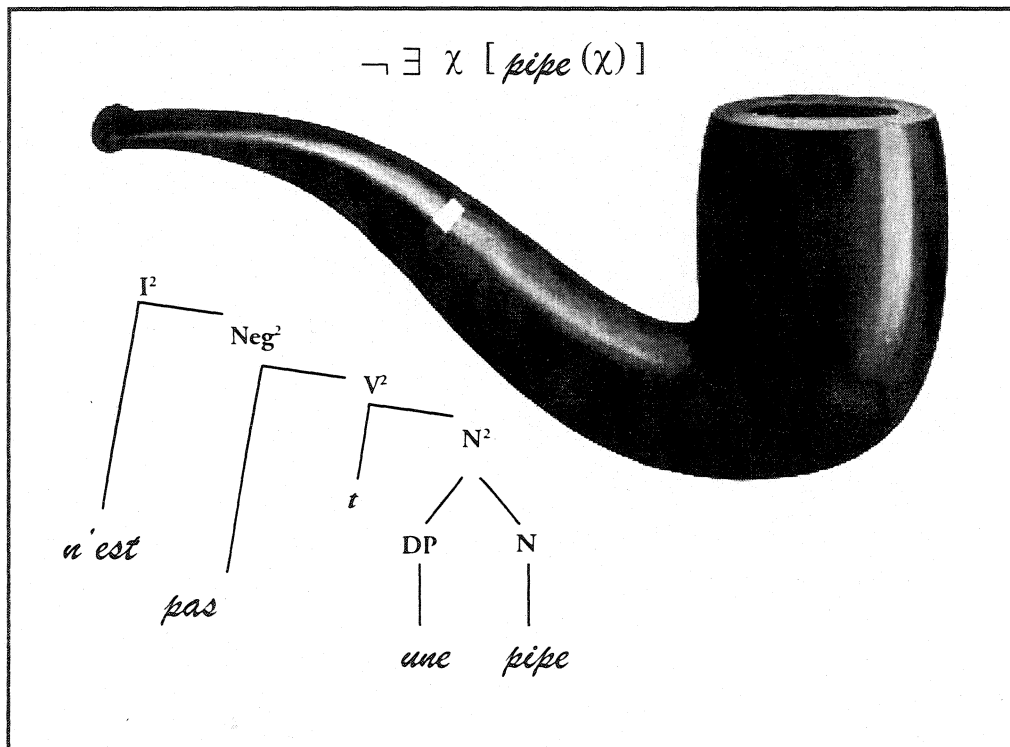


# DIALANGUE

BULLETIN DE LINGUISTIQUE  
volumes 8 et 9, avril 1998

Unité d'enseignement  
en linguistique et en langues modernes  
Université du Québec à Chicoutimi

## QUESTIONS DE SENS SYNTAXE, SÉMANTIQUE, LEXICOGRAPHIE



*Mais, au fait, qu'est-ce qu'une pipe?*

- ARTICLES • COMPTE RENDU • ACTUALITÉS LINGUISTIQUES
- PROJETS DE MÉMOIRES DE DEUXIÈME CYCLE

# LEXICOGRAPHIE ET CONSTRUCTION DU SENS

## Du dictionnaire considéré comme hypertexte

Georges VIGNAUX  
(CNRS -INaLF<sup>1</sup>)

Cette recherche<sup>2</sup> s'inscrit dans la perspective de l'informatisation du *Trésor de la langue française*, grand dictionnaire de langue en dix-neuf volumes produit par le CNRS et en voie d'achèvement. L'objectif est de concevoir les fonctionnalités d'une interface de lecture et de parcours d'un fonds lexicographique de cette ampleur. Une telle couche fonctionnelle, qui donne au lecteur la possibilité de circuler dans un ensemble des données suivant un système de renvois non directifs (ou réseau de liens), est communément appelée «hypertextuelle». Elle est envisagée comme *système d'aide au parcours* dans des dictionnaires informatisés en vue d'offrir à l'utilisateur plusieurs types d'opportunités :

- (i) des appréhensions de faits de langue ;
- (ii) des exemplifications de la pratique du linguistique ;
- (iii) des mises en relation et par suite, des rapprochements ou des regroupements entre faits de sens ;
- (iv) des constructions de domaines cognitifs, *i.e.* de connaissances et de savoirs.

### LA PROBLÉMATIQUE GÉNÉRALE

Les études déjà accomplies en la matière nous ont permis de définir un certain nombre de positions problématiques sur la question de l'hypertextualité et des dictionnaires.

En premier lieu, il convient de préciser le point de vue sur les dictionnaires électroniques. Le fonds informatisé est une base de données lexicales qui se prête à des investigations à travers des procédures de recherche multi-critères. Le niveau de l'hypertextualité apparaît lorsque ce fonds informatisé n'est plus considéré comme un contenu dont il s'agit d'extraire des données vérifiant des propriétés choisies, mais comme une somme textuelle qu'un lecteur souhaite parcourir. Le dictionnaire n'est plus alors pensé comme un simple objet de consultations, mais comme un lieu de lecture à part entière,

---

1. Institut National de la Langue Française, E.N.S., Grille d'Honneur - Le Parc, F-92211 Saint-Cloud cedex.

2. «Lexicographie et hypertextes», en collaboratin avec David Piotrowski (CNRS-INaLF), Ouyen KIEU (Lab. d'Informatique de Paris 6) et Éric Bruillard (Informatique, IUFM Créteil). Programme « Production, validation et mise à disposition de données et d'outils linguistiques », CNRS, Paris.

et il s'agit de déterminer les fonctionnalités qui permettent d'investir le fonds lexicographique suivant l'optique d'un travail interprétatif.

Pour définir un tel dispositif, il est nécessaire d'identifier les principaux types d'activités de lecture auxquelles se prête un fonds lexicographique. En effet, pour garantir la cohérence des pratiques de lecture hypertextuelle d'un dictionnaire, les fonctionnalités doivent favoriser l'appropriation du contenu lexicographique conformément à ses capacités d'instruire et de signifier, *i.e.* à ses potentialités internes de sens et de savoir.

En résumé, l'enjeu consiste à construire une interface fonctionnelle fournissant au lecteur les outils automatiques nécessaires à élaborer des parcours de lecture « consistant », c'est-à-dire des modes de lecture qui visent à agir sur les savoirs et les sens qui contiennent les dictionnaires.

Il s'agit pour ce faire, de construire à la fois :

- (1) un modèle théorique des organisations du lexique et de leur intrication au sémantique ;
- (2) une maquette informatisée favorisant dans un premier temps, l'exploration de contenus linguistiques informatisés, et dans un second temps, la génération d'*indexations* stratégiques lorsque l'utilisateur se place en perspective de recueil personnel de connaissances, c'est-à-dire d'*appropriations cognitives* déterminées.

#### LE MODÈLE THÉORIQUE

Il s'agit donc de prendre en compte les pratiques lexicographiques et de formaliser ces pratiques en ce qu'elles ont d'exemplaires pour une modélisation des générations de sens à partir du lexique en termes de processus architecturant « les passages du mot au sens ».

Concrètement, et à partir des observations expérimentales que nous avons déjà conduites, cela signifie que tout parcours de lecture à partir d'une entrée de dictionnaire pourra emprunter les étapes suivantes dont l'ordre n'est pas nécessairement chronologique :

- (i) *des extractions d'objets-mots (relations de type 1)* : ce qui va être immédiatement perceptible au lecteur du dictionnaire, ce sont les différenciations constitutives du « système langue » ; ce sont celles entre verbes et substantifs, entre articles et prépositions, entre adverbes et verbes, entre places des mots et fonctions de ces mots dans l'énoncé. Autrement dit : tout ce qui relève de la catégorisation descriptive et fonctionnelle du lexique.
- (ii) *des localisations de significations (relations de type 2)*, c'est-à-dire des ancrages de contenus dans des classes qui vont distribuer le lexique selon des similitudes physiques ou socio-culturelles *i.e.* des insertions dans des situations d'usage ou de connaissance ; par exemple : les différentes entités « meubles » qui vont composer la classe du mobilier.
- (iii) *des indexations sémantiques (relations de type 3)*, qui vont correspondre à l'inscription de ces classes dans des *domaines* spécifiés et de là, dans des *univers*.

#### EXEMPLE :

L'entrée « métier » dans le petit Robert électronique va correspondre en langue à deux types de désignations : (i) « l'occupation » ; (ii) « la machine ». À ce premier niveau de définitions (R1), nous

avons donc une relation de différenciations/opposition fondée sur une complémentarité originaire : l'occupation suppose la machine, celle-ci fait le métier.

À partir de cette double entrée, explorons le premier parcours : celui de la machine. Là sont exemplifiés des types de métiers datés comme « métier à tisser », « métier à tricoter ». L'entrée « métier » ne s'emploie plus que pour des machines appartenant aux débuts de l'industrialisation (les filatures) et liées au processus physique et manuel. On sait qu'il existe bien d'autres classes d'activités où il y a des machines qu'on n'appelle plus « métier ». L'ordinateur a généré ainsi les « métiers de l'informatique » ; il n'est pas lui-même un métier. La relation de *définition* (R1) aboutit très vite, on le voit, à une fermeture partielle.

En revanche, la considération du terme « métier » sous l'angle de « l'occupation » conduit à un parcours beaucoup plus riche :

- (i) la localisation (R2) sous forme de situations professionnelles (métiers manuels/intellectuels) conduisant à la différenciation entre « état » (être du métier) et « carrière » (avoir un bon métier) ;
- (ii) l'indexation sémantique en termes de domaines (R3) où les métiers sont spécifiés par types d'activités ou types d'objets (les métiers techniques, les métiers de l'hôtellerie ; par suite (R3), l'ancrage de ces domaines au sein d'univers, qui vont permettre de décliner des réseaux de savoirs et de savoir faire (cuisine, logement, tourisme, etc.).

On voit que les parcours du lecteur peuvent être nombreux d'une entrée à l'autre et que le dictionnaire se constitue de la sorte comme instrument approprié aux extensions cognitives, à l'encontre des représentations qui lui sont communément associées.

#### LA REPRÉSENTATION COMMUNE : LES DICTIONNAIRES COMME SOURCES NORMATIVES

Pour le grand public en effet – et c'est un reliquat des enseignements scolaires, la langue est avant tout perçue comme un système normatif, une ensemble de prescriptions sur les façons d'écrire et d'assembler les mots, dont la maîtrise atteste d'une certaine position dans la hiérarchie sociale. Dans cette perspective, un ouvrage dont la matière est la langue (dictionnaire ou grammaire) constitue une somme de référence, une sorte de « table de lois », à laquelle les locuteurs vont se rapporter pour évaluer la justesse de leurs tournures et de leurs écritures, ou encore pour les guider dans leurs formulations et leurs rédactions. L'orthographe, qui est le parangon de la normativité linguistique, constitue dans cette optique un lieu de préoccupation majeur du commun des locuteurs, et les dictionnaires, dépositaires de la norme orthographique, seront le plus souvent consultés à cet effet.

Le second usage dominant des dictionnaires, à savoir la recherche du sens des mots, est aussi le corrélat d'une conception populaire du lexique : le lexique comme nomenclature. Cette conception s'exprime dans l'opinion que la richesse du vocabulaire témoigne de l'ampleur du savoir. S'il en est ainsi, c'est que les objets du savoir progressivement mis au jour, objets du monde ou de pensée, se sont vus en tant que tels attribuer des dénominations. La connaissance des vocables nommant des objets et des sens qui en donnent la description, équivaudrait alors à la stricte connaissance des dits objets. Confusion, donc, entre l'ordre des mots et l'ordre des choses, entre l'ordre des signifiés et celui

des concepts, qui amène les locuteurs à rechercher dans le dictionnaires, sous la rubrique *définition*, de quoi combler leur ignorance lorsqu'ils rencontrent un terme supposé désigner quelque chose qui est ambigu, voire inconnu.

En d'autres termes, les représentations sociales de la langue confinent les dictionnaires dans le statut de base de données. Un dictionnaire est ainsi regardé comme un ouvrage dont les contenus sont des informations sur le fonctionnement et la forme des unités de la langue – et l'usage des dictionnaires va consister à y localiser les éléments répondant aux questions que se posent les locuteurs. On est aux antipodes des pratiques interprétatives où les données textuelles sont investies par un agent cognitif qui s'implique dans l'élaboration d'une représentations sémantique « travaillant » ces données.

L'usage consultatif des dictionnaires ne procède pas cependant, de la seule logique socioculturelle ; il provient aussi de l'idéal de l'analyse lexicographique qui ambitionne de donner pour chaque mot de la langue une caractérisation univoque réglée suivant un système de catégories descriptives. De nouveau, le texte lexicographique se retrouve pensé comme une somme complexe d'informations susceptibles d'être alors atteintes par l'application de procédures d'interrogation croisées.

Dans leurs utilisations les plus rudimentaires, issues d'une représentation socioculturelle de la langue, comme dans leurs emplois les plus savants, qui prennent appui sur les structures fines de l'analyse lexicographique, les dictionnaires se présentent donc comme des bases de données : ensembles structurés d'informations à consulter. Et on est en droit de penser que l'information des fonds lexicographiques ne va rien changer à cet état de fait. En effet, l'application des structures et des opérations de l'informatique aux données lexicographiques, à première vue, semble se situer en prolongement direct des usages consultatifs traditionnels. L'informatisation d'un recueil lexicographique n'aurait pour conséquence que de faciliter l'accès au fonds sans modifier qualitativement des attitudes de lecture.

#### LES CONSÉQUENCES DE L'INFORMATISATION DES DICTIONNAIRES : L'HYPertextUALITÉ

Cela est sans doute vrai si l'on réduit l'informatisation à la mise en œuvre de procédures de recherches puissantes et à la gestion d'importants volumes structurés. Mais cette vision est par trop réductrice. Car le transfert sur ce nouveau support qu'est un ordinateur ouvre, à travers les conditions et formes de l'hypertextualité, un nouvel espace de lecture et de travail sur les textes qui va permettre d'aborder les dictionnaires comme des livres à part entière, c'est-à-dire, comme des textes faisant l'objet d'une lecture à visée interprétative – et non plus comme des collections d'informations à consulter.

Dans Piotrowski (1994) et Bruillard, Piotrowski, Vignaux (1996) nous avons ainsi montré que l'information a pour effet de scinder l'unité du texte imprimé en trois modules fonctionnels interdépendants. En effet, alors qu'un texte sous format « papier » fusionne jusqu'à les confondre les dimensions :

- de l'enregistrement (ce qui est « enregistré » sur le support papier l'est sous forme de marques typographiques) ;

- de la présentation (ce qui est montré au lecteur, ce sont les marques typographiques);
- et de la consultation (qui suit, sur le texte imprimé, l'enchaînement plus ou moins linéaire des marques typographiques).

Un texte sous format électronique se compose de structures d'enregistrement, d'affichage et de consultation autonomes : l'enregistrement (module de présentation) et des moyens qui permettent de le parcourir pour retrouver des informations (module de consultation).

Dans la logique d'une lecture consultative traditionnelle du dictionnaire (considéré comme base de données), le module de l'enregistrement (du fonds lexicographique) est organisé suivant les catégories et les relations spécifiques aux informations qu'il recèle. Le module de la consultation a alors pour unique fonction de permettre d'accéder, suivant des modalités de requêtes complexes, aux données intangibles stockées en mémoire. Dans ce cas de figure, le module de consultation est entièrement contraint par les formes internes de la structure d'enregistrement et le module de présentation a une fonction purement passive : à savoir rendre manifeste les données enregistrées que les procédures de consultation ont atteintes.

#### L'ARCHITECTURE HYPERTEXTUELLE

La *logique hypertextuelle* apparaît en revanche, lorsque le volume de l'enregistrement n'est plus pensé comme un ensemble de données dont la distribution relationnelle et les contenus sont des invariants (parce qu'ils correspondraient à des spécifications ou des informations fixes et univoques), mais comme une « matière » textuelle, présentant un certain niveau d'organisation dont l'enregistrement rend compte, et que les lecteurs se proposent d'investir à différentes fins. Autrement dit, dans la perspective hypertextuelle, le module de l'enregistrement est une collection, au besoin structurée, de données que les lecteurs vont parcourir à l'aide des modules de consultation et de présentation dans l'objectif de s'approprier ces données, c'est-à-dire de leur apporter une qualification particulière qui aura pour conséquence d'aider à élaborer une certaine représentation sémantique et cognitive. Le système hypertextuel doit donc fournir les outils permettant de réaliser cette reconstruction du fonds d'origine à des fins particulières.

Dans cette perspective :

- (i) le module de l'enregistrement renferme une matière textuelle à qualifier (matière munie d'une structure minimale d'origine sur laquelle les procédures de consultation et de présentation peuvent prendre appui),
- (ii) le module de consultation, outre les procédures de recherche dans le fonds d'origine, comporte :
  - (a) des outils de spécification (annotation, catégorisation...) permettant de qualifier le fonds d'origine et (b) des outils d'aide à la lecture (pour guider les parcours de l'utilisateur en lui signalant des postes possibles de lecture),
- (iii) le module de présentation ne se borne plus à la fonction d'affichage des résultats de la consultation mais a pour but de rendre manifestes au lecteur des possibilités d'interprétation et de prolongement de ses parcours de lecture.

C'est sur cette base que nous visons à développer des outils informatiques susceptibles de se constituer comme systèmes d'assistance au lecteur dans :

- (i) la construction de mises en relation et d'indexations,
- (ii) la génération de domaines cognitifs fondés sur ces indexations.

Deux principales sortes de fonctionnalités sont ici envisagées, à savoir :

- d'une part, des fonctions de collecte destinées à effectuer des confrontations (comparaisons et oppositions) de données et des fonctions d'indexation permettant une qualification sémantique explicite en vue d'un traitement automatique ultérieur,
- et, d'autre part, des fonctions de parcours qui mettent l'accent sur la construction de représentations du monde en domaines d'expérience cohérents.

#### **PUBLICATIONS RELATIVES AU PROJET**

- PIOTROWSKI, D., 1994, « Dictionnaires électroniques, Questions d'architecture », *Le Français Moderne*, LXII, n° 2.
- PIOTROWSKI, D., 1995, « Les dictionnaires informatisés », in *Les concepts opératoires de l'INaLF*, INaLF, p. 72-75.
- BRUILLARD, E., PIOTROWSKI, D., VIGNAUX, G., 1995. *Du dictionnaire considéré comme hypertexte*, Rapport d'Étape, INaLF, (120 p.).
- PIOTROWSKI, D., 1996, « Opérations hypertextuelles et formes lexicographiques », in D. PIOTROWSKI, (éd.), *Lexicographie en Informatique. Autour de l'informatisation du Trésor de la Langue Française*, Actes du Colloque International de Nancy, Didier Érudition, Paris.
- PIOTROWSKI, D., 1997, « Formes opératoires de l'hypertextualité » (à paraître) in *Savoirs lexicaux et savoirs encyclopédiques*, D. PIOTROWSKI et P. STROCKINGER (éds), *Sémiotiques*, n° 12, juin 1997.
- VIGNAUX, G., 1996, « Hypertextes, dictionnaires: approche sémantique, perspective cognitive », in D. PIOTROWSKI (éd.), *Lexicographie en Informatique. Autour de l'informatisation du Trésor de la Langue Française*, Actes du Colloque International de Nancy, Didier Érudition, Paris.