : «ndiyan» (littéralement *eau* et signifiant *lèpre*), «hendu jinna» (littéralement *vent du diable* et signifiant *maladie infectueuse*), «keku», «baadi» et «sondu» (littéralement *tortue, singe et oiseau*, dans les trois cas, *maladies de nourissons*).

Pour tous ces mots tabous inspirés par la peur, nous allons donner, dans le tableau suivant, les euphémismes qui leur correspondent, suivis de leur signification en français :

**Tableau 1**

tabous (pular)	euphémisme (pular)	français
fowru	bonooru	hyène
buutoori	ladoori	panthère
ngayuuri	piyoori	lion
seedā	barke	un peu
lannugol	jinnugol	finir

tabous (pular)	euphémisme (pular)	français
siugol	jiidagol	suinter
maafugol	wasagol	utiliser le lait comme sauce
saɲalde	yaɲaare	porc-épic
girayngi	kullii jemma/mbabba ladde	sanglier
ñ iiwa	mawba	éléphant
mboddi	βoggol leydi	serpent
cula	raddo	fourmis magnans
ɲdiyan	damajan	lèpre
hendu jinna	wifuru	maladie infectueuse
keku	ko yaadata e suudu mun kon	maladies de nourisson
baadi	ko diwata ka hoore le dde kon	maladie de nourisson
sondu	ko wiirata kon	maladie de nourisson

## 2.2. — Les mots indécents ou déplacés et les euphémismes correspondants

Par pudeur, les locuteurs du pular, évitent certains relatifs aux fonctions biologiques naturelles, notamment l'évacuation de déchets (excréments, urines, etc.). Ces mots sont considérés (dans la plupart des langues d'ailleurs) comme impudiques et indécents et on les remplace généralement par d'autres tournures moins choquantes (euphémismes). C'est le cas des mots «bu'ugol» ou «beebugol» ou «beppugol» (déféquer), «beppe» (excréments), soofugol» (uriner), «coofe» (urine) et «fuytugol» (péter). Pour les bébés ces mots sont tabous parce qu'ils peuvent leur porter malheur.

Les euphémismes correspondant à ces mots sont très nombreux et varient souvent selon qu'on a à faire à un adulte ou à un bébé :

**Tableau 2**

mot indécent (pular)	euphémisme adulte (pular)	euphémisme bébé (pular)	Français
bu'ugol beebugol beppugol	- yaagol βaawa hoggo - selugol - yaagol hurgo - yaagol kabine	- raahagol - resagol	déféquer
beppe	- soβe - ko hawkoytee ka βaawa hoggo kon - seel	- raaha - resa	excréments
soofugol	- yaagol βaawo suudu - bawlugol - takkitagol - yaagol hurgo - firtugol duhol - ÿettugol salligi - hibbugol salligi	- bunnugol - lewnugol	uriner
coofe	- bawle - ko hibboytee ka βaawo suudu kon	bawle	urine
fuytugol	accutugol hendu	accutugol hendu	péter

Enfin par respect pour certaines catégories sociales, comme les vieux (parents ou non), les chefs, les marabouts, les notables, certains mots sont, non pas tabous, mais déplacés et même injurieux. C'est le cas pour certaines parties du corps, les habits ou certains verbes relatifs à des activités quotidiennes (aller, venir, manger, dormir, etc.).

Avant la suppression de la chefferie traditionnelle, enfreindre à cette règle du langage pouvait se traduire par le paiement d'une forte amende. Et c'est peut-être pour apprendre cette règle à tous qu'il existe un conte où le lion, roi des animaux, menace de dévorer le singe s'il ne répond pas correctement aux questions concernant les noms réservés aux différentes parties de son corps de roi.

**Tableau 3**

mot «déplacés» (pular)	euphémisme (pular)	français
hoore	sala	tête
gite	kolongaaje	yeux
noppi	gelordi	oreilles
hunduko	karaho	bouche
juude	sookeeje	bras
rendu	teege	ventre
pooβe	toolorde	fesses, anus
koyde	tundaaje	pieds
kufune	salaalaa	bonnet, coiffure
dolokke	boloojan	boubou, chemise
tuuba	nomoran	pantalon
yahugol, arugol	seenagol	aller, venir
artugol	sentagol	revenir
haalugol, wowlugol	maakugol	parler, dire
yiugol	tinugol	voir
joodagol	toolagol	s'asseoir
immagol	βaμagol	se lever
ñ aamugol	nafagol	manger
daanagol	huylagol	dormir

Toujours dans le cadre des euphémismes inspirés par le respect, il faut signaler que :

— pour le mot *dormir*, le bébé bénéficie du même respect que la catégorie sociale indiquée ci-dessus. En parlant du bébé on utilisera donc «huylagol» au lieu de «daanagol»;

— certains aliments socialement valorisés parce que liés à la pratique de l'islam, partagent avec le lait les mêmes euphémismes

«barke» (un peu de...) et «jinnugol» (finir). C'est le cas du miel (pour lequel il faut ajouter «jiidagol» = suinter) et du pain blanc toujours présents dans les cérémonies religieuses;

— pour toutes les activités religieuses, on évite les mots ordinaires «fuddagol» (commencer) et «lannugol» ou «gaynugol» (finir). On utilise généralement les euphémismes «bismagol» quand il s'agit de commencer à égrener le chapelet, «uuzagol» pour la lecture du Coran, «darnugol», pour la prière et on utilise «timmingol» pour exprimer la fin de toutes ces activités.

### 3. — ANALYSE DES TABOUS ET EUPHÉMISMES

On peut facilement élucider le phénomène des mots tabous en s'inspirant de la théorie du signe de Saussure et de ses continuateurs. Pour la conception saussurienne, le signe comporte deux faces : le signifiant et le signifié, et la relation qui existe entre les deux est arbitraire, immotivée et conventionnelle. De même que le caractère arbitraire du signe est limité pour ce qui concerne les onomatopées, l'aspect immotivé est, partiellement il est vrai, remis en cause par l'existence des mots tabous.

En effet, pour ce genre de mots, la motivation est si forte que non seulement les deux faces du signe semblent inséparables, mais aussi elles finissent par se confondre avec le référent. Pour les réalités qui engendrent la peur par exemple, ce n'est plus tout à fait le référent qui fait peur, puisque par l'intermédiaire de l'euphémisme on arrive à l'évoquer quand le besoin s'en fait sentir. Ce qui fait peur, c'est essentiellement le signifiant qui a fini par se confondre au référent.

C'est pour cette raison qu'on cherche constamment des signifiants de remplacement pour tenter d'opérer une certaine démotivation, et ce sont ces nouveaux signifiants qui constituent les euphémismes. Au bout d'un certain temps l'euphémisme perd son caractère immotivé et la nécessité de trouver un autre signifiant s'impose. Il s'établit ainsi que course constante entre motivation et démotivation qui se traduit par un renouvellement lent mais constant du vocabulaire. Les exemples de mots tabous donnés ci-dessus sont peut-être d'anciens euphémismes qui ont dû suivre le processus ainsi décrit.

Il suffit, pour s'en convaincre, d'examiner les tabous concernant les maladies : «ndiyan» (eau), «hendu» (vent), «keku»

(tortue), «baadi» (singes), «sondu» (oiseau) qui sont vraisemblablement d'anciens euphémismes devenus tabous à leur tour.

Pour créer de nouveaux signifiants (euphémismes), la langue fait souvent appel à ses ressources internes. Mais elle utilise également des mots étrangers qui ont l'avantage de la neutralité et de la démotivation. Les langues dans lesquelles le pular puise ces mots étrangers semblent être essentiellement l'arabe (langue religieuse), les langues mandé et le français (à cause du contact entre ces langues et le pular).

### 3.1. — Euphémismes internes (pular)

La plupart des euphémismes sont ainsi d'origine pular. C'est le cas pour presque tous ceux qui sont inspirés par la peur ou par la pudeur :

Certains sont des dérivés :

— «bonooru» (hyène) vient de la racine «bon» qui, avec le suffixe verbal de l'actif «-ugol» donne «bonugol» qui signifie : être mauvais, être mauvais, être méchant. Donc «bonooru» («-ru» = morphème de classe permettant de ranger le mot dans une des classes du pular) = l'animal mauvais;

— «ladoori» (panthère) provient du même modèle : «lad-» + «ugol» donne «ladugol» = ramper. Donc «ladoori» («-ri» = morphème de classe) = l'animal qui rampe (avant de bondir sur sa proie);

— «piyoori» (lion) = «piy-» + «ugol» = «piyugol» (battre), donc «piyoori» («-ri» = morphème de classe) = animal qui bat (ou abat) sa proie;

— «yuµaare» (porc-épic) = «yuµ-» + suffixe verbal de la voix moyenne «agol» = «yuµagol» (se harnacher), donc «yuµaare» («-re» = morphème de classe) = l'animal harnaché (de ses piquants);

— «wifuru» (la ventilation) = «wif-» + suffixe verbal «-ugol» donne «wifugol» (ventiler), donc «wifuru» («-ru» = morphème de classe) = la ventilation.

D'autres sont des composés :

— «kullii jemaa» (sanglier) = «kullii» (gros animal, monstre) + «jemaa» (nuit) = monstre nocturne;

— «mbabba ladde» (sanglier) = «mbabba» (âne) + «ladde» (brousse) = âne sauvage;

— «βoggol leydi» (serpent) = βoggol leydi» (terre) = la corde de la tdrre;

— «yaagol βaawo hoggo» (déféquer) = «yaagol» (aller) + «βaawa» (derrière) + «hoggo» (clôture) = aller hors de la clôture (hors du village);

— «yaagol hurgo» (uriner, déféquer) = «yaagol» (aller) + «hurgo» (enclos, latrine, w.c.) = aller dans l'enclos;

— «firtugol duhol» = «firtugol» (défaire un nœud) + «duhol» (ceinture) = défaire le nœud de sa ceinture;

— «accutugol hendu» (péter) = «accutugol» (laisser échapper, libérer) + «hendu» (vent) = libérer du vent.

D'autres sont de courts énoncés :

— «ko yaadata e suudu mun kon» = ce qui se déplace avec sa case (la tortue);

— «ko diwata ka hoore le dde kon» = ce qui saute dans les arbres (le singe);

— «ko wiirata kon» = ce qui vole (dans les airs) (l'oiseau);

— «ko hawkoytee ka βaawo hoggo kon» = ce qu'on jette hors du village (les excréments);

— «ko hibooytee ka βaawo suudu kon» = ce qu'on verse derrière la case (l'urine).

Enfin signalons des cas divers comme ceux de :

— «mawba» (éléphant) qui semblent être une contraction de «maw(nu)mba» provenant de la racine «mawn-» (grandir) et «mba», substitut de l'une des classes nominales du pular;

— «raddo» (fourmis magnans) semble provenir du radical «radd-» (du verbe «raddingol» = chasser) par un procédé économique de nominalisation par adjonction d'un «o» au radical et qui donne souvent des génériques;

— «Damajan» (lèpre) provenant de la combinaison de «Dama» = l'autre (en parlant de la classe des liquides et de «jan» = morphème de classe «-an» et consonne de soutien «-j-»;

— «jiidagol» (comme euphémisme = suinter) mot du vocabulaire courant, signifie *avoir des enfants* avec une connotation méliorative.

Ce qu'il faut retenir de tous ces cas, c'est que tous les euphémismes nominaux provenant des ressources internes de la langue sont de la même classe nominale que le tabou correspondant. Cela permet ainsi, sans ambiguïté, d'évoquer la réalité dangereuse en utilisant un substitut pronominal du nom tabou ou euphémique.

C'est ainsi qu'on pourra parler de l'hyène en utilisant simplement «ndu» (elle) puisque «fowru» et «bonooru» appartiennent tous les deux à cette même classe «ndu». Il en va de même pour «buutoori», et «ladoori» (panthère) de la classe «ndi», «girayngii» et «kullii jemma» (sanglier) de la classe «ngi», etc. Même quand le nom tabous est à la forme générique, l'euphémisme apparaît aussi à la forme générique, l'euphémisme apparaît aussi à la forme générique. C'est le cas par exemple, de «cula» et «raddo» (fourmis magnans) qui sont tous les deux à la forme générique. Dans ce corpus nous ne voyons qu'une exception à cette règle : «mboddi» de la classe «ndi» a pour euphémisme «βoggol leydi» de la classe «ngol».

### 3.2. — Euphémismes externes (emprunts)

Le recours aux mots étrangers pour contourner les mots tabous ou indécents est un phénomène apparemment répandu. En français par exemple l'utilisation des mots *w.c.*, *vécés*, *water*, entrerait dans ce cadre d'après Pierre Guiraud (1971).

Parmi les euphémismes recueillis ci-dessus, nous reconnaissons des mots arabes :

— «barke» (*un peu de...* en parlant du lait, du miel et du pain blanc) emprunt qui signifie également *bénédictio* dans la langue courante;

— «bawle» (urine) et «bawlugol» (uriner);

— «bismagol» (commencer) parce que pour le musulman, toute activité doit commencer par la prononciation de la formule «bismillaahi» (au nom de dieu);

— «uuzagol» (commencer la lecture du Coran) parce que toute lecture du Coran doit commencer par la formule «a uuzubillaahi minas'saytaanirrajiimi»;

— «timmingol» (finir) provient du mot arabe «tammat» (finir, fin).

Par ailleurs la plupart des euphémismes en rapport avec la vie sexuelle sont d'origine arabe. Le sexe (masculin et féminin) se dit «awra», la *fornication* se dit «jeeno» ou «zina» ou «jina» qui donne «jeenugol» (forniquer).

Certains autres mots de notre corpus proviennent de langues mandé :

— «salligi» (dans «hibuugol» ou «yettugol» «salligi») semble venir du maninka «sali ji» (littéralement *eau de prière*) désignant l'eau qu'on utilise pour faire les ablutions;

— «bolloojan» (boubou, chemise) provenant également du maninka signifierait dans cette langue *grande manche*.

Et, de façon générale la majorité des euphémismes inspirés par le respect proviendraient des langues mandé pour des raisons bien simples : la plupart des esclaves étaient de langues mandé et les mots qu'ils utilisaient à l'égard de leurs maîtres pour marquer leur déférence se seraient répandus et auraient été adoptés par tous les locuteurs de la langue pular du Fouta Jalon.

Enfin la langue français a fourni également des mots pour contourner la forte motivation qui caractérise les mots du pular pour les locuteurs de cette dernière langue. Dans notre corpus ci-dessus nous pouvons citer :

— «kabine» (emprunt au français) dans «yaagol kabine» qui signifie *aller au cabinet* (aller au w.c.);

— «seel» emprunt du mot français *selles* très fréquent dans les hôpitaux avec la nécessité pour les malades de procéder aux examens de selles.

Par ailleurs beaucoup de mots français sont occasionnellement empruntés dans le même but de neutraliser les connotations fâcheuses ou indécentes.

Enfin il faut dire que tout le phénomène décrit ici a de plus en plus tendance à appartenir au passé. Ces «finesses» du langage ne sont plus tout à fait connues que par les vieilles générations. Avec le recul des activités traditionnelles (comme l'élevage et l'agriculture) ainsi que les progrès de l'urbanisation, les jeunes générations ne savent presque plus qu'un vocabulaire particulier était réservé à certains domaines de la vie. Qui utilise encore dans les centres urbains le vocabulaire spécial réservé au lait, au miel et au pain blanc ?

C'est donc une situation en plein changement que nous avons tenté de décrire. Et nous espérons avoir réussi à montrer comment des mots arrivent parfois à se confondre avec leurs référents, et comment le phénomène des tabous et euphémismes lexicaux contribuent à provoquer un renouvellement constant du vocabulaire ne serait-ce que dans un domaine restreint.

## REFERENCES

- BAH, N.Y. (1974). — Etude sémantique, phonétique et intégration morphologique des emprunts du pular aux langues étrangères, Conakry, IPGAN, mémoire de D. S.
- DIALLO, A. (1989). — Etude de l'intégration morphologique des emprunts linguistiques arabes en pular de Guinée, Rabat, Université Mohamed V, thèse de doctorat 3e cycle, 331 pages.
- GUIRAUD, P. (1971). — *La sémantique*, coll. «Que sais-je ?», Paris, PUF.
- REY, A. (1976). — *Théories du signe et du sens*, I, II, Paris, Klincksieck
- TAMBA-MECZ, I. (1994). — *La sémantique*, coll. «Que sais-je ?» Paris, PUF.

**Signifiante et fonction culturelle  
des emprunts et des détournements  
sémantiques dans la poésie épique sénégalaise,  
d'expression wolof et d'alphabet arabe**

L'exemple du *diaza'ou chakor* de Moussa Kâ

**Ibra DIENE**

*Université Cheikh Anta Diop, Dakar, Sénégal.*

L'œuvre littéraire de Moussa Kâ est un immense réservoir de conflits de langues mais aussi une forme d'exploitation judicieuse du plurilinguisme. Et dans les textes comme le *Diaza'ou Chakor* (*La Récompense du mérite*) qui relate le voyage d'exil, par la mer, de Serigne Touba, marabout sénégalais et résistant pacifique à la pénétration coloniale, le plurilinguisme ne se traduit pas par des performances de langue autonomes les unes, les autres mais par une interdépendance qui laisse observer, au plan du lexique, la coprésence et le mélange de plusieurs langues. Celles-ci sont de nature et, dans le texte, de fonctions différentes.

Ainsi, le premier problème posé est celui de la transcription d'un texte oral, parce qu'en wolof, par un alphabet étranger et celle de certains mots français empruntés, souvent déformés par la pratique wolof, et de signifiant, eux-aussi en lettres arabes. D'autre part, quand l'emprunt est contourné par la traduction, l'auteur use de métaphores et de périphrases, de détournements sémantiques qui accueillent des réalités françaises dans une langue wolof également transcrite en lettres arabes.

La signifiante de cette alternance de langues et de textes ou des détournements sémantiques n'est pas une simple insuffisance du wolof pour la traduction de réalités et de contextes complexes ou étrangers. Elle touche des réalités partagées et spécifiques permettant aux mots de véhiculer la mémoire de l'époque, des lieux et des faits.

Ces mots portent aussi la poéticité et l'écriture devient, avant tout, un travail lexicologique, l'emprunt et le détournement facilitant la poétisation mais enrichissant aussi la langue d'accueil.

### **LES FORMES DE L'ALTERNANCE DES LANGUES. TRANSCRIPTION ET MÉMOIRE DES MOTS**

Le *Diaza'ou Chakor* est un texte de sources diverses. Tout d'abord il s'inspire du *Diaza'ou Chakor Diawabou Abdou Latif*, journal de voyage où Serigne Touba a retracé, à l'intention de Abdou Latif qui le lui avait demandé, son itinéraire d'exil lors de sa déportation par les autorités coloniales. Ce document en arabe est donc une source écrite tout autant que les morceaux du Coran et des Qacida de Serigne Touba cités ou collés dans le livre de Moussa Kâ. Les mots français sont eux aussi d'une langue écrite mais le contact qu'en a Moussa Kâ est oral, d'après une prononciation souvent déformée par les habitudes phonologiques des Wolofs. D'autres renseignements contenus dans cette épopée sont de source orale, des conversations tenues par Serigne Touba, reprises directement par Moussa Kâ ou à partir de fidèles proches du marabout.

La trace des sources écrites est plus vive parce qu'au lieu d'emprunts de mots, le poète fait appel à de véritables textes. Cette forme d'intertextualité est en même temps une alternance de langues, la plupart des morceaux empruntés étant en arabe dans le texte ou une traduction de l'arabe ou vers l'arabe. De ce point de vue, pour qui parle les deux langues, l'alternance des textes arabes et wolof est presque toujours une mise en parallèle des significations. Mais les redondances ne sont perceptibles que chez le lecteur plurilingue, l'auditeur exclusivement wolophone ou le lecteur uniquement arabophone ne pouvant les découvrir. C'est pourquoi elles (les redondances de significations) ne sont pas des formes d'insistance mais des facteurs de lisibilité quand bien même la compréhension d'une seule langue ne donne accès, surtout pour l'arabe, qu'à l'esprit du texte et non aux détails. D'ailleurs, du wolof à l'arabe, l'alternance n'est souvent que le résultat d'une traduction au plan de la signification. Quand on regarde la signifiante, les choses deviennent plus complexes et nous y reviendrons ultérieurement.

En dehors des textes arabes provenant d'une traduction du texte wolof, il y a des passages en arabe qui montrent beaucoup plus l'intertextualité puisqu'ils sont des citations ou des collages. Là aussi, il y a des significances complexes et des significations rendant compte des contextes, de la pensée des protagonistes ou des analogies avec d'autres situations; analogies faites par le poète souvent dans le sens d'un rappel de l'histoire de l'islam et dans la perspective de rendre le discours épique ou de manifester la spiritualité du sujet. Ainsi dans les réponses de Serigne Touba aux discours profanes, le Coran est souvent cité pour montrer la conformité de la pensée du marabout avec le livre saint. À défaut de cela, Moussa Kâ fait dire à son héros une déclaration en arabe et respectant dans son contenu les préceptes de l'islam. À Madiakhaté Kala, qui lui propose d'assouplir sa conduite spirituelle et de collaborer avec les rois païens, Moussa Kâ fait dire à Serigne Touba en incise dans son texte :

[\*sɛ : x \*bamba na : kɔ asbina : lila : ho / dahran wala : ila : ha? ila : la : ho ] (\*)

[man jal : a doj na ma ta dojlo : na kɔ / ma hindakum jan fadu faŋg na : kɔ] (1)

La citation ou le collage de textes oraux, presque toujours en wolof sauf certaines réponses du marabout, sont moins nets parce qu'ils ne contrastent pas avec le noyau du texte et s'y intègrent parfaitement. Même l'énonciation, la source d'énonciation, ne changent pas ici, le poète se contentant d'attribuer, par incise, les propos repris. Dans ce cas aussi, il y a un problème de traduction et de détournement de sens révélant une grande maîtrise de la rhétorique quand le poète rend compte de paroles dites par des français. Les services des interprètes sont certes, parfois, signalés mais ici la mémoire des mots est moins fidèle, sans doute à cause de la faible maîtrise du français chez le poète mais aussi et surtout à cause des visées ironiques et de lisibilité puisque Moussa Kâ choisit fréquemment le parler de son temps et non du temps ou du contexte de l'aventure. Ainsi des expressions comme : [ka : bal], [nda : na : n], [pa : ŋk], [wa : nɛ], en vogue à l'époque, sont mis dans la bouche du colon ou dits pour le traduire et cherchent d'abord la lisibilité plutôt qu'à reproduire la signification d'un message ou la mémoire des mots.

---

\*) Tous les o qui se présentent sont ouverts

1) Kâ, M.S. (s.d., vers 46-47) : *Djaz'ou Chakor*, Taïf, éd. Serigne Mbaye Diop, p. 4

Les emprunts de mots français et arabes sont eux plus marqués au regard de la signifiante mais n'en sont pas moins des indices d'enrichissement linguistique. Là, il est possible d'observer que, pour le français, il s'agit surtout de traduire des réalités absentes de la culture wolof, plutôt que d'indices d'identification du colon. D'ailleurs le refus de ces vocables débouche sur des périphrases et des métaphores wolof qui opèrent le travail de traduction par détournement sémantique. Le plus souvent, les mots français empruntés relèvent du lexique militaire, administratif ou de la navigation maritime. Les mots arabes visent, eux, beaucoup moins à suppléer un manque et sont donc plus pertinents au plan de la signifiante.

D'ailleurs là, le poète recourt le plus souvent à l'emprunt et pratique rarement la traduction, même métaphorique.

## **DU DÉTOURNEMENT À L'EMPRUNT : UNE SIGNIFIANTE HIÉRARCHISÉE**

Le détournement sémantique par trait d'esprit, par périphrase ou par allégorie est à la fois désignation de réalités empruntées et refus, au plan du signifiant linguistique, de cette réalité étrangère.

Dans les cas de l'emprunt et du détournement, il y a toujours une volonté de suppléer à un manque mais les signifiantes ne sont pas au même niveau. Dans l'emprunt, le sujet parlant se laisse assimiler par l'autre, son fournisseur, alors que, dans le détournement sémantique, il tente de l'assimiler; d'assimiler ce qu'il offre par une sorte d'intégration linguistique. Ainsi, quand Moussa Kâ, par emprunt d'un vocabulaire militaire français, accepte de s'aliéner dans la réalité et le monde occidental, il n'en refuse pas moins, prenant la parole pour Cheikh Ahmadou Bamba, cette manière de se dissoudre en l'autre et en son monde. Cela explique que, connaissant le mot *gouverneur* pour l'avoir emprunté, même formellement déformé, il le remplace par [borom ndarab ndaka : ru] ou par [kilifag ndaka : ru]. Il y a là une bipartition qui cache une réalité culturelle complexe. Parce que, pour le mouridisme, l'administration coloniale est l'ennemi de l'islam. Par conséquent, le français devient une langue négativement marquée. Voilà pourquoi le Cheikh n'emploie jamais cette langue, même pas le plus petit mot. Et quand il y a nécessité, au lieu de recourir à l'emprunt, le texte le fait procéder au détournement sémantique ou à la traduction périphrastique.

Le poète Moussa Kâ, lui, dans sa relation des faits, a souvent recours à l'emprunt de mots français. On pourrait évoquer l'extrême nécessité si les emprunts ne touchaient que le lexique de réalités inconnues. Mais la signifiante est ici plus large puisqu'elle peut être, même, de fonction référentielle, l'emploi des mots permettant de convoquer le contexte et la réalité où le Français est protagoniste ou acteur, où son monde est évoqué. Par exemple, dans le passage ci-dessous, la technicité occidentale se traduisant par la rapidité dans le voyage est évoquée avec des mots français empruntés :

[kero : g la bamba romb wa : endalu : s / masin ba rule na ba we : su mandalu] (2)

Pourtant Moussa Kâ ne recourt pas au français aussi volontiers qu'à l'arabe. Même employant des mots français, il n'en manifeste pas moins, par la déformation, sa résistance et son refus de se laisser intégrer dans la réalité linguistique française, puisque les déformations phonétiques et phonologiques opérées ne sont pas seulement l'indice de ce qu'on pourrait appeler un accent étranger. Elles procèdent parfois d'une domestication, d'une wolofisation du mot français, le poète, réaffirmant ici, par le refus de la bonne prononciation française, l'importance des traits d'oralité et du phonétisme au delà même de l'identification du signifiant des mots.

Ainsi Lieutenant devient [jɛtnaŋ], Libreville [limbarwali]. Transformé de la sorte, le phonétisme de ces mots cesse d'être étranger et l'auditeur a l'impression de rester dans sa langue. Un travail semblable est fait à propos de la plupart des mots arabes empruntés quand bien même ici les déformations sont moins importantes.

L'emprunt peut donc manifester une volonté de distinction surtout quand le poète reprend le discours de l'ennemi et sans préjudice à sa fonction d'enrichissement de la langue d'accueil. Ce réalisme est plus net avec les mots et textes arabes empruntés. Au contraire des mots français mis dans des discours profanes, les mots arabes connotent ici la spiritualité et le mysticisme. Mais ces mots arabes ont aussi une valeur dénotative qui enrichit le wolof inapte à traduire les concepts du discours religieux musulmans, surtout quand il s'agit d'évoquer les traits du mystique, la pratique religieuse, l'histoire de l'islam. La fonction référentielle est ainsi mise en exergue l'intertextualité. De ce fait, les mots et textes arabes renvoient au Coran, à la religion et partant à la figure du héros qui représente cette religion.

---

2) *Ibidem*, p. 14, v. 241.

Une étude comparative des emprunts et des détournements sémantiques traducteurs révélerait ainsi deux attitudes contradictoires vis-à-vis de l'autre, autre représenté par sa langue. Compte tenu de cela aussi, les emprunts et les métaphores traductrices n'ont pas le même statut en français et en arabe ni les mêmes sujets emprunteurs.

D'ailleurs, une même périphrase, née d'un détournement sémantique traducteur, peut être, selon les contextes textuels, de signifiante et de valeur diverses. L'expression [borom ndar] est plusieurs fois employée dans le texte. Dans un premier sens qui est une traduction «propriétaire de Ndar (Saint-Louis)» acception du mot propriétaire consacrée en wolof pour désigner les chefs coutumiers et, partant, tout chef et seigneur. Cette signification comporte un aspect connoté, la puissance attachée au seigneur et souverain. Mais d'un autre point de vue, cette appellation laisse transparaître la vanité du pouvoir temporel et, partant, le péché blasphématoire qui lui est attaché, étant donné son ambition de concurrencer Dieu, le seul seigneur admis chez les musulmans.

L'emploi de cette expression porte une autre signifiante, une signifiante procédant de l'ironie et soulignant l'inappropriation de l'expression puisque son réemploi, sa commutation, débouche sur un discours absurde : [borom ndarab ndaka : ru] (= propriétaire de ndar de Dakar).

Il y a là un intertexte évident avec les chansons mourides, populaires et résistantes de l'époque, chansons en wolof et bâties sur la vanité et l'irréalité du pouvoir colonial.

[borom ndar mo : muɔ : ndar daʃu·laj faj / ki la bind bind ndar mo mo : m ndar ba :ri saha : jɛ] (3)

Ainsi, même par un simple détournement sémantique qui apparemment ne fait pas intervenir la valeur du signifiant, la mémoire des mots est toujours convoquée, au même titre que quand il y a emprunts, dans le livre de Moussa Kâ. Cela ne se perçoit pas seulement au plan des fonctions dénotatives et connotatives du discours, au plan des fonctions référentielles ou métalinguistiques mais aussi poétiques et culturelles.

---

3) Traduit en français, ce morceau de chanson wolof a le sens suivant : «Propriétaire de Ndar ne t'appartient pas. Tu es victime de louanges trompeurs. Celui qui a décidé de ton existence et de l'existence de Ndar, c'est celui-là le propriétaire de Ndar».

## FONCTIONS CULTURELLES ET LITTÉRAIRES DE LA MÉMOIRE DES MOTS

L'emprunt de mots et le détournement de sens sont des formes d'ouverture et même de culture. Dans le poème de Moussa Kâ, cela est souvent vérifié. Si l'arabe y est le plus souvent, du moins sous sa forme de texte, l'indice d'une érudition du sujet, le français apparaît, lui, comme celui d'une ouverture, la preuve d'un contact avec des réalités, et donc avec une langue, étrangères. De ce point de vue, le classement isotopique des termes français empruntés suggère une signification politique mais surtout culturelle, la culture étant posée comme un facteur de civilisation. Ainsi, les termes quartiers, soldats, gendarmes, armée, lieutenant, capitaine et canon connotent la violence et le bellicisme mais surtout la rigueur et l'organisation du système colonial, sa hiérarchie.

Le formalisme bureaucratique et administratif est certes souligné à travers le vocabulaire administratif comme [tampɛ] (tampon), [garantɛ] (garanti), [siɲɛ] (signer) mais ces mots, étant l'expression d'un pouvoir et de l'organisation sociale, sont prêtés à l'administration de Dieu et à la hiérarchie divine.

[ba mu binde : sali bana : wasalim / ɲfɔ : l ma : ka tontul k o qalalmukalim]

[fibri : lu tɛktal k o wa d o : rijaqsim / bu : r jal : a tampɛl k o fuma : ji b opa : m] (4)

[mbɛ : d ma n o : n ja daɲu k o karantɛ / bisub kɛr o : g ru : ga la bu : r garantɛ] (5)

[bu : r jal : a daldi siɲɛ fɔx j onɛn ba / mu siɲɛ ga : ja daldi fital \*bamba] (6)

L'organisation religieuse et le rôle de l'église sont également suggérés par les termes labé [lab : ɛ] (abbé), [rom] (Rome), »fa : p] (Pape) et leur pratique sert de vitrine à la civilisation européenne et chrétienne. La guerre des langues signifie ici une guerre de religion et partant de culture. C'est à ce titre que la figure du marabout défenseur de l'islam se double de celle d'un défenseur du noir et protecteur des déportés noirs comme Samory Touré et sa famille. Mais la technicité

4) Kâ. M. S. (vers 406-407) : *Diaza'ou Chakor*, p. 22.

5) *Ibidem*, p. 11, v. 185.

6) *Ibidem*, p. 4, v. 79.

et partant le matérialisme occidentaux sont également suggérés à travers la machine administrative et les communications. Le mot [kabra] (cabler) est souvent cité dans ce sens.

[kɛr o : g la fa : p bu mag ba daldi kabla / b or om tugal na : k o s ə ri ʎ bu mag la] (7)

L'ouverture que traduit l'emploi de mots français a aussi des indices et des incidences littéraires. Ainsi les réticences devant la langue du colonisateur et même le refus de celui-ci par le Cheikh neutralisent difficilement les attraits du français chez le poète Moussa Kâ. A ce titre le sémantisme devient moins important, le versificateur s'attachant à la mémoire sonore des mots. Les termes français «merci», «boulangier» permettent ainsi de construire l'écho des rimes tout autant que d'autres mots comme [armɛ] (emprunté à l'expression militaire «présentez arme») et [k ontɛ] (pour décrire la cadence sonore des pas).

[ gi : r b o : ba wu : r jaŋg k o j xec o : k s ri si / kaŋga : m ji ak tuba : b ji na : k o mersi] (8)

[ u f ol : i kapte : n ba jabal k o dawmɛ / karɛ ga a : ndak m o : m ta d o : tu armɛ] (9)

[ca takusa : n la ga : l ga daldi k ostɛ / dun bu u na : \*tab o : k mara jaj k ontɛ] (10)

Comme dans l'exemple précédant ceux-ci, l'emprunt joue ici un rôle versificateur, c'est-à-dire rimique et partant rythmique comme d'autres mots français ou arabes dont la fonction peut être en plus prosodique.

Quand, en plus de ce rôle rythmique et sonore, le sémantisme du mot emprunté est convoqué, le poète cède le plus souvent à une ambition d'harmonisme, à une imitation recréant les discours des protagonistes ou leur univers culturel de référence. D'ailleurs l'emprunt de texte en arabe qui veut révéler l'érudition du poète ou du sujet joue également la même fonction. C'est par la citation et le collage que Moussa Kâ révèle la culture de Serigne Touba et la sienne

---

7) *Ibidem*, p. 19, v. 350.

8) *Ibidem*, p. 4, v. 44.

9) *Ibidem*, p. 22, v. 415.

10) *Ibidem*, p. 14, v. 244.

propre. Montrant sa maîtrise des textes de Serigne Touba et les citant, ou collant abondamment, Moussa Ka fait de son épopée une biographie inspirée par l'œuvre du sujet ou plutôt une autobiographie de Serigne Touba médiatisée par Moussa Kâ.

Du point de vue littéraire et poétique, l'emprunt des mots français et celui des mots arabes sont de fonctions similaires. Parce qu'autant que les termes français, les mots arabes jouent un rôle rythmique, prosodique et de référence culturelle, de renvoi à une civilisation et à un univers.

D'autre part, l'emprunt de mots arabes est parfois au second degré. Par exemple, voulant expliquer et traduire une expression française, Moussa Kâ, buttant devant l'insuffisance du wolof, adopte un mot arabe, parfois wolofisé et qui, traduisant le mot français, explique ainsi un emprunt par un autre. Cela justifie parfaitement la richesse plus grande du plurilinguisme et de l'association des langues sur le simple *code switching*. Ainsi, un même vers peut contenir des termes relevant de trois langues différentes.

[là amba wax jɛtnaŋ ba wax k o du : du / antalparɛ : t ba m o j  
b or om tawhi : du] (11)

L'étude de cette poétique révélerait une sorte de supra langue poétique qui, réhabilitant l'oralité dans la poésie, construit sa poéticité à partir de signifiants qui sont d'origines différentes mais qui sont combinables dans un texte mixte et véritablement polymorphe. La preuve en est que le tissu linguistique n'est souvent arabe que de par son signifiant visuel, les réalisations orales produisant des performances françaises, wolof ou arabes. L'instrument linguistique est ainsi modelé selon la langue orale utilisée et en cela les poètes wolof utilisant l'alphabet arabe procèdent à un travail phonétique et phonologique riche d'innovations.

Au regard de ces observations, il devient évident que l'emprunt a, ici, comme ailleurs de façon générale, une fonction communicative et linguistique, celle de suppléer à une insuffisance du wolof. Jacques Mouradian a pu, très tôt, dans un article de 1940, constater cela à propos du Wolof devant le français et l'arabe :

«Sollicités par deux civilisations qui s'opposent, les Wolofs empruntent plus volontiers aux Français les choses et les mots de la vie moderne; Mais, irresistiblement attirés par l'Islam qui

---

11) *Ibidem*, p. 9, v. 147.

gagne chaque jour davantage le monde noir, non seulement en surface, mais déjà en profondeur, c'est du monde musulman que ces noirs islamisés reçoivent les choses et les mots de leur vie spirituelle.» (12)

Mais l'emprunt, comme le montre ce poème de Moussa Kâ, va aller très vite au delà de cette première fonction. Parce que si l'alphabet arabe a une valeur essentiellement phonétique, en plus du rôle de construction du signifiant visuel et scripturaire, il reste évident que les mots et textes arabes et français, quand bien ils suppléent à un manque, n'en sont pas moins et surtout des indicateurs culturels, des instrument de poétisation. De même, l'ersatz de l'emprunt, le détournement sémantique, joue un rôle similaire. Ainsi plus que les mots wolof qui les intègrent, dans le poème de Moussa Kâ, ces termes, textes et expressions sont la mémoire d'une époque, d'une pensée et d'une civilisation.

---

12) MOURADIAN, J. (1940) : «Notes sur quelques emprunts de la langue wolof à l'arabe», *Bulletin de l'I.F.A.N.*, vol. 2, n° 3-4, p. 284.

## RÉFÉRENCES

- DUMONT, P. (1983) : *Le français et les langues africaines au Sénégal*, Paris, A.C.C.T., Khartala, 380 p.
- BA, O. (1972) : *Ahmadou Bamba face aux autorités coloniales 1889-1927*, Abbeville, F. Paillart, 250 p.
- DIEYE, Ch. A. (1985) : *Sur les traces de Cheikh Ahmadou Bamba : l'exil au Gabon: période coloniale : 1895-1902*, s.l. Ndigël, 119 p.
- SAMB, A. (1972): «Essai sur la contribution du Sénégal à la littérature d'expression arabe», *Mémoires de l'I.F.A.N.*, n° 87.



# Intégration des francicisms <sup>(1)</sup> dans les langues bantu

Zachée Denis BITJAA KODY

*Université de Yaoundé I, Yaoundé, Cameroun*

Des suites de leur cohabitation séculaire avec la langue française, les langues bantu (LB) <sup>(2)</sup> des pays africains francophones empruntent régulièrement à cette langue de nombreux termes liés surtout aux domaines de la technologie et des produits de consommation courante connus grâce à la civilisation française. Ces francicisms ou mots d'emprunt au français fréquemment utilisés dans

---

1) **Francicisme**: en l'absence d'un terme adéquat désignant les emprunts au français présents dans les langues africaines et autres, nous initions ce terme sur le modèle des mots : *anglicisme, germanisme, latinisme, hellénisme*, etc., courants dans les dictionnaires. Pour nous, le terme *francicisme* renvoie non seulement à tout mot ou expression d'origine française utilisé comme emprunt dans une autre langue, mais également à toute forme étrangère transmise à une autre langue par l'intermédiaire de la langue française.

2) Abréviations :

LETAC	Lexiques Thématiques de l'Afrique Centrale
CELTA	Centre de Linguistique Théorique et Appliquée (Zaire)
LD	Langue de Départ (le français dans ce cas précis)
LA	Langue d'Arrivée (l'une des LB du corpus)
LB	Langues bantu
BS	basaa
CB	ciluba
Ex.	exemple
FR	français
KR	ikinyarwanda
KS	kiswahili
MK	monokutuba

les langues du corpus (3), subissent de nombreux processus linguistiques visant leur intégration. Nous examinerons d'abord le traitement phonétique et phonologique des sons du français dans les LB. Nous décrirons ensuite les processus morphologiques, la classification nominale et les considérations sémantiques qui accompagnent souvent ce processus d'intégration, afin de dégager les principes linguistiques généraux qui gouvernent l'emprunt lexical des LB au français.

Nous prendrons appui sur des sources écrites, constituées essentiellement des emprunts lexicaux consignés dans les *LETAC*. En plus du fait que ces ouvrages montrent que les LB empruntent, presque toutes, les mêmes substantifs au français (4), ces ouvrages exposent clairement ces francismes dans leurs formes phonétique, phonologique et morphologique. Les langues de référence dans cette étude seront l'ikinyarwanda (KR) du Rwanda, le basaa (BS) du Cameroun, et le kiswahili (KS), le monokutuba (MK) et le ciluba (CB) du Zaïre.

## 1. — CARACTÉRISTIQUES GÉNÉRALES DES LANGUES BANTU

Les LB retenues dans cette étude partagent plusieurs caractéristiques. Sur les plans phonétique et phonologique, elles comportent à quelques exceptions près les mêmes systèmes de consonnes et de voyelles dont nous élaborons une synthèse comme suit :

### Phonèmes consonantiques

p	t	c	k
b	d	j	g
m	n	ny	
mb	nd		ŋg
f	s		
v	z		
	l		
		y	w

### Phonèmes vocaliques

i	u
e	o
ɛ	o (*)
	a

3) Voir le corpus de base en annexe.

4) Voir le corpus de base en annexe.

\*) Il s'agit dans la suite d'un o ouvert.

Quelques variations phonétiques existent d'une langue à l'autre et chacune possède quelques phonèmes non attestés dans les autres, c'est le cas des consonnes suivantes par langue :

MK :		- c, j, g, ny
CB :	+ sh	- g, j, v
BS :	+ h, nj, ŋ, ɓ	- z, f, v
KS :	+ h, nj	- d, y
KR :	+ ts, sh	

Toutes ces langues ont une nasale syllabique /N̄-/ homorganique à la consonne consécutive. Elles connaissent aussi les processus phonologiques de la labialisation, de la palatalisation et de la prénasalisation, même si ceux-ci n'affectent pas les mêmes consonnes d'une langue à l'autre.

En ce qui concerne les voyelles, contrairement au basaa qui possède les sept voyelles du tableau, les langues ciluba, monokutuba, kiswahili et ikinyarwanda ne distinguent pas /e/ de /ɛ/, ni /o/ de /ɔ/; elles présentent par conséquent un système réduit à 5 phonèmes vocaliques.

Les LB sont en majorité tonales; chaque syllabe y est prononcée avec une hauteur relative mais significative de la voix, et les tons y ont une valeur distinctive au même titre que les consonnes et les voyelles. Cependant, certaines d'entre elles à l'instar du kiswahili et du monokutuba ont perdu cette caractéristique et tendent à présenter un accent lexical démarcatif.

Sur le plan morphologique, toutes les LB sont agglutinantes et utilisent la préfixation comme processus morphologique le plus courant. L'une de leurs caractéristiques majeures au sein de la catégorie des substantifs est leur système de classes nominales; chaque thème nominal est rangé dans un et un seul genre constitué d'une paire de classes singulier/pluriel marquées par des préfixes indépendants spécifiques. Les emprunts au français n'échappent pas à cette taxinomie.

## 2. — CARACTÉRISTIQUES GÉNÉRALES DU FRANÇAIS

La langue française comporte sommairement les caractéristiques phonologiques et morphologiques suivantes :

**Phonèmes consonantiques** : /p, t, k, b, d, g, m, n, ny, f, s, S, v, z, J, R, l, y, w/

**Phonèmes vocaliques** : /i, ü, u, e, ø, ɛ, o, , œ, ɛ, œ, ɔ, ɔ, a, A, ã/

La labialisation et la palatalisation affectent fréquemment les consonnes françaises. Les sonantes /l/, /R/ et /s/ s'associent aisément à d'autres consonnes pour former des complexes CC, qui fonctionnent comme marges syllabiques : /pl-, pR-, sp-, -Rp, -lp,.../.

Sur le plan suprasegmental, à l'opposé des LB tonales, le français comporte un accent démarcatif fixe et non distinctif. Celui-ci est marqué au sein du mot par une forte intensité articulatoire sur la dernière syllabe articulée. Cet accent est suivi d'une chute de l'intensité qui signale la fin du mot.

Quant à sa morphologie, le français est une langue analytique qui utilise abondamment la préfixation et la suffixation dans la dérivation nominale.

Avec autant de divergences structurelles, il y a bien lieu de se demander comment certains mots français ont pu franchir toutes ces barrières phonologiques et morphologiques reconnues comme des freins à l'emprunt, pour être couramment utilisés dans les LB.

### 3. — RÈGLES D'ADAPTATION ATTESTÉES

#### 3.1 — Les facteurs

L'un des phénomènes les plus courants dans l'intégration d'un mot étranger est l'adaptation de ses structures phonologiques et morphologiques à celles de la langue d'arrivée. Si le lexique est reconnu comme le domaine le plus perméable aux échanges interlinguistiques, les systèmes phonologique et morphologique qui fonctionnent sur la base d'un nombre limité de règles sont très résistants aux changements (Coetsem, 1988 : 7-10). Cependant, malgré la rigidité de ces systèmes, tous les locuteurs d'une langue ne prononcent pas forcément le mot d'emprunt de la même manière. Les mots d'emprunt sont ainsi l'un des domaines du lexique où la variation idiosyncratique est des plus importantes. Cette variation dans l'articulation résulte de plusieurs facteurs.

##### 1 — *La distance linguistique*

Plus les divergences phonologiques et morphologiques sont grandes entre la langue de départ (LD) et la langue d'arrivée (LA), plus les locuteurs de la LA auront des difficultés à articuler le mot d'emprunt et à l'intégrer dans leur langue (Weinrich, 1979 : 1). Moins ces divergences sont importantes, plus les emprunts auront de la facilité à s'implanter. Cependant, ce préalable a plus ou moins de force selon le degré de bilinguisme des locuteurs.

## 2 — *Le bilinguisme*

Il est reconnu que les bilingues tendent à articuler les emprunts tels qu'ils sont prononcés dans la LD qu'ils maîtrisent relativement (Weinrich, 1979 : 26-27; Coetsem, 1988 : 107, 112), alors que les monolingues constituent la classe pour laquelle une bonne intégration phonologique de chaque mot d'emprunt est nécessaire, habitués qu'ils sont à l'utilisation d'un seul système phonologique. On peut donc comprendre qu'ils soient les principaux agents de l'adaptation phonologique des emprunts linguistiques. Une fois que le mot est adopté et introduit dans la langue par les bilingues, les monolingues s'emploient inconsciemment à le mastiquer et à le digérer jusqu'à ce qu'il prenne la «couleur locale», c'est-à-dire l'ensemble des caractéristiques structurelles de la LA.

## 3 — *Le temps*

On ne le mentionne pas souvent, mais le processus d'adaptation peut être long et laborieux. En effet, ce n'est pas au premier jour de son emploi dans une langue que le mot d'emprunt est débarrassé de tous les traits qui sont incompatibles avec ceux de la LA. Il peut en garder bien des décennies après sa première utilisation dans la LA, tout ceci dépendant du degré de complexité de la règle à contourner; car dans tout processus d'adaptation, il s'agit bien de quitter un système de règles, celui de la LD, pour un autre système, celui de la LA.

## 4 — *La fréquence d'emploi*

Les emprunts désignant des objets d'usage courant interviennent quotidiennement dans la LA; ils sont ainsi régulièrement exposés à l'adaptation linguistique par opposition aux termes techniques qui, confinés à une utilisation dans un domaine précis et par une catégorie de personnes (généralement bilingues), prennent plus de temps à se conformer aux normes structurelles de la LA.

### 3.2 — **Adaptation de l'accent aux tons**

Si nous représentons les syllabes non accentuées d'un mot de la langue française par des traits horizontaux afin de les distinguer de la syllabe accentuée du même mot que nous matérialisons par une courbe ascendante puis descendante, la courbe accentuelle canonique des mots français serait la suivante :

— — — ^  
1 2 3 4

Les syllabes 1, 2, et 3 sont facultatives. La syllabe 4 est obligatoire, car lorsque le mot est monosyllabique, la tension et la détente accentuelles restent observables au cours de son articulation comme dans les exemples suivants :

Ex.1	en-va-hi-sseur	ci-tron-nier	ci-tron	lame
	- - - $\wedge$	- - $\wedge$	- $\wedge$	$\wedge$
	[ã va i sœR]	[si tR o nje]	[si tR õ]	[lam]

Les LB tonales qui empruntent ces mots convertissent les courbes accentuelles ci-dessus en schèmes tonals. Les syllabes non accentuées du mot français prennent automatiquement un ton bas dans la LB, alors que la syllabe accentuée y prend soit un ton HB final, soit un ton H sur la pénultième et un ton B final, ces deux mores résultant d'un éclatement retraçable du ton HB sous-jacent, suivant les dispositions tonales régulières dans la LA.

En guise d'illustration, la langue basaa admet un ton HB sur les syllabes finales, alors que l'ikinyarwanda et le ciluba n'acceptent que des tons bas sur celle-ci. Devant un mot d'origine française chaque langue réagit selon ses règles tonales internes :

Ex.2	FR	accents	BS	KR	CB
	pompe	$\wedge$	p ombɛ	ìpómpò	mpómpì
	cassette	- $\wedge$	kàsɛt	kàsètì	kàsétè
	salade	- $\wedge$	sàlàt	sàlàdì	nsàláádè

Grâce à cette règle de traitement tonal, il est aisé de repérer tous les francismes par leur schème tonal dans une LB.

### 3.3 — Adaptation des phonèmes

Tels que présentés aux paragraphes 1 et 2, le français et les LB partagent plusieurs sons qui, sans être identiques de par leurs définitions et les rapports internes qu'ils entretiennent dans leurs systèmes respectifs, sont cependant suffisamment proches pour que nous puissions les assimiler dans le cadre de cette étude. Il s'agit des consonnes /p, t, k, b, d, g, f, v, s, z, m, n, ny, l/; et des voyelles /i, e, o, u, a/.

En dehors de ces sons analogues, qui ne posent aucun problème d'adaptation (Coetsem, 1988 : 22), nous assistons à diverses simplifications phonologiques lorsqu'un mot français s'implante dans une LB.

### 1 — *Voyelles orales du français*

Les voyelles orales du français sont adaptées de la manière suivante dans les LB :

français		bantu à 5 voyelles		bantu à 7 voyelles
/i, ü/	→	/i/, /u/		/i/
/e, ø, ε, œ, ɛ̃/	→	/e/		/e/, /ɛ/
/o, o/	→	/o/		/o, o/
/a, A/	→	/a/		/a/

Ex.3	FR	KS	MK	CB	KR	BS
	bureau		bilo	bìiló	ìbirò	bìrò
	rhubarbe	rubarbe		lùbáábà	rùbàrúbè	
	dindon	nden	doden	dondèndù	dèndò	

### 2 — *Voyelles nasales du français*

Les voyelles nasales /ɛ̃, ã, õ, œ̃/ perdent leur nasalité et obéissent aux règles d'adaptation décrites pour leurs correspondantes orales.

Cependant, cette perte de la nasalité vocalique est compensée par la prénasalisation de la consonne subséquente. La nasalité se perd souvent lorsque la voyelle nasale apparaît en finale du mot français.

Ex.4	FR	MK	KS	CB	KR
	dindon [dɛ̃dɔ̃]	dendo	ndendo	ndèndù	dèndò
	client [kljã]	kidiya	kilia	kìliyâ	ùmìkìliyâ

### 3 — *L'hiatus*

La plupart des LB ne présentent pas cette situation définie comme la rencontre de deux voyelles à l'intérieur du même mot. Lorsqu'un mot d'emprunt comporte un hiatus, les LB introduisent souvent une semi-voyelle épenthétique entre les voyelles en contact. La semi-voyelle épenthétique sera /y/ avant une voyelle antérieure, et /w/ avant une voyelle postérieure.

Ex.5	FR		MK	CB	KR
	protéine	[pr otein]	polotoyini	pòlòtèyínà	pròtèyìnè
	knock-out	[kao]	kawo	kàwó	

#### 4 — Consonnes simples du français

On assiste généralement aux adaptations suivantes :

Ex.6	FR		MK	BS
	/ʃ/ → /s/	machine [maʃin]	→ masini	màsîn
	/ʒ/ → /z/ ou /s/	garage [gaʁaʒ]	→ galazi	gârâs
	/ʀ/ → /l/ ou /r/	bureau [byʀo]	→ bilo	bîrô

#### 5 — Syllabes fermées et complexes consonantiques français

Les consonnes françaises labialisées et palatalisées ne posent aucun problème d'adaptation aux LB qui connaissent aussi ces processus. Par contre, les séquences de consonnes à l'intérieur d'une syllabe ou entre deux syllabes du mot français constituent une pierre d'achoppement. Pour toutes les LB, ces séquences de consonnes sont des aspérités articulatoires qui doivent être laminées au cours de l'implantation de l'emprunt. Le cas échéant, elles recourent au processus phonologique de l'épenthèse et insèrent une voyelle d'appui entre les consonnes contigües. La voyelle épenthétique est très souvent une copie de l'une des voyelles qui entourent le complexe, ou sa correspondante dans la LA.

Ex.7

FR	KS	MK	CB	KR	BS
client	kilia	kidiya	kìliyâ	ùmìkìliyâ	kìliyân
diplôme	diploma	dipolomi	dìipòlòmà	dìipòlómì	lìpòlòm
carbone	karboni	kalaboni	kàrbónè	kàrùbónè	

L'insertion de la consonne ou de la voyelle d'appui est guidée par la structure syllabique de base CV des LB. Toutes les syllabes des francismes qui ont une structure autre que CV sont sujettes à cette simplification syllabique. Grâce aux règles d'autorégulation inhérentes à chaque langue, les structures de la langue d'arrivée finissent toujours par l'emporter sur l'organisation initiale des éléments. Dans les LETAC, seuls les francismes jugés d'emprunt récent (et marqués L) par les auteurs résistent encore à cette règle de l'épenthèse et

présentent des séquences de consonnes ou des hiatus; preuve, s'il en faut, qu'ils sont en cours d'implantation.

6 — *Intégration de nouvelles structures et de nouveaux phonèmes ?*

La non-adaptation de certains sons et la non-application occasionnelle de la règle de l'épenthèse entraînent l'utilisation de certains sons et complexes sonores irréguliers dans les LB.

Ex.8

[ʃ] dans chou [ʃu]/[shu] utilisé comme tel dans les LB;

[ʒ] dans garage [gaRaʒ] noté /j/ dans les lexiques (5);

[pR] et [ei] de protéine [pR otein] encore prononcés /protein/ en kiswahili.

Cependant, malgré le fait que ces sons et complexes sonores soient transitairement fonctionnels et distinctifs dans les LB, ils ne peuvent pas être considérés comme INTÉGRÉS dans leurs systèmes phonologiques parce qu'ils ne remplissent pas le critère de productivité propre aux phonèmes d'une langue. En effet, pour qu'un son initialement introduit dans une langue par le biais d'un emprunt lexical soit considéré comme phonème de cette langue, il doit pouvoir intervenir dans la formation d'autres mots natifs. Pour Coetsem (1988 : 106) :

*«If the phonological loan is integrated into the rl [recipient language], i.e. when it is part of the rl code, it loses viewed synchronically-its status as a loan; the phonological loan has evolved from a nonnative element to a native constituent. [This] integration directly affects the structural make-up of the rl.»*

Ainsi, une fois qu'un emprunt phonologique est intégré dans le système d'une langue, il cesse du même coup d'être un emprunt et devient un phonème à part entière. Les emprunts phonologiques contenus dans les francicisms n'ont pas encore atteint ce stade de l'intégration phonologique dans les LB.

---

5) Voir le corpus de base en annexe.

### 3.4 — Traitement morphologique des francicismes

La structure canonique des nominaux dans les LB est la suivante :

(préfixe indépendant) + RADICAL + (suffixe non grammatical)

Le radical est le seul morphème obligatoire, ce qui signifie que certains mots seront formés sans marques apparentes soit du préfixe, soit du suffixe, soit des deux.

#### 1 — *Le radical*

Tout francicisme, quelle que soit sa morphologie d'origine et sa longueur, est globalement considéré comme un radical susceptible d'accueillir un préfixe et un suffixe propres à la LA, même si ces derniers ne sont pas toujours apparents.

Ex.9 FR	KS	MK	KR
fromage		Ø-fomaz-i	ì-fóròmāj-ì
ananas	Ø-nanas-i	Ø-nanas-i	ì-nānās-ì

#### 2 — *Le préfixe indépendant*

Il est la marque morphologique de la classe nominale dans laquelle le francicisme est rangé. Sa structure syllabique est l'une des suivantes : | Ø-, Ñ-, V- ou CV- |. Il gouverne la classification qui résulte de plusieurs facteurs dont :

##### a) le signifié

Lorsque le signifié du francicisme est humain, les LB ont tendance à le ranger dans le genre I.

Ex.10

FR	KS	MK	CB	KR
client	kilia	kidiya	kìliyâ	ùmìkìliyâ
(+ humain)	(cl.1,2)	(1a,2a)	(1a,2a)	(cl.1,2)

##### b) l'initiale du francicisme

Lorsque la syllabe initiale du francicisme correspond phonologiquement à un préfixe de classe quelconque dans la LA, elle est souvent prise pour un indice de classification. Le francicisme concerné est ainsi automatiquement rangé dans la classe correspondante quelle que soit sa signification.

Ex.11	FR	CB	KR	BS
	cacao	kà-kàó cl.12/13		
	café	kà-àfê cl.12/13		
	carte	ka-làtà cl.12/13		
	équipe		ì-kipì (cl.5/6)	hì-kîp (cl.11/12)

c) l'affectation d'un indice de classe

En dehors de ces contingences, la langue peut fournir un indice de classe propre, afin de ranger le mot d'emprunt dans une classe donnée.

Ex.12

FR	CB		KR	
pompe	m-pómpì	cl.9/10	ì-pômpò	cl.5/6
fromage	m-fòlòmáàjè	cl.9/10	ì-fòròmāJì	cl.5/6
garage	n-gàláájè	cl.9/10	ì-gárāJì	cl.5/6

d) la dérivation nominale

Lorsque le francicismisme est à l'origine d'une dérivation nominale dans la LB, il est conservé comme substantif, mais son dérivé acquiert le préfixe d'une autre classe nominale. C'est ce qu'on observe souvent dans les relations fruit/arbre fruitier en kiswahili et en monokutuba :

Ex.13

FR	KS		MK	
cacao	Ø-kakao	cl.9,10	Ø-kau kau	cl.9,10
cacaoyer	m-kakao	cl.3,4	mu-kau kau	cl.3,4
papaye	Ø-papai	cl.5,6	Ø-payi-payi	cl.9,10
papayer	m-papai	cl.3,4	mu-payi-payi	cl.3,4

Malgré ces différents processus morphologiques visant une parfaite intégration des emprunts, la grande majorité des francicismes est rangée dans le genre V (cl.9/10) qui semble être un «fourre-tout», voire une porte d'entrée aux emprunts de toutes origines. Le préfixe de ces deux classes est souvent |Ø-|, ce qui permet aux francicismes de

conserver des formes analogues à celles qu'ils avaient en français (6). Dans chacune des langues du corpus, la distribution des francicisimes dans les classes nominales obéit approximativement aux pourcentages suivants, obtenus sur la base des données du corpus : classes 1/2 = 5 %, classes 3/4 = 7 %, classes 5/6 = 10 %, classes 7/8 = 4 %, classes 9/10 = 68 %, autres classes = 6 %.

### 3 — *Le suffixe non grammatical*

C'est une voyelle finale non porteuse de sens et sans fonction grammaticale que l'on trouve toujours en finale de mot dans celles des LB qui n'admettent pas de structure syllabique fermée. Elle y apparaît donc comme une nécessité phonologique alors qu'elle ne l'est plus en basaa, LB qui perd progressivement la syllabation ouverte et présente des syllabes fermées. Ainsi, dans les langues à syllabation ouverte unilatérale, les francicisimes qui s'achèvent phonologiquement sur une consonne prennent obligatoirement le suffixe non grammatical matérialisé par /-i/ dans la plupart des cas.

Ex.14

FR	KS	MK	CB	KR	BS
hectar(e)	hektari	ekitari	héktáàrè	hègitàrì	èktâr
musiqu(e)	muziki	miziki	mìzíkì	mùzìkà	nsík

### 3.5 — *Sémantique des francicisimes*

Compte tenu de la différence des écosystèmes, quelques signifiés sont légèrement différents de ceux courants en France. C'est ainsi que le francicisme *prine* du monokutuba, qui provient du mot français *prune*, désigne une baie typiquement africaine (*assa* au Cameroun), inconnue en France. On assiste aussi à quelques glissements sémantiques comme dans les cas où *franc* signifie à la fois monnaie, argent et franc dans certaines langues (7).

Mais d'une manière générale, les emprunts lexicaux au français gardent leur sens dans les LB. Ils sont alors considérés comme un enrichissement du vocabulaire permettant aux locuteurs de ces langues de nommer de nouveaux objets et de nouvelles perceptions qui n'existaient pas antérieurement dans leur environnement. Les

---

6) Voir le corpus de base en annexe.

7) Voir le corpus de base en annexe.

francismes cités en exemple dans cet exposé et repris en annexe, témoignent incontestablement du riche apport de la culture et de la langue françaises aux peuples africains en général et aux LB en particulier. Comme le pense Brunot (1928), les mots sont des «témoins de l'histoire». Nous convenons ainsi avec Nyrop (1934 : 68) :

«[qu']il est toujours intéressant d'étudier dans une langue les éléments étrangers. Ils nous révèlent tant de choses à la fois sur la nation qui fournit et sur celle qui emprunte. On n'exagère pas beaucoup en disant qu'on pourrait écrire de grands chapitres de l'histoire du développement culturel d'un peuple rien qu'en se servant des mots étrangers auxquels il a donné droit de cité.»

### Annexe : Corpus de base

Dans ce corpus nous n'avons retenu que les francismes apparaissant dans au moins 3 des 5 langues.

FR	KS	MK	CB	KR	BS
ananas	nanasi	nanasi		ìnânāsì	
argent	franga		mfwàlángà	àmàfàrángà	
autobus	bisi	bisi	bísè	òtòbìsì	bìs
avocat	avokati	savoka	àvòkâ	mvòkà	
bureau		bilò	bìiló	ìbìrò	bìrô
cacao	kakao	kaukau	kàkàó	kàkàwò	kàkáo
cacaoyer	mkakao	mukaukau	kàkàó	kàkàwàyè	kàkáo
café	kahawa	kafi	kààfê	ìkâwá	
caféier	mkahawa	mukafi	kààfê	ìkâwá	
caméra	kamera		kàmèrà	kàmèrà	
cassette	kaseti	kaseti	kàsétè	kàsètì	kàsEt
carbone	karboni	kalaboni	kàrbónè	kàrùbónè	
carotte	karoti	kaloti	kàrótè	kàròòtì	kàrôt
carte	kadi		kálàtà	ìkàrítã	kààt
cérise	serize	selize	nsèdíízà		
chou	shu	shu	shû	ìshù	shû
cinéma	sinema		sèndèmá	sìnèmã	sìnimá
client	kilia	kidiya	kìliyâ	ùmìkìliyà	kìliyâe

cocotier		mukokoti	kòkó	kòkò	
coupe		kupu	kúpà		kúp
coupe-coupe		kupekupe	kupukupu	kùpèkùpè	kùpàkùpà
démocratie	demokrasi		dèmòkràsì	dèmòkàràsí	
dindon	ndendo	dendo	ndèndù	dèndò	
diplôme	diploma	dipolomi	dìipòlòmà	dìipòlómì	lìpòlóm
écologie	ekolojia	ekoloji	èkòlòjì		
épinard		pinadi	pínààlà	pìnààrì	
équipe		ekipi		ìkìpì	hìkìp
facture		fakitidi	faktífè	fàgìtìrì	
film	filamu	filimi	fílmè	filímì	
fromage		fomazi	mfòlòmáàjè	ìfòròmâjì	
garage	garaji	galazi	ngàláájè	ìgàràjì	gàràJ
hectare	hektari	ekitari	héktààrè	hègìtârì	èktâr
hôtel	hoteli		òtélé	hòtèlì	òtel
kilo..	kilo	kilo	kìlò	kìló	kílò
knock-out	kao	kawo	kàwó		kā-ó
licence	leseni		lìsânsè		lìsâns
macabo		makabo	màkàbó	màkàbò	origine
machine	mashini	masini	màshínyì	màshìnì	màsín
microbe	mikrobi	mikolobi		mìkòróbù	
milliard	miliardi		mìliyáálè	mìliyààrì	
million	milioni		mìliyô	mìliyòònì	
musique	muziki	miziki	mìzíkì	mùzìkà	nsík
navette	naveti		nàvétè		nàvet
pamplemousse		pampelemusu	pàmpàlàmusà		pàmpèlèmús
papaye	papai	payi-payi	cìpàpááyì	ìpápāyì	
papayer	mpapai	mupayipayi	cìpàpááyì	ìpápāyì	
pénalty	penalti	penati	pènà	pènàlìtì	pènàtì
photogr..		foto	fòtó	ìfòtò	fòtò
politique		politiki	pòlitíkè		pòlitík
pomme		pome	pómè	pómè	pôm
pompe		pompi	mpóm̀pì	ìpóm̀pò	pOmbe

protéine	proteini	polotoyini	pòlòtèyínà	prètèyìnè	
quittance		kitansi	kìtánsè	gìtânsì	
radio		ladyo	ràdyó	rààdìyó	ràdìò
raquette	raketi		làkétè	ràkètì	ràkÈt
rhubarbe	rubarbe		lùbáábà	rùbàrúbè	
safari	origine		sàfàrì	ísàfàrì	
salade	salade	saladi	nsàláádè	sàlâdì	sàlât
taxi	taksi	takisi	tàksì		tàksî
technique	tekniki	tekiniki	tèkníkè	tèkìnikì	tegnîk
télév..	televisheni	televizyo	tèvê	tèlèviziyó	tèlê
télex	teleksi		tèléksè	tèlégìsì	
tennis	tenisi	tenisi	tènìsè		tènîs
tomate	tomate	tomati	kàmátà	itómàtí	tòmât
tonne	toni	toni	tónè	tónì	tôn
vitamine	vitamini	vitaminu	vitàmínè	vitàmìnì	vitàmîn

### Remerciements

Nos remerciements vont au Professeur André Clas de l'Université de Montréal, à M. Nazam Halaoui du GRESLET, Université de Montréal et à M. Maxime Da Cruz de l'Université Nationale du Bénin, pour leurs remarques et suggestions pertinentes.

## RÉFÉRENCES

- ACCT/CERDOTOLA/Équipe Nationale Zaïroise (1983) : *Lexiques Thématiques de l'Afrique Centrale : Zaïre, ciluba : Activités économiques et sociales*, Paris, 311 p.
- ACCT/CERDOTOLA/Équipe Nationale Zaïroise (1983) : *Lexiques Thématiques de l'Afrique Centrale : Zaïre, kiswahili : Activités économiques et sociales*, Paris, 267 p.
- ACCT/CERDOTOLA/Équipe Nationale Zaïroise (1983) : *Lexiques Thématiques de l'Afrique Centrale : Zaïre, monokutuba : Activités économiques et sociales*, Paris, 287 p.
- ACCT/CERDOTOLA/Équipe Nationale du Cameroun (1983) : *Lexiques Thématiques de l'Afrique Centrale : Cameroun, Jasaá : Activités économiques et sociales*, Paris, 318 p.
- ACCT/CERDOTOLA/Équipe Nationale du Rwanda (1983) : *Lexiques Thématiques de l'Afrique Centrale : Rwanda, ikinyarwanda : Activités économiques et sociales*, Paris, 263 p.
- BELAND, R. et C. PARADIS (1997) : «Principled syllabic dissolution in a primary progressive aphasia case», *Aphasiology*, Cambridge University Press.
- BITJAA KODY, Z. D. (1997) : *Le contact des langues : Superstrat et enrichissement du français au Cameroun*, Rapport de mission FICU, AUPELF-UREF, Montréal, 32 p.
- BRUNOT, F. (1928) : *Les mots témoins de l'histoire*, Paris.
- COETSEM, Frans Van (1988) : *Loan phonology and the two transfer types in language contact*, Dordrecht, Foris Publications, xvii, 207 p.
- NYROP, K. (1934) : *Linguistique et histoire des mœurs*, traduit par E. Philippot, Paris.
- WEINRICH, U. (1979) : *Languages in contact, Findings and Problems*, 8<sup>e</sup> édition, The Hague/ Paris/ New York, Mouton, 148 p.

# Choix de la charge théorique et mots fonctionnels

Sylvie PORHIEL

*Université Paris 13, Villetanneuse, France*

## INTRODUCTION

Les mots ont une mémoire<sup>(1)</sup> (étymologique, ethnique, historique, théorique) qui constitue leur histoire (étymologie, sens ancien, tradition), leur «biographie», leur définition notionnelle : ils sont le réceptacle des différentes conceptualisations de la pensée. Les mots et leur histoire sont fortement ancrés dans la société tant du point de vue de l'individu que de son appartenance à un groupe professionnel. Les notions, élaborées par les savants, les théoriciens, et représentées par des termes, peuvent être définies de différentes manières. Quand on utilise un terme, cette définition doit être explicite. Considérons ce que propose le rapport interne de l'INaLF pour *concept* :

«La notion de concept connaît diverses acceptions tant en IA, qu'en logique, en philosophie et en linguistique. *En philosophie et en logique*, le concept est une représentation mentale, générale et abstraite d'un objet. *En linguistique*, le concept est un universel de représentation qui appartient au langage mais qui n'est dépendant d'aucune langue déterminée.

---

1) Cf. le livre intitulé *La mémoire des mots*. (Pillon, 1993) ou le numéro 114 de *Langages*.

*Dans une approche saussurienne, le concept correspond au signifié, le signifiant étant son image acoustique. Il en résulte qu'en règle générale un concept donné a pour pendant dans une langue donnée un signifié qui en tient lieu dans les limites de cette langue. En terminologie, le concept est un sémème construit dont la définition est stabilisée à l'intérieur d'un domaine ou d'une discipline de sorte que ses occurrences soient identiques à son type».* (Humbley, Otman, 1995 in INaLF : 25)

Chaque fois que les auteurs changent de discipline, de domaine pour définir le *concept*, ils utilisent des moyens langagiers, en l'occurrence la préposition *en* suivie d'un nom et l'expression lexicalisée (2) *dans une approche* suivie d'un adjectif (il peut aussi être suivi d'un complément nominal) pour baliser leur propos. Nous appelons ce genre de mots et d'expressions des **marqueurs d'intérêt discursif** : ces mots fonctionnels ont pour fonction de pointer, de sélectionner une information pertinente relative aux propos du locuteur, information sélectionnée dans la mémoire des termes pour ce qui nous concerne ici.

Nous nous intéressons ici à la «mémoire d'une notion (3)», en tant qu'il s'agit d'un acquis cognitif auquel on peut faire référence, dans un cadre théorique. A moins de vouloir retracer tout le parcours d'une notion dans l'histoire d'un domaine, d'une discipline, de la science (pour être encore plus vaste), celle-ci n'émerge que par séquences. Le locuteur, lorsqu'il parle, ne donne jamais accès (Ariel, 1988) à toute la mémoire cognitive du terme; il peut sélectionner une partie de cette mémoire au moyen d'outils langagiers (**dans le cadre** (*de (Dét+E) N+A*), (**au+du) point de vue** (*de (Dét+E) N+A*(4))). Ceux-ci font alors émerger un faisceau de traits caractéristiques, ce qui conduit à opérer un choix et à évoluer dans le cadre circonscrit au préalable.

En effet, les notions élaborées par des théoriciens, s'utilisent dans des disciplines différentes et leur définition n'est bien souvent pas la même d'une théorie à l'autre, et ce, tout particulièrement dans le domaine des sciences humaines. Nous évoquons parmi d'autres les trois notions suivantes : *connecteur, prédicat, pragmatique*. Elles

---

2) Nous préférons cette appellation à celle de *locution* étant donné que cette dernière recouvre des acceptions fort différentes.

3) Nous ne faisons pas de différence entre *notion* et *concept*.

4) N : nom; Dét : déterminant; A : adjectif.

possèdent indéniablement une charge cognitive théorique dans la mesure où elles ont été définies maintes fois et que les définitions sont parfois incompatibles. Dans cette communication, notre hypothèse est la suivante : les notions possèdent une charge cognitive théorique (5) importante qui fait partie intégrante du poids cognitif du terme en question. Lors d'une intervention, le locuteur ne sélectionne qu'un pan de la charge théorique d'une notion grâce à des outils linguistiques qui jouent le rôle de signaux. Ces séquences orientatrices sont formées d'expressions suivies de leur complément nominal et/ou adjectival.

### LES NOTIONS THÉMATIQUES ET LES NOMS DE THÉORICIENS POSSÈDENT UNE CHARGE COGNITIVE THÉORIQUE

La question qui se pose à présent est la suivante : comment indiquer le choix de telle notion sans reprendre la démonstration telle qu'elle a été entreprise par le chercheur ? Nous limitons notre analyse aux sciences humaines que nous restreignons encore à la linguistique. Dans ce cadre théorique, lors d'un échange, les protagonistes — le locuteur et les lecteurs-auditeurs — sont des spécialistes du domaine ou, à tout le moins, des personnes averties. Dans une situation de communication orale ou écrite, il y a sollicitation des connaissances théoriques de part et d'autre (les pré-requis), ce qui suppose un substrat de connaissance commun du fait d'une même culture professionnelle. Ces connaissances présentent, par conséquent, un caractère anaphorique et nous parlerons d'anaphore culturelle (6) à la

---

5) Tout comme on peut parler de *charge culturelle*. Cf. l'introduction de l'ouvrage dirigé par Remi-Giraud et Retat (1996). On peut aussi parler de *charge* en termes de coût comme le fait B.N. Grünig (1993) pour la *charge mémorielle*.

Les mots abstraits, des lexèmes, possèdent une charge mémorielle plus ou moins importante. Elle existe au niveau de la société comme diverses études l'ont souligné. On pourra consulter à ce propos les communications rassemblées dans les ouvrages *Les mots de la nation* (1996) et *Mots, représentations - Enjeux dans les contacts interethniques et interculturels* (1994), par exemple, dirigés respectivement par S. Remi-Giraud, P. Retat, et K. Fall, D. Simeoni, G. Vignaux.

6) L'anaphore culturelle — comme toute forme d'anaphore (du grec *ana-* «à rebours», «en arrière» et *pherein*, «porter») — est régie par un principe d'économie et de cohésion. Économie car elle dispense l'énonciateur de longs développements et autorise maint raccourcis [...] Cohésion car elle jette une sorte de «pont de connaissance» entre ce même énonciateur et son partenaire, ce qui les rapproche.

suite de R. Lapaire et W. Rotge (1991 : 119). Ces anaphores culturelles apparaissent dans le cadre de la présentation d'une recherche bien souvent sous la forme de noms de théories (ou de cadre théorique) et de noms de théoriciens (7), par exemple : *Ducrot, Chomsky, fonctionnalisme, etc.*

Il est facile de constater que l'approche d'une notion déterminée est variable selon l'univers de croyance (Martin, 1992) des chercheurs, la citation ci-dessous (une parmi d'autres) étant là pour le rappeler :

«Il semble en effet que chaque théorie linguistique qui existe, si développée soit-elle, ne s'occupe que de certains aspects de cette problématique, ce qui explique la confusion terminologique et notionnelle qui est si caractéristique de la linguistique, dès qu'on compare les théories».

Comme nous nous trouvons dans un contexte de pluridisciplinarité, il s'est créée une situation de contact où les différentes théories se côtoient les unes les autres. Cela a pour double conséquence de mettre en relation des théoriciens travaillant dans des domaines fort différents du vaste domaine de la linguistique et *a fortiori* des sciences humaines et de constater qu'un terme «transdisciplinaire» ne renvoie pas exactement à la même connaissance cognitive et culturelle qui est relative à chaque discipline (8). Il apparaît, alors, que les notions sont des espaces non stabilisés, faisant preuve d'une certaine plasticité (9), d'où des divergences quant aux valeurs assignées aux mots (termes), en fonction des points de vue des chercheurs, des théories, des cadres d'étude, etc. En conséquence, il est indispensable de trouver un moyen pour se positionner à des fins de communication efficace et de stabiliser les espaces notionnels.

---

Nous serions tentés de parler ici de cohésion intersubjective. L'expression employée par les grammairiens anglophones de *shared knowledge* est à ce titre très révélatrice : partager (*share*), c'est être uni/rapproché temporairement, pendant l'échange verbal. D'où un climat de connivence mentale entre les sujets parlants qui, même s'ils s'affrontent, sont tous deux liés par un acquis culturel commun.» (Lapaire, Rotge, 1991 : 119)

7) D'autres cas sont possibles mais nous ne les prenons pas en compte ici.

8) Cf., par exemple, le cas de certaines écoles qui ont développé un langage particulier, un langage propre à la théorie en question : e.g. la grammaire générative, les tenants de Culioli, Guillaume, etc.

9) Cf. la citation de l'INaLF proposée dans l'introduction.

Le scientifique utilise son pouvoir performatif pour introduire le sens dans lequel il définit un terme grâce à des marqueurs d'intérêt discursif suivis d'un nom de théorie ou d'un nom de théoricien :

- (1) «**Dans la terminologie de Rosch** <sup>(10)</sup> (1977), la structure de la catégorie du Npr dénominatif est digitale, alors que celle du Npr métaphorique est analogique». (Jonasson, 1994 : 236)
- (2) «**Dans une perspective essentiellement sémantique**, il est vrai; mais la prise en compte des aspects pragmatiques de la langue peut ouvrir des perspectives fécondes dans l'étude des organisations discursives (...)» (Caron, 1983 : 20)

Le locuteur introduit par conséquent une contrainte de contenu reconnue par une communauté à une époque donnée et qui fait appel à la mémoire cognitive. Soulignons que cette contrainte, tacitement admise par les pairs, ne vaut que dans le cadre de la situation d'énonciation. Par ailleurs, les expressions utilisées pour indiquer à l'auditeur le cadre choisi sont très souvent ce que I. Fónagy (1982) appelle des *expressions liées* (**dans la terminologie** (*de (Dét+E) N+A*), **au sens** (*de (Dét+E) N+A*)).

### **CETTE CHARGE THÉORIQUE EST MISE EN AVANT PAR LES MARQUEURS D'INTÉRÊT DISCURSIF QUI SÉLECTIONNENT DES INFORMATIONS**

Notre deuxième point concerne le caractère sélectif des marqueurs : ce sont des unités grammaticales simples ou complexes (conjonctions, prépositions) qui constituent des repères dans les textes du fait de leur caractère récurrent. Les marqueurs d'intérêt discursif sont des expressions plus ou moins figées dont les combinatoires ne sont libres ni sur le plan lexical, ni sur le plan syntaxique, ni sur le

---

10) Le principe de notation est le suivant : pour les suites mixtes, c'est-à-dire, celles qui acceptent comme complément un nom introduit par une préposition et un adjectif, la préposition du complément ne fait pas partie de l'expression lexicalisée. Ex. **dans le cadre** (*de (Dét+E) N+A*) : elle n'apparaît pas en gras. En revanche, quand seul le complément nominal est possible, alors la préposition fait partie de l'expression lexicalisée : **à propos de** (*(Dét+E) N*) et apparaît en gras dans les exemples.

plan sémantique. Étant donné cette interdépendance, nous appelons l'ensemble constitué du marqueur et de son complément une *suite* (11). Dans les exemples ci-dessous, elles sont soulignées; l'expression lexicalisée est en gras, le complément adjectival ou nominal en italique :

(3) «(...) en fait, deux «articulations» **au sens de Martinet.**»  
(Berrendonner, 1990 : 26)

(4) «**Dans l'approche des actes de langage**, la fonction des énoncés est purement instrumentale : ils permettent au locuteur d'agir avec ou sur son interlocuteur; **dans la perspective gricienne** (et a fortiori celle de Sperber et Wilson), elle est essentiellement cognitive : le langage, et donc les énoncés dans la communication, ont pour fonction de permettre la connaissance du monde, et la connaissance résultant de l'interprétation des énoncés fait intervenir des mécanismes de type autant inférentiel (pragmatique) que linguistique (syntaxique et sémantique). **Dans la perspective d'Anscombe et Ducrot**, la fonction des énoncés est liée à leur orientation, i.e. le fait qu'ils permettent de servir telle ou telle conclusion. On ne peut donc plus réduire la fonctionnalité du langage à l'instrumentalité ou à la connaissance, mais bien plutôt à l'argumentation». (Moeschler, 1986 : 153-154)

(5) **Dans la perspective chomskyenne**, il s'agit d'une structure *syntaxique*, c'est-à-dire d'un système de relations fonctionnelles entre éléments grammaticaux, qui donnera lieu *par ailleurs* à une interprétation sémantique. (Caron, 1983 : 29)

Dans les exemples ci-dessus, nous notons que les choix effectués dans la mémoire cognitive le sont à visée argumentative (12).

---

11) Parler de *suite* nous semble important car, lorsque l'on procède à des transformations, c'est bien la préposition et son complément qui commutent et non pas l'un ou l'autre.

12) La visée de discours, ou intension de signification, est l'étape liminaire de l'acte d'énonciation, celle dans laquelle aucun dire n'est mis en oeuvre. Le locuteur a la possibilité, grâce à certains adverbes en particulier, de caractériser ou de commenter sa visée de discours. Plusieurs types doivent être distingués, selon l'aspect de la visée de discours qui est en jeu. Celle-ci comporte tout d'abord un

Le locuteur indique au récepteur — en l'occurrence le lecteur-auditeur —, au moyen de repères (c'est-à-dire de marqueurs comme **dans la terminologie** (*de (Dét+E) N+A*), **au sens** (*de (Dét+E) N+A*), **dans l'approche** (*de (Dét+E) N+A*), **dans la perspective** (*de (Dét+E) N+A*)), les éléments à sélectionner dans ses connaissances pour comprendre l'énoncé.

Les marqueurs d'intérêt discursif sont donc des outils fonctionnels dont le rôle est d'orienter le processus de sélection et, partant, de construire quelque chose d'après le cadre de cette base. Ces outils constituent, en quelque sorte, des signaux servant à identifier une théorie. Dans les exemples ci-dessous, les séquences de mots soulignées sont des suites qui restreignent le contenu propositionnel qu'elles introduisent :

- (6) «**Du point de vue diachronique**, les termes homonymes ont des étymologies distinctes (...)

**Du point de vue théorique en synchronie**, on parle de plusieurs unités homonymes lorsque les représentations sémantiques que l'on construit pour rendre compte des différentes significations sont totalement disjointes (...)

**Du point de vue épi-linguistique en synchronie**, sont traitées, de fait, comme des homonymes les formes dont les sujets parlants s'accordent spontanément à considérer les significations correspondantes comme étrangères les unes aux autres (...)» (Fuchs, 1996 : 26-27).

Un contraste distinctif est clairement mis en relief : dans le paradigme possible, l'information introduite écarte les autres. De toutes les références potentielles, seules certaines sont actualisées dans

---

référent, une donnée événementielle à traduire sous forme linguistique. (Guimier, 1996 : 141)

Tout énoncé construit présuppose donc, de façon liminaire, l'existence de ce que, à la suite de Joly et Roulland (loc. cit.), on appellera une *visée de discours*. La visée de discours peut être définie comme l'intention de donner une forme linguistique à un fait appartenant à l'extra-linguistique en vue de communication à un destinataire. (Guimier, 1996 : 105)

La *visée constructive de discours*, qui se situe hors langage, dans l'univers d'expérience. Cette visée que, depuis 1939, Guillaume a successivement appelée «but de pensée», «sens d'intension», «visée d'effet», «visée de discours», habite tout sujet parlant en instance de parole. Elle rencontre la langue construite - construite en lui - qui lui offre des moyens d'expression plus ou moins élaborés. (Joly, 1997 : 208)

le contexte de la communication (voire même simplement dans un paragraphe). Grâce aux marqueurs d'intérêt discursif, le locuteur introduit une contrainte de contenu reconnue par une communauté à une époque donnée et qui fait appel à la mémoire cognitive : ils constituent ainsi des instructions interprétatives.

Résumons-nous. Il apparaît que la mémoire cognitive des mots (termes) est trop vaste. Des mots et des expressions (qui sont des outils langagiers) servent à l'appréhender, mais cela ne peut se faire que grâce à un découpage qui est alors par nature sélectif. Nous nous sommes limitée à la linguistique et les expressions métalinguistiques utilisées pour se «situer» par rapport aux notions présentent la particularité d'introduire des noms propres (théoriciens) et des noms abstraits (théories). Les exemples (4), (5) et (6) introduisent (sous leur forme nominale ou adjectivale) des noms de linguistes connus ou supposés tels et reconnus dans le monde des linguistes. Nous considérons avec K. Jonasson (1994) et G. Kleiber (1996) que les noms propres jouent un rôle cognitif important. Par convention, dans le domaine de la linguistique, on sait que tel nom de théoricien est apparié avec telle théorie, et quelle théorie a été élaborée par quel théoricien. Après l'acquisition de cette compétence référentielle, ces mots ou groupes de mots (*Martinet, actes du langage, gricienne, etc.*) servent d'instructions. Cela permet ensuite au locuteur de les utiliser comme invariants. En ce sens, les noms de théories et de théoriciens possèdent un caractère anaphorique.

### LA MÉMOIRE FONCTIONNELLE DES MARQUEURS D'INTÉRÊT DISCURSIF PERMET D'INTRODUIRE DES POINTS DE VUE DIFFÉRENTS

Le fait de trouver des compléments de nature anaphorique dans des suites constitue un sérieux raccourci épargnant toute nouvelle démonstration pour justifier l'emploi qui est fait du terme qui suit la suite : ce n'est pas la personne qui est considérée en tant qu'individu, mais sa théorie. De la même façon, à travers une théorie ou un courant de pensée on fait référence à un ou plusieurs auteur(s), tenant(s) de cette approche :

- (7) «(...) ainsi, ils s'approchent du statut de mot-outils, de «morphèmes» *dans le vocabulaire fonctionnaliste*, plus particulièrement de «quantifiants» *dans le vocabulaire de Wilmet*. (Forsgren, 1997 : 121)

- (8) «**Dans l'approche cognitive**, les entités réalisées par des expressions du texte ne doivent pas être recherchées ni dans le texte ni dans le contexte physique mais dans les représentations mentales du lecteur. (Dupont, 1996 : 222)

Les marqueurs d'intérêt discursif sélectionnent donc un aspect d'une notion, ce qui sous-entend que ladite notion peut être définie et interprétée selon des points de vue différents et donc qu'une même terminologie peut recouvrir des différences conceptuelles importantes. Cela permet de penser que nous avons affaire à un phénomène de polyphonie que nous définissons à la suite de O. Ducrot (1984 : 205) en ces termes :

«Le locuteur, responsable de l'énoncé, donne existence, au moyen de celui-ci à des énonciateurs dont il organise les points de vue et les attitudes».

En effet, le locuteur se fait l'interprète, le traducteur d'une personne extérieure à son propre dire. Ajoutons, pour finir, qu'il n'est pas toujours nécessaire d'indiquer à l'auditeur l'acception d'une notion, notamment dans un contexte où les protagonistes travaillent dans le même cadre théorique (13).

Il existe deux types de polyphonie, nous commençons par la polyphonie explicite. Examinons quelques suites : après l'élément introducteur (préposition ou expression lexicalisée prépositionnelle), le complément (nom ou adjectif) est un nom de doctrine ou de discipline :

- (9) Il s'ensuit que la dénomination **dans une telle optique lexicographique** est définie comme un processus de *condensation* qui s'oppose au processus inverse d'*expansion*, caractéristique de la définition (ou signification) (...) (Kleiber, 1984 : 83)

ou le nom d'un groupe de professionnels,

- (10) Sur notre vieux continent, nous savons qu'une frontière n'est jamais totalement naturelle **au sens des géographes**. (LM 1993)

---

13) Étant entendu que son sens a été au préalable défini.

ou un nom propre ou un adjectif relationnel dérivé d'un nom propre.

- (11) «**Dans le vocabulaire kantien, régulateur** exclut donc *transcendental*». (Granger, 1976 : 378)
- (12) «**Dans la théorie de Ducrot** ces instructions sont associées en majeure partie à des connecteurs». (Nef, 1986 : 71)
- (13) «Donc, d'ores et déjà précisons que les transferts de ce type sont, **chez Bally** comme **chez Tesnière**, considérés comme de simples variantes d'un autre modèle» (...) (Noailly, 1985 : 6)

Dans les exemples (10), (11), (12), (13), le locuteur situe explicitement le cadre propositionnel de sa prédication, la référence de la source étant mentionnée, située à travers un nom : le lecteur-auditeur est renvoyé à une doctrine, à un courant de pensée. Le locuteur fait intervenir des personnes extérieures, des points de vue sont confrontés ce qui montre, nous semble-t-il, l'existence d'une polyphonie explicite.

A l'inverse, dans les exemples (9) (14) et (15)

- (14) Si les notions de contraintes séquentielles (conditions d'emploi) et de contraintes interprétatives (instructions) peuvent être conservées **dans le cadre de la pragmatique de la pertinence**, on peut se demander ce qu'il advient des notions de RE et de RI. Comme la théorie de la pertinence est une théorie de l'interprétation, une telle distinction, élaborée **dans le cadre d'une problématique du discours**, disparaît. **Dans le cadre d'une théorie de l'interprétation**, il n'est plus possible en effet de définir une notion correspondant à celle de RE ou même à celle de RI. Mais cet abandon de concepts à priori opératoires ne doit pas nous empêcher d'envisager une articulation possible entre une théorie de l'interprétation comme la pragmatique de la pertinence et une théorie du discours pour les raisons suivantes. (Moeschler, 1989 : 78)
- (15) En grammaire générative, et notamment **dans la théorie du gouvernement et du liage**, c'est dans ce sens que l'on parle de phénomène de discours. (Charolles, 1993 : 301)

on observe le phénomène inverse. C'est à partir de l'évocation de la théorie que l'on est renvoyé à un/aux théoricien(s). Dans (14), il est question de plusieurs théoriciens la discipline étant devenue plurielle

(cf. *une*), dans (15) *dans la théorie du gouvernement et du liage* renvoie à N. Chomsky. Dans ce cas, la polyphonie est implicite.

## CONCLUSION

Cette analyse, a montré que des mots fonctionnels permettent de sélectionner une acception donnée dans la «mémoire des notions» en fonction du point de vue auquel on se place. Le problème peut s'envisager dans le sens contraire, c'est-à-dire qu'à une même notion correspondent des noms différents. Dans l'exemple (16) l'auteur mentionne les différents noms sans s'aider de marqueurs alors qu'en (17), le marqueur *au sens* fournit des indications :

(16) N'importe quelle phrase P est constituée d'une *énonciation* (i.e. *l'expressivité* de Guillaume 1919, les *voix* de Bakhtine 1929, le *modus* de Bally 1932, les *embrayeurs* de Jespersen 1924, de Jakobson, de Benveniste 1960,...) et d'un *énoncé* (i.e. *l'expression* de Guillaume, le *dictum* de Bally, le *propos* de Bonnard 1981, le *message* de Goosse 1986, etc.) Soit :

P = Énonciation + Énoncé (Wilmet, 1997 : 420)

(17) L'espoir de «bases de connaissances terminologiques» n'est pas chimérique lorsque leur matière est celle d'un domaine **au sens de la terminologie**, ou d'une «ontologie» **au sens de l'IA**. (Lerat, 1997 : 244)

Il importe de souligner que, dans une réflexion, on est toujours redevable à quelqu'un : on peut accepter son point de vue ou le refuser, mais notre réflexion s'appuie immanquablement sur des travaux antérieurs. Les mouvements de sens que l'on peut observer soulignent des points sensibles et indiquent que le sens se construit en fonction du chercheur, de l'époque, de l'avancement d'une théorie (14).

---

14) Je remercie toutes les personnes qui ont lu une première version de ce texte.

## RÉFÉRENCES

- ARIEL, M. (1988) : «Referring and Accessibility», *Journal of Linguistics*, vol. 24, pp. 65-87.
- DUCROT, Oswald (1984) : *Le dire et le dit*, coll. «Propositions», Paris, Éditions de Minuit, 233 p.
- FALL, Khadiyatoulah, SIMEONI, Daniel et Georges VIGNAUX (dir) (1994) : *Mots, représentations - Enjeux dans les contacts interethniques et interculturels*, Ottawa, Presses Universitaires d'Ottawa, 418 p.
- FÓNAGY, Ivan (1982) : *Situation et signification*, Amsterdam, John Benjamins, 161 p.
- GRÜNIG, Blanche-Noëlle (1993) : «Charges mémorielles et prédictions syntaxiques», *Cahiers de grammaire*, 18, pp. 13-29.
- GUIMIER, Claude (1996) : *Les adverbes du français - Le cas de l'adverbe en -ment*, coll. «L'essentiel français», Paris, Ophrys, 170 p.
- INaLF (Institut National de la Langue Française) (1995) : *Les concepts opérationnels de l'INaLF*, Document interne non publié, INaLF, 116 p.
- JOLY, André (1997) : «Autour d'un concept d'"opération" chez Gustave Guillaume», G. Kleiber, M. Riegel (dir) : *Les formes du sens - Études de linguistique française, médiévale et générale offertes à Robert Martin à l'occasion de ses 60 ans*, coll. «Champs linguistiques», Paris Éditions Duculot, pp. 203-218.
- JONASSON, Kerstin (1994) : *Le nom propre - Constructions et interprétations*, coll. «Champs linguistiques», Paris, Duculot, 256 p.
- KLEIBER, Georges (1996) :
- LAPAIRE, Jean-Rémi et Wilfrid ROTGE (1991) : *Linguistique et grammaire de l'anglais*, 2<sup>e</sup> édition revue et corrigée, Toulouse, Presses Universitaires du Mirail, 734 p.
- MARTIN, Robert (1992) : *Pour une logique du sens*, 2<sup>e</sup> édition revue et augmentée, coll. «Linguistique nouvelle», Paris, PUF, 319 p.
- NØLKE, Henning (1994) : *Linguistique modulaire : de la forme et du sens*, Louvain, Editions Peeters, 303 p.

PILLON, Agnessa (1993) : *La mémoire des mots - Ses unités, son organisation*, coll. «Psychologie et sciences humaines», Liège, Mardaga, 228 p.

REMI-GIRAUD, Sylvianne et Pierre RETAT (dir) 1996 : *Les mots de la nation*, Lyon, Presses Universitaires de Lyon, 321 p.

### Références des exemples

BERRENDONNER, Alain (1990) : «Pour une macro-syntaxe», *Travaux de linguistique* 21, pp. 25-35.

CARON, Jean (1983) : *Les régulateurs du discours - Psycholinguistique et pragmatique du langage*, coll. «Psychologie d'aujourd'hui 102», Paris, PUF, 255 p.

CHAROLLES, Michel (1993) : «Les plans d'organisation du discours et leurs interactions», Moirand *et al.*, *Parcours linguistiques de discours spécialisés - Colloque en Sorbonne les 23-24-25 septembre 1992*, coll. «Sciences pour la communication 41», Berne, Peter Lang, pp. 301-314.

DUPONT, Michel (1996) : «Le modèle des attentes du lecteur», *ILN* 96, pp. 221-229.

FORSGREN, Mats (1997) : «Un classique revisité : la place de l'adjectif épithète», G. Kleiber, M. Riegel (dir) : *Les formes du sens - Études de linguistique française, médiévale et générale offertes à Robert Martin à l'occasion de ses 60 ans*, coll. «Champs linguistiques», Paris Éditions Duculot, pp. 115-126.

FUCHS, Catherine (1996) : *Les ambiguïtés du français*, coll. «L'essentiel français», Paris, Ophrys, 184 p.

GRANGER, Gilles 1976 : «Syntaxe, sémantique, pragmatique», *Revue internationale de philosophie* 30, 3 / 4 (117-118), pp. 376-410.

KLEIBER, Georges (1984) : «Dénomination et relations dénominatives», *Langages* 76, pp. 77-94.

LERAT, Pierre (1997) : «La notion de connaissance est-elle un concept opérationnel en linguistique ?», G. Kleiber, M. Riegel (dir) : *Les formes du sens - Études de linguistique française, médiévale et générale offertes à Robert Martin à l'occasion de ses 60 ans*, coll. «Champs linguistiques», Paris Éditions Duculot, pp. 241-248.

LM (LE MONDE)

- MOESCHLER, Jacques (1986) : «Connecteurs pragmatiques, lois de discours et stratégies interprétatives : *parce que* et la justification énonciative», 3<sup>e</sup> Colloque de pragmatique, *Cahiers de linguistique française* 7, pp. 149-167.
- MOESCHLER, Jacques (1989) : «La problématique des règles d'enchaînement et d'interprétation revisitée», Ch. Rubattel, *Modèles du discours : recherches actuelles en Suisse romande*, Actes des Rencontres de linguistique française Crêt-Bérard, 1988, coll. «Sciences pour la communication 26», Berne, Peter Lang, pp. 61- 84.
- NEF, Frédéric (1986) : «Sémantique discursive et argumentation», 3<sup>e</sup> Colloque de pragmatique, *Cahiers de linguistique française* 7, pp. 69-92.
- NOAILLY, Michèle (1985) : *De l'adjectif à l'adjectivation*, Thèse, 449 p.
- PERRIN, L. (1994) : «Mots et énoncés mentionnés dans le discours», *Cahiers de linguistique française* 15, pp. 217-247.
- WILMET, Marc (1997) : *Grammaire critique du français*, coll. «Hachette Supérieur», Paris, Hachette, 670 p.

# Lexique et mémoire

**Pierre LERAT**

*LLI, Université Paris 13, Villetanneuse,  
France*

La part de la mémoire dans l'acquisition, la mobilisation et la perte des mots intéresse la psychologie et pourrait intéresser davantage la linguistique. Les psychologues distinguent une mémoire à long terme, qui est à la fois la mémoire des concepts et celle de leurs supports, les mots, et la mémoire à court terme, ou mémoire de travail, qui correspond au temps de la perception et de l'interaction. Il me semble que la lexicologie théorique a un parti à tirer de cette distinction pour comprendre le jeu du sens lexical et de la référence discursive en termes de coopération entre les deux formes de mémoire. Quelques suggestions à ce propos vont être faites ici, à l'occasion d'un bref panorama des concepts clés de la lexicologie générale telle que je la vois. Il va s'agir successivement de la compositionnalité, de la référentialité, de l'intensionnalité, de l'extensionnalité, de la représentabilité et de la normativité. Le choix de mots abstraits, pour chacune des six notions cruciales, veut dire qu'il s'agit de principes applicables aux unités lexicales quelle que soit la langue. L'ambition est d'esquisser la confrontation de deux logiques : celle du système et celle de l'usage réel.

## 1. — COMPOSITIONNALITÉ ET MÉMOIRE

Le principe de compositionnalité est l'application aux phrases, aux groupes de mots et aux mots du principe de Frege. De façon générale, «une construction donnée est dite compositionnelle quand

on peut déduire son sens de celui de ses éléments composants reliés par une relation syntaxique» (G. Gross, 1996 : 154).

En matière de morphologie lexicale, les règles ne sont applicables qu'à des morphèmes mémorisés. L'exemple classique de néologisme virtuel, celui d'*indécorable* chez Saussure, illustre suffisamment un mécanisme où le mot construit «évoque le terme dont il se compose et d'autres qui lui sont associés» (Saussure, 1978 : 181). Ce que l'auteur appelle ici terme est un morphème lexical, à partir duquel le mot est interprétable. Cette base lexicale peut être un sigle, comme dans le récent *SGMLiser*, ou une racine empruntée, comme dans les mots acadiens *framage*, *déloader* et *cleaner* (voir Péronnet, à paraître). Saussure parle en pareil cas de motivation morphologique, ce qui est correct mais insuffisant. Il n'y a bonne formation lexicale dans ces divers exemples que parce qu'il y a application de règle, et non pas seulement enchaînement de morphèmes. C'est l'apport du principe de compositionnalité de fournir un principe explicatif. Ainsi, *indécorable* suppose d'abord dérivation suffixale puis dérivation préfixale (voir notamment Guilbert, 1975).

Une autre illustration du principe de compositionnalité en matière de morphologie lexicale peut être tirée de nomenclatures mixtes, mi lexicales mi numériques. *Article 50* comporte un numéro d'ordre, comme *Louis XVI*, *dose léthale 50* une indication de pourcentage, comme *aspirine 500*; si l'on ne connaît pas la règle de codage des connaissances, la concaténation n'est pas éclairante par elle-même.

Les langues spécialisées abondent aussi en composés syntagmatiques au sens de Guilbert. Des unités lexicales acadiennes usuelles chez les pêcheurs, telles que *winch à bras* et *vent light*, sont compositionnelles selon la syntaxe du français, comme le montre l'ordre des mots.

L'ennemi de la compositionnalité n'est pas l'emprunt, car elle s'applique aussi à la perception et à la compréhension de mots étrangers, comme le montre l'anglicisation des langues romanes (voir Lerat 1995 : 65). C'est bien plutôt le figement, c'est-à-dire la mémorisation de séquences. Deux exemples suffiront à montrer que le figement est affaire de mémoire à long terme. Le premier est celui de *maladie de Parkinson*, composé qui s'annexe un nom propre, c'est-à-dire les connaissances encyclopédiques afférentes; notons au passage qu'il s'agit d'une description définie dégénérée, lexicalisée comme nom commun, mais je reviendrai sur ce phénomène qui est de

grande ampleur dans les langues spécialisées. L'autre cas est celui de *l'appel du 18 juin*, expression connotée à laquelle renverrait implicitement l'évocation, dans le discours, d'un appel du 19 juin, par exemple.

## 2. — RÉFÉRENTIALITÉ ET MÉMOIRE

Le vocabulaire référentiel, en dehors des indexicaux, qui ne sont que des auxiliaires de la référenciation, et des noms propres, qui ont une vocation particulière à référer, est fait de noms communs concrets qui servent à désigner des objets particuliers, avec ou sans déterminants selon les langues. Une remarque de Searle (1972 : 66) fournit un critère sûr : les expressions référentielles «répondent aux questions *qui ?*, *que ?*, *lequel ?*». Une autre observation du même auteur explique pourquoi la référence n'est pas une propriété sémantique, mais une propriété logique : l'identification. «La référence, dit-il, est un acte de langage, et les actes de langage sont accomplis non par des mots, mais par des locuteurs qui prononcent des mots» (Searle, 1972 : 64). La référence est donc tout entière dans le temps de la mémoire de travail, qui est celui de l'énonciateur.

Une bonne illustration de cette particularité des expressions référentielles est fournie par le mode d'existence langagier des descriptions définies. On sait qu'une description définie est «une expressions qui pointe sur un objet individuel déterminé en le décrivant comme l'objet qui a telle et telle propriété» (Allwood *et al.*, 1989 : 152). Elle est définie en ce sens qu'elle identifie un référent, mais c'est une description pour autant qu'elle utilise un vocabulaire qui «décrit un aspect de cet objet» (Searle, 1972 : 216). Ainsi, en emploi spécifique, *le Milanais* pourra désigner selon les univers de discours aussi bien Stendhal qu'un joueur de football ou que n'importe quelle personne de sexe masculin connue à la fois de l'énonciateur et du récepteur.

La description définie participe des deux mémoires, et à mon avis cette spécificité correspond au double mécanisme analysé par les logiciens et, depuis Kleiber, par les linguistes. L'identification, qu'elle soit déictique ou anaphorique, relève d'une temporalité brève, celle de la mémoire de travail; la description, au contraire, fait appel à la mémoire à long terme, où sont stockés à la fois ce que Sperber et Wilson appellent les entrées encyclopédiques des concepts et les entrées lexicales qui leur correspondent dans une langue donnée

(Sperber et Wilson 1989 : 86). Tel est le cas des descriptions définies dans le discours.

Il importe d'en distinguer les expressions définies lexicalisées, qui donnent naissance à des noms propres comme *Le Havre*, ainsi que les «noms propres dégénérés» au sens de Searle (1972 : 227). Il s'agit d'expressions telles que *La Banque d'Angleterre*, comme on en rencontre tout particulièrement dans la terminologie des institutions. Leur statut bâtard embarrasse lexicographes et terminographes, car ce ne sont pas vraiment des noms propres comme les toponymes et les patronymes, mais on refuse aussi de considérer qu'ils appartiennent à la langue au sens des linguistes. Nomenclature, disent les semi-habitués, pour se débarrasser d'un vocabulaire atypique, vraiment peu saussurien et peu chomskyen. Soit, mais le stockage dans la mémoire à long terme le qualifie pour l'encyclopédie.

### 3. — INTENSIONNALITÉ ET MÉMOIRE

Alors que la référence est une propriété logique d'expressions essentiellement nominales, l'intension est le contenu sémantique de concepts véhiculés par des substantifs, des verbes et des adjectifs. Pour que la valeur de vérité des propositions soit calculable, les logiciens ont besoin de considérer l'intension des concepts comme stable, ce qui n'est pas entièrement conforme au résultat des enquêtes des psychologues ni aux observations des linguistes, comme celles de Martin sur la relativité du concept de jeunesse (1992 : 26). Comme les expressions prédicatives se caractérisent par leur malléabilité en fonction de leurs arguments, les logiciens illustrent plus volontiers l'intension par des exemples de noms concrets; de fait, il est plausible de dire que l'intension du mot *arbre*, nom d'un objet naturel de grande évidence empirique, bien connu et rassurant, est «le concept créé par un acte d'abstraction concernant les propriétés que tous les arbres ont en commun» (Allwood *et al.*, 1989 : 126, n.1). Ce qui est constant pour des mots comme *jeune*, *arbre* ou *rêver*, c'est-à-dire pour la très grande majorité des mots, c'est l'existence d'une intension; l'intension est relative, mon concept d'arbre est influencé par la familiarité avec certains arbres plutôt que d'autres, et ma conception de la jeunesse est plus floue que quand j'avais vingt ans, mais d'une part l'intensionnalité est le lot commun d'«arbre» et de «jeune», d'autre part la communication prouve empiriquement que les mots ont un sens et la définition permet d'explicitier des traits dont une partie fait l'objet d'un consensus large.

L'intension étant ce qui légitime ou invalide les inférences, on peut la confondre non seulement avec le concept logique mais aussi avec le stéréotype social (voir Kleiber 1990 : 68). Celui-ci n'a aucune pertinence dans le cas des verbes supports comme *faire* ou *porter*, il en a une maximale dans les terminologies techniques où il se confond avec des conditions nécessaires et suffisantes, dans l'intervalle il y a du plus ou moins plausible (un lion craintif, un corbeau de bon augure, un vêtement de deuil fluo, une sonate techno etc.). On n'avancera qu'en traitant des types de mots de façon systématique, non en mêlant l'abstrait et le concret, le descriptif et l'idéologique etc.

L'intension est aussi à la base de la métaphore. Certes, le stimulus de la métaphore vive est nécessairement présent dans la mémoire de travail, mais ce qui est activé par ce stimulus dans la mémoire à long terme est une propriété saillante du référent partagée par un concept lexicalisé. Le cliché est une métaphore dégénérée parce qu'il est lexicalisé, comme la description définie dégénérée; la métaphore vive résulte au contraire d'un choc perceptuel ou émotionnel : bien métaphoriser, c'est voir le semblable, comme disait déjà Aristote.

#### 4. — EXTENSIONNALITÉ ET MÉMOIRE

L'extension, ensemble des objets logiques auxquels s'applique normalement un mot, est une propriété réservée aux classes majeures de mots, celles qui ont une intension, c'est-à-dire les noms, les verbes et les adjectifs. De fait, «une intension est quelque chose qui relie une expression linguistique à son extension. Elle détermine l'extension d'une expression linguistique» (Allwood *et al.*, 1989 : 128). C'est l'extension qui, dans un grand magasin, guide le consommateur vers les escalopes de veau à partir du fléchage du rayon boucherie; la signalétique est extensionnelle : elle ne remplace pas la connaissance mais au contraire la suppose.

Une expérience permet d'illustrer le rapport de l'extension avec la mémoire. Soit la description définie imaginaire *l'invention du glotton*. L'acte de référence présuppose l'existence de quelque chose qui soit un glotton. Le principe de compositionnalité permet des hypothèses, mais pas de certitude tant que je ne suis pas familier du concept de glotton; si mon sentiment linguistique m'oriente vers *glotte*, il peut s'agir de médecine, de pharmacie ou de phonétique, mais on peut aussi penser à *glotto-*, qui connote la linguistique; dans chaque cas, le mot nouveau ne peut s'analyser que sur fond de connaissances lexicales, mémorisées dans la mémoire à long terme.

## 5. — REPRÉSENTABILITÉ ET MÉMOIRE

Si l'on met à part la représentabilité des formes, qui relève principalement de la première articulation, niveau des transcodages et des translittérations, les unités lexicales se prêtent à une double représentabilité : iconique, quand il s'agit de leurs référents ; interconceptuelle, quand il s'agit des contenus.

La représentabilité iconique est inégale selon les parties du discours et même au sein des substantifs, puisque seuls des humains, des concrets et des abstraits privilégiés font l'objet de figurations dans les encyclopédies. Pasteur y devient un portrait chez Larousse, *pasta* un assortiment chez Zingarelli, le tennis un schéma de court dans un dictionnaire d'Oxford, l'archéologie une photo de fouille gallo-romaine chez Hachette : tout ce qui doit être visualisé a besoin d'être réduit à du concret. Ce concret est sélectionné en fonction de propriétés logiques culturellement saillantes : un visage de savant exemplaire, des aliments caractéristiques, une application typique, un lieu consacré (voir Lerat 1994).

L'exemple des illustrations dans les encyclopédies montre que les choix ne portent pas sur des référents de signes linguistiques mais sur des connaissances attachées à des mots. De même, la représentation des connaissances attachées aux mots en intelligence artificielle et aux termes en terminologie ne relève pas de champs lexicaux mais de champs conceptuels. Il ne s'agit pas seulement, pour autant, de figurer uniquement des connaissances sur les choses, mais bien de rendre compte de relations entre des concepts lexicalisés qui peuvent prendre la forme de verbes, d'adjectifs et de noms de toutes sortes. Ce sont effectivement des mots qui sont en jeu, non des morphèmes reliables entre eux uniquement en termes de signifiés saussuriens, c'est-à-dire de valeurs purement distinctives. Les connaissances lexicales classiques, comme celle des synonymes et des hyponymes d'un mot, ne sont pas étrangères à l'expérience des choses et des gens. Bien au contraire, les liens pertinents sont les plus fonctionnels au vu de la mémoire à long terme associée aux mots.

## 6. — NORMATIVITÉ ET MÉMOIRE

La dénomination, imposition sociale d'un nom pour désigner conventionnellement au moyen de substantifs, de verbes et d'adjectifs (voir Lerat 1995 : 49), est un cas typique de norme linguistique. La banalité du phénomène a été un peu occultée par des décennies de mise entre parenthèses de la référence, en dehors des travaux de

Benveniste, de Kleiber et de quelques autres, si bien que c'est plutôt chez les philosophes que l'on trouve des rappels au bon sens comme celui-ci : «manifestement, le fait que le mot *dog* signifie "chien" ou désigne des chiens a des conséquences normatives. L'une d'entre elles est qu'il existe des circonstances dans lesquelles il est approprié ou correct d'utiliser le mot *dog*» (Jacob, 1997 : 187).

La notion d'appropriation convient aussi aux collocations lexicales. Un chien courant aboie, il donne de la voix, il mène. Un humain parle, crie, poursuit. De même, la considération de domaines en terminologie permet jusqu'à un certain point de filtrer les équivalences plausibles en langue étrangère. C'est que le terme est la dénomination de connaissances spécialisées (voir Lerat 1995 : 45). Relève également de normes le recours au niveau de base plutôt qu'aux autres niveaux en matière de catégorisation (voir Kleiber 1990) dans la communication ordinaire, probablement aussi l'adéquation plus grande du bas niveau chez le spécialiste. La valeur par défaut des descriptions définies dégénérées est elle aussi affaire de norme : la nuit de la Saint Barthélémy évoque définitivement un massacre survenu en 1572, ce qui rend décodable l'allusion dans un syntagme comme *une nouvelle Saint Barthélémy*.

## CONCLUSION

1) La compositionnalité n'exclut nullement la mémorisation, qu'il s'agisse des morphèmes, des quasi morphèmes (comme les sigles) ou des composés.

2) La référence n'est possible que par la coopération de la mémoire à court terme (dont les objets constituent l'univers du discours) et de la mémoire à long terme (où sont stockées les connaissances lexicales et encyclopédiques).

3) Le sens lexical est fait d'un jeu de connaissances encyclopédiques assez stéréotypé pour autoriser des inférences. Il relève donc de la mémoire à long terme.

4) L'extension est prévisible pour autant que les propriétés intensionnelles sont des conditions nécessaires et suffisantes sélectionnées parmi d'autres qui sont également mémorisées.

5) La représentabilité est une propriété des unités lexicales qui fait entrer en jeu des connaissances encyclopédiques saillantes.

6) La normativité concerne tout particulièrement les dénominations et les collocations. En ce sens, les classes d'objets relèvent autant de la norme que du système.

## RÉFÉRENCES

- ALLWOOD, Jens, ANDERSSON, Lars-Gunnar et Östen DAHL (1989) : *Logic in linguistics*, Cambridge University Press.
- GROSS, Gaston (1996) : *Les expressions figées en français*, Paris, Ophrys.
- GUILBERT, Louis (1975) : *La créativité lexicale*, Paris, Larousse.
- JACOB, Pierre (1997) : *Pourquoi les choses ont-elles un sens ?*, Paris, Odile Jacob.
- KLEIBER, Georges (1990) : *La sémantique du prototype*, Paris, PUF.
- LERAT, Pierre (1994) : «Vocabulaire et connaissances. À propos d'"archéologie", "pasta" et "tennis"», *Studi Italiani di Linguistica Teorica e Applicata*, XXIII-3, pp. 501-508.
- LERAT, Pierre (1995) : *Les langues spécialisées*, Paris, PUF; trad. esp. Albert Ribas, *Las lenguas especializadas*, Barcelona, Ariel, 1997.
- MARTIN, Robert (1992) : *Pour une logique du sens* (1983), Paris, PUF.
- PERONNET, Louise (à paraître) : «Étude d'un type particulier d'emprunts utilisés en Acadie, les emprunts de nature hybride (français / anglais)», 5ème colloque international «Français du Canada - français de France», Bellême, 3 juin 1997.
- SAUSSURE, Ferdinand de (1978) : *Cours de linguistique générale* (1916), T. de Mauro ed., Paris, Payot.
- SEARLE, John R. (1969) : *Speech Acts*, Cambridge University Press, trad. fr. *Les actes de langage*, Paris, Hermann, 1972.
- SPERBER, Dan et Deirdre WILSON (1986) : *Relevance*, Blackwell, Oxford; trad. fr. *La pertinence*, Paris, Éditions de Minuit, 1989.

# Génération automatique de mots arabes à partir de racines et de schèmes

Hussein HABAILI et Slim BEN HAZEZ

*Universités Tunis I, et II  
& Laboratoire RIADI, Tunis, Tunisie*

## 1. — INTRODUCTION

Parmi les caractéristiques des langues chamito-sémitiques, telles que la forme à préfixe de la morphologie verbale, la prévalence des consonnes sur les voyelles (1); l'existence des consonnes pharyngales /h/, /ʔ/ et des emphatiques, la vraie caractéristique reste la notion de *racine* et de *schème*.

Ce système à double paramètre a été déjà mis en lumière par Jean Cantineau dans son article «Racines et schèmes» (2), étude dans laquelle il a revendiqué la réalité linguistique et de la racine et du schème (3). Ainsi, les rapports entre les racines et les schèmes dans les

---

1) Selon les statistiques d'Ullendorff le rapport consonnes-voyelles est de cinq à quatre. «Qu'est-ce-qu'une langue sémitique», *Orientalia*, 27, 1958, pp. 66-75; (Communication au *Congrès des Orientalistes* de 1957).

2) Jean Cantineau, «Racines et schèmes», *Mélanges William Marçais*, 1950, pp. 119-124.

3) Racine et schème sont deux entités formelles discontinues, la première constituée par une succession d'éléments phoniques dont la nature, le nombre et l'ordre sont constants, la seconde par une sorte de 'moule' de forme également constante, mais admettant comme ses éléments constitutifs n'importe quelle suite ordonnée de phonèmes qui définit une racine. Ainsi les trois consonnes (k. t. b.), dans cet ordre, sont une racine qui peut entrer dans divers schèmes (qu'elle relie donc en une

langues chamito-sémitiques peuvent être illustrés à l'aide de la citation de Jean Cantineau relative à l'arabe, une des langues la plus représentative du groupe sémitique :

«Chaque mot a sa racine et son schème; on pourrait comparer le vocabulaire à un tissu dont la trame serait l'ensemble des racines attestées dans la langue et la chaîne l'ensemble des schèmes existants. Chaque point d'intersection de la chaîne et de la trame serait un mot, car tout mot est entièrement défini sans ambiguïté par sa racine et son schème, tout schème de son côté fournissant des mots à différentes racines et la plupart des racines fournissant des mots de différents schèmes». (4)

## 2. — RACINES ET SCHÈMES

Cette organisation en racine et en schème qui est à la base de la dérivation lexicale est soumise à des contraintes tant au niveau des racines qu'au niveau des schèmes. En effet, la plupart des langues sémitiques et même chamito-sémitiques présentent des limitations quant à la nature et à la quantité des schèmes et des racines. On compte cent cinquante schèmes pour le sémitique, on constate une prédominance des racines trilitères sur les bilitères et les quadrilitères (80% de racines trilitères pour l'arabe, 65% pour l'égyptien, 52% pour le berbère, etc.). Il y a donc restriction sur la forme des schèmes et sur les consonnes qui constituent la racine.

Outre ces restrictions sur la forme des schèmes et sur la structure de la racine, la racine trilitère des langues sémitiques connaît un autre type de restriction qui régit la cooccurrence des consonnes au sein de cette racine. En effet, un examen attentif des racines trilitères révèle le phénomène de l'incompatibilité entre consonnes apparentées phonétiquement. D'une façon générale, on définit la

---

'famille' sémantique), tandis que la structure  $ma+R_1R_2uR_3$  forme un schème dans lequel  $R_1$ ,  $R_2$  et  $R_3$  représentent les phonèmes radicaux de n'importe quelle racine triconsonantique.

Soit alors  $(r_1, r_2, r_3, \dots, r_n)$  la suite des racines connues et  $(s_1, s_2, s_3, \dots, s_n)$  la suite des schèmes possibles dans la langue, les diverses combinaisons  $(rs)$  définiraient la totalité du lexique potentiel de l'arabe dont le lexique réel ne constitue qu'une partie. Un *mot arabe* n'est donc abstraction faite des marques flexionnelles dont il est obligatoirement muni, que l'intersection en un point de la série des racines et de la série de schèmes.

4) Jean Cantineau, «La notion de schème et son altération dans les langues sémitiques», *Sémitica*, 1950, p. 74.

racine sémitique comme une suite de trois consonnes différentes. Et en étudiant les consonnes incompatibles, on observerait que les langues sémitiques ne se servent que d'un très petit nombre de combinaisons possibles. À titre d'exemples, nous pouvons citer des cas présentés dans le cadre des conférences du GLECS <sup>(5)</sup> notamment celui «du verbe amharique, cité par Marcel Cohen, et dont la forme à préfixe /an-/ : (causatif) ou /tan-/ : (réfléchi) n'a jamais une première consonne radicale liquide. L'accadien tend à différencier les trois consonnes de la racine en ce qui concerne l'emphase et la sonorité. L'hébreu et l'araméen ont gardé l'empreinte de la loi de l'incompatibilité jusque dans la morphologie (...)» <sup>(6)</sup>.

En étudiant ces incompatibilités dans deux langues sémitiques, l'hébreu et l'araméen, A.Z. Ascoly arrive à trois distinctions :

1. «Distinction entre l'incompatibilité des consonnes quelles que soient leurs voyelles, et entre celle qui dépend plutôt de la voyelle dont est pourvue la consonne (...);
2. Distinction entre les positions des consonnes dites incompatibles, si elles forment les deux premières radicales d'une racine trilitère ou ses deux dernières radicales (...);
3. L'incompatibilité de certaines consonnes dépend de leur position réciproque (...)» <sup>(7)</sup>.

Les grammairiens arabes ont remarqué que certaines consonnes sont incompatibles. Maxime Rodinson souligne que «d'après le *Muzhîr* de Djalâl Ad-dîn As-suyutî les racines à trois consonnes radicales semblables sont très rares. Les consonnes contenant à la fois les consonnes /b/, /q/, /m/ ou /n/, /r/, /z/, /n/ n'apparaissent que dans des mots empruntés» <sup>(8)</sup>.

L'essentiel de la question a été noté par deux philologues arabes de la fin du IV<sup>e</sup> siècle de l'Hégire (X<sup>e</sup> siècle de l'ère chrétienne) : Ibn Farîs (dans *As-sâhibî*, éd. du Caire, 1910, pp. 75-88) et Ibn Jinnî (dans *Al - Hasâ'is*, éd. du Caire 1913, pp. 53-54). Dans son livre : *Sirr-Sinâ'at Al - I'râb*, Ibn Jinnî traite dans la conclusion de cet

5) Séances du 22 février, 24 mai 1939, et du 28 avril 1948.

6) A. Z. Ascoly, «Incompatibilité des phonèmes en hébreu et en araméen», *Comptes rendus du GLECS*, t. III, 1939, pp. 54-56.

7) A.Z. Ascoly, *Op. Cit.*, p. 55.

8) Maxime Rodinson *et al.*, «Entretien sur les incompatibilités de phonèmes dans les racines», *Comptes-rendus du GLECS*, t. III, 1939, p. 49-50.

ouvrage de la théorie de la racine. Il distingue certains facteurs phonétiques qui limitent les possibilités de combinaison des «*hurûf*» (lettres).

Cette théorie se base sur un principe qui consiste à éviter la succession, dans un même mot, de «*hurûf*» ayant des points d'articulation communs ou voisins. Ibn Jinnî précise que la meilleure combinaison est celle où les «*hurûf*» ont [des points d'articulation] loins les uns des autres. Ibn Jinnî établit un lien qui unit restrictions séquentielles et phonétique articulatoire : «lorsque le son est acheminé vers le point d'articulation d'un '*hurûf*' (lettre) et utilisé pour la prononciation de celui-ci et qu'on veut ensuite le tirer vers une articulation, il est plus normal de le diriger vers un point plus éloigné pour qu'il ait une différence entre les deux échos [ainsi produits]»<sup>9</sup>. Il faut préciser que ce principe se trouve lié d'autre part à ce que Ibn Jinnî appelle «*quwwat al-hurûf*» (la vigueur des lettres), c'est-à-dire leur force ou énergie articulatoire. Certains «*hurûf*» sont en effet, selon Ibn Jinnî plus «vigoureux» que d'autres.

C'est à partir de ces trois principes qui régissent la distribution des «*hurûf*» au sein de la racine trilitère de l'arabe, qu'Ibn Jinnî établit une sorte de hiérarchie en fonction de la capacité des «*hurûf*» à se suivre les uns des autres. Ce qui l'amène à conclure que «les incompatibilités sont uniquement la conséquence des points d'articulation», et que le «mot idéal» est celui qui est constitué de «*hurûf*» ayant des points d'articulation éloignés les uns des autres, c'est aussi celui dont les éléments constitutifs ont des caractéristiques phonétiques similaires». Ce faisant, Ibn Jinnî distingue trois catégories de *mots*, à savoir :

1. «Ceux dont les éléments proviennent de points d'articulation éloignés les autres;
2. Ceux comportant deux lettres identiques;
3. Ceux dont les éléments ont des points d'articulation voisins».

La catégorie la meilleure est la première, la dernière est défectueuse et il n'existe d'ailleurs que de rares spécimens» (10).

Dans son «Esquisse d'une phonologie de l'arabe classique», Jean Cantineau a formulé certains principes généraux qui régissent

---

9) Abdelkader Mehiri, *Les théories grammaticales d'Ibn Jinnî*, Tunis, Publication de l'Université de Tunis, 1973, p. 205.

10) *Idem*.

l'incompatibilité entre consonnes d'une même racine. Selon Cantineau :

«La plupart des incompatibilités portent sur la localisation de deux consonnes voisines. On peut poser en principe que les deux premières consonnes de la racine ne peuvent être ni semblables, ni de même classe de localisation, ni même souvent de classes de localisation voisine; quant aux deux dernières consonnes de la racine elles peuvent être semblables, mais si elles ne le sont pas, elles ne peuvent pas appartenir à la même classe de localisation voisine» (11).

Après avoir examiné l'ensemble des règles de Cantineau concernant les différentes séries de localisation, nous avons constaté que ces règles (12) sont assez vagues, car elles ne mentionnent pas d'une façon précise les séquences admises et celles qui sont exclues. D'autre part, ces règles n'expliquent pas les exceptions, telles les racines :  $\text{Jbm}$ , ?  $\text{bm}$ ,  $\text{lfm}$  et  $\text{lqm}$ . En plus, elles ne tiennent pas compte de la distribution des consonnes en position 1, 3 de la racine trilitère de l'arabe.

Dans l'étude de Joseph Greenberg «*The Patterning of Root Morphemes in Semitic*» (13), nous avons rencontré un type de règles assez semblables à celles de Jean Cantineau. Cependant, les règles de Joseph Greenberg qui régissent les occurrences de consonnes à l'intérieur du morphème-racine de l'arabe classique sont formulées d'une façon un peu plus générale. Elles peuvent être résumées de la façon suivante :

«Règle 1 : En position 1, 2, non seulement les consonnes identiques sont exclues, mais aussi les consonnes homorganiques non identiques.

Règle 2 : Les consonnes homorganiques non identiques sont aussi exclues en position 2, 3, mais pas avec la même rigueur que dans la position 1, 2. Cependant cette règle n'exclut pas les consonnes identiques en position 2, 3.

---

11) Jean Cantineau, «Esquisse d'une phonologie de l'arabe classique», *Bulletin de la Société Linguistique de Paris (BLS)*, t. 43, 1946, pp. 165-204.

12) A titre d'exemple nous citons la règle de Cantineau relative à la série labiale :  
«Il y a en général incompatibilité entre les phonèmes /b/, /f/, /m/, aussi bien dans la position 1, 2 que dans la position 2, 3; les exceptions sont fort rares : on peut tout au plus citer les racines :  $\text{Sbm}$ ,  $\text{çbm}$ ,  $\text{lfm}$  (à côté de  $\text{lqm}$ )».

13) Joseph Greenberg, «The Patterning of Root Morphemes in Semitic», *Word*, vol. 6, 1950, pp. 162-181.

Règle 3 : En position 1, 3, il y a une exclusion marquée mais non rigoureuse des consonnes identiques, et homorganiques non identiques.»

Après avoir révisé les travaux antérieurs, nous avons constaté que ni l'étude de Cantineau, ni celle de Greenberg ne rendent compte d'une façon précise et générale des contraintes de structure morphématique de la racine de l'arabe classique. Les deux études donnent l'impression qu'il s'agit plutôt de tendances qu'ont certaines consonnes de l'arabe à ne pas se combiner entre elles. Les deux auteurs se contentent de mentionner les exceptions sans pour cela arriver à expliquer ces dernières.

L'apport fondamental de notre recherche, réside dans la solution que nous avons apportée aux problèmes rencontrés par Cantineau et Greenberg à savoir celui des exceptions, des généralisations et de la formulation en termes de conditions de structure morphématique et de traits binaires. En effet, c'est dans les rapports entre les lieux d'articulation et les modes articulatoires que réside la solution à ces problèmes (14).

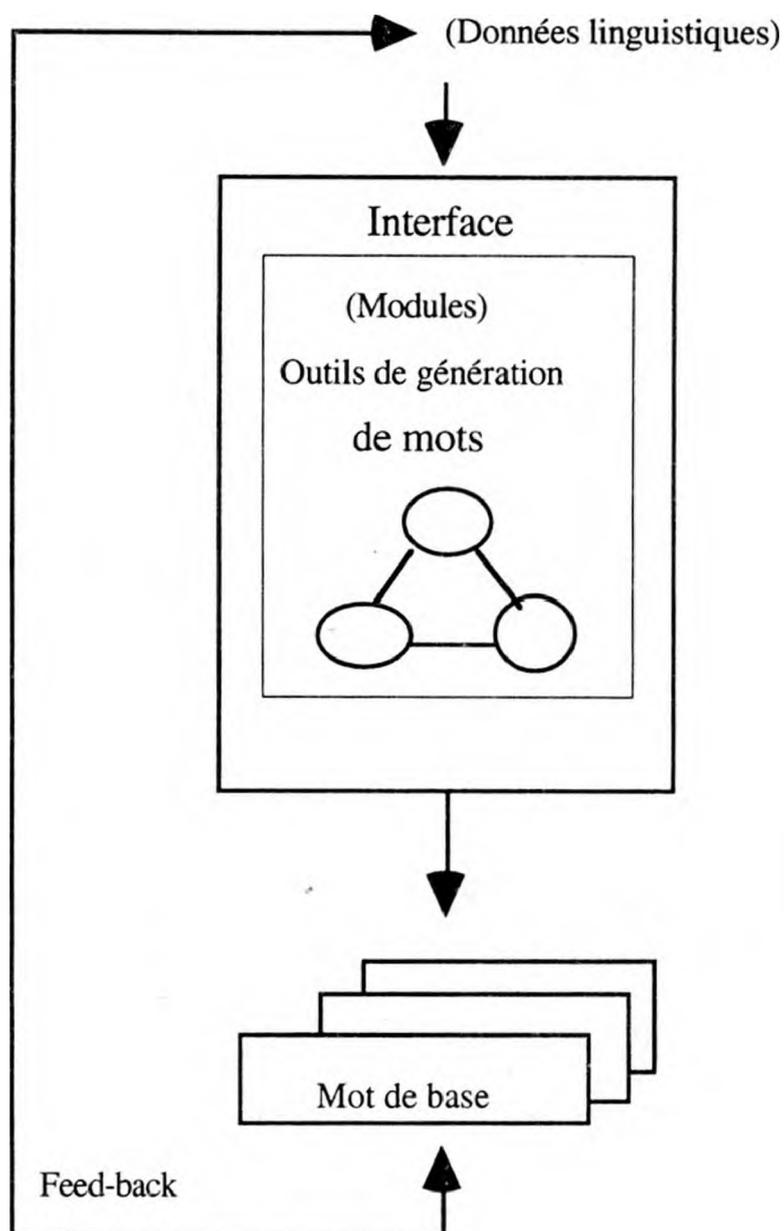
### 3. — GÉNÉRATION AUTOMATIQUE DE MOTS DE BASE

La conception du mécanisme de génération se base sur trois principes fondamentaux : *la racine, l'ordre des consonnes et le schème*. La langue arabe est caractérisée par la possibilité de création de mots par analogie selon des schèmes fixes. Ce procédé de dérivation est donc une voie très importante pour l'enrichissement du lexique arabe.

L'outil à développer doit être interactif et permet d'exécuter plusieurs scénarios de génération. Un scénario est un processus qui permet de créer automatiquement une liste de mots à partir de deux items lexicaux (une racine et un schème). Un *feed-back* s'avère nécessaire pour permettre à l'utilisateur de modifier ou d'ajouter certains paramètres à l'input pour affiner les résultats :

---

14) Pour de plus amples détails voir Hussein Habaili, «Principes directeurs pour l'élaboration d'une banque des morphèmes-racines de l'arabe standard», *Actes des troisièmes Journées scientifiques de l'AUPELF-UREF*, Montréal, Université de Montréal, A. Clas et P. Bouillon (dir.), TA-TAO : recherches de pointe et applications immédiates, coll. «Universités francophones», Beyrouth, AUPELF-UREF-FMA, pp. 457-476, 1993.



**Figure 1 : Système d'aide à la génération de mots de base**

Si on prend l'exemple du terme «ordinateur», «calculatrice», le scénario qui consiste à trouver l'équivalent de ce terme est constitué des entrées et des sorties suivantes :

— Contexte en entrée : <Racine = /hs b/; Catégorie = Nom d'instrument>

— Mots générés en sortie : /haasuub, hassaabat, haasib, haasibat/.

Le choix des schèmes est effectué à partir de la catégorie morphologique donnée. Dans ce cas, on peut choisir les schèmes exprimant un nom d'instrument : /faa uul, fa aalat, faa il, faa ilat/.

Pour généraliser le système selon cette démarche, il est possible de définir un mécanisme qui permet de choisir une racine (attestée ou admissible) et un ou plusieurs schèmes. Pour ce faire plusieurs processus ou modules peuvent être intégrés dans le système : classification sémantique des racines et des schèmes...

Pour satisfaire nos objectifs, il est intéressant de définir une architecture *modulaire* et *extensible* du système. Pour concevoir un tel système, il s'avère, nécessaire d'organiser de façon modulaire les connaissances linguistiques fournies afin de concevoir un ensemble de *modules* indépendants et autonomes de dérivation de mots. À partir des données initiales, ces modules doivent coopérer pour engendrer une liste de mots de base candidats. Ces modules doivent être aussi réutilisables dans plusieurs applications de traitement automatique de la langue arabe (traitement de texte, correction orthographique, traitement de la parole, traduction assistée par ordinateur, analyse morpho-syntaxique etc.).

Cette architecture générale du système d'aide à la dérivation de mots de base, s'intègre d'une façon naturelle, dans le cadre de la composante morphologique de l'arabe standard. Une telle étude de l'aspect génératif et dérivationnel du système morphologique arabe s'avère nécessaire pour analyser les processus de génération et de dérivation de mots de base de la langue arabe ainsi que la structure interne du mot arabe.

### 3.1 — Structure du mot en arabe

Généralement, on considère que les éléments «atomiques» ou de base pour la formation des mots de la langue arabe sont : la *racine*, le *schème*, ou le *mot de base* lui-même. Pour avoir une classification des mots adaptée pour le traitement automatique, nous distinguons deux classes initiales de *mots de base simples* (mot minimal sans ajout de préfixes et suffixes) :

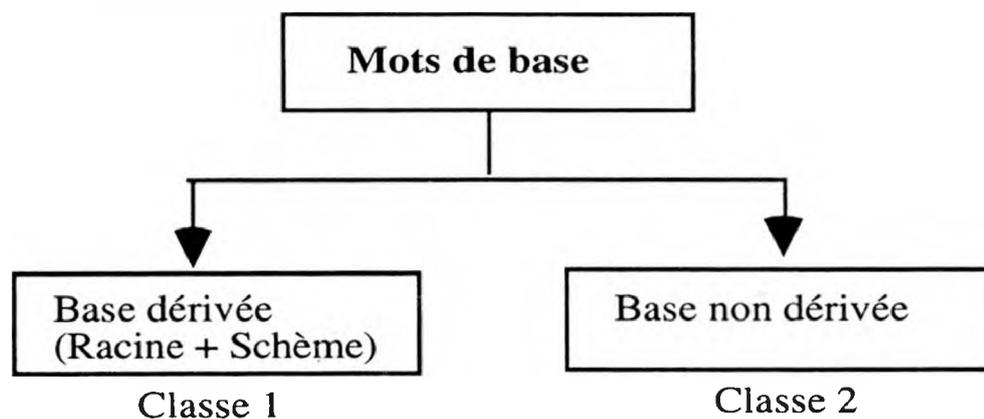


Figure 2 : Les classes initiales de mots de base

— **La classe 1** : concerne les bases représentables en racines et schèmes (base = racine + schème). Il s'agit des mots assimilés par le système morphologique dérivationnel de la langue arabe. À partir d'un schème et d'une racine on peut engendrer une base. Cette base peut être accompagnée d'un préfixe et d'un suffixe qui peut être vide et donc constituer un mot de base du lexique (Ex. : «télescope» : /mirqaab/, dérivé du schème /mif aal/ et de la racine /r q b/. Ce processus constitue un premier niveau de dérivation qui permet d'engendrer de nouveaux mots.

— **La classe 2** : regroupe les mots de base non dérivés comme /fiizja/ : «physique». Il y a plusieurs catégories de mots non dérivés : des substantifs non représentables en racines et schèmes, des emprunts, des noms propres, des particules, etc. Un second processus de dérivation de mots nouveaux consiste en un jeu d'affixation, de composition de mots à partir de mots de base.

Un *mot de base* est soit une base dérivée à partir d'une racine et d'un schème soit un *mot non dérivé*. Partant de cette hypothèse, nous considérons qu'un mot dérivé simple est composé de trois segments : [préfixe(s) + mot de base + suffixe(s)]. La figure ci-dessous décrit la structure interne d'un mot simple de l'arabe standard et montre les deux principaux aspects du traitement automatique du mot arabe : *la génération et l'analyse* :

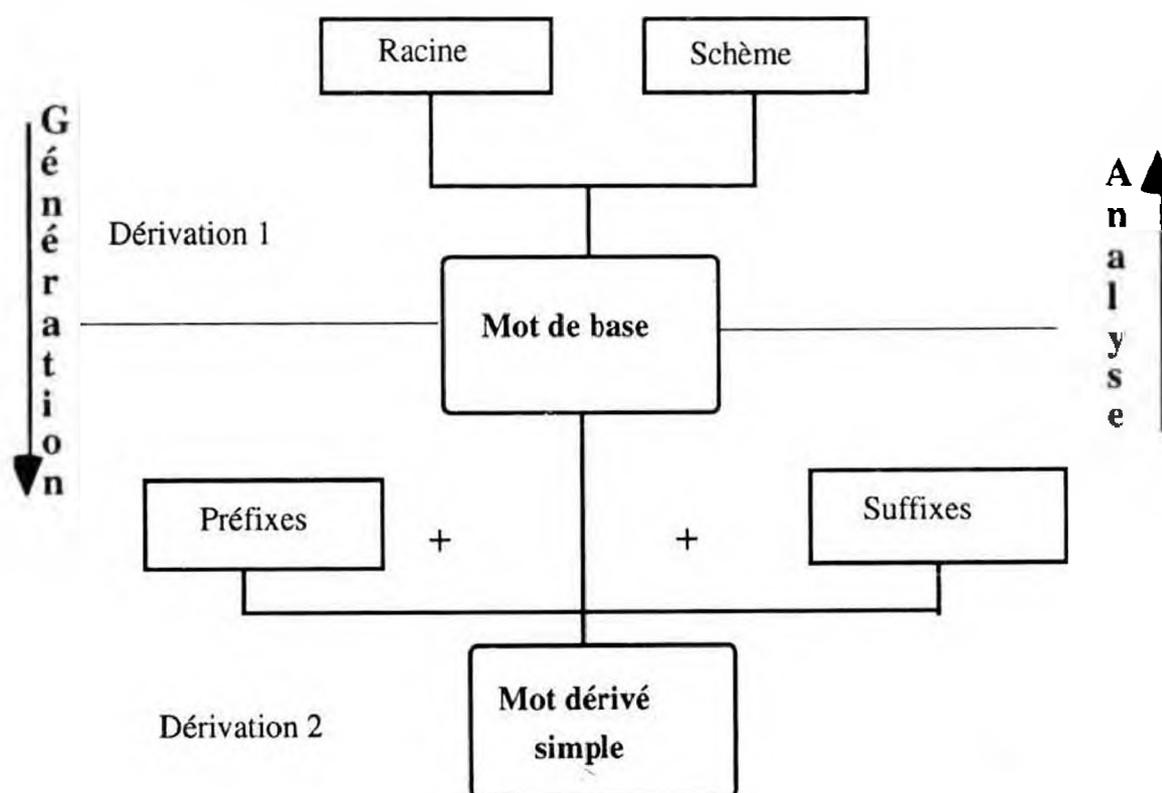


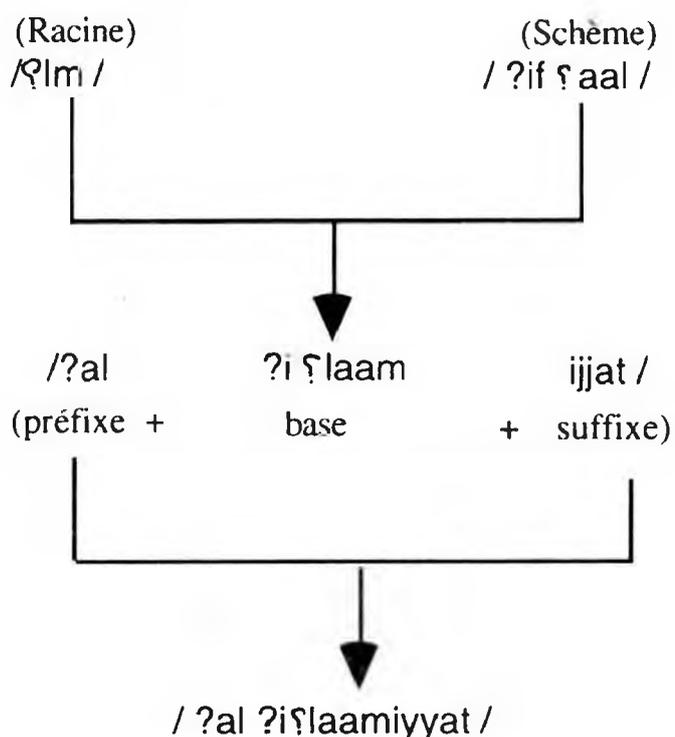
Figure 3 : Structure morphologique du mot arabe

On distingue donc deux niveau de dérivation :

— **Dérivation 1** : dérivation de mots de base (bases dérivées) par association d'une racine et d'un schème;

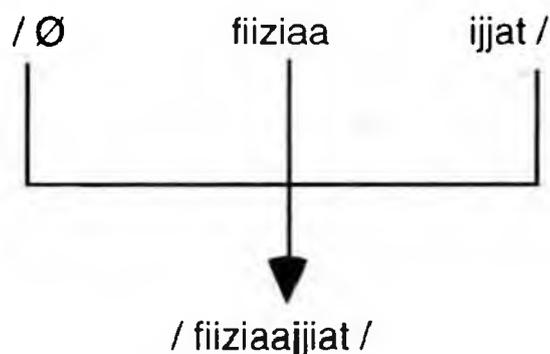
— **Dérivation 2** : dérivation de mots nouveaux par affixation aux mots de base.

La structure interne du nom abstrait /ʔal ʔiʕlaamyja/ : «informatique» dérivé de la racine /lm/ et du schème /ʔif ʕaal/ est la suivante :



**Figure 4 : Structure morphologique du mot : /ʔal ʔiʕlaamiyyat/**

La structure du mot : /fiiziaaʔijjat/ «physique» dérivé à partir d'un mot non dérivé est représenté comme suit :



**Figure 5 : Structure du mot /fiiziaaʔijjat/**

Dans cette conception de la structure du mot arabe, il faut tenir compte de plusieurs relations de compatibilité entre les différents constituants du mot :

— Les relations de compatibilité et d'incompatibilité entre les consonnes d'une même racine (voir plus haut les contraintes de structure morphématique);

— Les relations de compatibilité entre préfixes et suffixes d'un mot dérivé;

— Les relations de compatibilité de la base (racine / schème) avec les couples (préfixes, suffixes);

— La relation de compatibilité entre la racine et le schème : cette relation représente la relation la plus importante pour la formation des mots de la langue arabe. En effet, n'importe quelle racine ne peut pas être instanciée à n'importe quel schème. L'association d'une racine et d'un schème doit obéir à des contraintes phonétiques, syntaxiques et sémantiques.

### 3.2 — Structure d'un mot composé

Un mot composé de la langue arabe est formé de deux ou plusieurs mots simples possédant leur propre structure interne. Les éléments de composition peuvent être des mots dérivés nouveaux ou des mots déjà existants dans le lexique. La langue arabe peut former, par exemple, des mots composés à l'aide de deux mots simples de la façon suivante :

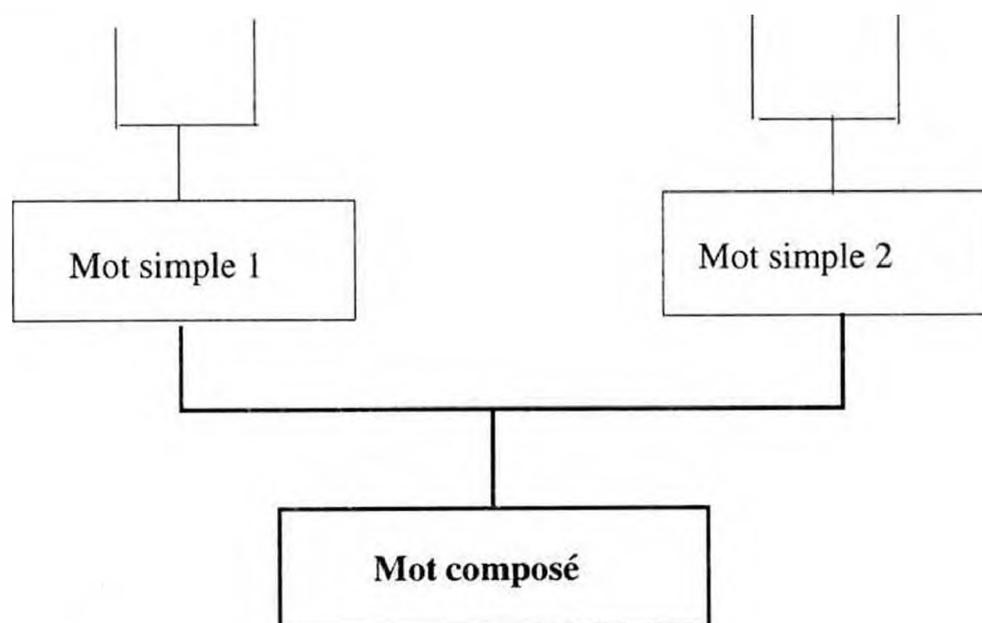


Figure 6 : Structure d'un mot composé de deux mots simples

Un mot composé peut être créé par l'adjonction à droite ou à gauche d'un mot bref ou d'une particule. C'est la manière dont la langue arabe prend en compte les préfixes de composition étrangers (*hyper, sous, auto, non*). Chaque préfixe peut être exprimé en arabe par un ou plusieurs mots brefs (Ex. : la particule /laa/ peut correspondre aux préfixes français /in-, a-, non-, an-/). Par exemple, le mot : /laa mihwarii/, «anaxial» est composé de la particule /laa/ qui désigne le préfixe /an-/ et de l'adjectif relatif /mihwarii/.

D'autres mots peuvent être obtenus par juxtaposition de deux mots simples (adjectifs relatifs, noms abstraits, substantifs...). Ex. : le mot «téléinformatique» est rendu par : /ʔal maʕluumaatijja ʔal bu dijja/; le mot «facsimili» par : /ʔal ʔistinsaax ʔal buʕdii/.

#### 4. — CLASSIFICATION DES SCHEMES DE LA LANGUE ARABE

Les bases nominales et les noms d'action (*Masdar*) constituent la base de formation du vocabulaire technique et scientifique. Dans le contexte de la néologie lexicale, une classification des schèmes est nécessaire pour la génération automatique des mots selon des catégories de schèmes fixes ou spécialisés. Nous présentons dans cette section une étude sur les schèmes nominaux arabes de formation de mots.

En arabe standard, les noms sont connus comme des substantifs et des adjectifs, la figure ci-dessous illustre les différentes catégories nominales des noms dérivés :

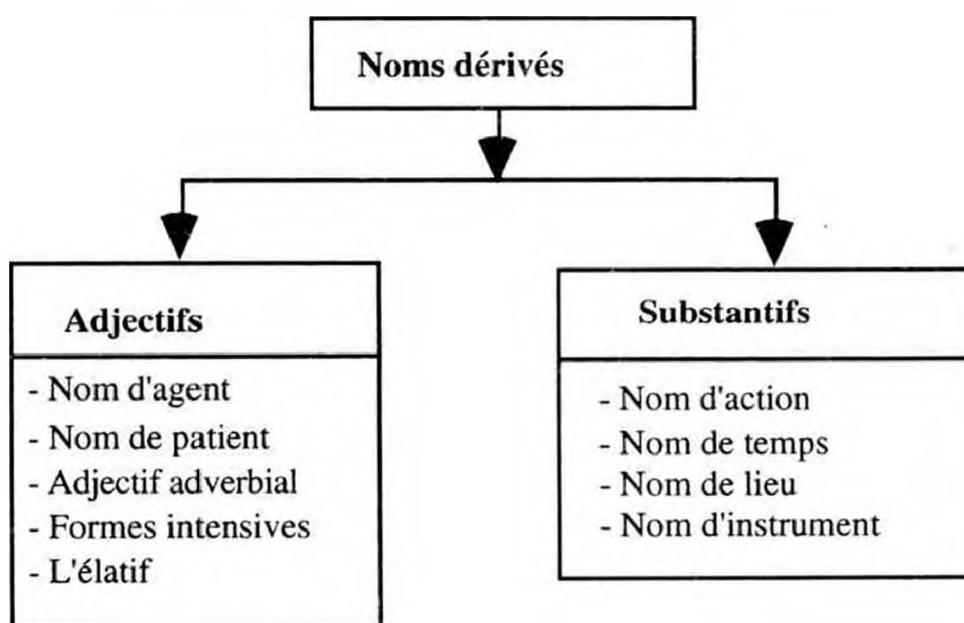


Figure 7 : Les catégories des formes nominales

#### 4.1 — Nom d'action (*Masdar*)

La langue arabe moderne utilise massivement des noms d'action des formes dérivées des verbes dans la terminologie technique et scientifique. La figure ci-dessous représente les différentes classes de noms d'action :

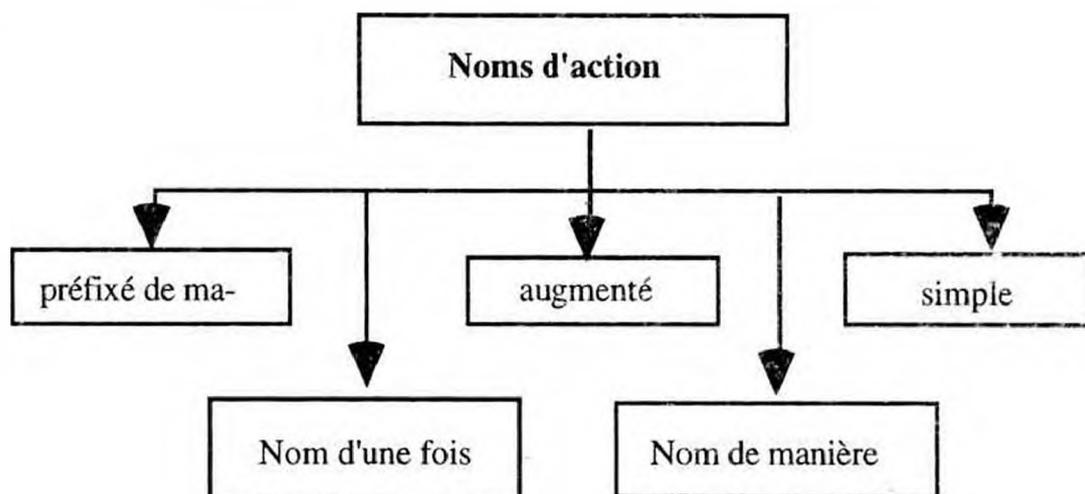


Figure 8 : Les classes de noms d'action (*Masdar*)

Parmi les *formes* les plus utilisés dans la formation de nouveaux termes, nous pouvons citer les schèmes des noms d'action simples :

Schème	Signifié (sens)	Exemples	Glose
/fiʕaalat/	Nom de métier	/zihaarat/	floriculture
/fuʕaal/	Cri, maladie	/Iumaad/	acidose
/faʕiil/	Cri, bruit	/zafiir/	expiration
/faʕlaan/	Mouvement	/nabadaan/	pulsation
/faʕaal/	Nom de métier	/zahaar/	fleuriste
/taf ʕaal/	Intensité	/taImaad/	hyperacidité

#### 4.2 — Noms d'instrument

Les noms d'instrument sont construits sur le type /mif ʕal/ parfois /mif ʕaal/ ou /mif ʕalat/ :

/mif ʕal/	/mirqab/	«télescope»
/mif ʕalat/	/mi3marat/	«creuset»
/mif ʕaal/	/minsaax/	«pantographe»
/fa ʕaal(a)t/	/nassaafat/	«torpilleur»
/faaʕilat/	/haasirat/	«fiche électrique»
/faaʕuul/	/haasuub/	«ordinateur»

### 4.3 — Noms de temps et de lieu

Ces dérivés expriment l'endroit ou le moment où une action a eu lieu et parfois les deux ensemble. Les noms de lieu sont généralement du type :

/maf ʕal/	/ma3maʕ/	«académie»
/maf ʕil/	/ma3lis/	«conseil»
/mif ʕalat/	/mintaqat/	«zone, région»

### 4.4 — Formes intensives

Nous allons citer seulement les formes les plus utilisées :

Les schèmes : /faʕil, faʕuul, fuʕula, faʕal(a)t, fiʕiil, mif ʕaal(a)t, mif ʕiil, mif ʕal, fuʕaal, faa ʕuul, faaʕilat, faʕuulat/ expriment l'intensité.

— Le schème /fa ʕaalat/ est utilisé dans la formation des noms d'instruments et il s'applique aux moyens de locomotion moderne : /darraa3at/ «bicyclette». Le même schème peut former certains noms d'objets à certaine intensité d'action : /3ammaaʕat/ : «accumulateur»; /sammaaʕat/ «écouteur».

### 4.5 — Élatif

C'est un aspect de l'adjectif qui exprime une valeur supérieure. Il recouvre le comparatif et le superlatif. Par exemple : /ʔakbaru/ «plus grand».

### 4.6 — Adjectif adverbial

Le schème : /ʔaf ʕalu/ est utilisé dans la formation des adjectifs indiquant les couleurs ou les particularités physiques : /ʔasmaru/ «brun».

Le schème /faʕlaan/ (féminin : /faʕlaa/, /ʔadbaan/ : «en colère». Les autres formes sont du type : /fuʕlaan, faʕl, fuʕl, faʕal, faʕaal, fuʕaal, faʕiil/, etc.

### 4.7 — Noms d'agent

Le nom d'agent (ou participe actif) est également utilisé dans la langue arabe moderne. Il peut prendre plusieurs formes :

— À la forme simple (I<sup>re</sup> Forme) : le schème est /faaʕil/ : /daaʔin/ : «créancier»; /raahin/ : «constituant ou débiteur».

— Aux formes dérivées :

II<sup>e</sup> Forme :        /mufaʕil/    /muʔarriʕ/ : «législateur»

Ce schème peut désigner encore les noms d'instrument :  
/murakkim/ : «accumulateur», /muka θ θ if/ «condensateur».

III <sup>e</sup> Forme :	/mufaaʕil/	/muqaawil/	«entrepreneur»
IV <sup>e</sup> Forme :	/muf ʕil/	/muʕrim/	«assassin»
V <sup>e</sup> Forme :	/mutafaʕil/	/mutasallim/	«accipens»
VI <sup>e</sup> Forme :	/mutafaaʕil/	/mutakaamil/	«isobare»
VII <sup>e</sup> Forme :	/munfaʕil/	/munʕakis/	«anatrope»
VIII <sup>e</sup> Forme :	/muftaʕil/	/multazim/	«engagé»
X <sup>e</sup> Forme :	/mustaf ʕil/	/mustaʕmir/	«colonisateur»

#### 4.8 — Noms de patient

Les noms de patient (ou formes passives) sont fréquemment employés (souvent au pluriel avec /+aat/ comme substantifs inanimés). Ils peuvent prendre plusieurs formes :

— À la *forme simple* dite «nue» ou (I<sup>re</sup> Forme) : schème /maf ʕuul/ :

/maʕduud/	«calculable»
/makruuh/	«détestable»
/maqbuul/	«acceptable»

— Aux *formes dérivées* :

II <sup>e</sup> Forme :	/mufaʕal/	/musabbaʕ/	«heptagone»
III <sup>e</sup> Forme :	/mufaaʕal/	/mufaaʕal/	«réacteur»
IV <sup>e</sup> Forme :	/muf ʕal/	/muʕʕam/	«dictionnaire»
V <sup>e</sup> Forme :	/mutafaʕal/	/mutabaddal/	«variable»
VI <sup>e</sup> Forme :	/mutafaaʕal/	/mutataabaʕ/	«séquentiel»
VII <sup>e</sup> Forme :	/munfaʕal /	/munʕataf/	«méandre»
VIII <sup>e</sup> Forme :	/muftaʕal/	/muhtamal/	«probable»
X <sup>e</sup> Forme :	/mustaf ʕal/	/mustanqaʕ/	«marécage»

## 5. — CONCEPTION DU SYSTÈME

À partir de l'étude des différents types de processus de formation de mots nouveaux et des caractéristiques phonologiques, morphologiques (structure interne), lexicales, syntaxiques et sémantiques des mots arabes, nous allons définir le principe de fonctionnement du système d'aide à la génération automatique de mots arabes et élaborer une architecture basée sur les principales caractéristiques de modularité des connaissances linguistiques traitées.

### 5.1 — Présentation générale du système

Généralement, les paramètres du système à l'entrée (input) sont représentés par un ensemble de données lexicales, morpho-syntaxiques ou/et sémantiques selon lesquelles doivent être formés le ou les mots nouveaux.

Ces données sont représentées par le contexte suivant : <Racine - schème / mot de base, catégorie grammaticale>. La génération peut se faire soit à partir d'une racine et d'un schème, soit à partir d'un mot de base. Le terme à construire doit avoir la même catégorie grammaticale donnée. La catégorie grammaticale peut constituer en fait une donnée sémantique puisqu'elle peut représenter le signifié d'un schème à choisir (nom d'instrument, forme d'intensité, nom d'agent, nom d'action...). Elle permet de délimiter l'ensemble des mots possibles qu'on peut générer à partir d'une racine et d'un schème.

Le système proposé est composé d'un *générateur* permettant d'engendrer une liste de mots nouveaux pour plusieurs catégories grammaticales (substantifs, adjectifs, noms d'action, noms abstrait...) à partir des *données linguistiques* fournies au système. Il permet d'exploiter un ensemble de connaissances linguistiques représentées par un ensemble de règles de formation de mots (RFM) <sup>(15)</sup> et de règles de transformation (RT) <sup>(16)</sup>, ainsi que de *procédures* permettant de mettre à jour et d'enrichir le lexique arabe par des mots et des néologismes.

---

15) Hussein Habaili et Mohamed Ben Ahmed, «Génération automatique de néologismes arabes à partir des règles de formation de mots», *Lexicomatique et dictionnaires*, Actes des 4es Journées Scientifiques du réseau LTT, Lyon, Université Lyon-2, A. Clas, Ph. Thoiron et H. Béjoint (dir.), coll. «Universités francophones», Paris/Beyrouth AUPELF-UREF-FMA, pp. 173-188, 1995.

16) Slim Ben Hazez, *Système d'aide à la génération automatique de néologismes arabes*, Mémoire de D.E.A. en Informatique, Université de Tunis II, École Nationale des Sciences de l'Informatique (ENSI), Tunis, 1996.

Le générateur est relié à un lexique (nous pouvons dénommer ce concept par un terme plus général *hyperlexique*) contenant : une *base des préfixes, des suffixes*, une *banque des racines*<sup>(17)</sup> : (dictionnaire des racines attestées, dictionnaire des racines admissibles), une *base de schèmes* regroupés par catégories, un *dictionnaire de mots de base attestés* avec des informations grammaticales associés aux mots.

Un *feed-back* peut être utilisé permettant à l'utilisateur de modifier ou d'ajouter certains paramètres à l'input pour affiner les résultats. L'acquisition des données, l'affichage des résultats, ainsi que les modifications des entrées sont réalisés de façon interactive avec le système :

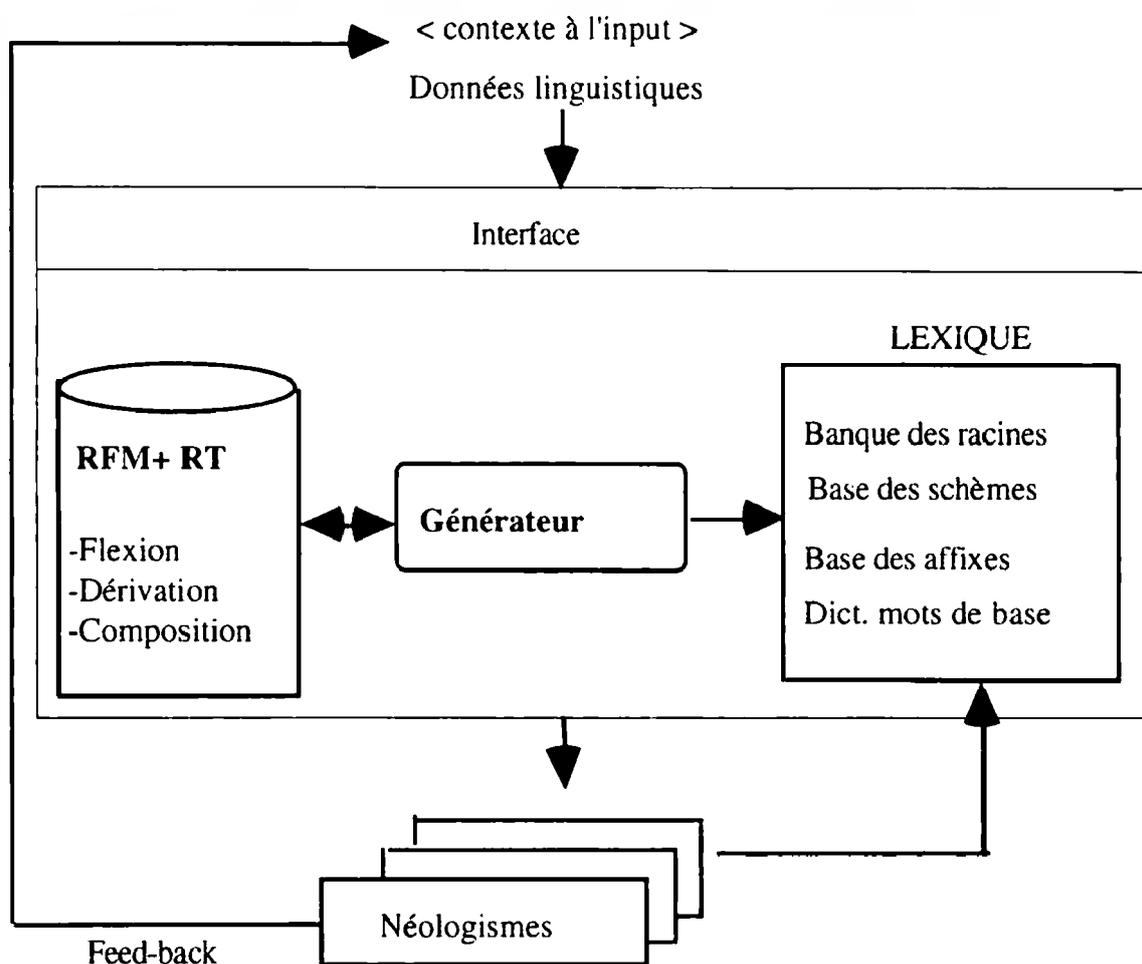


Figure 9 : Système d'aide à la génération de mots arabes

17) Hussein Habaili, «Principes directeurs pour l'élaboration d'une banque des morphèmes-racines de l'arabe standard» *TA-TAO : recherches de pointe et applications immédiates*, Actes des troisièmes Journées scientifiques du réseau LTT, Montréal, Université de Montréal, A. Clas et P. Bouillon (dir.), coll. «Universités francophones», Beyrouth, AUPEF-UREF-FMA, pp. 457-476, 1993.

Pour trouver l'équivalent du terme «calculatrice», il est possible d'effectuer plusieurs scénarios de génération. Dans chaque scénario on peut spécifier une racine différente qui porte une signification de la notion qu'on veut dénommer : /h s b/, /r t b/, /n m/. Chaque scénario permet d'engendrer une liste de mots dérivés de la racine et la catégorie donnée :

Scénario 1 : Données linguistiques : <Racine = /n m/ >  
<Catégorie = Nom d'instrument >

Exemple de mot généré : /naas aamat/

Scénario 2 : Données linguistiques : <Racine = /r t b/ >  
<Catégorie = Nom d'instrument >

Exemple de mot généré : /rattaabat/

Scénario 3 : Données linguistiques : <Racine = /h s b/ >  
<Catégorie = Nom d'instrument >

Exemples de mots générés : /hassaabat, haasibat, haasuub.../

Dans certains cas, il est possible à travers un *feed-back* de filtrer l'ensemble des mots générés et d'affiner la solution. Dans le cas du troisième scénario, il est possible, par exemple, de spécifier la propriété «intensité» (désignant la notion d'intensité de calcul) comme un trait sémantique caractérisant le terme nouveau. Puisque le schème /faa uul/ désigne une forme d'intensité, alors il est possible de sélectionner le mot /haasuub/.

Comme second exemple, nous pouvons désigner le terme «informatique», on peut donner un mot de base en entrée :

Contexte en entrée : <mot de base = /maʕluumaat/ >  
<Catégorie = nom abstrait >

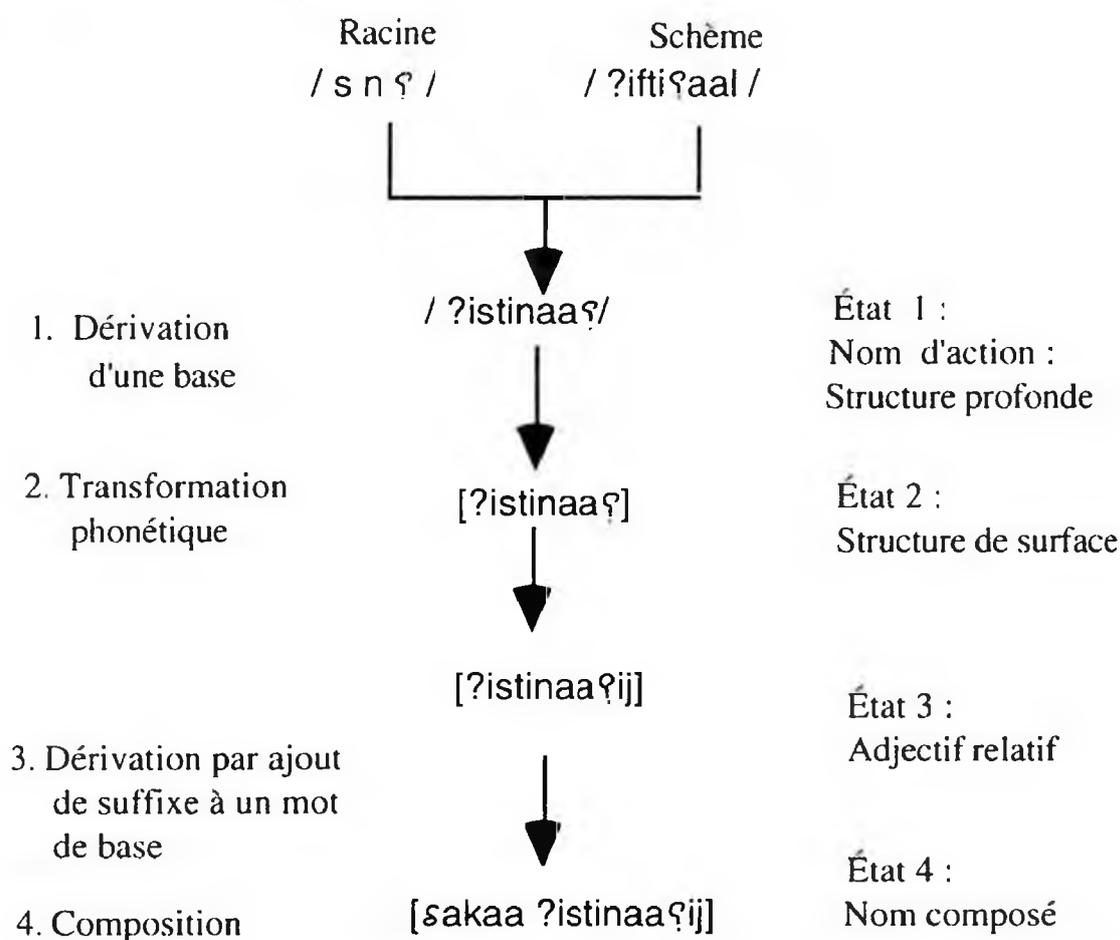
Néologismes générés : /ʔiʕlaamijjat, ʔiʕlaamijjaat/ /maʕluumaatijjat, maʕluumaatijjaat/

## 5.2 — Architecture générale du système

Généralement, l'ensemble des connaissances du système est constitué par les données initiales d'entrée, l'ensemble des règles lexicales de formation de mots et le lexique avec toutes ses composantes. On peut distinguer deux types de connaissances : des connaissances déclaratives (règles de formation de mots et de

transformation) et des connaissances procédurales (procédures de gestion du lexique). Nous rappelons que l'ensemble des règles qui régissent la génération de nouveaux mots sont classées par type : flexion, dérivation, composition, transformation morpho-phonologique, transcription orthographique.

Le problème de génération des mots est décomposable. En effet, la génération d'un mot peut être effectuée en plusieurs étapes. Chaque étape peut être réalisée en appliquant un ensemble de règles ordonnées de la base des connaissances. À titre d'exemple, la génération du terme /s akaa ?istinaa?ii/ : «intelligence artificielle» est réalisée selon les étapes suivantes :



**Figure 10 : Les différentes étapes de génération d'un mot**

L'architecture adoptée est composée d'un ensemble de *modules* dont chacun est spécialisé dans une étape particulière du processus de génération. Un module peut prendre la forme d'un système à base de connaissances classique, avec sa propre mémoire de travail, ou celle d'une procédure écrite dans un langage de programmation. Ces

modules manipulent une base de faits globale et sont activés par un générateur :

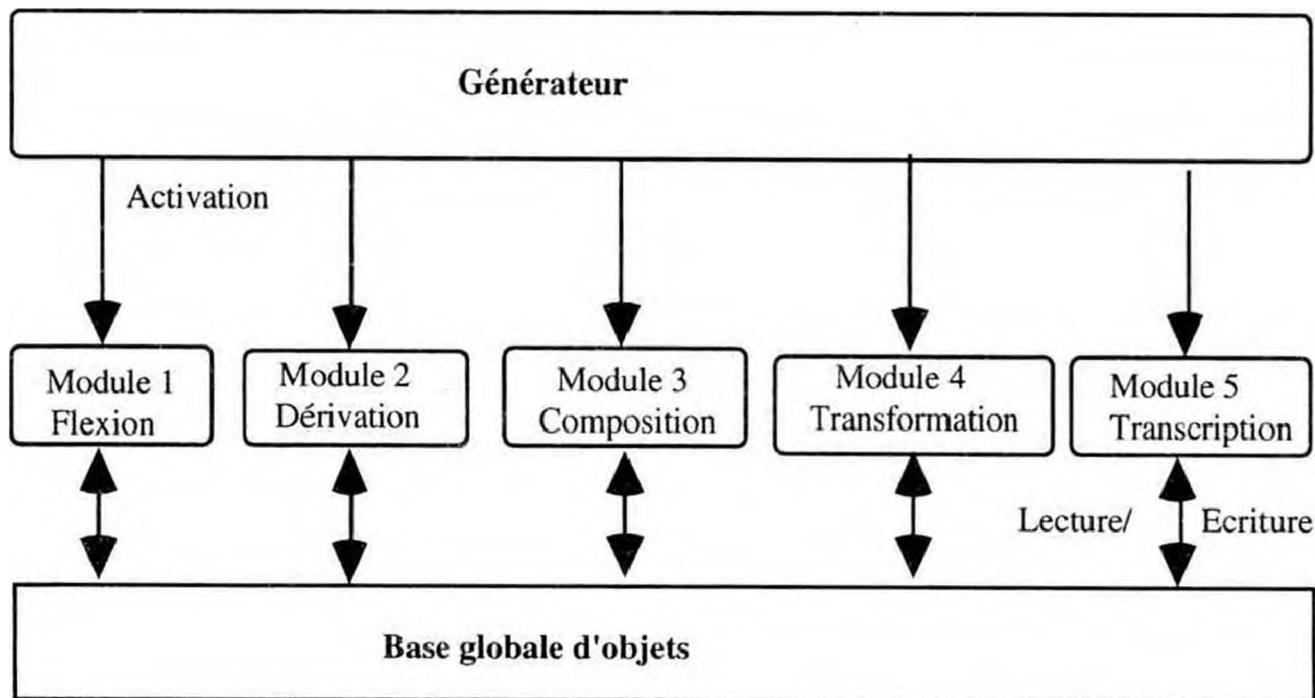


Figure 11 : Architecture générale du système

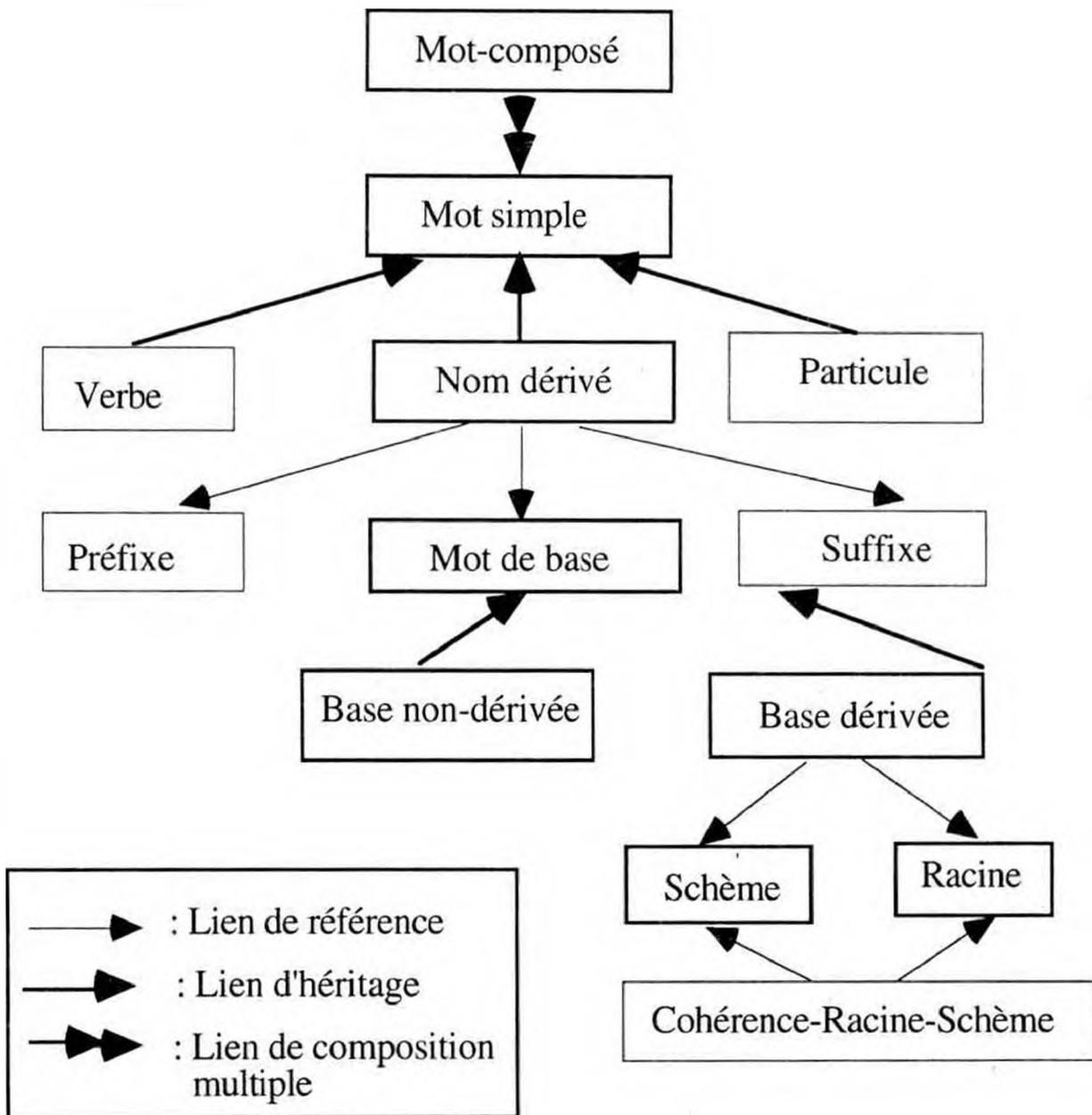
### 5.2.1 — Structure de la base globale d'objets

La base d'objets est une zone d'échange commune à tous les modules. Tous les items lexicaux (racine, schème, base dérivée, mot composé...) engendrés par les modules sont stockés dans la base de faits. Elle est composée d'objets en liaison avec les mots de base et les néologismes à construire. Ces objets représentent des données linguistiques initiales, des solutions partielles alternatives ou des solutions finales.

Nous avons essayé par l'utilisation de l'approche objet, de concevoir des objets linguistiques réutilisables et extensibles. La conception des objets exploités par les différents modules du système permet de fournir une organisation modulaire de la base d'objets. Ce formalisme permet aussi une représentation optimale des objets manipulés dans la base globale.

Pour définir la structure de la base de faits, nous avons suivi le formalisme de la méthode O\* (Brunet, 1991). Cette dernière est une méthode *orientée objet* pour la conception de systèmes d'information. Elle fournit deux structures décrivant *l'aspect statique et l'aspect*

*dynamique*. Elle repose sur un modèle combinant les avantages des modèles orientés objets et des modèles sémantiques. Le modèle conceptuel O\* décrit des phénomènes réels en terme d'objets. Cette notion d'objets facilite la conception du système dans son ensemble en réunissant les données et les traitements, les aspects statiques et dynamiques :



**Figure 12 : Graphe des relations statiques entre les classes**

Pour concevoir les différents objets linguistiques du domaine d'application, nous avons essayé d'analyser les différentes relations et types de liens (sémantiques) qui peuvent exister entre ces objets. Les différents liens entre les classes sont visualisés sur le graphe statique

du modèle conceptuel (Fig. 12). Trois types de liens peuvent exister entre les objets d'un système d'information : le *lien de composition, de référence et d'héritage*. Ces trois types de liens ont une sémantique proche du monde réel pour représenter les relations entre les objets du système.

### 5.2.2 — *Spécification des modules du système*

Le rôle de ces modules est de construire de façon incrémentale un ou plusieurs mots. Ils peuvent être coopératifs puisque les résultats d'un module peuvent servir comme données pour un autre module. Chaque module représente un outil autonome, indépendant qui dispose d'une mémoire de travail locale et participe à la création de mots nouveaux par accès en lecture / écriture à la base de faits. Les changements effectués par les différents modules dans la base de faits sont de type : création, modification ou suppression d'un objet (item lexical).

Dans cette section, nous spécifions un ensemble de modules qui peuvent être intégrés dans un système d'aide à la génération automatique de mots de base et de néologismes arabes.

#### *a) Modules de gestion du lexique :*

Ces modules représentent l'ensemble des outils de gestion du lexique permettant :

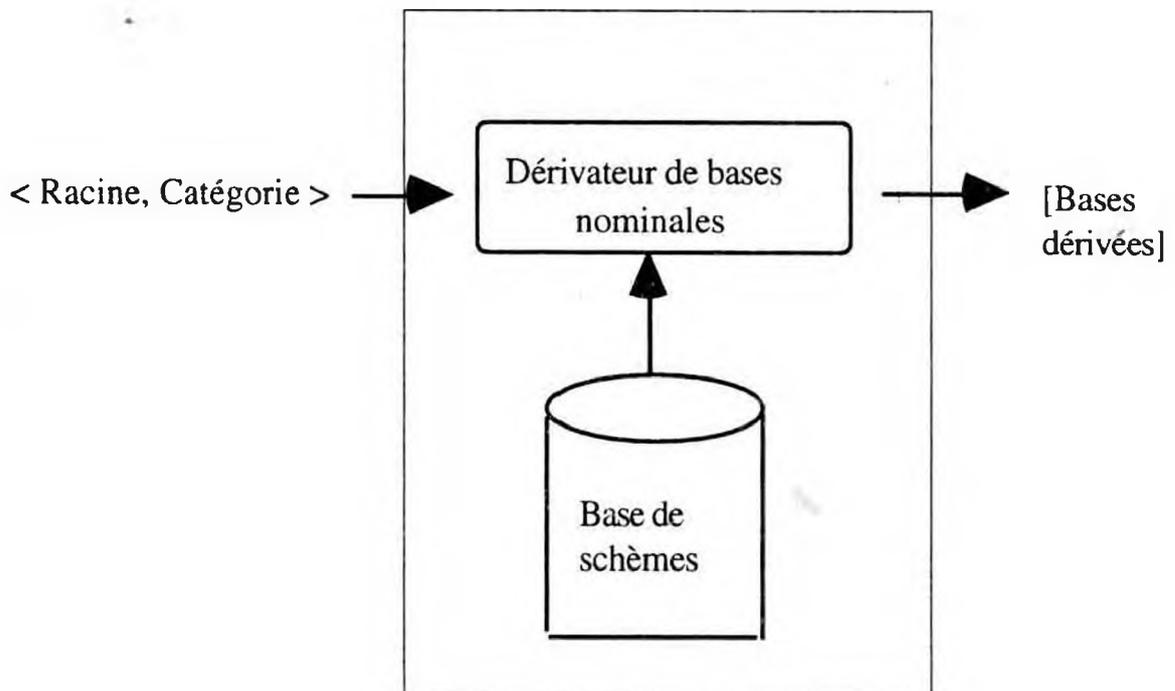
- La création des éléments du lexique;
- La mise à jour du lexique (banque des racines, dictionnaire de mots de base...);
- La consultation en accès direct et séquentiel des composantes du lexique.

#### *b) Dérivateur de bases nominales :*

Ce module permet d'engendrer une ou plusieurs bases nominales qui peuvent être retenues comme néologismes ou être utilisées comme données pour d'autres modules. La figure ci-dessous présente le schéma fonctionnel d'un dérivateur de bases nominales. Ce dérivateur accède à la base de schèmes par la catégorie donnée afin de sélectionner le ou les schème(s) correspondant(s). Cette base est composée d'un ensemble de schèmes permettant un accès direct par *l'attribut* catégorie. Les schèmes sont regroupés par catégorie. À partir d'un objet racine et d'un objet schème, le dérivateur permet d'engendrer une base dérivée par application de la méthode : *Créer-*

*base (racine, schème)*. Ce constructeur substitue les symboles abstraits /F?l/ du schème par les consonnes de la racine concernée. À titre d'exemple, étant donné le contexte suivant en entrée :

< Racine = /h s b/; Catégorie morphologique = Nom d'instrument >, le dérivateur permet de générer parmi la liste des bases possibles les termes : /hassabat, haasuub, haasib, haasibat.../ :



**Figure 13 : Schéma fonctionnel d'un dérivateur de bases nominales**

**c) — Analyseur morphologique :**

L'intégration d'un analyseur morphologique comme étape de pré-traitement s'avère nécessaire pour le processus de génération à partir d'un mot de base donné. L'analyseur doit permettre l'extraction de la racine à partir d'une base dérivée pour appliquer ensuite le processus classique de dérivation selon les modèles de schèmes. En effet, la manière de remonter à la racine permet d'élargir les possibilités de dérivation.

Si nous reprenons le même exemple du terme : /haasuub/, il est possible de fournir comme données de base un contexte formé par le mot : /hissab/ désignant la notion de «calcul» : <catégorie = nom d'instrument>. Un premier processus d'analyse morphologique peut être appliqué sur le mot de base /hissab/ pour extraire la racine : /h s

b/. Un second processus de dérivation classique à base de schèmes permet par la suite d'engendrer parmi les bases possibles le mot : /haasuub/ en fonction de la catégorie du schème donnée. En effet, la génération d'un nom d'instrument doit être effectuée à partir de la racine.

**d) Module de flexion :**

Le module de flexion permet de générer de nouvelles formes en appliquant des règles de la morphologie flexionnelle, à partir des données flexionnelles : <genre, type-pluriel, détermination>.

La partie *condition* des règles vérifie certaines valeurs de l'état du nom qui va subir la déclinaison (nom. genre, nom. nombre, nom. schème...). La partie *action* déclenche par exemple, des méthodes de génération des pluriels : *Générer - pluriel - sain; Générer - pluriel - brisé :*

Objet 1 (base dérivée) : État 1

- Nom : /qaa?imat/ Générateur-pluriel-sain (Objet 1)
- Catégorie : substantif ----->
- Genre : féminin
- Nombre : singulier
- etc.

Objet 1 (base dérivée) : État 2

- Nom : /qaa?imaat/
- Catégorie : substantif
- Genre : féminin
- Nombre : pluriel
- etc.

Si le type de pluriel est brisé, il est possible dans certains cas de choisir le schème du pluriel en fonction du schème de la forme du singulier : /nawaafi s / : «fenêtres»; /hawaassiib/ : «ordinateurs».

**e) Module de dérivation par affixation :**

Ce module consiste à générer des mots nouveaux par un jeu d'affixation aux mots de bases (généralement par suffixation), selon des règles de formation de mots de type dérivationnel.

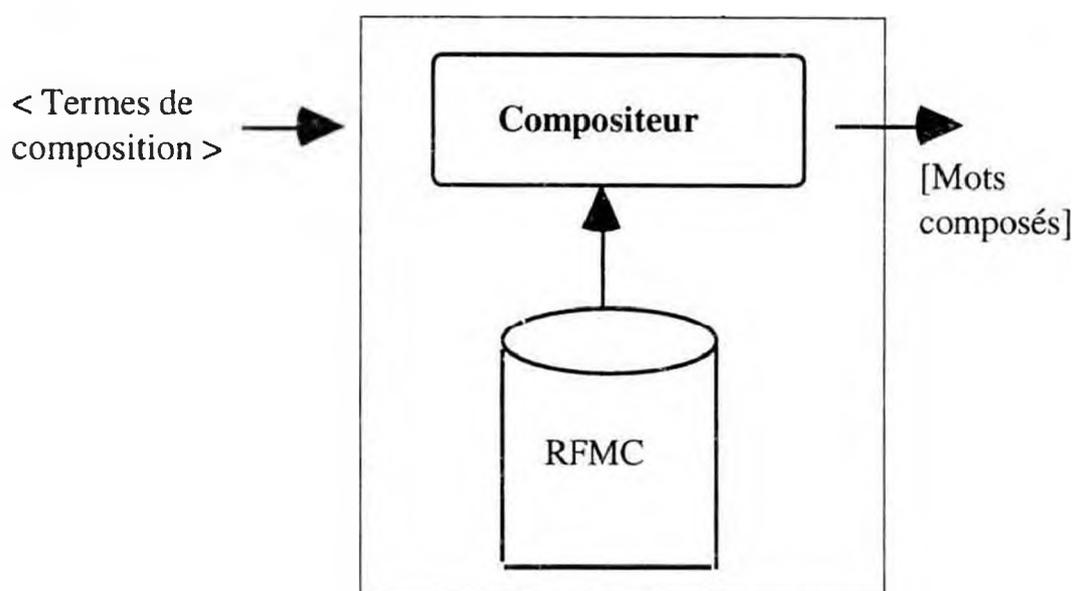
La règle : [t] A → [ [t] A + u ] B peut se traduire, par exemple, par la représentation suivante : Si (t. catégorie = A) alors Ajout-suffice (t, u). Où t représente un (mot de base) qui va subir la suffixation. La partie *condition* de la règle teste la présence de certains *attributs* du mot de base en entrée : (mot. catégorie, mot. genre, mot.nombre...). La partie *action* déclenche les processus de suffixation du type : ajout-suffixe (mot de base, suffixe).

Les adjectifs relatifs et les noms abstraits sont dérivés par ajouts respectifs des suffixes : /+ii/ et /+ijj/ à un mot de base :

— Le mot /haasubijja/ : «computationnelle» est dérivé à partir de la base : /haasub/ : «ordinateur»; l'adjectif relatif : /raqmii/ : «numérique» est dérivé à partir du nom : /raqm/ : «numéro».

**f) Module de composition :**

Le module de composition est constitué d'un ensemble de règles de formation de mots composés (RFMC) exploitées par un compositeur de mots. Les termes à composer sont des mots existants dans le lexique ou des mots nouveaux générés au cours du processus de génération :



**Figure 14 : Schéma fonctionnel du module de composition**

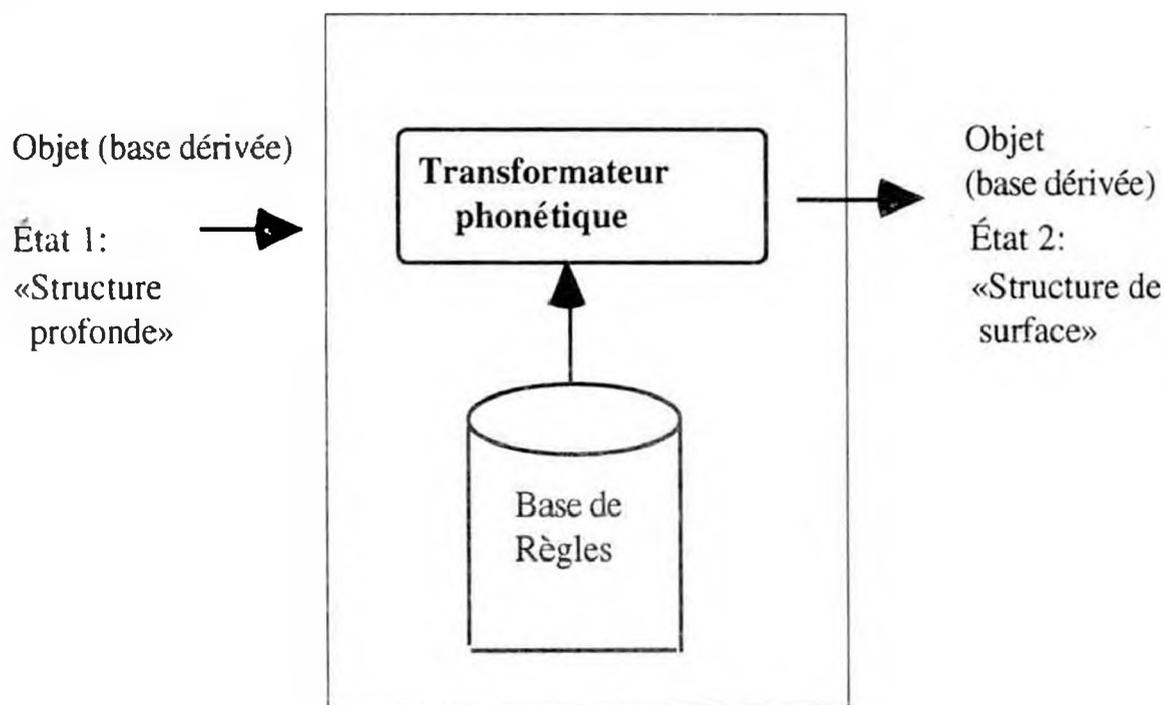
Un mot composé est formé de deux mots simples réunis selon des règles de composition de type :

$$[ [ \text{Mot 1} ]B + [ [ \text{Mot 2} ]A ]MC.$$

La partie *condition* des règles de composition spécifie les relations morpho-syntaxiques (catégorie, genre, nombre, détermination...) et les contraintes sémantiques que doivent vérifier les mots à composer. La partie *action* crée un objet composé à partir de deux mots simples. Par exemple le mot composé : /waa3iha bajnijja/ : «interface» est formé selon la règle générative : *Substantif + Nom abstrait (en +jja) fi Mot composé (Nom)*.

**g) Module de transformation phonétique :**

Le dérivateur de bases nominales permet de générer une forme dans sa structure profonde. Pour le cas des racines faibles et des formes dérivées de /ʔiftaʔala/, il est nécessaire d'appliquer les règles phonologiques pour générer des représentations phonétiques de surface. Cet outil s'avère nécessaire pour plusieurs systèmes (traitement de la parole, conjugueur automatique des verbes arabes...). La figure 15 montre un schéma fonctionnel du module de transformation phonétique. Ce module est composé d'un ensemble de règles de production et d'un transformateur phonétique. Le transformateur exploite cette base de règles phonologiques afin de générer la structure de surface de la base à partir de sa structure profonde :



**Figure 15 : Schéma fonctionnel du module de transformation phonétique**

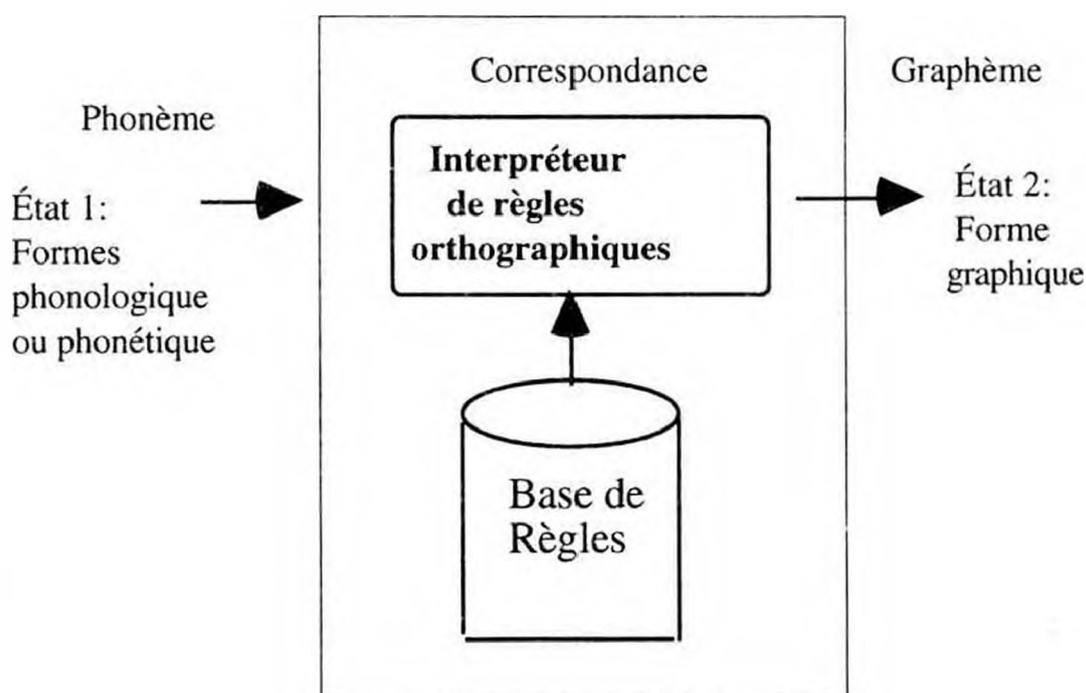
Les règles de production sont de la forme : Si (*contexte phonétique*) Alors *action*. Le contexte phonétique est de la forme : <CG, phonème, CD >. Il contient le phonème (ou le caractère) qui va subir la transformation avec son contexte gauche (CG) et son contexte droit (CD). Le contexte gauche et le contexte droit représentent, respectivement, les phonèmes précédents et les phonèmes suivants du phonème à transformer. Ce module procède par l'analyse du mot (base-dérivée) pour chercher un phonème qui

peut subir une transformation et déterminer son contexte phonétique. Ces données constituent avec d'autres propriétés de l'objet à générer la base de faits initiale (contexte de travail) à laquelle se réfèrent les procédures d'interprétation des règles.

La partie *action* des règles consiste à déclencher des opérations de type : élision, assimilation, substitution, épenthèse... Le nouveau mot généré peut constituer le contexte d'activation d'autres règles de transformation phonétique. Le processus s'arrête quand il n'y a plus de transformation à effectuer.

***h) Module de transcription orthographique :***

Ce module est nécessaire pour transcrire correctement chaque mot généré plus particulièrement les mots contenant les archigraphèmes : «*Hamza, Alif-final, Tä-final...*». Cet outil peut être réutilisé pour implanter des parties d'autres applications sur la langue arabe (correction orthographique, traitement de la parole...). Cet outil est composé d'un ensemble de règles orthographiques gérées par un interpréteur permettant d'inférer la forme graphique à partir de la forme phonologique ou phonétique :



**Figure 16 : Schéma fonctionnel du module de transcription orthographique**

Les règles de transcription orthographiques représentent des connaissances de *correspondance* : *phonème / graphème*. Dans ce type de correspondance, la difficulté réside dans le cas où un phonème peut avoir plusieurs transcriptions graphiques (c'est le cas de la transcription des archigraphèmes : /ʔ/, /t/, /aa/). La règle suivante montre un exemple d'une règle de transcription de l'archigraphème «Hamza» /ʔ/ :

Si (CG = /a/ et (Phonème = /ʔ/) et (CD = /a/) Alors (Graphème =)

Exemple : /saʔala/ →

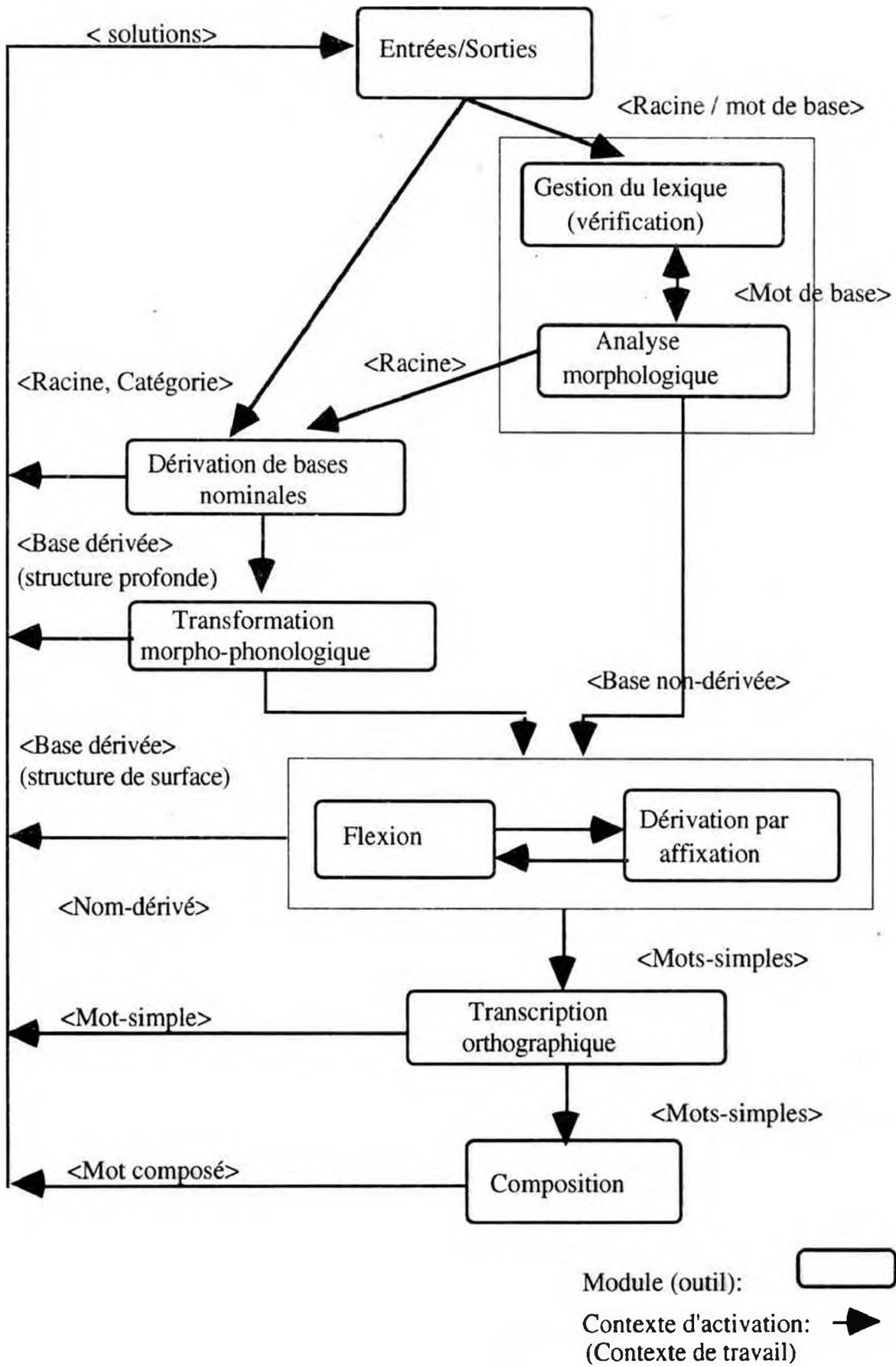
La partie *condition* de la règle de transformation spécifie le contexte du phonème à transcrire <CG, phonème, CD> et certaines caractéristiques morpho-syntaxiques de la forme à transcrire. La partie *action* applique l'opération de transcription de chaque phonème par son graphème correspondant. Pour faire la correspondance entre une forme phonologique (ou phonétique) et une forme graphique, il est nécessaire d'élaborer un dictionnaire des exceptions, ainsi qu'un ensemble de règles de traitement des exceptions (Ex. : /ʔuulaaʔika/ =).

Il est plus intéressant de réaliser un outil plus large permettant d'assurer à la fois les passages : graphème-phonème et phonème-graphème. Cet outil pourrait être utilisé à la fois pour la synthèse de la parole arabe à partir du texte écrit, et pour la reconnaissance optique des caractères arabes.

### 5.2.3 — Un générateur

C'est un programme gérant les différents modules du système. En fait, le rôle du générateur dans cette architecture consiste simplement à activer des séquences de modules selon un plan composé d'un ensemble d'étapes à réaliser permettant la génération de termes nouveaux.

Les différents modules du système possèdent des interactions fonctionnelles fortes dans la création des unités lexicales. Les modules sont généralement complémentaires, du fait qu'un module peut produire des termes servant comme données d'entrée pour un autre module. Chaque module possède un contexte de travail (données en entrées) et génère des résultats qui peuvent servir comme contexte d'activation à un autre module. Les étapes effectuées au cours de la résolution constituent le chemin de dérivation :



**Figure 16 : Processus de génération de mots**

A partir d'un ensemble de données acquises par un module, il est possible de remonter à la racine et au schème à l'aide d'une analyse morphologique et de déclencher le dérivateur de bases nominales. Cette étape peut être suivie d'une étape de transformation phonétique. Un processus de flexion et/ou de dérivation par affixation peut être appliqué sur des mots de base générés ou donnés à l'étape d'initialisation. Des mots composés peuvent être ensuite générés à partir de mots simples formés au cours de la résolution du problème. Une étape de transcription orthographique des mots générés est nécessaire. Chaque étape du plan peut produire des mots qui peuvent être retenus comme solutions :

Le processus de création de mots arabes est un processus variable qui peut ne pas s'adapter à un plan rigoureux de génération fixe. En effet, on remarque que :

- Chaque mot généré peut avoir un chemin de dérivation créé dynamiquement au cours de la résolution;
- Un module peut participer plusieurs fois dans le processus de génération;
- Les mêmes données en entrée peuvent être exploitées de différentes façons permettant d'avoir des résultats différents.

Exemple : Le même mot de base /ma luumat/ peut être exploité de plusieurs manières produisant plusieurs chemins de dérivation :

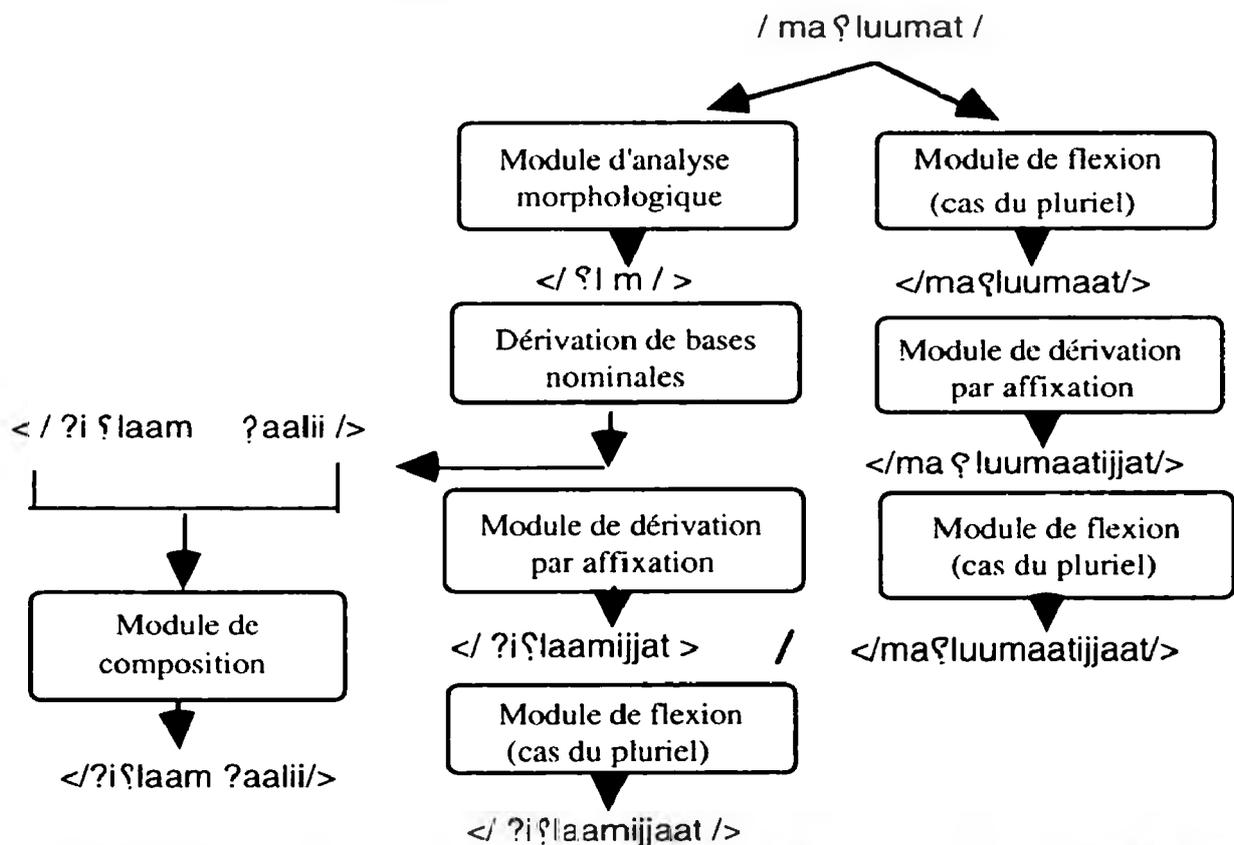


Figure 17 : Processus de génération établi à partir du mot /maʕluumat/

## 6. — CONCLUSION

Au début de cette étude, nous avons révisé les travaux antérieurs sur la nature de la racine et du schème et plus particulièrement sur les incompatibilités entre segments non-syllabiques à l'intérieur du morphème-racine du sémitique en général et de l'arabe standard en particulier. Notre but était d'arriver à partir de cette révision à une formulation logico-mathématique et à une généralisation des contraintes de structure morphématique de l'arabe. Notre but ayant été atteint, nous pouvons maintenant synthétiser les résultats obtenus, afin d'évaluer la place et le rôle de nos conditions de structure morphématique dans une grammaire de l'arabe standard.

En évaluant la place des contraintes de structure morphématique dans la grammaire de l'arabe, nous constatons qu'elles étendent le critère de simplicité au lexique. En effet, elles nous permettent de distinguer entre matrices admissibles et matrices non admissibles (mots possibles et impossibles) d'une manière qui paraît naturelle. Ainsi, lorsque nous ajoutons au lexique de l'arabe une condition spécifiant les coefficients de traits dans certaines configurations, les valeurs prédites peuvent rester non spécifiées dans les entrées lexicales. Comme le nombre de coefficients prédits est supérieur au nombre de traits spécifiés dans nos conditions, l'addition de ces conditions à la grammaire de l'arabe standard représente une vraie généralisation. Et comme il reste encore un nombre de configurations non attestées qui sont compatibles avec l'ensemble des conditions profondes de l'arabe standard, ce sont les *lacunes accidentelles*, qui constituent les matrices admissibles et non réalisées, qui sont d'une grande importance pour la néologie lexicale. Ainsi, les conditions de structure morphématique nous permettent de distinguer les configurations admissibles des non admissibles, sur la base d'une extension de la méthode d'évaluation du lexique de l'arabe standard.

L'innovation fondamentale qu'apporte ce type de conditions réside dans l'interprétation des contraintes de structure morphématique comme des conditions sur le lexique plutôt que comme des règles ordonnées qui s'appliquent à la manière des règles phonologiques.

Quant à l'intégration de ces conditions à la composante morphologique de l'arabe, elle se fait selon un processus analogue à celui qui définit le fonctionnement des trois types de conditions de structure morphématique (positive, négative, conditionnelle et biconditionnelle). Une fois incluse dans la composante morphologique, les conditions fonctionnent alors comme un ensemble de filtres ne laissant passer que les matrices acceptables.

Dans la seconde partie de notre recherche, nous avons essayé de concevoir un système d'aide à la génération automatique de mots de base à partir de racines et de schèmes et des néologismes à partir de mots de base. Ce système a pour tâche d'engendrer différents types de mots pour plusieurs catégories grammaticales (substantif, nom abstrait, adjectif, nom d'action...) et de permettre par conséquent l'enrichissement et la mise à jour du lexique arabe.

Cette étude a permis :

— l'élaboration, la formalisation et la modélisation d'un *ensemble de contraintes de structure morphématique* régissant la structure interne de la racine et du mot arabe, et qui sont indispensables pour la génération automatique des différents dictionnaires arabes : dictionnaire des racines attestées, des racines admissibles...;

— l'implémentation et l'application d'un *ensemble de règles de formation de mots* déjà élaborées pour l'arabe dans un travail antérieur sur la génération automatique de néologismes arabes à partir des règles de formation de mots <sup>(18)</sup>;

— la formalisation et l'implémentation d'un ensemble de règles morpho-phonologiques et des règles de transcription orthographique (correspondance phonème-graphème);

— la conception d'une architecture modulaire du système composée d'un ensemble de modules indépendants et autonomes de génération de mots. ces modules peuvent être exploités dans d'autres applications de traitement automatique de la langue arabe.

Le système proposé contient cinq modules opérationnels : flexion, dérivation, composition, transformation morpho-phonologique et transcription orthographique, avec une interface graphique en arabe sous WINDOWS 95.

Ce système a été implémenté avec un outil de développement de systèmes experts : Clips 6.0 (C Language Integrated Production System), intégrant un langage de programmation orienté objet COOL (Clips Object Oriented Language). Cet outil contient trois modules de représentation de connaissances : fonctionnel, déclaratif, orienté objet. L'interface du système a été réalisée avec l'outil XW clips.

---

18) Hussein Habaili & Mohamed Ben Ahmed, «Génération automatique de néologismes arabes à partir des règles de formation de mots», *Lexicomatique et dictionnaires*, Actes des IVes Journées Scientifiques du réseau LTT, Lyon, Université Lyon 2, A. Clas, Ph. Thoiron et H. Béjoint (dir.), coll. «Universités francophones», Beyrouth, AUPELF-UREF-FMA, pp. 173-188, 1995.

## RÉFÉRENCES

- BEJAOUI, Moncef (1985) : *Étude et réalisation d'un système expert à l'analyse morpho-syntaxique des phrases en langue arabe : Méthode ascendante*, Thèse de Doctorat 3<sup>e</sup> cycle en Informatique, U.P.S. Toulouse III, Laboratoire C.E.R.F.I.A.
- BEN HAMADOU, Abdelmajid (1993) : *Vérification et correction automatique par analyse affixale des textes écrits en langue naturelle : cas de l'arabe non voyellé*, Thèse d'État en Informatique, Tunis, Faculté des Sciences de Tunis.
- BEN HAMADOU, Abdelmajid (1993) : «Un traitement causal des erreurs de transcription des archigraphèmes dans les textes arabes», *Actes du colloque de Casablanca*, pp. 361-370.
- BEN HAZEZ, Slim (1996). *Un système d'aide à la génération automatique de néologismes arabes*, Mémoire de D.E.A. en Informatique, dirigé par Mohamed Ben Ahmed et Hussein Habaili. Tunis : Université de Tunis II, École Nationale des Sciences de l'Informatique.
- BOUBAKER, Ourida (1986) : *Al-SEMESS : un système expert en morphologie, syntaxe et sémantique pour l'étude de la langue arabe*, Thèse de Doctorat 3<sup>e</sup> cycle en Informatique, U.P.S., Toulouse III, Laboratoire C.E.R.F.I.A.
- BOUZGHOUB, M., GARDARIN, G. et P. VALDURIEZ (1994) : *Objet : concepts-langages-bases de données-méthodes-interfaces*, Paris, Édition Eyrolles.
- BRUNET, Joel (1991) : «Modeling the world with semantic object», *Objected Oriented Approach in Information systems*, North Holland, Elsevier-science Publisher B.V., pp. 361-379.
- CHAKER, S. et E. HAOUET (1993) : *Réalisation d'une base de données lexicales arabes par dérivation*, Projet de I.D.I., 5, Tunis, Faculté des sciences de Tunis.
- GARGOURI, Bilel (1995) : *Une expérimentation de l'approche orientée-objet dans le traitement des langues naturelles : le cas de la correction orthographique*, Mémoire de 3<sup>e</sup> cycle, DEA en modélisation et informatique de gestion, Tunis, Université Tunis III, Institut Supérieur de gestion.

- HABAILI, Hussein (1976) : *Contraintes de structure morphématique en arabe*, Mémoire de Maîtrise ès art en linguistique, Montréal, Université de Montréal.
- HABAILI, Hussein (1990). *Phonologie et Morphologie flexionnelle et dérivationnelle de l'Arabe*, Thèse pour le Doctorat d'État, Paris, Université de la Sorbonne-Nouvelle, Paris III.
- HABAILI, Hussein (1993) : «Principes directeurs pour l'établissement d'une banque des morphèmes-racines de l'arabe standard». *TA-TAO : recherches de pointes et applications immédiates*, Actes des Troisièmes Journées scientifiques du réseau LTT, Montréal, Université de Montréal, Beyrouth, AUPELF-UREF-FMA, pp. 457-476.
- HABAILI, Hussein (1995) : «Génération automatique de néologismes arabes à partir des règles de formation de mots», *Lexicomatique et dictionnaires*, Actes des IV<sup>es</sup> Journées scientifiques du réseau LTT, Lyon, Université de Lyon-2, Beyrouth, AUPELF-UREF-FMA, pp. 173-188.
- HATON, J.P.; BOUZID, N. *et al.* (1991) : *Le raisonnement en intelligence artificielle : modèles, techniques et architectures pour les systèmes à base de connaissances*, Paris, Éditions Inter Édition.
- LAASRI, H. et B. MAÎTRE (1989) : *Coopération dans un univers multi-agents basée sur le modèle du blackboard : Études et réalisation*, Thèse de Doctorat, Nancy, Université de Nancy.
- LAATAOUI, Mourad (1994) : *Compression de données : Approches par analyse morphologique*, Mémoire de DEA, Tunis, Université Tunis III, Institut Supérieur de gestion.
- LALANDA, Philippe (1992) : *Conduite du raisonnement dans un système à base de tableau noir en temps réel*, Thèse de Doctorat. Nancy, Université de Nancy.
- LAMBOLEZ, P., QUEILLE, J. *et al.* (1995) : «EXREP : a Generic Rewriting Tool for Textual Information Extraction», *Ingénierie des systèmes d'information*, vol. 3, n° 4, pp. 471-487.
- PUPPE, Frank (1993) : *Systematic introduction to expert systems : knowledge Representation and problem-solving methods*, Berlin, Springer.
- PUSTEJOVSKY, J. et B. BRANIMER (1993) : «Lexical knowledge representation and natural language processing», *Artificial intelligence*, vol. 63, Elsevier, pp. 193-223.

## **Les procédés linguistiques d'actualisation et de virtualisation et la production du mot**

**Kamel GAHA**

*Université de Tunis I*

Dans son allocution d'ouverture, le professeur Quemada a fait allusion aux débuts difficiles de la lexicologie; difficiles à cause de l'incompréhension de certains, de l'hégémonie de la syntaxe et des syntacticiens et de l'antimentalisme farouche des héritiers de Saussure, confortés dans leur refus du sens par le moment de gloire du distributionnalisme américain et bientôt planétaire.

Cette situation frustrante, mais peut-être aussi stimulante, se traduisait, et se traduit toujours, par la place qui est réservée au vocabulaire dans les manuels de grammaire destinés aux différents niveaux de l'enseignement, ainsi que par la conception des programmes d'enseignement.

A vrai dire il n'y a pas que cela. Les développements de la linguistique à partir de Saussure ont imposé la démarche systématique et la définition d'une unité de référence minimale comme conditions nécessaires de la pertinence et de l'efficacité de l'approche adoptée. Il fallait à l'image de la phonologie avoir une unité minimale insécable pour pouvoir prétendre au statut de science. Vous connaissez tous les retombées terminologiques de cet impératif de l'unité minimale sur les domaines de la lexicologie et de la sémantique.

De quoi disposait-on comme définition de cette entité minimale discrète; il y avait d'une part la définition saussurienne du signe linguistique, de l'autre le mot, autant dire le sphinx.

Or le signe chez de Saussure est une véritable utopie linguistique (1), ou plutôt une fiction linguistique d'un très haut rendement théorique; mais ce signe parfait isomorphe et se suffisant à lui-même n'existe nulle part, la plus simple des particules, le radical le plus nu, impliquent un nombre impressionnant d'informations associées au même signifiant et sont solidaires de tellement de formes, par contraste ou par ressemblance, qu'on comprend très vite que cette définition demeure très peu utile pour la description et l'analyse des mots.

Qu'est-ce qu'un mot ? Un inventaire des tentatives de définition aboutit très rapidement à quelques conclusions qui relèvent du bon sens. La diversité terminologique — monème grammatical, monème lexical, lexème, syntème, sémantème, groupe locutionnel, périphrase,... — signale une gêne, disons plutôt un certain nombre de problèmes. Nous ne retiendrons de ces problèmes que deux aspects qui peuvent éclairer la problématique plus générale de la définition du mot.

Le premier aspect consiste dans l'instabilité de la forme ainsi désignée, instabilité dont on peut certes rendre compte en termes de figement, de dérivation et de composition. Nous proposons de voir dans cette instabilité si gênante pour l'approche terminologique l'un des traits définitoires du mot qui, comme ces entités physiques étranges que sont les particules élémentaires de la matière n'ont d'existence que dans le mouvement et le cinétisme.

Le mot n'est, en quelque sorte, que par la tension qui le porte vers d'autres mots qui l'assimilent ou qu'il assimile pour constituer ces macro-séquences — ou syntagmes — actualisées. L'inachèvement morphologique et sémantique du mot — qui constitue l'un des traits définitoires du mot, ainsi que nous l'avons déjà laissé entendre, — n'est que l'anticipation du procès de l'actualisation dans la forme encore virtuelle.

Cette capacité qu'a le mot d'avalier, ou d'assimiler, les éléments constitutifs de son contexte, soit directement en maintenant la forme de ces éléments inchangée (2), soit indirectement en condensant les éléments de la séquence assimilée en des morphèmes qui reconduisent le contenu sémantique des séquences transposées (3),

---

1) *Cours de linguistique générale*, éd. Tullio de Mauro, Payot, 1972, p. 146.

2) Comme cela se passe dans la composition.

3) Comme dans toutes les formes d'affixation.

concerne bien évidemment la morphologie (4) et la lexicologie, mais elle implique également la syntaxe. On peut même dire que le principe qui permet à un mot d'intégrer un contexte étendu — syntagme ou énoncé — est le même que celui qui permet d'agglomérer des morphèmes en unités lexicales dérivées (5) ou composées (6), la seule différence qui sépare les deux modes d'existence et de fonctionnement du mot réside dans la suspension ou l'effectuation du procès de l'actualisation.

L'instabilité en question n'est donc pas un enjeu de simple délimitation. Grâce à elle, le mot semble nous indiquer une direction, nous signaler une tension qui le porte littéralement à l'extérieur de ses frontières conventionnelles. Le mot tel que le livrent les inventaires lexicographiques est par définition en quête d'un contexte à assimiler, à *avaler*. C'est ce mouvement qui lui donne une signification, plus exactement qui le prédispose à en acquérir une.

On pourrait vérifier le bien fondé d'une approche comme celle que nous proposons à partir d'un rapide examen des parties du discours (7). Ce paradigme des catégories de la langue est lui aussi hétérogène et pose de nombreux problèmes aux chercheurs et aux étudiants.

L'un de ces problèmes est celui que pose l'intégration de l'interjection au paradigme des parties du discours ou catégories grammaticales — ce que les manuels de grammaire et les dictionnaires désignent comme *nature du mot* —. En tant que séquence graphique compacte et facilement identifiable, l'interjection semble constituer un cas idéal dans le répertoire des différentes classes de mots. Mais il y a à cela une objection majeure que l'on peut déduire de l'observation immédiate de cette classe, la neuvième, puisque le paradigme est constitué de huit classes plus l'interjection.

---

4) Le cas particulier de la morphologie verbale peut être envisagé dans les mêmes termes, la désinence verbale qui actualise le procès est une véritable *détermination du radical verbal*. Bally étudie le rapport entre les affixes et les bases auxquelles ils s'appliquent selon le même principe.

5) Voir l'analyse de la dérivation par Louis Guilbert, dans: *La créativité lexicale*, Larousse, 1975, IIème partie, «La dérivation».

6) « [...], il faut à notre avis envisager les composés non comme des espèces morphologiques, mais comme des organisations syntaxiques. La composition nominale est une micro-syntaxe.», Emile Benveniste, *Problèmes de linguistique générale*, 2, Gallimard, 1974, p. 145.

7) E. Benveniste, «Catégories de pensée et catégories de la langue», dans *Problèmes...*, I, Gallimard, 1966 pp. 63-74

L'interjection exige une intonation ou une mélodie avant même toute actualisation dans le discours. Elle apparaît nécessairement avec le signe de ponctuation spécifique, le point d'exclamation, dans les inventaires lexicographiques, ce qui n'est le cas d'aucune autre catégorie de mots. Cela suffit à en faire une classe à part, une classe de mots qui emportent avec eux leur contexte d'actualisation. En fait ce sont des mots-énoncés, et peuvent difficilement constituer une partie du discours ou une classe de mots au même titre que les autres catégories. Ces proto-énoncés, éminemment synthétiques, désignent probablement un état primitif du langage encore proche du cri et du corps, un langage qui superpose les différents niveaux de l'énonciation en une forme syncrétique dense.

La restriction que nous apportons à l'analyse habituelle des parties du discours n'a rien d'original, et elle peut se réclamer aussi bien de Charles Bally que de P. Lerat. Le premier considère que «Les interjections constituent un cas-limite des catégories lexicales. D'une part, en effet, elles expriment des notions virtuelles de nature verbale : *patatras!* rend l'idée d'une chute, *gare!* celle d'un avertissement [...] En revanche, les interjections ne sont employées que sous la forme de monorèmes, revêtus d'intonations qui confirment leur caractère de phrases.» (8)

Pierre Lerat fait une analyse légèrement différente, mais aboutit à la même conclusion; selon lui «l'interjection relève plus de la pragmatique que de la syntaxe et de la sémantique, ...» (9). Pour nous, ce qui importe le plus, c'est moins le problème de classification lui-même que la confirmation de notre première hypothèse. Cette classe hybride superpose en fait les deux étapes, plus exactement les deux limites qui circonscrivent l'espace d'effectuation ou d'actualisation du mot. Ces deux limites fusionnent dans cette forme synthétique primitive qu'est l'interjection. On comprend mieux de la sorte le deuxième aspect ou la deuxième difficulté que nous signalions au début de notre propos.

Pour ce qui concerne l'analyse proprement grammaticale, une fois le problème de l'interjection traité à part, il ne nous reste plus

---

8) *Linguistique générale et linguistique française*, éd. Francke Berne, 4ème éd., 1965, p. 114-115.

9) Pierre Lerat, *Sémantique descriptive*, éd. Hachette Université, 1983.

que huit <sup>(10)</sup> catégories, classes ou parties du discours. Malgré les multiples réserves <sup>(11)</sup> qu'on peut émettre à propos de telle ou telle catégorie, on peut dire que la pratique, mais aussi la recherche grammaticales, confirment l'importance de ce critère de classement. En laissant de côté le problème posé par le comportement non systématique de certaines classes — celle des articles ou des déterminants par exemple — on peut à partir des huit catégories disponibles faire un certain nombre d'observations.

**N.B.** Le reconstitution du paradigme est une opération facile qu'on peut effectuer en deux temps; dans un premier temps une approche purement empirique permet de classer les mots d'un texte en séries distinctes en tenant compte de leur distribution. Le deuxième, temps plus technique, peut consister en une application de l'analyse de la phrase et de l'énoncé en niveaux par Benveniste <sup>(12)</sup>.

Nous pouvons maintenant travailler sur nos classes. Là aussi l'observation empirique et l'étude de la distribution des classes conduisent à une constatation que tous ceux qui s'intéressent à la langue ont faite. Certaines classes de mots se distinguent par une force d'attraction, une sorte de magnétisme qui attire d'autres classes. Les classes subissant cette force constituent des satellites — pour reprendre la métaphore de Tesnières — du noyau autour duquel elles gravitent. L'application de cette règle de l'attraction nous permet de réorganiser le paradigme des classes en structures ou formes, et cela avant toute actualisation. Un simple coup d'œil suffit à montrer que les structures ainsi dégagées correspondent à deux schèmes que nous connaissons très bien le syntagme nominal et le syntagme verbal.

---

10) Revenons à ces fameuses parties du discours, pour signaler tout d'abord qu'elles ont une histoire et que cette histoire mérite d'être écrite.. Nous nous nous contenterons ici de rappeler qu'elle ont commencé par être dix, les dix fameuses catégories d'Aristote:

Le substance / La quantité / La qualité / La relation / L'Agir / Pâtir / Ou / Quand / La situation / Avoir /

Nous les présentons dans la formulation que leur donne *La Logique* de Port Royal. Ceux qui ont lu l'ouvrage connaissent le jugement très sévère qui les accompagne: «Voilà les dix Catégories d'Aristote, dont on fait tant de mystères, quoiqu'à dire le vrai ce soit une chose de soi très-peu utile, et qui non seulement ne sert guère à former le jugement, ce qui est le but de la vraie logique, mais qui souvent y nuit beaucoup...» pp. 77-78

11) Comme exemple de cette relecture du paradigme des classes, on peut citer le traitement de l'article dans la *Grammaire Transformationnelle* de G. Gross (Larousse, 1977)

12) « Les niveaux de l'analyse linguistique », dans: *Problèmes*, ouvrage cité, pp. 119-131.

ARTICLE

NOM

ADJECTIF

Séquence nominale

VERBE

ADVERBE

Séquence verbale

Mais ces deux séquences ne totalisent, comme on peut facilement s'en rendre compte, que cinq classés syntaxiques, que faire alors des trois classes restantes ?

Les classes encore disponibles ne peuvent être employées ni dans le rôle de noyau, ni dans celui de satellite; elles ont par contre des fonctions spécifiques qu'une étude attentive de la distribution des mots permet de circonscrire avec assez de précision. Deux de ces classes, les prépositions et les conjonctions ont un rôle évident de mot de liaison; elles servent à relier des mots à l'intérieur d'une séquence, ou encore à étendre un noyau.

Reste la catégorie des pronoms qui n'assume qu'un rôle de représentation, et qui n'influe en rien sur la structure de la séquence. Il existe cependant certains pronoms dits relatifs qui constituent une catégorie mixte de représentants-transpositeurs, ou de mots de liaison qui sont en même temps des représentants.

Cette analyse révèle la solidarité du syntaxique et du sémantique et la nécessité d'envisager le problème de la délimitation des unités significatives en partant des macro-unités syntaxiques que sont les phrases. L'approche des niveaux de l'analyse linguistique chez Benveniste satisfait à cette exigence structurale. Elle montre en outre qu'il est dangereux de parler de délimitation d'un mot ou de frontière d'un mot, parce que le mot existe d'une certaine manière comme intégrant potentiel d'une structure plus complexe; il existe littéralement à l'extérieur de ses frontières; ce qui confirme les conclusions dégagées à partir de la première difficulté.

Signalons que ce travail effectué sur les classes de mots doit être situé dans le contexte d'une réflexion historique philosophique et historique que nous n'avons pas le temps d'entreprendre ici. Très succinctement, cette réflexion se réclame aussi bien d'Aristote que du travail effectué par les philosophes du dix-huitième siècle sur le langage et des contributions de Charles Bally et d'Emile Benveniste à la question, la liste ne prétendant aucunement à l'exhaustivité.

Les dix catégories d'Aristote (13) suscitent les commentaires les plus critiques dans *La Logique* d'Arnaud et Nicole :

«Voilà les dix catégories d'Aristote, dont on fait tant de mystères, quoiqu'à dire le vrai ce soit une chose de soi peu utile, et qui non seulement ne sert guère à former le jugement, ce qui est le but de la vraie logique, mais souvent y nuit beaucoup.» (14)

Dans la *Grammaire* de Condillac (15) les parties du discours des grammairiens, les catégories syntaxiques des mots, se réduisent à quatre, le nom, l'adjectif, le verbe et la préposition; les verbes se divisant à leur tour en «verbe adjectifs», comme vivre, aimer que le philosophe décompose en verbe être suivi d'un adjectif : être vivant...et en verbe substantif comme dans «Corneille est un poète», le verbe substantif type c'est justement le verbe être. Nous allons citer Condillac parce que l'idée, dans la formulation qu'il lui donne est encore plus parlante, et nous permet de réduire la catégorie du verbe à un simple opérateur, un ligament comme dira Bally plus tard.

«Les éléments du discours se réduisent à quatre espèces de mots. Il ne faut que des substantifs pour nommer tous les objets dont nous parlons; il ne faut que des adjectifs pour en exprimer les qualités; il ne faut que des prépositions pour en indiquer tous les rapports; enfin il ne faut que le verbe être pour prononcer tous nos jugements» (16).

Bally pour sa part opère une lecture du paradigme des classes syntaxiques qui lui permet de dégager la spécificité de certaines de ces classes qu'il appelle catégories lexicales.

«Nous appelons catégories lexicales les classes de signes exprimant les idées destinées à se combiner dans le discours au moyen des ligaments grammaticaux.

Les catégories lexicales comprennent, dans nos langues, les mots, ou plus exactement les sémantèmes virtuels désignant des substances..., des qualités, des procès et des modalités de la qualité et

---

13) Arnaud et Nicole, *La Logique ou l'art de penser*, éd. Flammarion, 1970, pp. 77-78.

14) *Ibid.* p. 78.

15) Cette *Grammaire* fait partie du *Cours d'études pour l'instruction du prince de Parme*, publié par Leroy dans le corpus des Oeuvres philosophiques, T. XXIII, V. 1, p. 395 et sq, V.2. Elle y occupe les pages 427-513.

16) Charles Bally, *Linguistique générale et linguistique française*, ouvrage cité, p. 456.

de l'action, autrement dit: des substantifs..., des adjectifs..., des verbes,... et des adverbes.

...les signes grammaticaux ou ligaments, qui expriment des relations, ne forment pas une catégorie lexicale...» (17)

Le paradigme des catégories grammaticales de base masque une structure syntagmatique que l'analyse peut facilement retrouver; le syntagme est l'horizon nécessaire du mot, et cette nécessité est inscrite aussi bien dans la forme du mot que dans son sémantisme. On comprend que des générations de grammairiens et de linguistes aient éprouvé les plus grandes difficultés à définir le mot en lui-même; cette entreprise, qui montre l'importance de l'Institution scolaire dans l'histoire des théories de la langue, est en réalité un non-sens flagrant, le mot ne peut être défini que par l'aptitude qu'il a de se combiner avec d'autres mots pour actualiser une représentation donnée.

Cette subordination nécessaire du mot au procès de l'actualisation est également visible dans les mécanismes linguistiques de *transposition*. C'est Lucien Tesnière qui, le premier, a attiré l'attention sur l'importance des mécanismes de *translation* et sur le rôle spécifique des *translateurs*, mettant en évidence une propriété linguistique qui allait s'avérer des plus productives. Les translateurs, qui sont des ligaments ou mots de liaison, permettent de transférer un mot relevant d'une catégorie syntaxique donnée dans une catégorie syntaxique différente. Le cas le plus facile et le plus typique est celui de la préposition *de* dans des contextes comme «le recueil *de* Mallarmé» ou le «livre *de* géographie», où un substantif est transposé en adjectif. Mais à côté de la classe des mots de liaison, nous trouvons une classe de transpositeurs, ou de translateurs affixaux, qui s'appliquent à une base lexicale pour en modifier la catégorie syntaxique, comme dans la série *interroger*, *interrogation*, *interrogé*, *interrogeable*...

Même si la nature du procès de transposition demeure la même dans les deux cas, le deuxième présente l'avantage de rendre sensible la disponibilité des catégories syntaxiques prégnantes pour la prise en charge d'une même représentation. La nominalisation, l'adjectivisation et la verbalisation ne constituent en fin de compte que des procédures de transposition définies à partir d'une base donnée (18).

---

17) *Ibid.*, p. 115.

18) Le phénomène de la transposition des catégories syntaxiques constitue le fondement de l'approche générativiste du lexique dans l'approche de Louis Guilbert (*La créativité lexicale*, Lib. Larousse, 1975). Mais il est également

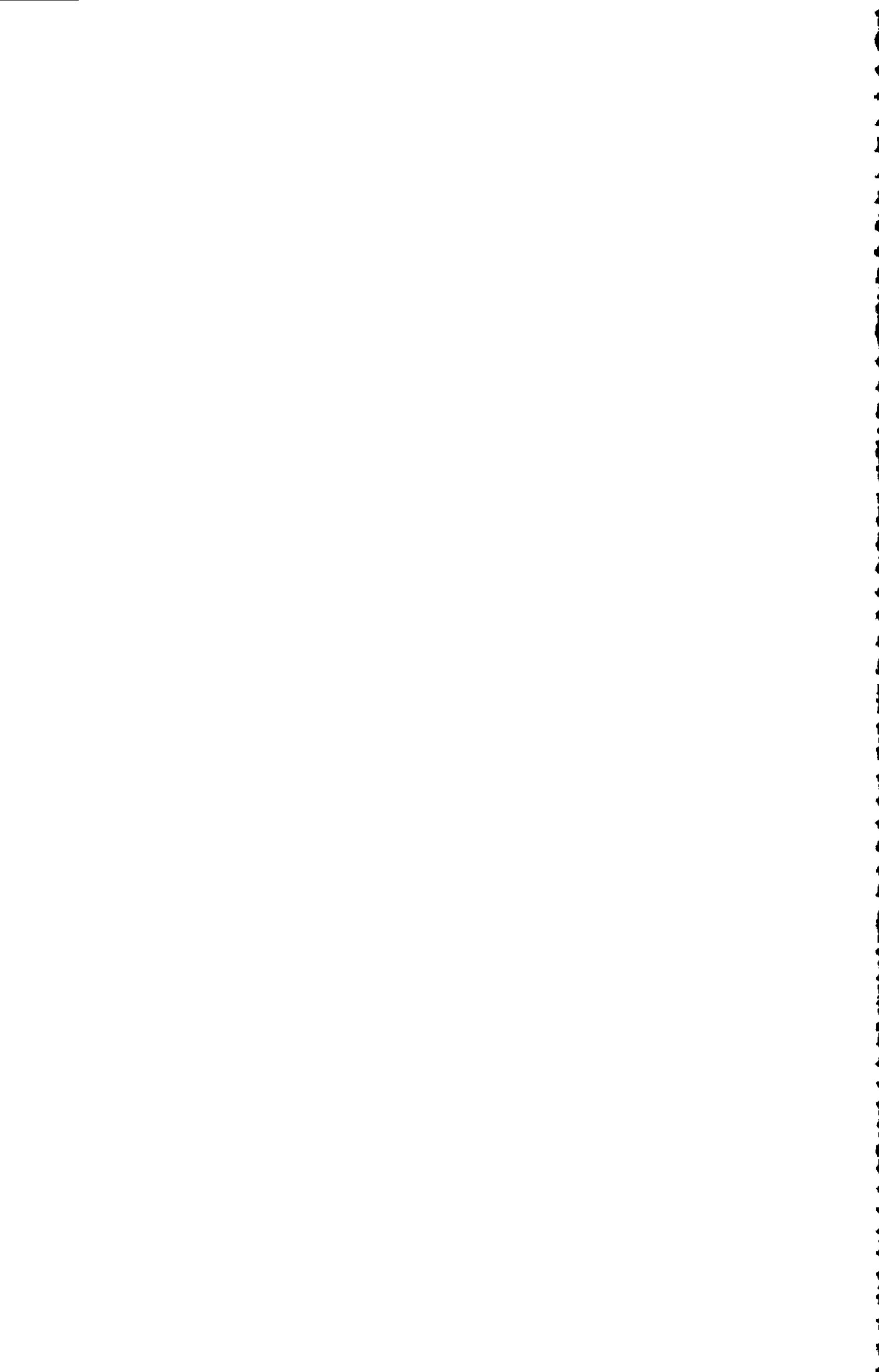
Non seulement le mot n'existe que pour devenir, mais on peut dire, au risque de heurter une tradition bien assise, celle de l'analyse grammaticale des langues analytiques, que le mot n'appartient potentiellement, *a priori*, à aucune catégorie syntaxique (19), et que les mécanismes de transposition ou de translation mettent à la disposition de l'utilisateur un système très fin qui lui permet d'envisager la même représentation, comme substance, comme qualité, comme procès, ou même comme modalité d'appréciation.

Même si le modèle proposé n'a pas la régularité qu'on peut escompter, les possibilités offertes par les mécanismes de transposition permettent de garantir cette mobilité exceptionnelle de l'esprit, nécessaire à la saisie instantanée des représentations mobilisées en temps réel par l'utilisateur. Se confirme de la sorte que ce qui pouvait apparaître comme inachèvement et imperfection du mot n'est au fond que la marque d'une disponibilité à laquelle l'actualisation donne tout son sens, en faisant de la langue, plus une intelligence exceptionnelle et une souplesse incomparable dans le maniement des concepts et des représentations acquises. Nous pensons même que dans cette imperfection, nous avons probablement le secret de l'une des fonctions les plus étranges et les plus stimulantes de la langue, la fonction heuristique.

---

abordé par Bally dans l'ouvrage que nous avons cité (pp. 116-127), ainsi que par Benveniste dans «Mécanismes de transposition», *Problèmes* 2, pp.113-125.

19) «Il n'y a donc pas de substantifs de «substance» voués par nature à une fonction grammaticale, ...», L. Guilbert, *La créativité lexicale*, ouvrage cité, p. 133.



# Une problématique de la dénomination technique dans deux langues à comparer : mémoire des notions et mémoire des termes

Raphaël RAJASPERA

*Université d'Antananarivo, Antananarivo,  
Madagascar*

L'objet de cette communication sera de mettre l'accent sur quelques problèmes de la dénomination technique comme «*lieu de mémoire*» et comme «*objet à mémoriser*» : lieu de mémoire d'abord, puisque chaque mot a son histoire et que l'appellation doit, en principe, évoquer immédiatement à l'esprit la chose désignée et réveiller des souvenirs qui s'y attachent; ensuite, objet à mémoriser, parce que les concepts et les notions ont besoin de noms spécifiques, qui les fixent dans la mémoire et dont il faut se souvenir pour communiquer vite et avec précision entre spécialistes d'un domaine donné.

Il s'agit alors de montrer, avec des exemples de deux langues (français, malgache), 1) dans quelle mesure l'évocation des termes (1), comme lieux de mémoire, peut rappeler des notions, des concepts et des valeurs d'une culture, et 2) dans quelle mesure l'évocation des notions et des concepts peut rappeler des termes comme objets à mémoriser (il s'agit en particulier des problèmes de terminologie et de dénomination en fonction de la capacité de notre mémoire).

---

1) En donnant à «terme» le sens de mot appartenant au vocabulaire particulier à une spécialité, celui de vocable technique comme élément lexical de toute une terminologie.

## TERMES ET MÉMOIRE DES NOTIONS

L'évocation des termes peut rappeler directement les notions et les concepts correspondants. Déjà, dans l'univers mental du locuteur, le mot par lui-même, en tant que masse sonore ou graphique du signe linguistique, ajoute de sa réalité propre au concept, et lui donne sa forme et son contour : tout se passe comme si le mot créait la chose. Par la même occasion, le concept trouve dans le mot et dans sa définition une fixation et une délimitation. À plus forte raison, quand un mot devient un terme, c'est là que, à l'abri de la polysémie, se fixent mieux dans la mémoire les notions et les concepts. Y jouent leur rôle, a) la motivation étymologique des termes, b) les emprunts motivés et, c) des facteurs socio-affectifs qui interviennent dans des associations mémorielles.

### La motivation étymologique des termes

L'étymologie des termes techniques sert quelquefois de point de repère pour se rappeler ou pour rappeler les concepts et les notions. A ce moment-là, c'est la motivation étymologique de ces termes, qui jouera un rôle déterminant dans l'évocation des notions.

On sait par exemple que des disciplines, comme la médecine, utilisent à profusion les étymons grecs pour opérer une taxinomie, une classification rigoureuse des notions et des concepts. Un terme médical comme «homéopathie» peut alors être considéré comme un dérivé fortement motivé par rapport à l'élément «homéo-», du grec *homoios* (analogue) et à l'élément «-pathie», du grec *pathos* (maladie) : en effet, il doit rappeler, d'après sa filiation historique et d'après le lien étymologique qu'il entretient avec son antonyme «allopathie», cette méthode thérapeutique connue depuis 1790 et consistant à traiter les malades (cf. : «-pathie») à l'aide d'agents qui déterminent une affection analogue (cf. : «homéo-») à celle qu'on veut combattre.

Sur un modèle semblable, on peut imaginer que se forme ainsi tout un vocabulaire technique motivé, comprenant de nombreux autres vocables qui se groupent spontanément dans l'esprit du locuteur. Les dérivés et les concepts qu'ils portent s'évoquent l'un l'autre par association. Que l'on songe, par exemple, aux séries étymologiques des termes français formés sur l'élément d'origine grecque «therm(o)-», et qui rappellent tour à tour des notions autour de l'idée de «chaleur» : thermal, thermochimie, thermodynamique, thermographie, thermomètre, thermonucléaire, thermos, thermostat,

thermotropisme, ainsi que d'autres dérivés comme «endo-» et «exothermie», etc.

Mais la relation protosémantique entre les termes du champ morpho-sémantique, fondée sur les ressemblances de leurs formes et sur la parenté de leurs sens, n'est évidente que pour l'usager, qui a une connaissance de la langue et de l'étymologie. C'est ce qui va lui faciliter leur emploi. Bien entendu, pour celui qui méconnaît ou qui oublie ces rapports étymologiques et dérivatifs, le vocable reste immotivé et ne rappelle pas toujours d'emblée des notions très précises. D'où l'effort particulier qu'il doit fournir pour assimiler les termes avec les définitions et pour les avoir disponibles dans sa mémoire : on sait que les vocables sans motivation doivent être appris un par un par l'opération d'un montage séparé; et il est évident que l'usager initié dispose de plus de mots motivés que le profane.

### **La situation des emprunts motivés**

Le même problème se pose pour l'usage des termes empruntés à une autre langue, à une langue étrangère et non apparentée à la sienne : les emprunts ne sont motivés que pour le locuteur bilingue. C'est la situation des emprunts malgachisés, par exemple.

Il existe, dans le parler quotidien des locuteurs malgaches bilingues, une certaine quantité de termes d'emprunt au français. Volontaires ou involontaires, ces emprunts, tels que «*atimosifera*» («atmosphère») et «*termometra*» («thermomètre»), n'évoquent évidemment les notions qu'ils désignent que pour les usagers connaissant et le français et le référent. Ils ne sont motivés ou ne rappellent les notions ou les concepts évoqués que par référence à la langue prêteuse. Pour un Malgachophone qui serait unilingue ou qui ne connaîtrait pas du tout le nom français du référent, ils restent immotivés : «Ils n'évoquent rien dans son esprit ni à son oreille», écrit un fervent défenseur de la langue malgache, le Pasteur Andriamifidy, dans un article sur la corruption de la langue malgache (2). Ce sont donc vraisemblablement de ces emprunts menacés d'être condamnés à l'oubli, surtout depuis la malgachisation de la langue d'enseignement. Puisque les notions qu'ils désignaient étaient censées

---

2) Nous l'avons traduit ici en français. Texte tiré d'un ouvrage intitulé : «Andriamifidy olo-mangam-pirenena malagasy» («Andriamifidy, une personnalité malgache»), février 1974, centième anniversaire du temple Ambonin' Ampamarinana, pp. 41-50.

être plus ou moins connues dans la langue maternelle, les locuteurs avertis les remplaçaient volontairement par des termes malgaches équivalents.

En fait, qu'évoquent pour les uns et pour les autres les termes d'emprunt ? Ici, les motivations d'accueil des termes étrangers dans la langue emprunteuse varient selon les réactions individuelles ou collectives des utilisateurs.

Pour les uns, le refus des emprunts (quand on en connaît les signifiés) est une réaction psychologique contre les abus qu'on en fait et qui font oublier l'existence de termes équivalents dans sa langue. C'est la raison pour laquelle, progressivement, un certain nombre d'emprunts malgachisés comme «*radiô*», «*termometra*», «*telefonina*», «*televiziona*» ont été remplacés par des syntagmes malgaches plus évocateurs, semble-t-il : «*fandefasam-peo*» «radio» litt. «transmission de voix ou de sons», «*faneho mari-pana*» «thermomètre» litt. «indicateur de température», «*fifandraisana an-tari-by*» «téléphone» litt. «communication par fil (métallique)», et «*fahita lavitra*» «télévision» litt. «vision-loin» (calque sémantique rappelant l'allemand «*fernsehen*»). L'enseignement et les media de masse, à force de répétition, ont joué un grand rôle dans la diffusion et dans la mémorisation de ces néologismes.

Pour d'autres, l'adoption de quelques emprunts, notamment dans la langue technico-scientifique, est devenue une nécessité selon les besoins de la dénomination : les objets, les opérations, les notions ou les concepts importés ne sauraient être évoqués de façon appropriée que par leurs noms d'origine. Par exemple, il n'y aurait pas d'autres termes que «*maotera*» et «*traktera*» pour désigner un moteur et un tracteur. Des termes d'emprunt comme «*alijebra*» «algèbre», «*kalera*» «scalaire», en mathématique, «*bailina*» «bail», «*dividandy*» «dividende», «*banky*» «banque», dans la langue des affaires (3), ont été retenus par les lexiques et dans la mémoire des usagers, parce que, monosémiques et monoréférentiels, ils ont été jugés intraduisibles et irremplaçables, donc préférables au choix discutable d'une traduction par tels ou tels termes indigènes. Il arrive même que certains emprunts «se remotivent» par leurs connotations particulières liées au prestige de la langue prêteuse !

---

3) RANDRIAMIZAH, Rabemananjara (1978).

## Les facteurs socio-affectifs et les associations mémorielles

Des facteurs socio-affectifs interviennent dans l'évocation des termes et, conjointement, dans le rappel des concepts, dont : les images poétiques et les métaphores actualisées dans la langue, les souvenirs laissés par les noms propres les valeurs sociales, nationales et culturelles que portent les termes créés, et les connotations.

Ils semblent précipiter un phénomène — certes, plus psychologique que linguistique, mais non négligeable — que d'aucuns appelleraient des associations mémorielles.

### *Les images*

Les lexiques techniques enregistrent un certain nombre d'images et de métaphores, qui établissent une relation, voire une analogie entre les termes et les choses nommées. Le terme «naviguer», par exemple, qui, d'après sa motivation étymologique, évoquait au départ le comportement d'un navire, puis s'applique successivement à l'avion et à la fusée dans l'espace, a pris un sens nettement métaphorique en informatique : «naviguer sur Internet» par Netscape évoque bien, à partir de la représentation du timon d'un bateau à voile ou d'une goëlette du XVI<sup>e</sup> siècle par un icône, l'image spatiale, adéquate pour la notion, d'un déplacement d'informations d'un point à un autre ou d'une orientation vers plusieurs points dans la conception d'un univers immense.

Que s'est-il passé ? Linguistiquement parlant, il s'est produit une série de transferts métaphoriques dans la langue technique. De métaphore en métaphore, on en est arrivé à créer des acceptions nouvelles de vocables déjà existants par le jeu des associations d'idées.

Même dans un langage aussi éloigné de la poésie que celui des mathématiques, on peut observer une utilisation de la métaphore en appliquant à un concept qui lui appartient un terme imagé emprunté à la langue commune.

Il s'agit, par exemple, de la traduction malgache *mampita* du qualificatif «transitif» appliqué à «une relation binaire R telle que  $x R y$  et  $y R z$  entraînent  $x R z$ ». En effet, *mampita* a originellement un sens bien concret, celui de «faire franchir la rivière». *Mampita* «transitif» (au sens mathématique) représente donc un cas métaphorique d'emprunts sociaux. Il y a eu spécialisation du terme, entraînant un déplacement de son aire sémantique, par changement et extension de sens.

Or, il semble que le sentiment et l'imagination créatrice soient intervenus dans ces associations d'idées pour contribuer à les cimenter et à les fixer dans la mémoire des usagers. Autrement dit, les métaphores, qui ont surgies dans l'esprit des terminologues à la suite d'associations d'idées, résultent essentiellement d'opérations mentales, lesquelles, en tant que telles, n'ont apparemment rien à faire avec le langage lui-même. C'est sans doute pour cette raison que Bachelard disait (paradoxalement) : «Toute langue scientifique est métaphorique».

À travers la métaphore et par le jeu des associations d'idées, il s'est alors effectué tout un montage de réflexes et d'habitudes mentales à identifier le mot et la chose, à associer le signifiant et le nouveau signifié du vocable : l'évocation du terme réveille dans l'esprit le souvenir de la notion ou du concept et, inversement, celle du concept l'image du vocable.

#### *Les noms propres*

Un deuxième facteur contribue à la mémorisation des termes et des notions : c'est le souvenir évoqué par les noms propres. Ces noms propres ont la propriété d'associer le souvenir de personnes aux notions désignées par leurs noms et de rendre ainsi présents dans la mémoire les signifiants et les signifiés voulus.

Ainsi, des concepts, des notions, des machines ou des ouvrages portent directement les noms de leurs créateurs : théorème de Pythagore, loi de Bernoulli, balance Roberval, véhicules de marque Peugeot, le grand Larousse Illustré, vernier, watt, faraday, newton, moteur Diesel, etc. Dans plusieurs langues, ces noms sont rendus soit sans modification de leur orthographe d'origine, soit translittérés selon le phonétisme de la langue d'accueil. Certains de ces noms propres sont même devenus des dérivés : en français, faradisation, galvanisation, pasteurien; en malgache, *fampimirana kariteziana* «équation cartésienne», *vory herimisiana* «groupe hermitien» (mathématique), etc.

Ce sont là des appellations acceptées internationalement, à quelques exceptions près. Elles ont été adoptées et gardées pour honorer et perpétuer la mémoire d'un personnage auquel on attribue la paternité de la notion ou de l'objet. Elles offrent l'avantage de ne poser aucun problème de traduction dans aucune langue, de ne rien apporter qui vienne déranger la notion abstraite ou le concept que ces termes supportent.

### *Les valeurs sociales et les connotations*

Le troisième facteur de facilitation du rappel des notions par l'évocation des termes, ce sont les connotations et les valeurs de la culture, au sens large, auxquelles sont liés les vocables, une fois enregistrés dans la langue technique.

Il y a des chances, en effet, que les connotations d'un terme permettent de mieux retenir un concept ou une notion. Un terme comme «tendon-d'Achille», par exemple, évoque suffisamment la notion correspondante par les connotations mythologiques qu'il comporte : il rappelle le souvenir du siège de Troie où Achille fut atteint mortellement au talon par une flèche empoisonnée, et aide à situer facilement ainsi l'organe désigné.

L'histoire des termes et des mots se reconstitue de cette façon par les témoignages des monuments linguistiques où ils sont attestés. C'est là aussi que des mots-clés ou des mots témoins, caractéristiques d'une société à une époque donnée, reflètent et rappellent des faits de civilisation.

A ce propos, le malgache usuel et technique offre un certain nombre d'images évocatrices du passé, successivement renouvelées par transferts sémantiques.

Un terme intraduisible en français, le mot *hasina*, par exemple, qui s'apparente de loin au «mana» identifié par les sociologues, pourrait donner une idée de ce que peut être le terme comme symbole matériel d'un fait spirituel mémorable dans les croyances profondes d'un peuple. Désignant tantôt un arbre sacré (*hasin'andriana*, espèce de dragonnier), tantôt des cadeaux offerts au souverain pour reconnaître sa suzeraineté, tantôt une vertu, une grâce surnaturelle, ce mot d'origine indonésienne a pris divers sens selon les époques et selon les contextes : c'est «l'obole» (*hasin'Andriamanitra* «le *hasina* de Dieu») dans les églises catholiques malgaches; c'est le caractère sacré, qui donne une force surnaturelle aux souverains, à l'époque de la royauté; c'est, appliqué aux hommes en général, une dignité personnelle ou une valeur, cette part de divin et de surnaturel que possède chaque être humain. Entrant en dérivation ou en composition syntagmatique, ce terme, fondé sur le protosémantisme «caractère sacré» ou «valeur», se trouve au centre d'un réseau d'associations différentes, prolongeables dans plusieurs directions. Par exemple, en linguistique malgache d'inspiration fonctionnelle, on en vient à traduire «monème» par *hasin-teny* (*hasina* + *teny* «mot») et

«phonème» par *hasim-peo* (*hasina* + *feo* «son») sans doute pour souligner leur caractère distinctif d'unités de première et de deuxième articulation.

C'est ainsi que des termes, porteurs de valeurs sociales, nationales ou culturelles, s'associent de façon naturelle dans la mémoire et, s'appelant les uns les autres, facilitent une assimilation rapide des notions. L'émission, l'audition ou la lecture de certains de ces mots à forte connotation historique ou de ces termes forgés par associations d'idées font surgir dans notre conscience, d'une façon plus ou moins précise, une partie de ces images qui se sont élaborées au cours de notre expérience. Le mot évoque l'expérience de certaines valeurs qui s'associent au signifié, et le souvenir de cette expérience évoque à son tour le mot ou le terme.

## NOTIONS ET MÉMOIRE DES TERMES

Cependant, l'évocation réciproque «terme/notion-concept» ne s'effectue pas toujours d'elle-même. Il arrive parfois que le concept ou la notion soient présents à notre esprit, mais que nous oublions les termes qui les désignent. Nous faisons alors, à la façon de l'ordinateur, une véritable recherche en mémoire des termes possibles correspondant au «signifié». C'est là que, en fonction de la capacité de notre mémoire des mots, les notions peuvent rappeler les termes qui les portent, sous certaines conditions : complexes ou nouvelles pour la langue qui doit se les approprier, elles ont besoin d'être dénommées par des termes qui se retiennent le plus facilement possible. Nous pensons ici aux unités syntagmatiques condensées en séquences figées et aux unités lexicales abrégées. C'est là aussi que se pose le problème de l'économie linguistique dans la désignation des notions rappelées.

### La mémoire des séquences figées

On dit souvent <sup>(4)</sup> que notre mémoire retient mieux les mots en groupes que les mots isolés : nous ne pourrions jamais ni conserver ni employer tous les mots que nous savons d'une langue, si nous devions les apprendre un par un, séparément.

---

4) Notamment, MARTINET, André (1970); GUIRAUD, Pierre (1964); MITTERAND, Henri (1960).

Or, il se trouve que beaucoup de notions ou de concepts techniques, plus ou moins complexes, sont désignés par des mots longs, ayant plus de deux syllabes : associations et groupements dans lesquels l'esprit fait entrer les mots. Ces groupements ont tantôt un caractère passager et se désagrègent presque aussitôt après s'être formés, pour être oubliés définitivement. Tantôt ils ont plus de fixité et sont d'autant plus stables qu'ils sont plus souvent répétés.

Les groupements les plus fugitifs semblent être surtout les syntagmes les plus longs, les syntagmes périphrastiques. L. Guilbert, par exemple, a montré comment le vocabulaire de l'aviation s'est sans cesse renouvelé de périphrases en périphrases jusqu'à ce que soit fixé et lexicalisé le terme «avion», à une époque où les objets n'avaient pas encore d'existence conceptuelle stable (5). En effet, la dérivation syntagmatique en périphrase n'est qu'une des premières étapes de la lexicalisation des termes techniques, bien connue également en malgache : certaines périphrases malgaches, qui s'appuyaient sur la traduction-calque d'une définition notionnelle ou d'une paraphrase du terme français (6) ont dû elles-mêmes être condensées en syntagmes moins longs pour être mémorisées et utilisables dans la pratique.

Les groupements les plus stables sont les unités complexes toutes prêtes, celles qui, formant des unités codées comme les proverbes, les locutions phraséologiques ou comme les collocations consacrées par l'usage, s'inscrivent indissociablement dans notre mémoire. Ce sont, par exemple, les lexies composées en français «langage-machine», «code-système», «carte de manipulation», «barre des tâches» (informatique), «machine à laver» (langue usuelle), et, en malgache, les composés amalgamés, résultats du figement des syntagmes, *savazoro* «diagonale» litt. «action de séparer en deux — angle», *haja mampitovy* «isomorphisme» litt. «respect (des structures) — qui compare, qui rend égal» (mathématique).

Ce sont des séquences figées, dont le signifiant et le signifié sont unis par un lien parfois arbitraire. Mais elles présentent l'avantage, une fois apprises, d'être inscrites dans le code de notre mémoire, qui doit les reproduire telles quelles dans le «discours» : les mots se

---

5) GUILBERT, Louis (1965).

6) Nous pensons ici à l'expression juridique «fanambadiana heverina ho ara-dalana kanefa tsy ara-dalana» litt. «mariage que l'on croit légal sans être légal» pour traduire «mariage putatif». Cette expression a été condensée en «anambadiana heverina ho ara-dalana» «mariage que l'on croit être (de bonne foi) légal».

regroupent dans notre esprit parce que leur association a en elle-même quelque chose qui nous frappe, qui nous a impressionné, soit par la sonorité esthétique, soit par le sens imagé, par exemple.

Du reste, les associations naturelles des mots semblent le mieux se graver dans notre esprit, sans qu'on ait besoin de faire l'analyse systématique de leurs composantes. Ainsi, dans une communication sur la malgachisation populaire (7), un lexicographe italien, le Révérend Père Profita, citait une foule de néologismes naturellement forgés par les paysans malgaches (8) pour désigner des notions qui leur étaient nouvelles : «bulldozer» *milina fitrongisana* litt. «machine-instrument pour fouiller» (évoquant l'action excavatrice du bulldozer), «thermomètre» *fangalan-kafanana* litt. «instrument pour prendre la température», «faucheuse» *hetim-bozaka* litt. «ciseaux des herbes», «antenne télescopique» (d'un poste récepteur) *fandri-peo* litt. «piège pour capter ou saisir - son», et «micro» *fanitaram-peo* litt. «moyen d'amplifier le son, la voix».

Il se trouve donc que, pour désigner une notion technique, souvent notre mémoire, selon sa capacité, semble emmagasiner mieux les unités syntagmatiques condensées en séquences figées que des mots courts et isolés.

### Les abrègements de syntagmes

Cependant, notre mémoire étant limitée et cherchant à économiser au maximum l'effort pour retenir les mots, on tend de plus en plus aujourd'hui à écourter la graphie des séquences figées et des unités syntagmatiques. Ainsi, pour les besoins d'une communication rapide, on désigne beaucoup de concepts ou de notions techniques en abrégeant leurs dénominations complexes : ce sont les procédés de troncation et de siglaison.

#### *La troncation*

Dans les langues techniques, l'économie linguistique réclame de plus en plus l'usage de vocables plus courts, par abrègement des séquences figées : on écourte les mots jugés trop longs en n'en retenant qu'une partie, juste ce qu'il faut pour que le signe ainsi

---

7) R.P. PROFITA (1977) : «À l'écoute des nouveaux mots créés par le peuple», communication à l'occasion du 75<sup>e</sup> anniversaire de l'Académie malgache.

8) Ces néologismes, étonnamment pertinents, n'ont jamais été enregistrés dans un dictionnaire et ils ont été recueillis, au hasard des conversations avec les paysans malgaches, par le R.P. Profita.

obtenu soit intelligible. La langue orale informatique, par exemple, comporte des abréviations, telles que «macro» (pour «macro-instruction») et «ordino» (pour «ordinogramme») (9). Et, probablement sous l'influence des progrès informatiques, on tend à abrégier la graphie des syntagmes figés comme pour les cas de «courriel» (courrier électronique), «télécel» (téléphone cellulaire) et «didacticiel», «minitel», «traducticiel», etc.

La brachygraphie, suivant la loi du moindre effort, permet alors d'obtenir un rendement plus économique dans la communication : des expressions de plus de sept syllabes et jugées trop longues encore sont réduites à deux, trois ou quatre syllabes, sinon elles seraient difficiles à retenir et à répéter dans les échanges langagiers, où l'on cherche surtout la rapidité et l'efficacité. Par exemple, en chimie, on combine «acetum» et «alcool» en «acétal». Dans le domaine de la pharmacie, un terme comme «calcivit» évoque bien, dans sa forme abrégée, la composition à forte teneur en calcium (gluconate, glycérophosphate de calcium) et en vitamines (B 1, D 2, C).

Il semble de plus en plus évident qu'actuellement, parmi les quantités considérables de vocables à emmagasiner dans notre mémoire, l'abrègement des unités syntagmatiques figées et la brachygraphie finissent par l'emporter sur les syntagmes longs.

Toutefois, dans le cas particulier du malgache, nous n'avons observé que rarement le phénomène de troncation des mots longs et des syntagmes dans les vocabulaires techniques et scientifiques. À notre avis, cela tient à deux raisons : d'une part, le malgache, comme toutes les autres langues austronésiennes, est une langue agglutinante et contient donc normalement beaucoup de formes parasythétiques; d'autre part, l'abrègement des unités lexicales complexes, fréquent dans la langue orale commune selon la loi du moindre effort, est rare dans la langue technique, qui est relativement récente, et qui, pendant la longue période d'essai dans la malgachisation, se voulait la plus explicite possible pour garder le contenu informationnel des termes.

### *La siglaison*

Toujours pour répondre au besoin de réduire l'énonciation linguistique à son volume le plus mince, et à celui de soulager notre mémoire, l'industrie, le commerce, l'administration et diverses

---

9) MARCELLESI, Christiane (1973).

technologies scientifiques produisent sans cesse des sigles, qui abrègent les expressions trop longues et trop complexes : S.A.R.L., S.N.C.F., T.V.A., H.T.M.L...

L'avantage des sigles et des acronymes, c'est sans doute l'économie de la langue, qui réduit le coût des expansions lexicales. Mais ils ne sont possibles que dans le milieu étroit des seuls «initiés», et ils perdent leur motivation initiale pour le grand public. En cas d'abus, la réduction des termes en sigles risque non seulement de rendre ces sigles ambigus, mais même totalement inintelligibles dans le discours.

### **Le problème de l'économie linguistique**

Se pose alors le problème de l'économie linguistique pour la mémorisation des termes : faudra-t-il conserver la forme longue des unités complexes pour ne pas surcharger la mémoire d'unités nouvelles à apprendre ? Ou faudra-t-il, au contraire, donner aux concepts une dénomination abrégée de la forme longue, afin de réduire au minimum l'effort physique et même mémoriel pour répéter le terme ?

Dans le premier cas, c'est la solution de l'économie paradigmatique, qui est proposée. On évite ainsi d'ajouter une nouvelle unité à la liste des termes que le locuteur doit retenir. Or, on sait que, souvent, les unités complexes, que sont les néologismes syntagmatiques dans les vocabulaires techno-scientifiques, sont des collocations, certes nouvelles, d'unités déjà anciennes : par exemple, en français, on peut supposer qu'il y a moins d'effort pour retenir l'unité complexe métaphorique «oreille-de-mer» que son synonyme plus court «haliotis» (zoologie), parce que les composantes du syntagme sont des mots usuels; en malgache, l'emprunt «*kômpa*» (compas) semble beaucoup moins motivé que son synonyme, l'unité complexe imagée *fanoritamparibolana* litt. «instrument pour tracer (*fanoritana*) un cercle (*faribolana*, lui-même dérivé de *faritra* «trace, limite» et de *volana* «lune»)), dont les éléments composants sont familiers aux usagers. Toutefois, il y a sans doute une limite à la longueur des unités complexes, que pourra supporter notre mémoire.

Dans le deuxième cas, l'économie syntagmatique devient une solution aux problèmes posés aux besoins communicatifs par l'usage des termes longs. En effet, on trouvera plus économique d'adopter une désignation courte (unités tronquées, sigles) : il est évidemment plus pratique de dire «courriel», comme «E-mail», que «courrier électronique» !

Mais il arrive que la forme abrégée de la forme longue perde sa motivation initiale : la forme tronquée comme «métro», qui subsiste de l'unité complexe «chemin de fer métropolitain», est souvent, en soi, une tranche antérieurement non significative de cette unité. Les termes abrégés par troncation ou par siglaison ne peuvent être motivés que pour ceux qui en connaissent la signification et l'usage dans un contexte donné et dans une situation particulière. En dehors de ce contexte, ou pour ceux qui ne connaissent pas ou ont oublié la signification initiale des abréviations, l'apprentissage devra se faire de nouveau par l'opération psychologique d'un montage séparé, mot par mot : comme pour le cas de tous les néologismes paradigmatiques courts, cela entraînerait alors un alourdissement du fardeau mémoriel, contrairement à l'effet escompté.

Il y a donc une alternative. Mais ce qui déterminera le choix final de la première ou de la deuxième solution sera la fréquence d'emploi du terme : si le concept ou la notion sont mentionnés avec une grande fréquence, il sera plus pratique d'adopter une dénomination abrégée; mais si, au contraire, il est plus rare qu'on désigne le concept ou la notion nommément, il sera plus économique de soulager la mémoire en utilisant la forme longue, en gardant le contenu informationnel des éléments de chaque lexie composée (car plus les unités du système sont nombreuses, plus chacune d'entre elles apporte d'information).

Il reste que, entre la solution d'une économie paradigmatique et d'une économie syntagmatique, il y a lieu de se demander comment sauvegarder l'efficacité ou le haut degré de disponibilité des termes dans telle ou telle situation particulière de communication. Car la notion de degré de disponibilité des termes correspond à la présence plus ou moins immédiate de ces termes et des concepts dans notre mémoire.

## CONCLUSION

Le problème de la mémoire des termes techniques revêt un double aspect : le mot est à la fois lieu de mémoire et objet de mémoire. En ce sens, les termes doivent rappeler à l'esprit les notions et les concepts, en même temps que, inversement, les notions et les concepts doivent rappeler les termes qui les désignent et qu'il faudra mémoriser.

D'une part, la dénomination peut permettre de se souvenir des notions, si elle est motivée étymologiquement, si elle est faite

d'emprunts complètement assimilés par les usagers, ou si elle est suffisamment évocatrice d'images, de valeurs sociales, culturelles ou historiques dans un contexte national où elle est profondément ancrée : c'est que «...tout passe par le signe linguistique, qui est symbole et mémoire» (Clas, 1993).

D'autre part, par la même occasion, la dénomination, une fois le concept fixé, devient un objet à mémoriser. Il reste à savoir sous quelle forme linguistique les notions et les concepts se présentent le mieux à l'esprit et quels termes se retiennent le mieux.

Ici, sont mis en relation la mémoire, la longueur phonique des mots, l'effort articulatoire pour les prononcer et pour les produire dans le discours : 1) les néologismes courts posent les problèmes du coût de l'emmagasinage dans la mémoire; 2) les lexies composées, groupements de mots formés par association mentale, semblent plus aisées à mémoriser que les mots courts immotivés; 3) on pense aussi à l'abrègement de ces syntagmes par souci d'économie; mais en cas d'abus, les abréviations comportent les mêmes inconvénients que les mots courts : la perte de la motivation.

On ramènera alors la possibilité de mémoriser les unités soit à leur répétition soit à leur degré de disponibilité : plus ces unités seront répétées, plus elles seront raccourcies pour des raisons de coût du point de vue de leur longueur phonique et de l'effort articulatoire ou mémoriel; et plus elles seront motivées et riches en information par leur composition originale, plus elles ont des chances aussi d'être disponibles dans notre mémoire.

Le problème de la mémorisation des notions techniques et de celle de leur support lexical amène donc à reposer le problème plus général du rapport entre fréquence, disponibilité et coût du vocabulaire.

## RÉFÉRENCES

- CLAS, André (1993) : «Formations et recherches en génie linguistique», *Universités*, magazine de l'agence francophone pour l'enseignement supérieur et la recherche, vol.14 n° 2, mai-juillet, p. 39.
- GOUGENHEIM, Georges, MICHEA, R., RIVENC, P. et A. SAUVAGEOT (1964) : *L'élaboration du français fondamental 1<sup>er</sup> degré*, Centre de recherche et d'étude pour la diffusion du français, Librairie Marcel Didier, Paris.
- GUILBERT, Louis (1965) : *La formation du vocabulaire de l'aviation*, Paris, Larousse, tome 1.
- GUIRAUD, Pierre (1964) : *L'étymologie*, Paris, PUF, coll. «Que sais-je?».
- MARCELLESI, Christiane (1973) : «Le langage des techniciens de l'informatique : quelques aspects de leur vocabulaire écrit et oral», *Langue française*, février, n° 17, pp. 59-71.
- MARTINET, André (1970) : *Éléments de linguistique générale*, Paris, Librairie Armand Colin.
- MEL'CUK, Igor, CLAS, A. et A. POLGUÈRE (1995) : *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve, coll. «Universités francophones», Éditions Duculot.
- RAJASPERA, Raphaël (1996) : «Pour la création d'une terminologie scientifique et technique en malgache», *Meta*, Montréal, Les Presses de l'Université de Montréal, vol. 41, n° 3, pp. 435-458.
- RANDRIAMIZAH, Rabemamanjara (1978) : *Dictionnaire des affaires : malagasy-français-english*, Publications INPF, Antananarivo, Nouvelle Imprimerie des Arts Graphiques.



## Les emprunts : des mots qui se souviennent ?

Ambroise QUEFFÉLEC

*Université de Provence, Institut National  
de la Langue française-CNRS, Aix-en-Provence, France*

*«La mémoire des mots, cette grande paresseuse qui  
refuse de rêver»*

*Gaston Bachelard, L'air et les songes (Essai sur  
l'imagination du mouvement)*

Les mots ont-ils une mémoire ? La réponse réservée sur leur pouvoir mnémonique que donne G. Bachelard, un des grands philosophes inspirés du XX<sup>e</sup> siècle, justifie qu'on puisse se poser la question dans un colloque qui, justement, est consacré à la mémoire des mots. Nous limiterons ici nos analyses aux emprunts c'est-à-dire aux «forme(s) d'expression qu'une communauté linguistique reçoit d'une autre communauté» (Deroy, 1956 : 18), et plus spécifiquement aux emprunts lexicaux puisque ceux-ci sont les plus emblématiques <sup>(1)</sup> de ces

---

1) Sans doute la notion d'emprunt dépasse-t-elle le cadre limité du lexique, conformément à la définition de Dubois *et alii* (1973 : 188) : «Il y a emprunt linguistique quand un parler A utilise et finit par intégrer une unité ou un trait linguistique qui existait précédemment dans un parler B et que A ne possédait pas; l'unité ou le trait emprunté sont eux-mêmes appelés emprunts». Deroy (1956 : 21) note cependant dans sa thèse : «On entend par "emprunt" le seul emprunt de mot ou emprunt lexical. Il est en effet le plus fréquent, le plus apparent, le plus largement connu. On ne doit pas cependant oublier que les langues s'approprient aussi des sons, des façons d'accentuer, des traits morphologiques, des sens, des tours syntaxiques. Mais c'est de l'emprunt de mot que procèdent, en réalité, tous les autres, sauf peut-être certaines influences syntaxiques.

transferts d'un système linguistique à un autre. Nous nous demanderons si dans le passage d'une langue-source L1 à une langue-cible L2 un lexème conserve dans son usage en L2 des marques de sa langue de provenance. Sans doute pourrait-on dire que dans la transplantation d'une langue à une autre il ne s'agit plus du même mot et notre questionnement serait sans objet. Si l'on pense, comme le sens commun et la tradition nous y incitent, qu'il y a bien continuité et identité, au moins partielles, du lexème translaté, on est en droit de se demander si le mot voyageur a une mémoire et porte trace de ses origines. Nous nous intéresserons ici plutôt au signe linguistique lui-même qu'aux représentations auxquelles il donne lieu. Par ailleurs, nous focaliserons notre attention sur l'emprunt dans les situations de contact de langue chez des sujets bilingues et nous questionnerons en particulier les variétés de français régionales du Sud (essentiellement africaines) sur lesquelles nous travaillons depuis de nombreuses années. En outre, nous n'envisagerons que les emprunts intégrés <sup>(2)</sup> que nous différencierons essentiellement sur des critères de fréquence d'emploi dans la langue d'accueil <sup>(3)</sup> des simples «xénismes» ou «périgrinismes», ces «mots sentis comme étrangers et en quelque sorte cités» (Deroy, 1956 : 224). Notre interrogation concernera la mémoire des mots empruntés aux différents niveaux où son intégration est censée s'effectuer : phonétique et phonologique, graphique, morphosyntaxique et sémantique.

---

Le mot emprunté apporte avec lui des éléments grammaticaux qui secondairement se développent de façon autonome. C'est pourquoi une étude de l'emprunt est et doit être, avant tout, lexicologique». Le même auteur, plus de vingt ans plus tard, réaffirme d'ailleurs «qu'on peut maintenir que tout emprunt est lexémique» (Deroy, 1980 : 12).

- 2) Nous sommes évidemment conscient du caractère arbitraire et largement subjectif de cette différenciation entre «véritables emprunts» et xénismes auxquelles se heurte tout linguiste décrivant le lexique d'une langue en contact. Il existe en fait un continuum entre emprunts «naturalisés» et périgrinismes qui rend toute distinction précaire et frappe de caducité toute critériologie permettant de les différencier. Ce flou distinctif explique partiellement les écarts que l'on observe chez les différents descripteurs dans le dénombrement des emprunts d'une langue à une autre. Pour ne prendre que le cas des emprunts du français central à l'arabe, il peut varier de 257 mots (Ben Smail, 1994 : 151) à 269 (Guiraud, 1965) ou 776 (Benhamouda, 1996 : 16).
- 3) Nous mettrons ici entre parenthèses le problème de l'homogénéité de la langue d'accueil et de sa «pureté» chez des locuteurs africains bilingues pour qui le mixage de codes est la modalité d'expression la plus naturelle. Comme le note Garafanga (1987 : 24) (et à sa suite Thiam, 1994 : 12) «... le code switching est le seul vernaculaire du bilingue, c'est-à-dire la variété dans laquelle le bilingue se trouve le mieux à l'aise, le passage à d'autres variétés, standards ou vernaculaires aux monolingues, étant dicté par la présence d'un élément étranger, *i.e.* la formalité de la situation ou la présence d'un monolingue». Cette existence prégnante de franwolof, franc-sango ou francarabe rend encore plus délicate l'identification des emprunts intégrés.

## 1. — LA MÉMOIRE PHONOLOGIQUE ET PHONÉTIQUE

Trois hypothèses sont possibles quant au traitement de la face sonore de l'emprunt :

— La première postule la totale intégration des emprunts par rapport au système phonologique de la langue d'accueil : cette intégration aux plans phonologique et phonétique constitue dans la lexicologie traditionnelle l'un des critères les plus pertinents pour différencier mots naturalisés (dont la forme sonore respecterait le système de la langue emprunteuse) et simples citations (conservant phonétiquement les traits de la langue d'origine). Ainsi, lorsqu'il étudie les emprunts au français du sango, Taber (1964) prend comme principal critère de sélection de ces emprunts leur degré d'intégration au système phonologique de cette langue nationale et officielle centrafricaine. La tradition lexicographique va d'ailleurs dans le même sens puisque les dictionnaires, sauf exception, proposent pour les emprunts des transcriptions conformes aux règles phonétiques du code d'accueil. Dans le *Dictionnaire sango-français*, L. Bouquiaux *et alii* (1978 : 40), partant du principe qu'«en ce qui concerne les emprunts à une langue étrangère, il est bien connu qu'un locuteur, à moins qu'il ne dispose d'une excellente connaissance de ladite langue [étrangère], les adapte naturellement à la phonologie de sa propre langue», transcrivent et graphient ces emprunts conformément au système phonologique du sango sans se préoccuper de leurs éventuelles variations de prononciation en fonction du contexte énonciatif et de la maîtrise que les locuteurs possèdent éventuellement du français (4). Tel est encore le cas du *Dictionnaire Universel de la francophonie* (à paraître prochainement) qui, pourtant novateur sur de nombreux points, adopte pour certains emprunts une transcription phonétique en tous points semblable au système du français central, et ce de façon souvent artificielle (5).

---

4) De manière un peu abrupte, les auteurs (1968 : 42) refusent «de s'aligner sur les propositions du *Mémoire des Sociétés Bibliques* qui établit une subtile distinction entre les emprunts qu'il considère comme totalement intégrés et ceux dont la prononciation varie en fonction du degré de connaissance du français». Pour justifier leur position, ils allèguent que dans ce dernier cas il s'agirait de simples citations de la langue d'emprunt («celui qui a une bonne connaissance de cette langue les citant en quelque sorte») et évacuent de façon quelque peu sommaire le problème des multi-prononciations possibles des mots empruntés.

5) Les incertitudes quant à la prononciation de mots comme *khalife* obligent les rédacteurs à proposer deux transcriptions [kalif] et [xalif], seconde transcription qui est plus conforme à l'usage africain mais qui introduit subrepticement un [x] étranger au français central (phonème non répertorié d'ailleurs dans les «principaux signes diacritiques» de la page XII de l'ouvrage).

— Une deuxième hypothèse, symétrique de la première, affirme la totale indépendance de la forme sonore des emprunts par rapport à la langue d'accueil. Les emprunts conserveraient leurs propres jeux de phonèmes et leur représentation phonétique resterait conforme à la langue d'origine. Cette idée de systèmes phonologiques coexistants a été soutenue par Fries et Pike qui, à propos des emprunts du matazec à l'espagnol, relèvent par exemple que le mot *siento* conserve une représentation phonétique [siento] alors que les règles phonotactiques de la langue d'accueil exigeraient une forme [siendo]. Ce respect absolu de la prononciation de la langue d'origine est un phénomène rare dans les corpus de français en Afrique. Il s'observe surtout au Maghreb pour les vocables concernant le domaine religieux où l'adaptation à la langue-cible ne se produit que rarement, la déformation sonore de termes sacralisés par le Coran paraissant constituer un traitement irrespectueux de la langue de la Révélation assimilable à un blasphème.

— Une troisième hypothèse, enfin, «affirme un chevauchement partiel entre les phonologies du vocabulaire original et celui de la langue d'accueil, en sorte qu'un certain nombre de règles de cette dernière sont activées ou bloquées. Cette hypothèse présente l'avantage de rendre compte du fait que tous les mots susceptibles d'être empruntés ne le sont pas, ou que à «égalité de signifié», seuls certains signifiants peuvent être incorporés dans une langue d'accueil» (Durand-Deska et Durand, 1994 : 80).

C'est cette dernière hypothèse qui fait une part assez belle à la mémoire du signifiant qui semble la plus conforme à la réalité de nos observations de terrain. Si l'on prend l'exemple des emprunts du français à l'arabe dans le contexte maghrébin, on observe que bon nombre de termes conservent dans le français local leur prononciation arabe. Ainsi Gaadi (1995 : 146) observe que certaines lexies empruntées «ont une seule prononciation, conforme au phonétisme arabe : ce sont surtout les lexies qui contiennent le phonème [x] orthographié *kh* comme *cheikh* [ʃex], *fekhar* [fexar], *khaddar* [xadar], *makhzen* [maxzen], *mokhzani* [moxazni]». D'autres vocables, selon le même auteur, adoptent «une double prononciation : l'une francisée, l'autre conforme au phonétisme de l'arabe» comme *casbah* [kasba] ou [qasba], *cadi* [kadi] ou [qadi]. Benzakour (1995 : 71), en partant des mêmes observations, note cependant que pour certains mots au moins, «les phonèmes étrangers au système français cèdent peu à peu le pas aux phonèmes voisins, plus proches de la langue d'accueil, du point de vue du mode et/ou du point articulatoire. La consonne postpalatale uvulaire [q] de [qasida] («poème chanté») ou de [moqadem] («chef de quartier») par exemple, commence à être remplacée par sa voisine postpalatale [k]. La consonne pharyngale

[ ' ] et/ou [c] laisse place à la voyelle [a] ou disparaît (cf. *adoul* «notaire traditionnel» ou *alem* «savant en théologie»).»

## 2. — LA MÉMOIRE GRAPHIQUE

Dans leur processus d'intégration dans la langue d'accueil, les emprunts se trouvent confrontés au choix d'une graphie : ici encore se pose la question de la mémoire de l'emprunt puisque la graphie adoptée peut témoigner ou non de sa forme première dans la langue d'origine. Lorsque L1 et L2 utilisent des systèmes de codage graphique sans rapport (par exemple, langues européennes et arabe) la mémoire ne peut jouer ou ne joue que très indirectement (cf. cependant *infra*) ; en revanche, lorsque les systèmes de transcription graphique de L1 et L2 recourent à un code commun ou partiellement commun, comme l'alphabet latin, l'adoption d'une graphie dans L2 implique une prise de position par rapport à la graphie en usage dans L1. Divers traitements graphiques se rencontrent.

— La graphie dans la langue d'accueil peut n'avoir aucun rapport avec son homologue dans L1 : tel est le cas, semble-t-il, pour l'orthographe officielle de la plupart des langues africaines de transcription graphique assez récente. Établi par des spécialistes et des linguistes, leur système orthographique est assez étroitement calqué sur leur système phonologique. Le rapport entre l'écrit et l'oral paraît surtout régi par la seule fonction phonographique puisque chaque lettre fonctionne comme un graphème phonogrammique représentant une réalisation phonique dans un rapport de bijection : la graphie est alors étroitement phonologique : comme le notent pour le sango Bouquiaux *et alii* (1978 : 40), «pour la notation des emprunts faits au français par le sango, il me paraît donc nécessaire de les transcrire en se conformant à la réalisation du locuteur sango». La même idée est partagée par Diki-Kidiri (1977 : 13-22) qui propose, pour les emprunts comme pour le reste du vocabulaire sango, de substituer une graphie phonologisante aux orthographes plus anciennes sous influence latine ou française instituées autrefois pour le culte par les clergés catholiques et protestants (qui adoptaient d'ailleurs des systèmes de transcription différents).

— La graphie de la langue d'origine peut inversement être intégralement respectée : c'est par exemple le cas en français d'un certain nombre de termes d'origine anglaise comme *week-end* ou *football* (alors que *foutebaul* ou *ouiquinde* auraient été tout aussi valides) ou d'origine allemande comme *schlague* (auquel on eût pu substituer *chlague*). C'est à cette thèse que se rattache N'Sial (1993 : 84) lorsqu'il affirme, de manière

un peu trop abrupte à nos yeux, que «la plupart de ces lexies [emprunts du français zairois aux langues locales] ne subissent ni des modifications de sens ni des modifications de forme».

— Le cas le plus fréquent est cependant celui d'une graphie mixte qui prend en compte à la fois la graphie de l'item dans la langue d'origine et le système de transcription de la langue d'accueil. Cette hybridation graphique est d'ailleurs le lot commun de la plupart des emprunts du français régional aux langues africaines en contact. Elle est à l'origine de la surabondance des variations observées dans la graphie en français de ces emprunts, variations qui ont été relevées par de nombreux commentateurs et dont une analyse fine a été amorcée dans le cadre du projet Orthaf (malheureusement avorté) par Delaporte, 1993, pour le français du Sénégal et par Edema et Nduku (1993) pour le français congolo-zairois. Cette multiplicité graphique, outre les problèmes de normalisation qu'il suscite pour la bonne intégration des emprunts, est parfois à l'origine d'erreurs et peut créer une insécurité linguistique chez le lecteur de L2 qui ne connaît pas le code L1 et ignore quel mode de lecture il doit adopter. Ainsi, après avoir relevé la multigraphie *tiédo*, *thiedo*, *tyédo*, *ceddo* du terme d'origine wolof passé en français, désignant le «guerrier, mercenaire, mécréant», Dialo (1990 : 65), note qu'il y aurait sans doute lieu de chercher à réduire ce foisonnement formel anarchique qui peut être une source d'erreurs pour les «non-initiés» et relève «chez un coopérant français, des aberrations —de lecture», telles que par exemple [ksesal] pour *xeesal* ou [sedo] pour *ceddo*».

Dans le cas des emprunts à l'arabe, Morsly (1995 : 43-44) constate que la pratique graphique de la presse algérienne de langue française vis-à-vis des emprunts «semble caractérisée par un manque de systématisme et une hésitation impressionniste commandée par le libre-arbitre des journalistes». Les graphies adoptées, dans leur asystématisme, tiennent compte à la fois de la prononciation réelle (avec des variantes et des approximations liées à l'instabilité de la forme sonore) et de conventions qui avaient plus été plus ou moins officiellement prescrites à l'époque de la colonisation. Adoptées sous l'influence des premiers arabisants, ces conventions qui se sont imposées avec le temps se retrouvent «dans les transcriptions utilisées aujourd'hui : *gh* pour la fricative uvulaire sonore [d], *kh* pour la fricative uvulaire sourde [x], *r* pour la vibrante apicale [r]». Les recherches effectuées sur les autres variétés de français utilisées au Maghreb (Gaadi, Benzakour pour le Maroc, Ould Zein pour la Mauritanie, Naffati et Gheribi pour la Tunisie) confirment qu'ailleurs une identique *chakchouka* graphique prévaut pour la graphie des mots d'origine arabe.

Pour les emprunts du français parlé en Afrique noire, les mêmes tâtonnements s'observent. Comme le montre le corpus de l'inventaire centrafricain de Queffélec (1997), les scripteurs hésitent entre les graphèmes *c*, *k*, voire *qu* pour transcrire [k] (d'où *coco* et *koko*, *cola* et *kola*, *kinkékiba* et *quinquéliba*, etc.), *u* et *ou* pour transcrire [u] (d'où *koundi* et *kundi*, *gbalukuma* et *gbaloukouma*, etc.), *s* ou *ss* pour transcrire [s] intervocalique (d'où *kissolo* et *kisolo*, *sissongo* et *sisongo*, etc.).

### 3. — LA MÉMOIRE MORPHOLOGIQUE

Le problème de la mémoire des emprunts se pose aussi au niveau morphologique dans le choix des catégories morphologiques adoptées par le terme voyageur dans la langue emprunteuse. Comme la très grande majorité des emprunts sont des substantifs ou des adjectifs, ce choix est particulièrement probant pour ce qui est des catégories du genre et du nombre.

Pour le genre, l'emprunt est doté d'une mémoire forte puisqu'il conserve généralement son genre d'origine; ainsi, pour les emprunts à l'arabe du français parlé en Algérie, Smaali (1994 : 31) relève que «le déterminant varie en fonction de la classe masculin/féminin à laquelle appartient le lexème en arabe : *le darki*, «gendarme», *la wilaya*, «département, préfecture», *la daïra* «sous-préfecture», *la qacida* «texte poétique ancien». Mêmes constatations chez Benzakour (1995 : 72) pour qui dans le français du Maroc, «le genre des emprunts correspond, dans la majeure partie des cas, à celui de la langue d'emprunt (cf. *le wali* «super-préfet», *la fetwa* «consultation juridique fondée sur la charia»). Cependant certains flottements sont possibles selon la même auteure qui relève des hésitations sur le genre (*djellaba* et *caïdat* employés indifféremment aux deux genres) (6). Gaadi (1995 : 148) confirme cette fidélité de l'emprunt au genre qu'il avait à l'origine et ne signale qu'une exception, *califat* qui «féminin en arabe devient masculin en français». Le chercheur note d'ailleurs que pour les référents sexués les deux formes arabes, masculin et singulier peuvent coexister : *chérif* masculin «descendant mâle du Prophète», *chérifa* féminin «descendant de sexe féminin du Prophète» (7).

6) Même constat chez Ould Zein (1995 : 171) qui note que certains emprunts sont employés aux deux genres : *le/la jihad*; *le/la quinze*.

7) On pourrait ajouter pour l'Algérie *cheb* masculin / *cheba* féminin «jeune chanteur / chanteuse de raï». Cette dualité de formes correspondant à ses différences de sexe chez les référents soulève des problèmes complexes : ainsi, pour des raisons culturelles, *cheikh* substantif polysémique au masculin qui possède au moins quatre

En ce qui concerne le nombre, le choix de la langue emprunteuse est plus aléatoire : trois possibilités existent :

— Elle peut respecter la variation morphologique de la langue d'origine : ainsi on relève des formes distinctes opposant singulier et pluriel pour les substantifs à genre unique : par exemple dans les français d'Afrique du Nord se rencontrent les oppositions empruntées à l'arabe du type *un chahid / des chouada* «martyr(s)», *une souika / des souikat* «petit(s) marché(s)». De même dans les français subsahariens, les affixes de classe spécifiques du singulier ou du pluriel de certaines langues africaines prêtes pour se maintenir : ainsi, pour le signifiant «Blanc(s)», on relèvera les oppositions : *un muzungu / des bazungu* (fr. du Burundi, cf. Frey, 1996 : 25), *un mundele / des mindele* (fr. du Congo, cf. Queffélec-Niangouna, 1990 : 219) *un toubab / des toubabou* (fr. du Sénégal), etc. Pour les substantifs qui connaissent déjà une opposition de genres, quatre formes distinctes pourront distinguer masculin singulier, féminin singulier, masculin pluriel : féminin pluriel : ce paradigme se rencontre par exemple pour *cherif* et *cheikh* (masc. sing.) / *cherifa* et *cheikha* (fém. sing.) / *chorfa* et *chioukh* (masc. plur.) / *cherifate* et *cheikhate* (fém. plur.).

— La langue emprunteuse peut faire fi de la langue source et distinguer singulier du pluriel conformément à ses propres règles : ainsi l'opposition *-/-s* servira à opposer singulier et pluriel dans certains mots empruntés à l'arabe dans les français du Maghreb : *une mélehfa / des mélehfas* «voile(s) traditionnel(s) des femmes maures»; *la médina / les médinas* «quartier(s) populaire(s)», *un tagine / des tagines* «préparation(s) culinaire(s)». L'oubli des règles de la langue d'origine peut conduire à prendre comme base la forme du pluriel de la langue d'origine, d'où l'opposition relevée en français de Mauritanie *un abide / des abides* «esclave(s)», alors que *abid* est le pluriel de *abd* (8). Cette mise en conformité de l'emprunt avec le système de la langue-cible est particulièrement fréquent chez les locuteurs qui ne connaissent pas les

---

sens dans le français marocain «chef religieux, savant» / «personne âgée vénérable» / «fonctionnaire du Ministère de l'Intérieur responsable d'une tribu sur le plan administratif» / «chef d'orchestre», ne connaît le féminin *chikha* que pour la dernière acception : «femme qui danse et chante avec un orchestre». Inversement *negafa*, substantif féminin arabe passé en français et évoquant un métier traditionnellement dévolu aux femmes qui font profession de maquiller et de vêtir les mariées pendant leurs noces, ne se masculinise dans la presse marocaine que sous la forme hybride (et peut-être humoristique) *négafa-boy* pour référer à l'homme accidentellement chargé des mêmes fonctions.

8) Le français standard commet une «faute» semblable lorsqu'il oppose *un Touareg / des Touaregs* en prenant comme base *touareg* qui est en fait déjà le pluriel de *targui*.

règles de la langue-source ou n'en ont qu'une connaissance réduite. Ainsi les emprunts à l'arabe relevés par Lanly (1960) ou Duclos (1991) dans leurs descriptions du français des Pieds-Noirs, adoptent une variation morphologique en genre et nombre ignorant presque toujours les pratiques de la langue d'origine.

— Enfin le pluriel de l'emprunt peut adopter une forme hybride, compromis plus ou moins arbitraire entre les systèmes morphologiques des langues en contact. Les emprunts d'origine arabe peuvent adopter le pluriel arabe (ou plus exactement une transposition approximative de celui-ci) affecté à l'écrit du -s caractéristique du pluriel français. On aura ainsi des couples :

*un alem / des oulémas / ulémas*

*un cheikh / des chioukhs*

*un chérif / des chorfas.*

Dans la pratique, certains mots peuvent combiner les trois possibilités de marquage du pluriel, en fonction de l'inspiration du locuteur ou du scripteur : Morsly (1995 : 42), note que dans le français de la presse algérienne bon nombre d'emprunts à l'arabe emploient les trois systèmes de marquage selon le modèle :

	<i>les + plur. ar.</i>	<i>les + plur. fr.</i>	<i>les + plur. ar. + plur. fr.</i>
<i>wilaya</i>	<i>wilay-ate</i>	<i>wilaya-s</i>	<i>wilay-ate-</i>
<i>smoudjahid</i>	<i>moudjahid-ine</i>	<i>moujahid-s</i>	<i>moudjahid-ine-s</i>
<i>ksar</i>	<i>ksour</i>	<i>ksar-s</i>	<i>ksour-s</i>
<i>souk</i>	<i>aswak</i>	<i>souk-s</i>	<i>aswak-s</i>
<i>fellah</i>	<i>fellah-ine</i>	<i>ellah-s</i>	<i>fellah-ine-s</i>

De même, en français d'Afrique subsaharienne, le pluriel de *taximan*, mot hybride <sup>(9)</sup> combinant le suffixe *-man* à la base *taxi-*, connaît les trois séries de formes *taximen*, *taximans*, *taximens*, tout comme *toubab* qui admet les trois pluriels *toubabou*, *toubabs*, *toubabous*.

9) On ne peut parler ici stricto-sensu d'emprunt puisque *taximan* n'existe pas en anglais (qui utilise *taxi-driver*). *-man* fonctionne bien comme un suffixe emprunté à l'anglais, et se trouve attesté dans d'autres lexèmes de formation voisine comme *mbaoman*, *roseauman* (fr. Centrafrique). On relèvera que ce suffixe *-man* entre en concurrence avec des suffixes français plus usuels comme *-eur* (*taxieur* courant en Algérie) ou *-iste* (*taxiste* bien attesté en Tunisie).

#### 4. — MÉMOIRE SYNTAXIQUE

C'est sans doute dans le domaine syntaxique que les emprunts se révèlent les plus amnésiques vis-à-vis de leur langue-source en adoptant prioritairement les règles d'accord de la langue d'accueil.

Par exemple, comme le note Gaadi (1995 : 148) les adjectifs d'origine arabe emprunté par le français parlé au Maroc «s'accordent en genre et en nombre selon le système français» ainsi que l'attestent les exemples suivants :

*un poulet beldi / une poule beldie / des poulets beldis;  
des genres badaouies; des performances gnaouies, etc.*

De même, les emprunts d'origine substantivale s'intègrent dans le syntagme nominal en suivant les règles de combinaison de la langue d'arrivée : ainsi la très grande majorité des substantifs arabes employés dans les français du Maghreb adoptent le schéma canonique du SN de la langue d'accueil : actualisateur fr. + substantif d'or. ar., d'où en fr. de Mauritanie *les moughataas, les kebbas, des hadiths, ces mélehfes, etc.* Lorsque le substantif entre dans un syntagme plus complexe où il est subordonné régi par un autre substantif, les règles de la langue d'accueil s'imposent encore au détriment de celles de la langue d'origine : *les enfants de chouhada* («de martyrs») *la femme du wali* («du préfet»), etc.

Cependant, il peut se faire que l'emprunt conserve quelques-uns des traits syntaxiques de la langue d'origine : par exemple, certains substantifs d'origine arabe peuvent conserver dans la phrase française les modalités (actualisateurs) de la langue de départ comme les articles *el* ou *al* : Gaadi cite des exemples de son corpus de presse francophone marocaine :

*... la jalabia, el balgha et la barbe sont de rigueur (L'Opinion, 4 : 4/92)*

*Les pauvres pèlerins ont dû dépenser l'argent qu'ils avaient mis de côté pour el héj (= «le pèlerinage») (L'Opinion, 19/5/92).*

Parallèlement dans son corpus de presse algérienne, Smaali (1994) relève :

*«El barani» est indésirable lorsqu'il n'a rien à proposer (= «l'étranger»).* (Algérie-Actualité, 13/7/93)

*il a commencé à travailler avec «el khaoua» (= «avec les frères»).* (El Moujahid, 21/4/93)

«*El daâwa*» prêche la bonne parole (= «l'appel à l'Islam»). (*Algérie-Actualité*, 13/7/93) (10).

Par ailleurs, certains emprunts peuvent concerner des unités supérieures au mot et conserver leur forme de la langue d'origine. Tel est le cas de synthèmes comme *souk-el-fellah*, littéralement le «marché du paysan» qui désigne dans la réalité algérienne un grande surface commerciale, de *aïd-el-kébir*, littéralement la «grande fête» ou «fête du mouton» et des autres termes du même paradigme (*aïd-el-adha*, *aïd-el-fitr*, *aïd-es seghir*), ou de *hizb franca* («parti de la France»). En Afrique subsaharienne, on rencontre aussi de tels synthèmes, comme, en français de Centrafrique, *koli-gala* (littéralement «homme du marché» d'où «commerçant sédentaire masculin») par opposition à la *wali-gala* (littéralement «femme du marché», d'où «commerçant sédentaire féminin»), *kota zo* («grand homme», d'où «personnage important»), *mama mokonzi* («femme chef», d'où «femme chef de quartier»), *kongo yasika* («nénuphar») etc.

Ces entorses au respect des règles de fonctionnement de la langue d'accueil restent cependant peu fréquentes, sauf peut-être pour les titres de journaux (11) : les emprunts dans l'ensemble se conforment aux règles de fonctionnement de la langue d'accueil.

## 5. — MÉMOIRE SÉMANTIQUE

Au niveau sémantique, les emprunts conservent dans les français régionaux d'Afrique une mémoire assez forte de leur sémantisme dans leur langue d'origine. Ce fort pouvoir de réminiscence tient au fait que les locuteurs de ces variétés sont dans leur grande majorité bilingues. La connaissance de la langue-source ne manque pas de donner aux emprunts

---

10) Ces trois dernières attestations posent cependant problème : si *barani*, *daawa* et *khaoua* sont par ailleurs entrés comme emprunts dans le français parlé en Algérie, comme le prouveraient d'autres contextes, les guillemets mis par les journalistes semblent signifier qu'il s'agit plutôt de citations de l'arabe dans un discours en français, procédés fréquents dans le discours mixte.

11) Beaucoup de journaux francophones africains ont des titres empruntés à des langues en contact, par ex. *El Acil*, *El Moujahid*, *El Watan*, *El Manchar* (Algérie), *Al Bayane*, *Al Maghrib*, *Al Asas* (Maroc), *Al Bayane*, *Al Joumhouriya*, *Al Moustaqbal*, *Chaab* (Mauritanie), *E-lè Songô*, *Be Africa Sango*, *Lisoro*, *Vouma* (Centrafrique), *Etumba*, *Mweti* (Congo-Brazzaville), *Elima*, *Salongo*, *Taïfa* (Congo-Kinshasa), *Jamana*, *Faso*, *Yeelen*, *Kabako* (Mali), etc. Il y aurait lieu de s'interroger sur ces les raisons de ces choix, révélateurs selon nous d'une insécurité linguistique sur le statut de la langue d'écriture. Quoi qu'il en soit, ces titres fonctionnent comme des noms propres et à ce titre s'intègrent *in extenso* — y compris avec l'article arabe le cas échéant — dans la phrase française.

une richesse sémantique qu'ils n'auraient pas chez des locuteurs monolingues. Ainsi, les emprunts à l'arabe dans les français d'Afrique du Nord conservent une partie au moins de la polysémie qu'ils possèdent en arabe. Un terme comme *baroud* garde par exemple en français du Maroc les différentes acceptions qu'il possède en arabe :

1. «poudre à canon ou à fusil»
2. «canon»
3. «combat»
4. «combat désespéré mené pour l'honneur».

De même *bled* cumule les divers sens de :

1. «village»
2. «zone reculée»
3. «terroir», «campagne», «région»
4. «pays», «nation».

Le bilinguisme des locuteurs confère à ces emprunts une richesse dénotative que ne leur connaît pas le français central qui les a empruntés à date plus ancienne à l'argot des troupes coloniales (cf. Christ, 1991 : 205-210 et 236-243) et a réduit leur polysémie à une simple monosémie.

D'ailleurs, même lorsque le terme à l'origine monosémique, conserve en français son unique sens, il possède chez les bilingues des connotations spécifiques propres à la langue d'origine. Les mots d'origine arabe comme *flouss* «argent», *kaouah* «café», *toubib* «médecin», *barda* «bagage, sac» ou *maboul* «fou», ne possèdent pas dans l'usage maghrébin les connotations péjoratives, familières ou ironiques qu'ils possèdent dans le français central. Ces différences de connotations rendent souvent délicat le maniement de ces termes par les locuteurs lettrés (en particulier journalistes) connaissant les connotations que leur attache le français de France. Ce double registre connotatif est souvent responsable du sentiment d'insécurité linguistique généré par leur usage.

## CONCLUSION

Notre interrogation sur le pouvoir de mémoire des emprunts dans les français régionaux d'Afrique montre qu'en dépit de leur intégration dans la langue d'accueil, ils conservent une partie des traits qui les caractérisaient dans la langue d'origine. Aux plans sémantique et morphologique, ces réminiscences restent assez fortes et prouvent qu'au moins chez des locuteurs bilingues les emprunts sont des mots qui se

souviennent. Ce constat nous amène à nous interroger sur la pertinence des critères mises en place par la lexicographie traditionnelle pour opérer la distinction entre le xénisme et le véritable emprunt que cette lexicographie a trop souvent tendance à évaluer du seul point de vue de la langue emprunteuse sans prendre en compte le sentiment linguistique des locuteurs. Il nous incite aussi à nous interroger sur les raisons de ces emprunts dont bon nombre ne correspondent pas à une nécessité absolue. Il nous invite enfin à nous interroger sur les perceptions que les usagers bilingues ont de ces emprunts puisque la mémoire des mots se confondant un peu avec la mémoire de ceux qui les utilisent, ce sont eux qui sont les vrais porteurs de mémoire.

## RÉFÉRENCES

- BENHAMOUDA, B. (1996) : *L'origine arabe de la langue française*, Paris, Dialogues Éditions, 130 p.
- BEN SMAÏL, M. (1994) : *Dictionnaire des mots français d'origine arabe*, Tunis, Société Tunisienne pour l'Éducation Récréative, 151 p.
- BENZAKOUR, F. (1995) : «Le français au Maroc. Processus néologiques et problèmes d'intégration», A. Queffélec, F. Benzakour et Y. Cherrad-Bencheфра (éds), *Le français au Maghreb*, Aix-en-Provence, Pub. Université de Provence, pp. 61-76.
- BOUQUIAUX, L. et alii (1978) : *Dictionnaire sango-français. Bakari sangofaranzi et Lexique français-sango. Ketebakari faranzi-sango*, Paris, SELAF, 666 p.
- CHRIST, G. (1991) : *Arabismen im Argot*, coll. «Europäische Hochschulschriften», Frankfurt am Main, Lang, 632 p.
- Collectif (1995) : *Dictionnaire Universel*, Vanves, EDICEF, 1503 p.
- DELAPORTE, G. (1993) : «Application du projet Orthaf : analyse graphémique, morphémique et lexémique des particularités du français au Sénégal», D. Latin, A. Queffélec et J. Tabi-Manga (éds), *Inventaire des usages de la francophonie : nomenclatures et méthodologies*, coll. «Universités francophones», Montrouge, AUPELF-UREF et John Libbey Eurotext, pp. 349-354.
- DEROY, L. (1956) : *L'emprunt linguistique*, Paris, Les Belles Lettres.
- DEROY, L. (1980) : «Vingt ans après *L'emprunt linguistique* : critique et réflexions», *Cahiers de l'Institut de Linguistique de Louvain*, 6, 1-2, pp. 7-18.
- DIALO, A. (1990) : «Le contact wolof-français au Sénégal», A. Clas et B. Ouoba (éds), *Visages du français. Variétés lexicales de l'espace francophone*, coll. «Universités francophones», Montrouge, AUPELF-UREF et John Libbey Eurotext, pp. 59-68.
- DIKI-KIDIRI, M. (1977) : *Le sango s'écrit aussi... Esquisse linguistique du sango, langue nationale de l'Empire Centrafricain*, Paris, SELAF, 187 p.
- DUBOIS, J. et alii (1973) : *Dictionnaire de Linguistique*, Paris, Larousse.

- DUCLOS, J. (1991) : *Les particularités lexicales du français d'Algérie (français colonial, pataouète, français des Pieds-Noirs) de 1830 à nos jours*, Thèse de Doctorat, Université de Tours, 672 p.
- DURAND-DESKA, A. et P. DURAND (1994) : «La forme sonore des emprunts : les mots anglais en polonais et en français», *Travaux du Cercle Linguistique d'Aix-en-Provence*, 12, pp. 79-107.
- EDEMA, A. B. et K. NDUKU (1993) : «Application du projet Orthaf : les variantes graphiques de l'IFA : un point de vue zaïrois», D. Latin, A. Queffélec et J. Tabi-Manga (éds), *Inventaire des usages de la francophonie : nomenclatures et méthodologies*, coll. «Universités francophones», Montrouge, AUPELF-UREF et John Libbey Eurotext, pp. 349-354.
- FRIES, C. et K. PIKE (1949) : «Coexistent Phonemic Systems», *Language*, 25, pp. 29-50.
- GAADI, D. (1995) : «Le français au Maroc. L'emprunt à l'arabe et les processus d'intégration», A. Queffélec, F. Benzakour et Y. Cherrad-Bencheffa (éds), *Le français au Maghreb*, Aix-en-Provence, Pub. Université de Provence, pp. 131-151.
- GARAFANGA, J. (1987) : «Code-switching ou le vernaculaire du locuteur bilingue au Rwanda», *Langage et société*, 41, pp. 243-25.
- GHERIBI, A. (1996) : *Les particularités lexicales de la presse tunisienne d'expression française*, Mémoire de maîtrise, Université de Provence, 108 p.
- GUIRAUD, P. (1965) : *Les mots étrangers*, coll. «Que sais-je ?», Paris, PUF, 123 p.
- LANLY, A. (1962) : *Le français d'Afrique du Nord*, Paris, Bordas, 367 p.
- MORSLY, D. (1995) : «El-Watan, El-Moudjahid, Algérie-Actualités, El-Djeich, Liberté, Le Matin... La presse algérienne de langue française et l'emprunt à l'arabe», *Plurilinguismes*, 9-10, pp. 35-53.
- NAFFATI, H. (1993) : *Particularités lexicales de la presse tunisienne de langue française*, Mémoire de maîtrise, Université de Provence, 194 p.
- N'SIAL, S. (1993) : *La francophonie au coeur de l'Afrique. Le français zaïrois*, Paris, Didier Érudition, 211 p.
- OULD ZEIN, B. (1995) : *Le français en Mauritanie. Étude morpho-syntaxique et lexicale*, Thèse de Doctorat, Université de Provence, 810 p.

- QUEFFÉLEC, A. (1997) : *Le français en Centrafrique. Lexique et société*, coll. «Universités francophones», Vanves, EDICEF, 304 p.
- QUEFFÉLEC, A. et A. NIANGOUNA (1990) : *Le français au Congo, (R. P. C.)*, Aix-en-Provence, Pub. Univ. Provence-A.E.L.I.A.- INaLF-CNRS, 333 p.
- SMAALI, D. (1994) : *Les particularités lexicales du français dans la presse algérienne actuelle*, Mémoire de maîtrise, Université de Provence, 203 p.
- TABER, C. R. (1964) : *French loan-words in Sango : a statistical study of incidence*, Hartford Seminary Foundation, 99 p. .
- THIAM, N. (1994) : «La variation sociolinguistique du code mixte wolof-français à Dakar : une première approche», *Langage et Société*, 68, pp. 11-34.

## Une mémoire au service de la créativité ?

Lina SADER FEGHALI

*Centre d'Études et de Recherches en Terminologie Arabe  
(CERTA), ETIB, ILT, Université Saint-Joseph, Beyrouth, Liban*

Un ordinateur est-il un حاسب, un حاسوب, une رتابة, une منظمة ou tout simplement un كمبيوتر ? Ne vous alarmez pas ! Tous les mots qui sembleraient pour certains incompréhensibles ne sont que les équivalents arabes d'ordinateur et une illustration classique de l'un des problèmes terminologiques de la langue arabe qui n'a pas encore pu se doter d'un lexique spécialisé capable d'exprimer les notions en constante évolution dans notre monde moderne. Par conséquent, les traducteurs, rédacteurs, journalistes, lexicographes et autres usagers de la langue arabe font face à maints problèmes quand ils tentent de traduire, de rédiger voire de lire un texte spécialisé.

Evoquer de tels problèmes étonnerait quiconque saurait que l'arabe moderne est une langue vivante vieille de 2000 ans et qui est parmi les langues officielles de l'ONU à côté du français, de l'anglais, de l'espagnol, du russe et du chinois. Il serait encore plus étonné s'il apprend que la langue arabe comporte, entre autres, 24 dénominations pour désigner l'année, 64 dénominations pour la pluie, 170 pour l'eau et 255 pour le cheval.

Tirailée entre les courants laxistes et puristes, la terminologie arabe demeure lacunaire. En vue de mieux aborder les problèmes rencontrés au quotidien par les chercheurs, nous essaierons de les placer d'abord dans le contexte de l'état actuel de la terminologie

dans le monde arabe en général et au Liban en particulier. Nous tenterons ensuite de répondre à la question suivante : à quel point la mémoire des mots peut-elle être au service de la créativité en terminologie ? Nous terminerons enfin par des perspectives et des conclusions qui, à notre avis, permettront d'augmenter les chances de réussite des recherches terminologiques en arabe.

## 1. — SITUATION ACTUELLE

### 1.1 — Dans le monde arabe

La langue arabe n'a intégré que 50 néologismes en l'espace de cinq siècles soit du début du XVI<sup>e</sup> siècle jusqu'au début du XX<sup>e</sup> siècle. Ce «marasme terminologique» est bien à l'origine des difficultés que rencontrent les locuteurs arabophones dans leurs communications quotidiennes d'autant plus que 80 % des langues des pays développés sont constitués de termes techniques. De plus, la réforme de l'arabe classique a remédié au problème de différenciations lexicales résultant d'influences colonialistes et ce par l'unification du vocabulaire existant et a créé de nouveaux mots pour désigner de nouvelles réalités. «L'arabe dit moderne est le produit des traducteurs de langues européennes vers l'arabe. Cet arabe moderne se trouve en conséquence dans les lexiques bilingues européens-arabes dont le vocabulaire est différent selon les auteurs». (Lakhdar Ghazal, 1994 : 46).

Dans un souci d'unification et dès les années 1980, des banques de terminologie ont été constituées dans le monde arabe, à savoir trois banques :

— La Banque LEXAR de l'IERA, démarrée en 1978 et pionnière dans le domaine de la constitution d'une banque de termes bi-alphabétiques, elle se caractérise par des clefs de recherche sophistiquées et par une structuration notionnelle au niveau de la consultation et de la mise à jour des données.

— La Banque de Terminologie de l'Arabie Saoudite (BASM) : commencée en 1983 en collaboration avec la banque de données TEAM de Siemens. Quadrilingue (arabe, français, anglais et allemand), elle comporte, d'après les statistiques de 1993, 300 000 données.

— La Banque de Terminologie de l'Académie de Jordanie, cons-tituée en 1988 et qui contient 430 000 termes (Ait Taleb, 1996 : 57-59).

Sur un autre plan, plusieurs Académies de langue arabe ont vu le jour notamment au Caire, au Maroc, en Jordanie, en Irak et en Syrie. Une activité lexicographique intense est enregistrée. Les lexicographes s'érigent en terminologues et créent un grand nombre de termes. Jusqu'aux années 1980, 570 dictionnaires bilingues avec l'arabe comme langue de départ ou d'arrivée sont recensés. N'oublions surtout pas l'activité de normalisation terminologique de l'ALECSO qui siège ici même à Tunis et celle du «Bureau de coordination de l'arabisation». Cependant, ces efforts sont demeurés sporadiques et n'ont fait qu'aggraver le problème. En effet, l'absence de coopération et d'échange, l'inexistence d'un organisme normalisateur unique et surtout la diversité des langues sources de transfert ont abouti à une terminologie incohérente qui se reflète dans les diverses publications en langue arabe.

Qu'en est-il de la situation spécifiquement au Liban ?

## 1.2 — Au Liban

### 1.2.1 — *Fonction des langues*

Pour mieux présenter la situation de la terminologie au Liban, il est nécessaire de prendre comme point de départ sa situation linguistique et plus spécifiquement la fonction des langues en usage. Dans la préface de son livre *Anatomie de la francophonie libanaise*, R. P. Sélim Abou, recteur de l'USJ, affirme que l'enquête sociolinguistique menée au Liban «accuse une nette tendance au trilinguisme chez les jeunes générations. En même temps elle dévoile les fonctions respectives spontanément assignées à chacune des trois langues en présence : l'arabe littéral, langue officielle et langue de culture; le français, langue de communication, de formation et de culture; l'anglais, langue de communication internationale et d'information. L'avenir est peut-être à ce type de trilinguisme» (Abou 1996 : 6).

Le français demeure donc une langue à usage professionnel parfois en concurrence avec l'anglais dans la communication des entreprises avec l'étranger. Ces deux langues s'additionnent de

surcroît à l'arabe qui de par son statut officiel ne se limite pas à l'usage personnel. Ce trilinguisme de facto a été à la base du choix des langues de l'École de Traducteurs et d'Interprètes de Beyrouth : l'arabe en tant que langue A, le français en tant que langue B et l'anglais en tant que langue C.

Notre programme d'enseignement de la traduction ne pouvait ignorer l'apport de la terminologie surtout quand il s'agit de traduire des textes spécialisés. Ce besoin est encore plus pressant quand l'arabe est une des langues de travail, l'arabe, langue de départ et d'arrivée, étant lacunaire dans les situations de communication scientifique et technique. En effet, les traducteurs, rédacteurs, journalistes, lexicographes et autres usagers de la langue arabe font face à maints problèmes quand ils tentent de traduire, de rédiger voire de lire un texte de spécialité. Les lacunes des ouvrages lexicographiques ne facilitant pas la tâche, il était donc nécessaire de créer une infrastructure susceptible de répondre aux besoins en matière de terminologie de nos étudiants en traduction d'abord et partant des diverses facultés de l'USJ et des entreprises.

### *1.2.2 — Naissance du CERTA et de TERMUS*

Depuis octobre 1996, les activités du Centre d'Études et de Recherches en Terminologie Arabe (CERTA) de l'Institut de Langues et de Traduction de l'Université Saint-Joseph de Beyrouth ont été lancées. Son équipe de recherche a mis au point une base de données terminologiques taillée sur mesure (TERMUS) tout en tentant en parallèle de recenser les besoins terminologiques les plus urgents.

De plus, le CERTA assure l'encadrement nécessaire à tous les travaux de recherche terminologique entrepris par les étudiants de 2<sup>e</sup> cycle de l'ETIB.

TERMUS est une base de données informative en ce sens qu'elle n'est pas liée à un organisme de normalisation terminologique; les données qui y sont répertoriées visent à mettre à la disposition des traducteurs un outil d'information et d'aide à la traduction terminologique. Les langues de travail y sont le français, l'arabe et l'anglais.

Notre base de données terminologiques TERMUS fonctionne dans un environnement Windows 3.11 for Workgroups version

anglais-arabe en mode mono ou multi-utilisateur. Elle a été conçue sur le Système de Gestion de Base de Données (SGDB) Access 2.0.

TERMUS offre à l'utilisateur des moteurs de recherche variés accessibles par un écran de recherche convivial qui lui permettent d'accéder facilement aux données saisies.

TERMUS est consultable en direct, seules les données suivantes peuvent être imprimées par l'utilisateur :

- les fiches de termes;
- les lexiques par domaine à entrée arabe, française ou anglaise;
- les listes bibliographiques par domaine.

TERMUS se présente sous forme de fiche composée de quatre blocs et de plusieurs champs.

Le BLOC GESTION comprend les informations suivantes :

Rédacteur - Réviseur - Organisme - Date rédaction - Date de mise à jour - Indicatif de fichier - Domaine générique - Domaine spécifique.

Le BLOC TERMINOLOGIE est le même pour chacune des langues de la base :

Vedette - Catégorie grammaticale - Cote de pondération - Définition - Contexte - Notes linguistiques - Notes techniques

Quant au BLOC SOURCES, il est propre à chaque fiche. Un bouton de renvoi permet de consulter la base de données documentaires.

Des tables séparées trilingues ont été prévues pour les champs relatifs aux synonymes, aux variantes et aux renvois.

## **2. — PROBLÈMES**

Bien avant l'achèvement de la mise au point de notre base de données terminologiques TERMUS, les recherches terminologiques étaient déjà lancées par nos chercheurs dans différents domaines. Malgré la variété des sujets, les problèmes qui freinaient l'avancement des recherches s'avéraient identiques :

— D'abord, si le chercheur décide de s'assurer de l'orthographe ou de la signification d'un équivalent proposé par un dictionnaire bilingue voire vérifier son existence, comme le ferait un locuteur français qui a recours au Grand Robert, il se rend vite compte que les dictionnaires généraux arabes ne répertorient pas toujours l'usage actuel des mots du lexique général ni ne mentionnent le vocabulaire technoscientifique qu'il soit ancien ou contemporain. En effet, leurs auteurs n'y ont répertorié que les «vocables attestés dans la poésie antéislamique, le Coran, les traditions prophétiques, la poésie et la prose classique» (Troupeau, 1986 : 13).

— Ensuite, étant donné que les «dictionnaires spécialisés où figurent l'arabe seul ou à côté d'autres langues et qui répondent aux critères internationalement admis (univocité sémique, délimitation des domaines et définition des notions) sont plutôt rares» (Hardane 1993 : 493), le chercheur n'a accès qu'à des dictionnaires bilingues qui lui offrent parfois, et surtout les dictionnaires spécialisés, en plus de l'équivalent recherché une définition en arabe qui a elle-même besoin d'explication. De plus, les équivalents qui y sont proposés ne répondent guère aux attentes de notre chercheur puisqu'il sont parfois une création propre au lexicographe et non l'attestation de l'usage.

— C'est alors que le chercheur décide d'aller à la recherche d'ouvrages non lexicographiques spécialisés rédigés en arabe. Il est rarement satisfait puisque la plupart des ouvrages sont traduits et les équivalents qu'ils proposent ne sont pas identiques à ceux qui sont proposés par les dictionnaires bilingues ou multilingues.

— Enfin, en panne d'instruments, le chercheur a recours aux sources orales. Ces dernières ne peuvent l'aider puisque dans l'enseignement supérieur la langue de communication est l'anglais ou le français. Les experts, ne connaissant donc généralement pas les termes servant à désigner la notion en question, n'hésitent pas à utiliser le terme tel quel dans sa forme étrangère sans se soucier de son inadaptation aux règles linguistiques arabes puisqu'ils sont soucieux uniquement d'assurer une communication rapide et efficace. Par ailleurs, si l'équivalent recherché existe et est largement répandu — cas certes très rare — ils considèrent par snobisme que l'usage d'un jargon technique étranger leur permet de se distancier du simple

profane. De là, d'ailleurs, surgit la difficulté de trouver les équivalents arabes dans des contextes réels d'utilisation.

Par conséquent, suite à cette recherche documentaire souvent frustrante, notre chercheur se trouve souvent face à deux extrêmes :

D'une part, si l'équivalent arabe existe, **plusieurs possibilités** s'offrent à lui :

— Chaque document consulté, voire un même document, propose un équivalent différent pour une même notion (synonymie pléthorique). À titre d'exemple, dans le domaine de l'informatique le terme disquette a deux équivalents = اسطوانة / قرص. Dans le domaine du génie génétique, alors que clonage a bel et bien un équivalent إستنساخ le terme clone n'a pas encore trouvé dans l'usage un équivalent établi, les usagers hésitent entre نسيلة / مثيل / نسخة etc.

— Dans un même ouvrage, la même étiquette désigne des notions différentes. Ex : dans le domaine de l'emploi le terme أجره est l'équivalent proposé pour traduire indifféremment les termes : rétribution / rémunération / salaire / paie / appointements.

— L'équivalent proposé revêt souvent la forme d'une périphrase ou d'une définition. Ex : l'acide désoxyribonucléique connu communément sous l'appellation A.D.N. est traduit sous forme de périphrase الحامض النووي الريبوزي الناقص الأوكسيجين ou plus simplement par l'utilisation du sigle A.D.N. ou D.N.A

— Dans des ouvrages arabes issus de pays arabophones différents, une même notion est désignée par une étiquette différente (variantes régionales); l'équivalent étant une translittération ou une traduction du terme anglais ou français dont l'orthographe n'est pas souvent uniforme. Ceci est surtout dû au fait que chaque pays arabe a adopté la langue source de transfert et la culture de la puissance mandataire qui le gouvernait. Ex : l'unité de mesure utilisé surtout dans le domaine de l'informatique, le pouce ou l'*inch* a comme équivalent = بوصة / إنش, selon la langue de transfert utilisé par l'utilisateur.

Soucieux de respecter le principe de l'univocité, le chercheur doit donc proposer un seul équivalent. Il tâchera, dans la mesure du possible, de se conformer aux cinq normes principales de l'ISO 704 :

proposer un terme correct linguistiquement, motivé, précis voire transparent, concis, réceptif à la formation de dérivés et en relation monosémique avec la notion. A ces normes s'ajouteront les dix-huit principes de base pour la création et le choix des termes scientifiques et techniques approuvés lors du Symposium sur l'unification des procédés de création néologique (Khatib, 1993 : 49-51).

Il n'oubliera surtout pas d'être à l'écoute de l'usage qui a le dernier mot. Nous pensons ici à l'équivalent du terme satellite : قمر صناعي. Ce dernier a eu le dessus sur l'une des propositions de Lakhdar Ghazal de l'IERA : أفلاك \ أفلك qui pourtant répond plus aux exigences linguistiques et terminologiques. «En effet, l'enrichissement du lexique spécialisé d'une langue vivante peut se faire, et se fait d'ailleurs souvent, sur le terrain de l'usage, indépendamment des réflexions théoriques menées par les savants» (Hardane, 1993 : 482).

D'autre part, à l'autre extrême, notre chercheur qui avait l'embaras du choix se retrouve face à un **vide notionnel et lexical**.

Pour résoudre ce problème, les Académiciens arabes, notamment ceux qui appartiennent au courant des puristes, soucieux de sauvegarder le génie de la langue, proposent une démarche en trois temps : d'abord, par ordre de priorité, puiser dans les anciens dictionnaires et ouvrages scientifiques en arabe l'équivalent recherché. Ensuite, face à une nouvelle notion, la traduction s'impose si le sens est traduisible. Sinon, il faut avoir recours aux méthodes de création néologique, à savoir : la dérivation, la composition, la néologie de sens. Enfin, l'emprunt demeure l'ultime recours. Malgré la diversité des moyens à sa disposition, le problème n'est toutefois pas résolu. Le chercheur ne dispose pas d'une référence fiable qui puisse orienter son choix définitif. Ce choix, quel qu'il soit, n'est pas assurément celui retenu par d'autres locuteurs ou ne correspond pas à leur goût linguistique. De surcroît, l'existence de plusieurs académies arabes (Caire, Maroc, Jordanie, Irak et Syrie) ne lui facilite pas la tâche. La langue arabe reste figée : il suffit de lire les comptes rendus des réunions de l'Académie arabe du Caire pour se rendre compte de la lenteur des décisions et de leur inadaptation aux réalités linguistiques actuelles. En l'espace de 15 ans (1973-1988), les congrès d'arabisation ont adopté 107 473 termes unifiés (Khatib, 1993 : 52-54). Constatation pertinente à ce sujet : ces termes ne sont

pas facilement accessibles et partant ne sont pas utilisés à l'unanimité par tous les locuteurs arabophones. À tout ceci s'ajoute le manque de coordination entre les divers organismes de recherche qui a abouti à la «babélisation» de la terminologie arabe. Voilà la fin du périple de notre chercheur qui n'a pas réussi à résoudre complètement les problèmes évoqués. Faire appel à la mémoire des mots lui suffirait-il ?

### 3. — APPORT DE LA MÉMOIRE DES MOTS

S'il veut respecter le souhait des académiciens arabes, notre chercheur doit veiller à sauvegarder le patrimoine linguistique et scientifique de la langue arabe donc faire appel à la mémoire des mots. Cette méthode a été fructueuse mais est-ce une source intarissable ? L'arabe dispose de 9273 racines et est une langue ouverte à la dérivation et il est possible d'obtenir de chaque racine 200 à 300 mots (Khatib, op. cit. : 15). De plus, que de mots arabes ont été empruntés par les langues occidentales ! Un dictionnaire publié en 1997 (Bacha, 1997 : 9) a recensé 803 mots français, 350 anglais et espagnols, 300 italiens, 250 portugais, 200 allemands, 100 grecs et bulgares et 300 mots dans les autres langues européennes. Fierté des académies arabes, ces emprunts se réduisent en nombre, jour après jour, en raison de l'évolution des langues emprunteuses qui éliminent les termes décrivant des concepts vieillissés. De plus, il est regrettable que beaucoup de mots français d'origine arabe ont généré des dizaines d'autres dans les langues occidentales alors que le mot arabe d'origine a été jeté dans les oubliettes. Par exemple, le terme polysémique *artichaut* a été emprunté à la langue arabe (خرشوف) qui nous est revenu après sous une autre forme (أرض شوكة).

N'oublions pas aussi un point important : la langue arabe étant une langue sémitique, une limitation d'ordre structurel l'empêche d'être ouverte à la suffixation et à la préfixation. De même, les éléments brachygraphiques (abréviatifs et idéographiques) propres aux langues de spécialité sont presque inexistantes en arabe. Vouloir à tout prix sauvegarder la pureté de la langue est un souhait certes louable mais avons-nous oublié que tout au long de son histoire, l'arabe, tout en puisant des termes dans son patrimoine culturel, a subi une influence étrangère tous azimuts en empruntant une grande partie de sa terminologie à l'araméen, au persan, au grec et au latin ?

Le facteur temps est aussi à prendre en considération. Une langue ne pourrait évoluer à la cadence désirée si les dizaines de termes techniques qui naissent tous les jours doivent se trouver une racine dans la langue cible. Si nous empêchons l'entrée de termes étrangers dans la langue arabe, juste par souci de purisme, nous empêchons par conséquent l'introduction des notions qui les accompagnent.

En mettant en évidence le danger de l'importation du lexique étranger, nous occultons un autre danger véritable : lorsque la langue d'arrivée emprunte des structures syntaxiques propres à la langue avec laquelle elle est en contact qui sont étrangères à sa propre syntaxe. Par ces calques, l'arabe perdrait sa propre syntaxe et ne sera arabe qu'en surface. En effet, ceci est surtout évident dans les ouvrages non lexicographiques traduits.

Sur le terrain, pour bon nombre de locuteurs arabophones, au Liban et peut-être ailleurs, communiquer en arabe des informations d'ordre technique et scientifique équivaut à parler une langue étrangère. L'usage des lexiques, des dictionnaires bilingues, des glossaires et des traductions demeurent l'apanage des autorités administratives officielles et des traducteurs. Souvent, en cas de litige, le texte technique rédigé en français ou en anglais fait loi. Donc, les véritables décideurs, ingénieurs, médecins, chercheurs, techniciens, etc., adoptent plutôt une attitude méfiante face à la capacité de la langue arabe de répondre à leurs besoins de communication ou considèrent cette situation avec une grande résignation; les langues étrangères, l'anglais ou le français étant perçus comme les langues de communication internationale par excellence.

#### **4. — PERSPECTIVES ET CONCLUSIONS**

Face à cet état de choses, notre objectif en créant TERMUS était de tenter de répondre aux besoins d'ordre terminologique de nos étudiants-traducteurs. Notre base ne devrait pas venir s'ajouter aux autres bases de données du monde arabe, elle doit jouer un rôle complémentaire, donc nous sommes tous appelés à coopérer pour :

— Explorer les domaines vierges afin d'éviter le dédoublement des efforts;

— Rechercher des moyens afin de faciliter l'échange d'informations;

— Puiser dans la langue arabe riche en ressources et rechercher une méthodologie de recherche terminologique uniforme;

— Choisir un seul canal linguistique de transfert en l'occurrence le français «...élément essentiel de l'identité lexicographique et terminologique arabe [...] pressé constamment de s'armer théoriquement, méthodologiquement et techniquement pour faire face au déferlement des nouveautés notionnelles et terminologiques apparues en anglais, [le français] se trouve le mieux rodé à l'actualisation terminologiques» (Hardane, 1993 : 496).

— Être à l'écoute des usagers. «Il faut aussi écouter comment les gens parlent, et se parlent, pour comprendre comment ils usent des mots, comment ils en créent, pourquoi ils s'y heurtent» (Gaudin, 1995 : 232).

En tentant de combler, chacun de son côté, les lacunes de la terminologie technoscientifique arabe, ne sommes-nous pas en train d'aggraver le problème de la synonymie pléthorique ? Quelle est la langue qui puisse prétendre au titre de pureté ?

Toutes les langues du monde sont en situation d'échanges continus et les emprunts, tant qu'ils ont la consonance de la langue d'arrivée, s'y intègrent facilement et l'enrichissent.

Un désir de coordination exprimé par tous les organismes de recherche dans ce domaine sous le contrôle d'une seule et unique autorité de normalisation serait susceptible d'assurer l'uniformisation de la terminologie arabe scientifique et d'éviter la babélisation qui la caractérise. Ces V<sup>es</sup> Journées scientifiques sont certes une occasion pour lancer un appel dans ce sens. Une langue ne peut vivre et résister si elle ne consolide pas sa crédibilité et si elle ne répond pas à un besoin d'expression et de communication. Les néologismes ne peuvent rester cloisonnés dans des académies. Il appartient aux usagers de les diffuser ou de les condamner.

## RÉFÉRENCES

- ABOU, Sélim *et al.* (1996) : *Anatomie de la francophonie libanaise*, coll. «Universités francophones-Prospectives francophones», Beyrouth, FMA pour l'AUPELF-UREF, 319 p.
- AIT TALEB, Saadia (1996) : «Le poste de travail du terminologue arabe», *Terminologies Nouvelles : Banques de Terminologie*, Rint, n° 15, Bruxelles, pp. 57-65.
- EL BACHA, Mohamed (1997) : *L'arabe émigré : dictionnaire des mots français dérivés de l'arabe*, Beyrouth, Ed. Dar El Ilm Lil Malayin, 359 p.
- GAUDIN, François (1995) : «Champs, clôtures et domaines : des langues de spécialité à la culture scientifique», *META*, 40 (2), pp. 229-237.
- HARDANE, Jarjoura (1993) : «Rôle du français dans l'élaboration terminologique arabe», *Francophonie différentielle*, Actes du Colloque International sur le français langue seconde (FLSH, USJ 20,21,22 mai 1993), Paris, l'Harmattan, 1994, pp. 481-499.
- KHATIB, Ahmad Shafic (1993) :  
الخطيب أحمد شفيق . حول توحيد المصطلحات العلمية . من محاضرات مجمع اللغة العربية في مؤتمره التاسع والخمسين في القاهرة (12 - 26 نيسان 1993) . دائرة المعاجم . منشورات مكتبة لبنان . بيروت . ص 54
- KHATIB, Ahmad Shafic (op. cit.) :  
الخطيب أحمد شفيق . المواصفات المصطلحية وتطبيقاتها في اللغة العربية . محاضرة أقيمت خلال ندوة حول اللغة العربية وتحديات القرن الحادي والعشرين . المنظمة العربية للتربية والثقافة والعلوم . ص 31
- LAKHDAR GHAZAL, Ahmed (1994) : «L'expérience de l'IERA en matière d'implantation et de diffusion de la terminologie scientifique et technique», *Terminologies Nouvelles : Implantation de termes officiels*, n° 12, Bruxelles, pp. 45-46.
- TROUPEAU, Gérard (1986) : «Les problèmes posés par la traduction de l'arabe médical ancien en français moderne», *META*, 31 (1), pp. 11-15.

## **La mémoire des mots, un problème fondamental pour la traduction**

**Antoine NOUJAIM**

*Université Libanaise, Jounieh, Liban*

La traduction est une opération complexe. Transfert de sens d'une langue à l'autre ou simple transcodage de mots ? En fait, dans le cas où elle serait une correspondance de mots, il est rare de trouver un mot qui soit l'équivalent parfait d'un autre, chacun ayant sa propre histoire philologique, étymologique, phonétique, sémantique. Dans le cas même où on la considère comme un transfert de sens d'une langue à l'autre, la difficulté persiste, car on ne peut dissocier ce sens du texte qui en constitue la trame, et des mots qui en constituent les éléments, chacun ayant évidemment un sens précis dans le contexte bien déterminé, tout en se profilant cependant sur un fond dont les racines remontent très loin dans le passé, aux moments germinatifs et nébuleux de la formation d'une langue et d'un peuple.

Le mot dans la langue n'est pas, en effet, une existence détachée. Sa mémoire peut parfois s'étendre à la mesure de l'histoire de la langue elle-même. Son étymologie peut évoquer des contacts lointains, sinon originaires, avec d'autres langues et d'autres cultures. Son évolution phonétique va de concert avec l'évolution de la langue elle-même dont elle reproduit les repères principaux. Son pouvoir magique de signification est fonction, lui aussi, de l'évolution culturelle de toute une société.

Les dimensions de la mémoire des mots s'étendent, en effet, aux mesures des dimensions d'une langue, d'une civilisation, d'une culture et d'une vision du monde en perpétuel mouvement conditionné par des mobiles internes et d'autres externes résultant du choc et de la rencontre avec d'autres langues, civilisations, cultures et visions du monde.

Si la traduction des mots peut rendre compte de leur état synchronique actuel, il est difficile qu'elle puisse envelopper toute leur diachronie et la sous-entendre comme un fond indispensable, indissociable du sens contemporain de ces mots, qui en prolonge le contexte de signification dans l'espace et dans le temps. Cette difficulté est d'autant plus marquée que les deux langues, source et cible, à savoir l'arabe et le français ou l'anglais, n'appartiennent pas à la même famille.

A cela s'ajoute pour certaines langues, comme l'arabe, que la diachronie elle-même est fonction d'une synchronie bien fixe qui introduit l'«a-temporel» comme une référence de base inaltérable, pour le temporel; d'où «l'intraduisibilité» de «l'a-temporel» dans des langues qui n'en respectent pas ce caractère. Quelle place jouent les mots dans l'opération de traduction ? Peut-on rendre compte, dans cette opération, de toute la mémoire des mots ? Peut-on considérer la traduction du sens comme un subterfuge valable à la difficulté de reproduire des mots équivalents à ceux de la langue-source, avec leur mémoire ?

Nous essayerons, à partir d'exemples tirés de la langue française et de la langue arabe, de montrer les difficultés inhérentes à la traduction des mots qui ont leur mémoire respective dans chacune de ces deux langues.

## ORIGINALITÉ DES MOTS ET DES LANGUES

Les langues ont leur système propre : leur grammaire, leur syntaxe, leur génie particulier; mais aussi leurs mots dont quelques-uns ont une saveur originale et unique, ce qui pose un problème certain pour la traduction. «Les langues [...] sont des systèmes originaux, conditionnés par l'histoire et qui réalisent des entités selon des universaux, mais aussi selon des valeurs propres» (Clas et Bouillon, 1994 : XV). Il est des mots quasi intraduisibles. Ainsi en est-il d'un *magari* ou d'un *aggiornamento* en italien, qui ont toute

une autre saveur et expressivité qu'un «ah si !» ou une «mise à jour». Ainsi en est-il d'une « نخوة : *nakhwa* » ou d'une « مروءة : *muruat* » en arabe dont s'approchent des termes comme générosité, honnêteté, courage, sans s'y superposer cependant.

Les exemples de ce genre abondent dans chaque langue, ce qui complique davantage le problème de la traduction, les mots ayant chacun son individualité propre à l'intérieur d'un système linguistique qui les recouvre, lui aussi, d'un cachet distinctif. Pour s'en rendre compte, il suffit de considérer comment la même langue est différemment employée par des peuples différents ou des collectivités distinctes à l'intérieur d'un même peuple. Sans aucune difficulté, on peut constater des divergences plus ou moins importantes entre l'anglais britannique et l'anglais américain, entre le français en usage au Canada et celui en usage en France (par ex. «chauffer un char» et «conduire une voiture»), entre les différents dialectes, voire les emplois littéraires variés de la même langue arabe, d'un pays à l'autre.

Il est évident qu'un tel phénomène de divergence crée des ambiguïtés à lever par le traducteur. Le problème n'en devient que plus crucial quand on voit comment des langues de la même famille, transforment chacune à son compte, le sens d'une même racine dont elles conservent les sons. Dans les langues sémites, une même racine comme *sabar* donne le *sabro* syriaque qui signifie l'espérance, et le *sabre* arabe qui signifie l'endurance; une racine *'ebed* donne le verbe syriaque *'bed* qui signifie «faire», «adorer», alors qu'elle donne le verbe arabe *'abada* qui signifie également «travailler» et «adorer», mais avec une notion de servitude. En hébreu, il semble que la même racine dans *Abed Yahvé* signifie : serviteur de Dieu. C'est la personne qui s'anéantit devant Dieu. Une autre racine comme *Imar* devient en syriaque *émar* et se traduit par «dire», alors qu'elle devient en arabe *amara* et se traduit par «ordonner». Une dernière racine de ce genre enfin comme *nassiya* donne le *nessiyono* syriaque qui signifie la tentation; et le *nissiyâne* arabe, qui signifie l'oubli.

On pourrait dégager, à partir de ces exemples, des divergences significatives quant à la notion de travail par rapport à des bédouins et des sédentaires; la notion de péché où la volonté joue un rôle important, et celle d'omission par un phénomène d'oubli; la notion

de parole libre et celle d'ordre dicté; la notion d'espérance eschatologique et celle d'endurance physique et psychique. Marqué par ses conditions géographiques, historiques, ethniques même, religieuses et culturelles, chaque peuple emploie les mots et leur confère un sens, à sa manière.

Chaque mot a certes sa dénotation, mais aussi une connotation sociolinguistique. Chien peut connoter la bassesse pour certains peuples ou collectivités, vache un caractère sacré, cochon la saleté, etc. «Les connotations constituent un fait linguistique collectif, ni purement individuel ni non plus totalement général ou universel, à vrai dire intermédiaire entre la parole et la langue, mais plus proche de cette dernière» (Ladmiral, 1994 : 145).

Même les nombres qui devraient être les plus impersonnels ou scientifiques, ont des connotations collectives et particulières : ainsi de 1, 3, 7, 13... «A l'intérieur même du langage scientifique, le recours à tel ou tel terme connote en général l'appartenance à une école déterminée et un investissement de la personne du chercheur, sujet de l'énonciation» (Ladmiral, 1994 : 153). Les langues s'enveloppent en outre de dimensions mythologiques qui s'attachent aux mots et que l'on ne saurait ignorer. «Le traducteur doit être un dragueur — qui fasse remonter à la surface, en langue-cible, du fond de la langue-source et en aval la substance de ces dépôts connotatifs, car ils peuvent receler des trésors dont viendra s'orner le langage poétique comme le langage courant» (Ladmiral, 1994 : 155)

Il s'ensuit que chaque langue a son génie et son univers mental; en somme, chacune exprime une vision particulière du monde. Ceci montre par le fait même qu'il est très difficile de pouvoir traduire les nuances contenues dans les termes de chaque langue dont il faut remonter l'histoire. Ceci étant vrai des langues appartenant à la même famille originelle comme l'arabe et le syriaque, combien plus vrai encore ne devrait-il être à propos de langues de familles différentes comme l'arabe et le français ! Le Français qui prononce le terme «Soudan» n'a pas présent à l'esprit ce que les Arabes ont senti pour la première fois en découvrant ce pays, ce qui les a entraînés à le dénommer «السودان : les noirs». Le même phénomène, mais inverse, se produit peut-être pour les arabes qui prononcent le terme «Nigeria» dont la racine est significative pour les occidentaux. Les

mots ont, ainsi, chacun son histoire et sa mémoire. Les correspondants qu'on leur trouve risquent souvent de manquer cette originalité propre; ils en traduisent le sens immédiat, actuel, et ratent les évocations lointaines. En termes de dénotation et de connotation, on pourrait dire «que les mots cheval, canasson et coursier (ou employeur, patron et chef d'entreprise, etc.) ont la même dénotation, le même dénoté, alors que chacun des trois est porteur d'une connotation différente» (Ladmiral, 1994 : 117), ce qui rend la tâche du traducteur plus difficile.

A la difficulté importante de l'originalité des mots, s'ajoute avec plus d'acuité encore la difficulté émanant de la divergence des attitudes vis-à-vis des mots dans la traduction.

### **Importance des mots**

Les mots dans un texte à traduire suscitent, en effet, chez les traducteurs et les théoriciens de la traduction, des attitudes variées, voire opposées. Certaines religions, notamment les religions «révélées», revêtent les mots d'un cachet sacré. Dans la Bible, Abram qui suit la consigne divine et se dénomme Abraham, n'est presque plus la même personne. Le changement de nom exprime un changement de condition, de profession ou de vocation. Ainsi en est-il des esquimaux qui changent de nom, en fonction de leur travail en été ou en hiver. La dénomination est un phénomène culturel et même existentiel, d'une importance capitale. Dans la même Bible, au livre de la Genèse, c'est à travers des mots que l'acte de création s'opère. «Dieu dit : que l'univers soit et l'univers fut». Davantage, en voyant l'homme mélancolique, Dieu lui conseilla de donner des noms aux créatures. «L'homme désigna par leur nom tout bétail, tout oiseau du ciel et toute bête des champs...» (Gn 2,20). Une manière de les recréer et de les faire exister pour lui-même ! Depuis les origines, les mots ont un caractère magique, sacré, créateur... Dans le christianisme, «le Verbe» désigne le Christ, qui n'a pas manqué de déclarer : «Le ciel et la terre passeront et mes paroles ne passeront pas» (Mt. 24,35). Dans l'Islam, le Coran est la parole éternelle de Dieu.

A un plan différent, des théoriciens modernes et certains praticiens de la traduction en parviennent au culte des mots. Peter Newmark (1988 : 73) affirme : «*We do translate words because there is nothing else to translate; there are only the words on the page;*

*there is nothing else there*» (1). Dans le même ouvrage, il déclare aussi : «*Many translators say you should never translate words, you translate sentences or ideas or messages. I think they are fooling themselves*» (2). Prise de position pour la «traduction sémantique» aux dépens de la «traduction communicative» ou pour une nouvelle forme de «traduction littérale» aux dépens de la «traduction libre», ces affirmations dénotent une grande confiance en l'originalité des mots et en leur pouvoir magique d'expression, ainsi qu'une croyance à la nécessité de l'exactitude et de la fidélité dans la traduction.

Quant à l'opposition du sens et des mots, c'est un faux problème. Tout en reconnaissant que l'accomplissement du sens ne se réalise qu'à travers une phrase dans un texte, on peut dire qu'il n'y a pas de mots sans significations, ni de sens qui ne soit pas véhiculé par des mots ou des signes quelconques. On peut certes adhérer à l'opinion de Jean-Paul Sartre qui écrit dans *Qu'est-ce que la littérature* : «Dès le départ, le sens n'est plus contenu dans les mots puisque c'est lui, au contraire, qui permet de comprendre la signification de chacun d'eux; et l'objet littéraire quoiqu'il se réalise à travers le langage, n'est jamais donné dans le langage; [...] aussi les cent mille mots alignés dans un livre peuvent être lus un à un sans que le sens de l'oeuvre en jaillisse; le sens n'est pas la somme des mots, il en est la totalité organique» (3). Mais si le sens des mots se précise à travers la phrase, et celui de la phrase à travers le texte, cela n'empêche que ces mots, pris isolément, sont des éléments de phrases en puissance, et des promesses de sens en germination. Les mots sont, en effet, de mystérieux êtres énigmatiques. Dépositaires de l'expérience intellectuelle séculaire des peuples, ils ont un pouvoir magique de représentation, de reproduction et même de création. À considérer leur capacité d'évaluation et les péripéties de leur histoire, on les prendrait pour des êtres vivants, sans que cela n'empêche certains de leur dénier tout genre de halo et de les accuser de trahison vis-à-vis des idées. Le problème n'en devient que plus complexe et grave, en traduction : quelle place accorder aux mots dans la

---

1) Newmark, Peter (1988) : *A Textbook of translation*, New York, London, Prentice Hall International, p.73, in Lederer Marianne, 1990, p. 62.

2) *Ibid.*, p. 63.

3) Sartre, Jean-Paul (1955) : *Qu'est-ce que la littérature?*, Gallimard, pp.50-51; in Lederer, Marianne, 1994, pp. 23-24.

traduction ? Doit-on opter pour une traduction de mots ou une traduction de sens ? Et dans le cas où l'on traduit le sens, peut-on détacher le sens des mots qui l'expriment; peut-on rendre toutes les charges culturelles, affectives et historiques dont ils sont porteurs ? En traduisant des mots ou des phrases, que fait-on de la mémoire des mots ? Si toute traduction est une trahison, ne le serait-elle pas encore plus lorsqu'on reconnaît aux mots des dimensions spatio-temporelles qui transcendent le «hic et nunc» de l'écrivain, du traducteur et des destinataires dans les deux langues ? Cette problématique est d'autant plus difficile que les mots, en tant que tels, sont insignifiants pour d'aucuns et riches pour d'autres, des éléments du langage, fortuits et contingents pour certains, essentiels et indispensables pour d'autres.

### Répercussion sur la traduction

Le mot traduction, dans son acception la plus générale, «désigne toute forme de «médiation interlinguistique», permettant de transmettre de l'information entre locuteurs de langues différentes. La traduction fait passer un message d'une langue de départ (LD) ou langue-source dans une langue d'arrivée (LA) ou langue-cible» (Ladmiral, 1994 : 11). Le message à transférer d'une langue à l'autre est évidemment exprimé par un texte; ce dernier est composé de phrases qui sont, à leur tour, composées de mots. Ceci situe la traduction, d'emblée, au coeur de l'univers des mots à comprendre, à reproduire et à respecter dans les deux opérations. «Le traducteur, dit Cary (1963b : 33), se doit de savoir — et de sentir — qu'à certaines époques des gens étaient convaincus que prononcer un mot équivalait à créer une chose, qu'une phrase pouvait déclencher des bouleversements dans l'ordre cosmique, que connaître le nom d'une bête donnait prise sur la bête, que pour évoquer le démon et l'obliger à agir il suffisait d'écrire son nom et qu'on n'avait pas le droit de prononcer le nom de Dieu» (in Larose, 1989 : 4). Certaines tribus arabes, avant l'Islam, croyaient qu'en donnant à leurs enfants des noms de bêtes féroces ou de carnassiers : «*Fahed* : panthère», «*Assad* : lion»; «*Sacre* : faucon», etc., elles les prémunissaient de leurs agressions.

Les mots ayant cette valeur sacrée ou magique, on comprend qu'il y ait deux approches pour les traduire. Si l'une est la traduction par équivalences, comme nous le verrons plus loin, l'autre est

caractérisée par la recherche systématique de correspondances; elle «s'efforce de conserver des signifiés en changeant de signifiants» (Lederer, 1994 : 67). Sans discuter de la première approche, pour l'instant, disons que la deuxième accorde aux mots une valeur primordiale.

Si, dans les temps modernes, l'homme ordinaire, à force de lire les mots sans attention et de les redire sans y penser dans les conversations, a oublié leur magie et leur charme et s'est laissé entraîner par un ensemble de mécanismes dépersonnalisants, le traducteur, contraint d'y faire attention et de les écouter, retrouve leur charme et leur goût. Le respect qu'il a des mots du texte de départ suscite parfois en lui l'obsession de l'équivalent et du terme adéquat à retrouver dans la langue-cible. Tout au moins, il se trouve en face de la très vieille problématique qui n'a cessé de se poser de diverses manières au cours des siècles, et qui réapparaît avec plus d'acuité, à cette ère de l'ordinateur : la traduction peut-elle se faire mot à mot ? Les théoriciens de la traduction et les traducteurs eux-mêmes, n'ont cessé d'opter chacun pour une solution propre qui concorde avec son système général. Pour Jean-René Ladmiral, «Le problème de la traduction est souvent posé dans les termes antinomiques d'un débat académique : traduction littérale ou traduction littéraire dite "libre", autrement dit la fidélité ou l'élégance, la lettre ou l'esprit. Ce sont ces deux pôles d'une même alternative, indéfiniment rebaptisés, qui scandent l'histoire de la traduction selon un mouvement de balancier entre «'équivalence formelle» et «l'équivalence dynamique» (Nida, 1964 : 159 sqq.), entre le mot-à-mot et les «belles infidèles» (cf. Mounin, 1955)...» (Ladmiral, 1994 : 14). Dans certaines théories, on sent le respect des mots, même si par ailleurs ces théories présentent des conceptions divergentes de la traduction.

### **Traduction des mots et rigueur des théories**

Larose (1989 : 13) décrit, parmi les théories contemporaines de la traduction, celle de Vinay et Darbelnet qui représente ce qu'on nomma naguère l'École montréalaise de stylistique : «la traduction est d'abord et avant tout «une discipline exacte» et, même s'il n'y a pas de traduction unique d'un passage donné, «il est permis de supposer que si nous connaissions mieux les méthodes qui gouvernent le passage d'une langue à l'autre, nous arriverions dans

un nombre toujours plus grand de cas à des solutions uniques» (Vinay Jean-Paul et Darbelnet Jean, *Stylistique comparée du français et de l'anglais*, Montréal, Beauchemin, 1968, p. 24). Ce postulat cependant amène les auteurs à privilégier une solution parmi de nombreux équivalents potentiels et donne ainsi l'impression au lecteur, à cause de l'accumulation d'exemples de rapprochements langue à langue, que pour chaque segment en langue de départ (LD), il n'existe qu'un seul équivalent en langue d'arrivée (LA), et ce malgré les nombreuses mises en garde qu'ils adressent audit lecteur». Bien que Vinay et Darbelnet ne soient pas partisans de la littéralité en traduction, cette ambition d'en faire «une discipline exacte» avec des solutions uniques, ne manque pas de respect vis-à-vis des mots et des segments de langue auxquels il faut trouver des segments équivalents dans la langue d'arrivée. Analysant la théorie de Darbelnet, Larose (1989 : 28) affirme : «Dans *Niveaux de la traduction*, Darbelnet définit la traduction comme étant «l'opération qui consiste à faire passer d'une langue dans une autre tous les éléments de sens d'un passage et rien que ses éléments, en s'assurant qu'ils conservent dans la langue d'arrivée leur importance relative, ainsi que leur tonalité, et en tenant compte des différences que présentent entre elles les cultures auxquelles correspondent respectivement la langue de départ et la langue d'arrivée». Par éléments de sens, l'auteur entend les unités sémantiques, c'est-à-dire les unités de pensée ou unités de traduction. Ces éléments de sens n'excluent pas les signes de ponctuation ou de graphie : par exemple, «la rue de Rivoli» est une rue de Paris, tandis que «la Rue de Rivoli» désigne le ministère des Finances». Il en va de même pour le Quai d'Orsay avec ou sans lettre capitale...

Le *Dictionnaire encyclopédique Quillet* (1968 : 6970) présente une conception analogue de la traduction, quant à la méthode qu'il semble prôner et au résultat obtenu : «Rejetant le système ancien qui, aux XVII<sup>e</sup> et XVIII<sup>e</sup> siècles, avait abouti aux «belles infidèles» qui prétendaient ajouter grâce et élégance à l'oeuvre originale, on considère aujourd'hui qu'une bonne traduction doit serrer le texte de l'auteur d'aussi près que possible, en conservant toutes les images, les tours de pensée, l'ordre des mots même, le ton et les caractéristiques du style propre de l'écrivain sans en rien retrancher, ni rien y ajouter : il doit s'établir ainsi une sorte de parenté spirituelle et artistique entre l'auteur et le traducteur, de telle sorte que la

traduction puisse paraître à la lecture une oeuvre originale, et que si, d'autre part, on la collationne avec l'original, on l'y retrouve intégralement».

De tendance plutôt linguistique, certains théoriciens, voire certains praticiens de la traduction, soucieux d'appliquer le positivisme dans tous les domaines de la connaissance et de l'action, prennent des attitudes, plus rigoristes que rigoureuses, et montrent un tel attachement aux mots, aux tournures stylistiques et grammaticales et à la forme d'une manière générale que leur traduction finit par être littérale ou mot à mot, un transcodage de mots ou de phrases. Il se produit une sorte de socialisation du texte-source qui empêche que l'on puisse s'en éloigner, ne serait-ce que formellement. La traduction des textes sacrés a longtemps suscité des craintes et des méfiances : il a fallu l'auréole des Septante pour que les juifs admettent la traduction de la Bible. Lorsque les textes-sources ont un cachet sacré, leurs traducteurs sont psychologiquement portés à en respecter la littéralité.

Il est clair que de telles attitudes, ne mésestiment pas l'importance des mots comme segments ou éléments de segments de phrases dans les deux langues de départ et d'arrivée. Vinay et Darbelnet, Newmark, les traducteurs des textes sacrés et les partisans de la littéralité n'ont pas la même conception de la traduction. On peut cependant, malgré les divergences parfois radicales entre leurs théories respectives, en dégager une attention marquée aux mots et à la nécessité de s'y astreindre pour que le texte traduit ne soit pas une adaptation de l'original.

En fait, quelles que soient les théories adoptées en traduction, elles ne peuvent, faire fi des mots que chaque langue utilise selon ses propres normes grammaticales, pour l'expression des idées. Il est vrai, comme nous l'avons déjà mentionné, que le mot : nom, verbe, déterminant, ne trouve la plénitude de sa signification que dans la phrase, et celle-ci dans le paragraphe et le texte. Il reste néanmoins admis par tout le monde que le mot, en tant que lexème, a sa signification en lui-même. Plus, il est soumis aux péripéties du temps; comme un être vivant, il naît, grandit, se développe, vieillit, devient archaïque, peut même mourir et ressusciter... Il a un passé, un présent et un avenir. Comme tel, le mot que la langue propose à l'emploi actuellement n'est pas «a-temporel»; s'il dépend des besoins de la

société de cette époque et de l'usage qu'on en fait hic et nunc, il n'en est pas pour autant dépourvu d'histoire. «Les mots sont ces lieux de mesure où s'enracinent une identité culturelle, un système politique, une époque, un homme, un type de société, une discipline (nouvelle). Ils véhiculent des valeurs de la culture, au sens large, à laquelle ils sont liés». Cela ne fait que rendre plus complexe l'opération de les traduire : si le traducteur connaît bien le sens actuel des mots, sa connaissance embrasse-t-elle toute l'histoire dont ces mots sont chargés, pour que sa traduction puisse être fidèle et reproduire tout le sens qu'ils expriment.

Les exemples de phénomènes propres à chaque langue, quant à la signification des mots et leurs dimensions socioculturelles, sont très nombreux. On peut se contenter de méditer sur certains parmi ces cas sémantiques originaux, choisis surtout dans le domaine de la dénomination.

*\* Les noms propres*

Chaque langue a ses noms propres. Dans les langues sémites, ces noms propres ont souvent une signification qu'il est difficile de pouvoir traduire dans les langues européennes, sans faire perdre à ces noms une part très importante de leur signification et de la vision du monde qu'ils renferment. Ainsi de tous les noms composés avec la racine «Ile» en araméen : Gibrâïl, Mikhaïl, Raphaïl... qui deviennent en français Gabriel, Michel, Raphaël, perdant beaucoup de l'ambiance religieuse, culturelle, sociologique, mythologique... qu'ils suggéraient dans la langue d'origine : Gabriel signifie la puissance du dieu des dieux, Raphaël = la miséricorde du dieu des dieux, Michel = la crainte du dieu des dieux. La même chose peut être dite des mots composés avec la racine «Allah» en arabe : Nasrallah = victoire de Dieu, Choukrallah = remerciement de Dieu, etc. Tous ces noms évoquent naturellement une atmosphère religieuse. En les empruntant tels qu'ils sont, le traducteur peut faire sentir à ses lecteurs cette ambiance particulière. Mais son texte reste imprécis : il devrait rendre compte précis de tous les implicites religieux auxquels les parents ou les grands-parents, voire les aïeux lointains, ont sans doute pensé en choisissant ces noms pour leurs enfants. Les dictionnaires monolingues rappellent parfois l'étymologie de ces mots. Les dictionnaires bilingues le font rarement; les traducteurs se contentent d'emprunter les mots tels qu'ils sont dans la langue d'origine, avec

quelques variantes parfois comme dans le cas de Michel (provenant de Mikhaïl), ce qui dérouté davantage les lecteurs. Le même cas peut se présenter pour la traduction en arabe des noms propres d'origine gréco-latine ou française comme Auguste (Augustus) qui devient «Ugustus» en arabe, sans évoquer tout le halo de dignité et de vénération implicite dans «Auguste» et que rendrait bien un terme comme Jalil جليل ou au féminin Jalila جليلة. Ainsi de Théodore qui devrait se traduire plutôt par *Atallah* = don de Dieu, de Théophile (Théophilos) qui devrait se traduire plutôt par *Hubballah* = amour de Dieu, tout en soulignant que la notion de Dieu n'est pas la même dans le contexte culturel des deux langues arabe et française.

Bref, les limites de cette étude ne permettent pas de citer un plus grand nombre de cas de traduction de noms propres par emprunt. Il est évident qu'un tel procédé respecte l'étymologie du mot et sa sonorité; mais en l'absence de notes explicatives du traducteur qui pourraient compenser la perte infligée au sens, ce procédé fait souvent perdre au mot une grande partie de sa mémoire. Pour sauvegarder cette mémoire il vaudrait mieux traduire le mot par un autre dont l'étymologie est analogue à la sienne, malgré les divergences qui peuvent exister entre les éléments qui composent les deux étymologies.

#### \* *Les noms communs*

La traduction étymologique des noms propres n'est en fait qu'un cas particulier de la traduction des noms communs. Ceci nous conduit à l'analyse du procédé de traduction par un transcodage de mots qui suppose qu'à chaque mot de la langue-source correspond un mot équivalent dans la langue-cible.

En fait, ce principe est très souvent aléatoire. L'équation de correspondance entre les termes qui désignent le même objet dans deux langues de familles différentes n'est pas absolue. La compréhension et l'extension des deux termes et des deux concepts peuvent ne pas être identiques : le terme homme qui désigne l'homme ou la femme est traduit par «انسان» (*'insān*) alors que le même terme homme qui désigne l'homme par opposition à la femme (le *vir* latin) est traduit par «رجل» (*rajul*). De la même manière le terme humanité est traduit tantôt par «انسانية» (*'insāniya*), tantôt par

(*baġariyat*) بشرية ce qui rappelle le *besro* syriaque, c'est-à-dire la chair par opposition à l'esprit pur; et la mémoire du mot nous plonge ainsi dans toute une philosophie de l'incarnation et de la dualité de la nature humaine qui ne signifient pas la même chose pour des sémites et des gréco-romains, des orientaux et des occidentaux, des arabes et des européens...

Par ailleurs, les mots, pris isolément, étant souvent polysémiques, leur sens est ambigu. Cette ambiguïté est certes levée par la phrase et le texte qui précisent le sens à choisir ou à conférer aux mots. Ceci ne permet pas cependant de résoudre le problème de la mémoire des mots à recréer par l'opération traductrice. Si l'on prend par exemple le terme «personne», mis à part le pronom indéfini qu'il désigne, il peut être traduit par «شخص» (*akhs*), «اقنون» (*Uqnum* = hypostase), «ذات» (*thāt* = un ipse d'où vient ipséité)... Même si le texte indique le sens à donner à ce mot, ce sens ne rend pas compte de la charge philosophique, théologique et culturelle dont chacun de ses emplois est chargé. La mémoire du mot personne renferme toute une théologie christologique et trinitaire qu'on ne trouve pas dans le terme *%akhs* qui rappelle plutôt le «personna latin : le masque» et le personnage ou la personnalité que l'on aperçoit dans «ما يشخص اليه». La mémoire du terme «الذات : *thāt*» arabe est, elle aussi, insuffisamment rendue par le terme latin *ipse* passé en français. Des notes explicatives s'imposent certes pour suppléer, d'une manière ou de l'autre, à ces imperfections de correspondance entre les mots, à moins de pouvoir créer des termes savants aussi expressifs.

Il est en effet une possibilité de forger des mots pour mieux rendre un sens auquel les termes existants ne correspondent pas. Le terme monothéisme que certains traduisent improprement par «وحدانية : *wahdaniyat*» devrait être plutôt traduit par «أحادانية : *ahadaniyat*» ce qui s'oppose mieux à polythéisme; et l'on devrait traduire وحدانية plutôt par unithéisme, ce qui correspond mieux à la notion de Dieu un et trine à la fois sans verser dans le polythéisme. La traduction nécessite ainsi la compréhension préalable de toutes les nuances contenues dans le mot et une capacité de création pour forger des termes correspondants dans la langue-cible. «La néologie est une dimension essentielle à tout langage vivant» (Ladmiral, 1994 : 156). Il reste néanmoins vrai que les termes forgés, quelle que soit

ur capacité de signification, n'arrivent pas bien souvent à reproduire toute la mémoire des mots originaires, ni à recréer l'atmosphère intellectuelle qui enveloppait ces mots. Pour Danica Seleskovitch (1975 : 35-36), «deux langues ont beau posséder un mot pour une même chose, ces deux mots correspondent rarement l'un à l'autre dans toutes leurs acceptions ou traits de signification».

### Des mots à équivalence partielle ou lacunaire

Par ailleurs, tout en reconnaissant qu'il est possible de trouver des équivalences complètes entre quelques mots de langues différentes, désignant les mêmes objets, il est plus courant de rencontrer des équivalences soit partielles, soit apparentes, ou de fausses correspondances, voire des lacunes et des surabondances de termes dans une langue par rapport à une autre. «Le vocabulaire français est riche d'une multiplicité de noms de fromages, de variétés de pain qui reflètent une réalité différente de la réalité anglaise... Ce sont des cas où, à notre avis, il convient de défendre le recours à l'emprunt et à la note en bas de page. Une autre solution consiste à recourir à l'équivalence ou au terme générique : «fromage», par exemple, au lieu d'une variété spécifique. Mais dans la mesure où le traducteur vise à être un médiateur de communication, ne convient-il pas qu'il permette au public d'accéder à une autre civilisation ? (Ballard, 1991 : 17). Dans la même perspective, des linguistes se sont enchantés à compter plus d'un millier de termes en arabe pour désigner des objets comme l'épée, le cheval, le chameau et le lion. Cela correspond certes à des particularités de la vie tribale des bédouins au désert, qu'il est facile de comprendre en les situant dans leur contexte social, géographique et économique. Mais ce phénomène représente de grandes difficultés pour le traducteur, par exemple : un terme comme «*al-yamani* اليماني» désigne «une» épée de fabrication yéménite et de forme particulière; un autre comme «*al-mouhannad* : المهند», une épée fabriquée avec du fer provenant des Indes... *Jawad* désigne le cheval rapide; *al-hisan* soit le cheval mâle, soit le cheval ancien; *al-adham*, le cheval noir, *al-faras*, soit la jument, soit le cheval... *Al-ward* désigne le lion qui s'abreuve; *al-qunnas*, le lion qui chasse; *al-hisabre*, le lion de forme massive; *al-gadanfar*, le lion de taille énorme; *al-laith*, le lion très puissant... Le chameau est, lui aussi, objet de phénomènes analogues; ce qui rend l'opération de la traduction très difficile, dans sa phase de compréhension d'abord,

ensuite dans sa phase de réexpression. On peut évidemment suppléer au transcodage des mots en traduisant des mots parents par des groupes de mots ou des périphrases ou par un recours à d'autres procédés... Ce que nous en retenons cependant, c'est la nécessité pour traduire avec fidélité le sens de certains mots, de recourir à plusieurs mots dans la langue-cible ou à une note explicative du traducteur. Lorsque ces mots sont véhiculés par des structures syntaxiques qui varient d'une langue à l'autre, les divergences sont encore plus marquées. «Les différences de structures syntaxiques des langues imposent leurs contraintes au traducteur de sorte que même le plus médiocre apprenti traducteur ne procède pas mot à mot, mais respecte la syntaxe de la langue d'arrivée» (Lederer, 1994 : 113).

### Contingence des mots

Parallèlement à l'attitude qui estime les mots, il est une attitude qui minimise leur importance. La prise de position pour la primauté absolue des mots ne manque pas de soulever des critiques. Jean Delisle en parvient à intituler son article dans les *Études traductologiques en hommage à Danica Seleskovitch* : «Le froment du sens, la paille des mots» (in Lederer, 1990 : 61-72). Il y déclare : «Le traducteur n'est pas un eunuque commis à la garde des mots [...]. Vue sous l'angle du procédé, la traduction interprétative évite le décalque servile des mots au nom de la fidélité au sens et du respect du caractère idiomatique de la langue d'arrivée [...]. En somme, du traducteur qui connaît son métier, on attend qu'il préserve le froment du sens sans nécessairement garder la paille des mots». Une phrase telle que : «Cette boisson me rafraîchit le coeur, ou me réchauffe le coeur» ne peut avoir le même sens pour un auteur vivant au Sahara et un autre vivant en Alaska. C'est pourquoi on peut la traduire soit par هذا الشراب ينعش قلبي ou هذا الشراب يثلج قلبي ou هذا الشراب يدفئ قلبي en fonction des conditions géographiques et climatiques.

La traduction ne peut se réduire en effet à un transcodage de mots. L'opération est en fait beaucoup plus complexe qu'un établissement mécanique de correspondances entre des mots. Le serait-elle même, le souci de fidélité exigerait que la traduction d'un mot par un autre puisse rendre compte de toute la mémoire inhérente

au mot source. Même les mots décalqués, voire empruntés, ont besoin de notes explicatives du traducteur qui les explicitent avec leurs charges culturelles, affectives, historiques, sociales.

Cette intervention de la subjectivité de l'auteur, du traducteur et des destinataires exprime le souci d'une plus grande fidélité au texte source et la volonté de reproduire tout le sens explicite et implicite contenu dans ses mots. Elle se justifie par l'exigence d'une plus grande objectivité ! La fidélité aux mots et à leur histoire contraint le traducteur à transcender leur immédiateté pour en saisir toutes les dimensions. Cela nous conduit naturellement à la théorie du sens qui, par un souci d'une plus grande rigueur scientifique, par une exigence d'une plus grande fidélité au texte source et par une insistance sur la liberté et la créativité du traducteur dans la langue-cible, professe la nécessité de traduire le sens du texte plutôt que les mots de ce texte. «La connaissance de tous les mots et des notions qu'ils recouvrent est une connaissance nécessaire, mais non suffisante, à la compréhension du sens d'un message» (Seleskovitch, 1975 : 45)

*\* Théorie du sens et mémoire des mots*

A l'antipode des théories qui s'attachent à la littéralité des mots et à la conception de la traduction comme transcodage de mots, de nombreux théoriciens et des praticiens, des courants modernes dans le domaine de la traduction refusent l'asservissement aux mots. «Toutes les recherches effectuées à l'ESIT tendent à prouver que la traduction par équivalences a une validité générale quelles que soient les langues ou les types de textes, littéraires ou techniques, textes de fiction ou de réalité» (Lederer, 1994 : 52)

Il suffit d'évoquer les figures de proue qui ont fait la théorie du sens, pionnière et représentative quant à la libération des traducteurs, de l'hégémonie des mots. Parmi les théories qui ont fait peu de cas des mots, au moins en apparence, la théorie du sens tient une place de prédilection. Soutenue par l'École Supérieure d'Interprètes et de Traducteurs (ESIT) de la Sorbonne Nouvelle-Paris III, notamment par Maurice Gravier, Danica Seleskovitch et Marianne Lederer, elle insiste sur un principe fondamental qui devrait régler le processus traductif, à savoir que «le sens est l'objet à saisir et à transférer». Dans un discours prononcé le 25 mai 1977 à la Sorbonne à l'occasion du 2<sup>e</sup> anniversaire de l'ESIT, Maurice Gravier affirme : «Traduire,

interpréter, ce n'est pas remplacer des mots par des mots, substituer une première mosaïque de mots à une autre mosaïque de mots. Il faut franchir la barrière des mots et de la syntaxe, il faut atteindre le sens, il faut comprendre, c'est le premier moment. Ensuite, il faut réexprimer, au besoin en oubliant les modes d'expression auxquels recourait l'auteur du texte primitif» (in Brauns, 1981 : 5)

Danica Seleskovitch (1975 : 69) affirme, pour sa part, «Lorsque, cessant d'examiner des mots isolés, on se penche sur la parole en action, sur le discours improvisé, on constate qu'il n'est plus possible de convertir une langue en une autre. Pour transmettre un message dans une autre langue, il ne s'agit plus de comparer des vocables, mais d'appréhender un sens et de l'exprimer «L'interprète ne répète jamais les mots de l'orateur, il reproduit ses idées», est le leitmotiv de ceux qui exercent le métier d'interprète de conférence». Pour illustrer ce point de vue, nous recourons à des exemples que n'approuverait pas l'auteur : la traduction d'expressions toutes faites. «La piété filiale» ne peut être traduite littéralement par «التقوى النبوية», mais plutôt par «البر بالوالدين», qui ne peut être traduit littéralement par «don ou aumône aux parents». Le sens d'un mot ne se précise, en effet, que lorsqu'il est employé avec un autre, ou mieux dans une phrase et un texte. Le terme pomme n'est pas le même dans pomme de terre, pomme d'Adam, pomme de la main, pomme du Liban... On ne peut le traduire de la même manière.

Dans la même perspective, on peut citer Marianne Lederer (1994 : 9-15) qui affirme : «Tout est interprétation. La traduction ne fait pas exception. Trop longtemps et trop souvent cantonnée dans la mise en oeuvre des langues, elle était censée être passage des mots ou des phrases d'une langue aux mots ou aux phrases d'une autre langue. Or, la voie menant à la traduction réussie suppose l'interprétation des textes et l'appel à des connaissances extralinguistiques [...] L'opération de traduction n'est pas la même selon que l'on traduit des mots, des phrases ou des textes [...] La décontextualisation produit une forme de traduction mot à mot que tous condamnent mais que beaucoup pratiquent [...] La tentation de traduire le mot se substitue [...] en la faussant à l'opération de traduction des textes. J'englobe sous l'appellation traduction linguistique la traduction de mots et la traduction de phrases hors contexte et je dénomme traduction interprétative, ou traduction tout

court, la traduction des textes.». Ainsi seulement, l'on peut comprendre que des proverbes peuvent être traduits. Ils renferment certes un bon sens universel; mais chaque peuple ou chaque civilisation les traduit à sa manière. L'important est de traduire ce bon sens et de suggérer cette manière propre. Ainsi de «Tel père, tel fils» qui devient «هذا الشبل من ذاك الأسد» ou «Qui se ressemble s'assemble» qui devient «ان الطيور على أشكالها تقع».

La théorie interprétative de la traduction et son corollaire, la théorie du sens refusent l'asservissement aux mots. «Situé dans la perspective de la théorie du sens, le processus traductif consiste non pas à rendre compte des mots ni à mettre au point un instrument de connaissance mais à recréer la magie qui se dégage de toutes les composantes affectives et notionnelles de l'oeuvre. Établie dans le respect de ses valeurs et des contraintes de la langue-cible, la nouvelle construction verbale est un objet autonome, mais qui lui reste assimilable par sa physionomie d'ensemble, par sa fonction esthétique et par l'effet produit».

Davantage, on peut admettre avec Ladmiral (1994 : 66) que «s'il n'y a pas en toute rigueur théorique d'authentique synonymie lexico-sémantique au niveau de la langue, il se trouve défini dans la pratique une synonymie contextuelle, situationnelle au sein de la parole-cible d'une traduction». Plus qu'une correspondance donc entre des mots chargés de mémoire, ce sont des correspondances entre des situations qu'il faut trouver. À ce niveau, les traducteurs pourront plus facilement éviter d'être des «traditeurs» ou des traîtres. S'ils ne reproduisent pas la mémoire des mots, ils ne manquent pas de le faire pour les situations.

En fait, l'ancienne dichotomie de la traduction littérale et de la traduction libre est stérile. Pour Jean Delisle, «L'essentiel est de traduire intelligemment en fonction d'une foule de paramètres, peu importe que ce soit en gardant intacte la forme du texte original ou en la modifiant. Préserver l'intégrité à la fois du sens du message et de la langue d'arrivée est l'idéal vers lequel tend le traducteur consciencieux. Et, à cet égard, le littéralisme n'offre aucune garantie de vérité et d'exactitude» (in Lederer, 1990 : 64-65). Si l'attachement à la littéralité émane d'une exigence de fidélité, on peut dire que cette exigence peut être respectée si l'on arrive à reproduire le même sens dans la langue-cible. Or la reproduction du même sens ne refuse pas

*a priori* le transcodage des mots, là où il est possible. Plus, le respect des mots : de leur mémoire et de leur histoire, peut contribuer à une plus grande fidélité au sens lui-même.

Le transcodage des mots n'est pas forcément une atteinte au sens. Le fait de pouvoir trouver dans la langue-cible un mot qui corresponde parfaitement au mot de la langue source, avec sa charge d'histoire et de mémoire, rend un excellent service à la traduction du sens avec toutes ses dimensions. Dans cette optique, même un partisan de la théorie du sens, Amparo Murtado Albir, peut se permettre d'affirmer que «l'équivalence de traduction est une équivalence de sens; elle est dynamique et contextuelle par nature, car le sens est dynamique. Toutefois, on constate facilement que toute traduction est un mélange d'équivalences contextuelles, dynamiques et éphémères, et d'équivalences de transcodage» (in Lederer, 1990 : 80). Il en parvient même à admettre sous forme de concession «que la possibilité d'emploi des équivalences de transcodage augmente lorsqu'il s'agit de textes à caractère dénotatif plus marqué» (in Lederer, 1990 : 81)

La mémoire des mots peut constituer un obstacle à la communication de l'intention réelle de l'auteur. À moins de saisir la mémoire d'un mot et de lui trouver un équivalent suffisant, le traducteur ne sera pas fidèle à ce que l'auteur a exprimé consciemment ou même inconsciemment par son emploi de tel ou tel autre mot. Voilà pourquoi, on ne peut en définitive interdire ou légitimer, dans l'absolu, tout transcodage de mots. À chaque situation, s'impose une solution. Selon Ladmiral (1994 : 19), «Face à une lacune lexicale de sa langue-cible (un mot «intraduisible»), le traducteur peut avoir recours à la solution désespérée de l'emprunt, qui importe tel quel le terme-source étranger (signifiant et signifié), ou à cette importation plus discrète qu'est le calque (emprunt du signifié sans le signifiant). Dans les deux cas, c'est le plus souvent le mot, mais aussi la chose elle-même qui sont importés de la langue-culture-source».

Les islamologues sont unanimes à utiliser les termes arabes eux-mêmes, auxquels ils ne trouvent pas de termes français à correspondance parfaite. Malgré certaines concordances, la croisade et la guerre sainte ne sont pas les équivalents parfaits du «djihad» qui comporte d'autres dimensions qu'aucun terme français ne contient, à

lui seul. Il en va tout autant pour beaucoup d'autres termes, tels que le « الوقف = *waqf* » que l'on traduit par « bien de mainmorte », ce qui ne rend pas compte de toute la mémoire du terme *waqf*. Ainsi en est-il également du terme *ummah* auquel ne correspond pas tout à fait le terme « nation ». Le terme *ummah* englobe tous les fidèles de la religion quelles que soient leur race, leur origine, leur histoire ou leur culture. Ainsi en est-il du terme *chari'a* qu'il faut reproduire tel quel, plutôt que de le traduire par loi. Le *hadith*, dans la civilisation islamique ne peut être simplement traduit par « propos », car il englobe, en plus des propos, toute la tradition orale de l'Islam. Ainsi de *khalife* ou *calife*, de l'*imam*, de l'*intifada*... Autant de termes propres à la civilisation islamique et au monde arabe, qu'il convient de traduire par emprunt, avec des notes explicatives, pour que la traduction remplisse son vrai rôle de communication entre des personnes, mais surtout entre des civilisations différentes. Les occidentaux à l'époque de la renaissance n'ont pas manqué de recourir à ce même procédé pour remplir des vides dans leur civilisation ou leur langue; aussi a-t-on pu introduire en français des termes comme *algèbre*, *algorithme*, *alcool*, *alchimie*, *alcade*...

Les échanges linguistiques sont des sèves nourricières et dynamisantes pour les peuples et leur culture; A l'instar des occidentaux, les arabes, malgré certaines réticences, n'ont pas manqué de recourir au même procédé d'emprunt qu'ils appellent « de manière significative » « arabisation : تعريب ». Comme il se produit avec d'autres langues, certains termes en français n'ont pas d'équivalent propre en arabe. C'est pourquoi les spécialistes recourent à l'emprunt pour les traduire. Ainsi de la « ديموقراطية » qui transpose la démocratie, de « تلفن » qui transpose téléphoner mieux que هتف, de راديو qui transpose radio mieux que مذياع, de تيفي pour télévision, etc.

Ces emprunts n'ont pas toujours la même valeur, ni la même capacité d'expression dans la langue-cible comme dans la langue-source, ni la même intégration dans la langue d'arrivée. Des termes comme *alleluia*, *kyrieleison* empruntés tels quels de l'hébreu et du grec évoquent certes, par le même effet de sonorités musicales, une ambiance de prière semblable à l'ambiance originaire dans la langue-source, sans exprimer toutefois, de manière explicite et précise le

même sens «louez le Seigneur dans la joie» et «Seigneur prends pitié» pour tous les priants chrétiens, qu'ils soient français, italiens ou arabes même. Un autre terme «*amen* = ainsi soit-il» traduit certes la fin d'une prière, sans pouvoir évoquer clairement la charge de sécurité qui caractérise la racine *amn* apparentée à sécurité, tranquillité et foi.

D'autres termes empruntés s'intègrent parfois si étroitement à la langue qu'il devient difficile d'en percevoir l'origine et d'en distinguer toute la portée. Des deux termes : arabe «طاولة : *tawila*» et latin *tabula* = *tavula* = table, il est difficile de percevoir celui qui est originaire et donc source. Le terme «فلسفة : *falsafat*» qui n'est autre qu'un emprunt de *philosophia* ou philosophie risque même de se dépouiller de ses origines étymologiques, pour des penseurs originaux comme Kamal El-Hage qui ont voulu y voir plus qu'un amour de la *sophia* grecque, «un assemblage d'analyse et de synthèse : «سَفّ et فَلَ». Le terme magasin, tout en provenant du terme «مخزن : *makhzan*» en arabe s'en est détaché pour avoir sa propre histoire.

Quoi qu'il en soit, le critère et le défi en même temps consistent à pouvoir trouver un équivalent au sens qui puisse rendre compte même de la mémoire des mots. À cela tient la difficulté de la traduction, mais aussi sa beauté. «Au lieu de rester en retrait du texte à traduire, de le considérer avec crainte et timidité, de le croire intouchable, le traducteur doit prendre conscience du fait que le simple phénomène de la traduction est déjà, en lui-même, une transformation de certains paramètres. L'intervention ne peut être invisible : autant qu'elle soit bénéfique, et qu'elle apporte au lecteur final, dans toute la mesure du possible, la totalité de ce dont bénéficie le lecteur premier». (Herbulot, *Le traducteur déchiré*, in Lederer, 1990 : 273). Et la mémoire des mots n'est pas un élément que l'on peut ignorer pour la recreation de cette totalité. Sans tomber dans une prétendue «scientificité» terminologique, ni dans un terrorisme «théoriciste», disons en définitive que la fidélité n'empêche pas la créativité. Le lien des deux fait le défi et l'originalité de la traduction. Même si l'«incrémentalisation paraphrastique», selon Ladmiral, fait éclater l'unité spécifique du mot-source et le prive de la fonction terminologique qu'il pouvait avoir dans le texte original, elle s'impose parfois au nom de la fidélité et de la créativité.

Bref, le problème que pose la mémoire des mots peut trouver des esquisses de solutions grâce à divers procédés : emprunts, décalques, création savante de termes équivalents, incrémentalisations paraphrastiques, notes du traducteur, équivalences de sens... On ne peut *a priori* admettre un seul procédé et refuser les autres. Il est des emprunts excellents : فلسفة, philosophie; d'autres, moins réussis linguistiquement mais plus révélateurs de la mémoire des mots : رديمق, démocratie. De même il est des décalques pertinents : محرك traduit moteur depuis le moteur premier en philosophie jusqu'au moteur de toutes sortes d'engins. وحدانية et أحداية traduisent réciproquement monothéisme et unithéisme. Ce qu'il ne faut jamais oublier, que l'on traduise des mots ou des textes, c'est que cette opération devrait nous conduire à «communiquer» avec l'autre : personne, peuple, civilisation; elle devrait nous conduire à «connaître» l'autre, c'est-à-dire naître avec lui, en assurant toute l'histoire et toute la mémoire des mots qu'il emploie, pour le faire «re-naître» tel quel, identique à lui-même, mais dans le contexte différent d'autres mots, d'une autre langue, d'une autre culture. Par la traduction, on peut réconcilier l'un et le multiple, le même et l'autre, et réaliser la diversité dans l'unité.

## RÉFÉRENCES

- BALLARD, Michel (1991) : *La traduction de l'anglais au français*, fac. Langues étrangères, Nathan universités, 268 p.
- BRAUNS, Jean (1981) : *Comprendre pour traduire — perfectionnement linguistique en français*. Paris, la Maison du Dictionnaire, 60 p.
- CHUQUET, Hélène et Michel PAILLARD (1989) : *Approche linguistique des problèmes de traduction*, Ophrys, 452 p.
- CLAS, André et Pierrette BOUILLON (dir) (1995) : *TA - TAO : recherche de pointe et applications immédiates*, coll. «Universités francophones», Paris/Beyrouth, AUPELF-UREF et FMA, 521 p.
- CLAS, André, THOIRON, Philippe et Henri BÉJOINT (dir) (1996) : *Lexicomatique et dictionnaires*, coll. «Universités Francophones», Beyrouth, AUPELF-UREF et FMA, 455 p.
- DELISLE, Jean (1994) : *La traduction raisonnée. Pédagogie de la traduction*, Ottawa, Les Presses de l'Université d'Ottawa, 484 p.
- Dictionnaire Encyclopédique Quillet* (1968) : Paris, Librairie Quillet.
- LADMIRAL, Jean-René (1994) : *Traduire : théorème pour la traduction*, Paris, Gallimard, 284 p.
- LAPLACE, Colette (1994) : *Théorie du langage et théorie de la traduction*, coll. «Traductologie», n° 8, Paris, Didier Érudition, 313 p.
- LAROSE, Robert (1989) : *Théories contemporaines de la traduction*, Québec, Presses de l'Université du Québec, 2<sup>e</sup> éd., 336 p.
- LEDERER, Marianne (dir) (1990) : *Études traductologiques en hommage à Danica Seleskovitch*, Paris, Minard, 286 p.
- LEDERER, Marianne (1994) : *La traduction aujourd'hui*, Paris, Hachette, 223 p.
- MOUNIN, Georges (1968) : *Clefs pour la linguistique*, Paris, Éditions Seghers, 189 p.
- SELESKOVITCH, Danica (1975) : *Langage, langues et mémoire*, Paris, Minard, 272 p.
- ZAKARIA, Michel (1997) : *Influence de la langue arabe sur l'apprentissage du français langue seconde*, Liban, 235 p.

IMPRIMERIE OFFICIELLE DE LA REPUBLIQUE TUNISIENNE

# المجلة التونسية للعلوم الإجتماعية

نشرية مركز الدراسات والأبحاث الإقتصادية والإجتماعية

عدد خاص : الأيام الخامسة العلمية لشبكة

LTT; AUPELF; UREF

ذاكرة الكلم

بإشراف

الطبيب البكوش ، أندري كلاس وصالح الماجري

## بمشاركة :

أندري رومان ؛ ليلي مسعودي ؛ محمد لزهر عباس ؛ حسن حمزة ؛

حبيبة نفاتي ؛ جوزف ديشي ؛ كريم شيبو وأن فلنا ؛ دوغلاس سكييس ،

إنقريد ماير وجودي كفناق ؛ صوفية بن يحي ؛ عبد الحميد كمون ؛

ميلود الطايقي ؛ إسماعيل تميمي وجنفياف لالكي-بوادن ؛ كرستيان

بواتاي ؛ ألفا مامادو ديالو ؛ إبراهيم ديان ؛ زكي دنيس بيجا كودي ؛ سيلفي

برهيال ؛ بيار لرا ؛ حسين هبايلي وسليم بن حزاز ؛ كمال قحة ؛ رفايال

رجسبرا ؛ أمبرواز كوفليك ؛ لينا صدر فغالي ؛ أنتوان نجيم

مركز الدراسات والأبحاث الإقتصادية والإجتماعية

23، نهج إسبانيا - تونس