



HAL
open science

Des textes au concept. Propositions pour une approche textuelle de la conceptualisation

Mathieu Valette

► **To cite this version:**

Mathieu Valette. Des textes au concept. Propositions pour une approche textuelle de la conceptualisation. Journées Francophones d'Ingénierie des Connaissances, Jun 2010, Nîmes, France. pp.5-16. hal-00491037

HAL Id: hal-00491037

<https://hal.science/hal-00491037>

Submitted on 10 Jun 2010

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

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

Des textes au concept

Propositions pour une approche textuelle de la conceptualisation

Mathieu Valette¹

¹ATILF, CNRS, Nancy-Université
mvalette@atilf.fr

Résumé : Il s'agit d'esquisser les conditions théoriques d'une approche textuelle de la conceptualisation. Nous souhaitons illustrer l'hypothèse qu'avant d'accéder au statut de signes dont les signifiés sont normés (*i.e.* les termes), les concepts émergents se manifestent dans les textes sous des formes hétérogènes telles que des groupements récurrents de traits sémantiques qui se stabiliseront – ou non – en unités lexicales nouvelles. Ces coalitions ont valeur de préconcepts ou de *protosémies*. L'enjeu est de décrire et de modéliser ce processus d'émergence pour lui donner, à terme, une place dans une théorie de la terminologie. Au plan pratique, l'objectif est de fournir à moyen termes des outils d'identification et de détection pour la veille et la constitution de terminologies.

Mots-clés : Linguistique, Sémantique textuelle, Formes sémantiques, Protosémie, Conceptualisation.

1. La fouille de données : métaphore minière ou archéologique ?

La terminologie textuelle (Slodzian 2000) s'est consacrée à l'extraction de candidats termes dans les textes pour les expertiser et le cas échéant les valider comme concepts termes. Elle a promu l'idée que les textes sont les lieux de production des termes mais ni les conditions de cette production ni l'éventualité d'une conceptualisation liée aux textes eux-mêmes ne semblent encore avoir retenu l'attention. C'est l'hypothèse générale que nous souhaitons défendre et illustrer, à contre-courant du postulat platonicien dont l'ingénierie des connaissances a hérité, selon lequel le concept préexiste au terme qui le désigne. Il n'est toutefois pas question d'adopter ici une posture logicienne et de considérer que la conceptualisation est un phénomène purement linguistique répondant à des règles de construction propres à une fonction du langage, ni de faire abstraction des conditions psychiques, sociales et interactionnelles de l'élaboration des concepts. Il s'agit d'une part de considérer que la textualité, c'est-à-dire les contraintes propres à la mise en texte, à la formulation linguistique, qu'il s'agisse de contraintes grammaticales, sémantiques, lexicologiques ou liées aux *traditions discursives* (parmi lesquelles les genres et les discours) (Koch 1997) joue un rôle clé dans la formation des concepts, et d'autre part d'appréhender les textes non plus seulement comme des ressources, des carrières qu'il



convient de forer pour en extraire la matière première terminologique ou conceptuelle, mais comme des *archives*, c'est-à-dire la trace objective du processus de création des concepts.

C'est donc dans une lignée davantage foucauldienne et archéologique que minière que nous souhaitons nous inscrire. Pour Foucault, il s'agit notamment d'identifier des *formations discursives*, c'est-à-dire des régularités entre objets, types d'énonciation, concepts et choix thématiques à l'origine de la production idéologique ou scientifique, autrement dit à l'élaboration de connaissances partagées. De là résulte l'importance d'en déterminer les règles de formation (Foucault 1969, 53). De fait, l'activité cognitive qui consiste à créer des concepts, les modifier, les ordonner et les articuler entre eux, se confond souvent avec l'activité de production de texte. On pourrait détourner pour la faire nôtre une proposition de G. Bachelard (1938, 61) et considérer qu'un concept est un groupement d'« approximations successives » – lesquelles sont, selon nous, observables dans des corpus diachroniques.

Nous souhaitons donc ici ouvrir un débat destiné à illustrer l'hypothèse qu'avant d'accéder au statut de signes dont les signifiés sont normés (*i.e.* les termes), les concepts émergents se manifestent dans les textes sous des formes hétérogènes, approximatives, telles que des groupements récurrents de propriétés sémantiques qui se stabiliseront – ou non – en unités lexicales nouvelles. Ces groupements ont valeur de préconcepts ou de *protosémies*, c'est-à-dire d'unités sémantiques antérieures à toute lexicalisation. L'enjeu est de décrire et de modéliser ce processus d'émergence pour lui donner une place dans une théorie de la terminologie. Au plan pratique, l'objectif sera de construire, à terme pour l'heure non échu, des outils d'identification et de détection pour la veille et la constitution de terminologies.

2. Signifiés et concepts, mots et termes

Le mot est un concept linguistique fragile. A la fois imprécis de par ses frontières théoriques et matérielles, et ethnocentrique parce que les langues sans mot sapent tout espoir d'en faire un concept universel, il demeure néanmoins un mode d'aperception du langage parmi les plus intuitifs. Très étudié par les linguistes, il est aussi très utilisé en ingénierie des connaissances. Une approche textuelle du mot pourrait, comme le suggère (Rastier 2001, 182-183) à propos d'une reconception possible du signe, s'inspirer d'un texte où Saussure écrit :

« [...] vous n'avez plus le droit de diviser, et d'admettre d'un côté le mot, de l'autre sa signification. Cela fait tout un. Vous pouvez seulement constater le kénôme  et le sème associatif  » (Saussure 2002, 93)

Le signe (*sème* signifiant *signe* pour Saussure) est donc « contextuellement défini » selon Rastier et peut-être vu comme un passage vide entre deux contextes, gauche et droite, autrement dit, le signe n'a de valeurs que celles que lui octroient le contexte. La radicalité de cette conception du signe nous agréé, tant elle semble aux antipodes de l'approche terminologique traditionnelle. En effet, si la lexicologie a

recours en maints lieux théoriques et pratiques au contexte pour définir le contenu sémantique du mot (par la collocation notamment)¹, la terminologie quant à elle, rapatrie nécessairement le sens dans le terme. Dans ce cadre général, nous présentons un ensemble de propositions adossé à la sémantique textuelle visant à situer l'étude du lexique dans le paradigme textuel. Plus précisément, notre projet est d'étudier les déterminations textuelles de la création et de la lexicalisation des concepts.

3. Forme sémantique et concept

Adoptons un empirisme de méthode : les mots et les termes, qui en sont un cas particulier, apparaissent dans deux types d'objet matériel ; le texte, objet construit de façon syntagmatique, où ils sont actualisés dans un état qu'on pourrait qualifier de dynamique et la ressource terminologique ou lexicale (thésaurus, lexique, dictionnaire, etc.), objet construit de façon paradigmatique, où ils sont dans un état passif, en attente d'une actualisation. Qu'est-ce qu'un mot, par-delà ces types d'objet ? Il est un signe constitué d'une forme et d'un contenu. La forme est acoustique ou graphique, c'est le signifiant, le contenu est, dans le paradigme structuraliste, une collection de propriétés sémantiques articulées entre elles, qui constituent le signifié. Ces propriétés sont des *sèmes*. Elles sont d'ordre métalinguistique et résultent concrètement d'une analyse ou d'une validation humaine effectuées par un expert. On peut distinguer deux types de collections de sèmes, selon que l'on se trouve dans une problématique du texte ou dans une problématique du dictionnaire.

3.1. Les signifiés comme concept ?

Composé d'un signifié et d'un signifiant, nous représenterons un mot de la façon suivante :

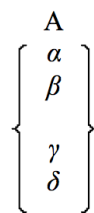


Fig. 1 – Un signifié composé des sèmes α , β , γ , δ associé à un signifiant A

A désigne le signifiant (c'est-à-dire, en pratique, le mot graphique) et α , β , γ , δ entre accolades désignent le signifié proprement dit (composé des sèmes regroupés). Tous ces traits n'ont pas nécessairement la même qualité ni la même valeur

¹ Lire (Blumenthal & Hausmann, éd., 2006).

sémantique, ce que figure le saut de ligne entre β et γ . Certains sèmes, par exemple, sont génériques d'une classe donnée. Ainsi, le signifié du mot « *chien* » peut être décrit comme la collection de sèmes présentée dans la partie gauche de la figure 2. Dans la classe sémantique des *Canidés*, le sème /canidé/ est générique, et /domestique/ est un sème spécifique de « *chien* » qui le différencie du renard ou du chacal par exemple. Si l'on construit une classe sémantique des *Animaux de compagnie*, /domestique/ sera, à l'inverse, un sème générique tandis que /canidé/ sera spécifique de « *chien* », puisque ni le chat, ni le poisson rouge ne sont des canidés. Il est à noter que l'hétérogénéité zoologique de la classe des *Animaux de compagnie* ne préjuge pas de sa cohérence : lorsqu'il s'agit de choisir un animal de compagnie, la sélection s'effectue sur des propriétés propres à la classe (l'encombrement, l'autonomie, etc.) et non sur des critères zoologique par exemple. En cela, ces classes sémantiques, qu'on appelle des *taxèmes*, ne sont pas assimilables à des catégories ontologiques classiques. Un taxème, en effet, est une petite classe sémantique correspondant à une situation d'usage précise et n'a vocation ni à l'universalité, ni à l'intangibilité. Un domaine, dans cette perspective théorique, est composé d'un ensemble de taxèmes correspondant à une pratique déterminée.

Dans la mesure où les sèmes sont des constructions, la liste n'en est ni exhaustive, ni fermée. Elle ne relève pas d'une quelconque valeur de vérité mais d'une *valeur d'usage*. Ainsi, le contenu sémantique de « *chien* » construit par un informateur âgé de quatre ans peut fort bien correspondre au signifié de droite, toujours sur la figure 2.

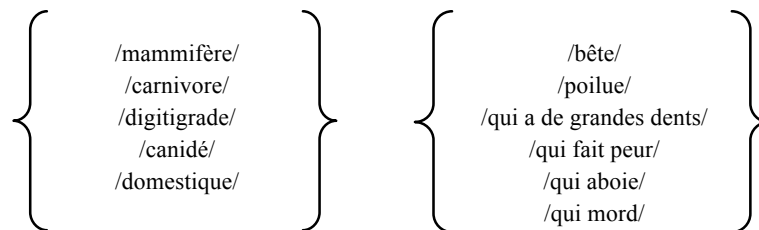


Fig. 2 – Deux signifiés possibles pour « *chien* ».

Aucun de ces signifiés, qui sous-tendraient deux définitions lexicographiques parfaitement distinctes, n'est meilleur qu'un autre. Les propriétés sémantiques diffèrent par leur seul contexte de construction. Le premier est savant, le second se fonde sur une ou plusieurs expériences sensibles ou fantasmées. Dès lors, dans une perspective pratique, on peut se poser la question de la permanence du concept de « *chien* ».

3.2. Les réseaux sémiques

Un texte est, formellement parlant, un alignement de signes (de mots) suivant des règles de construction syntaxiques. La mise en relation dans un texte de plusieurs signifiés donne lieu à de nouveaux regroupements de sèmes, syntagmatiques cette fois-ci, c'est-à-dire des groupements entre sèmes appartenant à des signifiés

différents. Ces groupements syntagmatiques sont occurrenceiels car spécifiques à un texte, ou à un ensemble de textes. L'interprétation d'un texte repose sur la reconnaissance et l'identification de ces groupements. On distingue deux types de groupements syntagmatiques, les fonds et les formes sémantiques.

3.2.1. Les fonds sémantiques (isotopies)

Lorsqu'un sème donné se retrouve à plusieurs endroits dans un même texte, il peut s'agir d'une isotopie. Les isotopies se structurent en réseaux et constituent le fond sémantique. Dans le texte qui suit, on observe une isotopie simple, par la récurrence du sème /tabac/ :

Le plus difficile dans l'arrêt du tabac^{/tabac/}, c'est la décision. Et cette décision doit venir d'une certitude que la vie sans tabac^{/tabac/} existe. J'étais un très gros fumeur^{/tabac/}. M'imaginer sans cigarette^{/tabac/} était un cauchemar².

Minimalement, cela signifie que ce texte traite de tabac. Mais les isotopies peuvent relever d'une description plus fine. Par exemple, elles peuvent être génériques, c'est-à-dire correspondre aux sèmes structurants de classes sémantiques telles que les domaines et les taxèmes. L'isotopie constitue le socle du parcours interprétatif, notamment par les échanges d'informations sémiques qui s'opèrent entre elle et les formes sémantiques. On en donnera une illustration dans le paragraphe 4.3.

3.2.2. Les formes sémantiques

Plusieurs sèmes distincts peuvent être instanciés ensemble dans des textes différents avec une certaine régularité. Ces groupements s'appellent des thèmes, ou des *formes sémantiques*. Nous les représenterons ainsi :

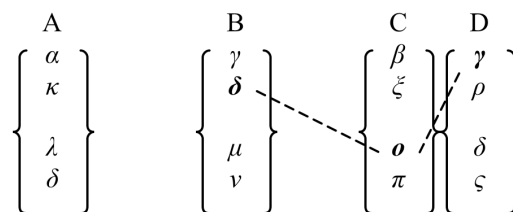


Fig. 3 – Une forme sémantique :
groupement stabilisé des sèmes δ , θ et γ dans différents textes

Par exemple dans les textes suivants :

Le citoyen est libre^{/liberté/} de fumer^{/fumer/} ou de ne pas fumer, de manger de la salade si ça lui chante et des rillettes s'il en a envie³

² <http://sante-medecine.commentcamarche.net/>

³ <http://www.le-tigre.net/>

Je suis contre l'interdiction de fumer^{/liberté/}, chacun est libre^{/liberté/} de son choix^{/liberté/4}

Les sèmes /liberté/ et /fumer/ sont en cooccurrence rapprochée. Ils constituent une petite forme sémantique, c'est-à-dire une unité sémantique dont la lexicalisation n'est ni stable ni synthétique, autrement dit, dont le signifiant est discontinu et variable. Cette unité sémantique correspond, par exemple, à l'unité lexicale « *liberté de fumer* » que l'on peut considérer comme un début de lexicalisation :

Ils ont la liberté de fumer^{/liberté//fumer/}, soit, je serai le dernier à la leur retirer, sauf dans les endroits publics⁵.

Les formes sémantiques ne sont pas des périphrases, elles importent en premier lieu pour les modalités qu'elle opèrent sur les signifiés des unités lexicales cooccurentes. La forme sémantique $\{/liberté//fumer/\}$ s'est particulièrement développée et tend à se stabiliser depuis la mise en application de la loi dite antitabac en janvier 2008. Elle constitue notamment un des arguments privilégiés de ses détracteurs (ce serait une loi liberticide). Parmi eux, les industriels du tabac ont beau jeu de l'exploiter, mais ils ont soin de restreindre cette liberté de fumer aux seuls adultes, la loi leur interdisant de faire la promotion du tabac auprès des enfants. Or, par un habile jeu rhétorique, leur discours peut fort bien avoir l'effet contraire, c'est-à-dire promouvoir le tabac chez les jeunes. Etudions ces quelques extraits de la documentation présente sur leurs sites Web :

Notre métier ne consiste pas à inciter des gens à fumer^{/fumer/}. Il consiste à offrir des marques de qualité à des adultes qui ont déjà pris la décision^{/liberté/} de fumer^{/fumer/} [...]. C'est pourquoi nous sommes convaincus que fumer^{/fumer/} devrait être le seul fait d'adultes conscients des risques de fumer^{/fumer/}. (JTI)⁶

Fumer^{/fumer/} repose sur une décision^{/liberté/} individuelle qui ne peut être que le fait d'adultes informés des risques liés au tabagisme. (ALTADIS)

Nous sommes convaincus que le choix^{/liberté/} de fumer^{/fumer/} doit être le choix d'adultes avertis et conscients des risques, un choix qui exclut de fait les jeunes, non adultes. (BAT)

Nous nous engageons ainsi à communiquer de manière responsable avec les adultes qui ont délibérément choisi^{/liberté/} de fumer^{/fumer/}. (ALTADIS)

JTI s'engage à fabriquer des cigarettes^{/fumer/} de qualité pour les adultes qui choisissent^{/liberté/} de fumer^{/fumer/} par plaisir. (JTI)

Les industriels du tabac associent à la forme sémantique $\{/liberté//fumer/\}$ une autre forme dont le pivot est « *adultes* ». Mais dans ces brefs énoncés, « *adultes* » subit un

⁴ <http://www.agoravox.fr/>

⁵ <http://www.philo5.com/>

⁶ Ces exemples et les analyses sont issus des recherches collectives de l'équipe linguistique (ERTIM-INaLCO, Paris) du projet C-MANTIC (ANR-07-MDCO-002).

certain nombre de modalités valorisantes, le plus souvent sous la forme d'adjectifs qualitatifs (« *conscients* », « *informés* », « *avertis* », etc.) de sorte que se construit un parcours interprétatif liberté de fumer + adulte^{/valorisant/}. Ce discours de valorisation de l'adulte libre de fumer, autrement dit, de l'adulte fumeur, rend possible une lecture spéculaire telle que le jeune, non adulte, est dévalorisé. C'est explicite dans le troisième exemple (« un choix qui exclut de fait les jeunes, non adultes »). On peut donc construire les deux formes sémantiques ci-dessous :

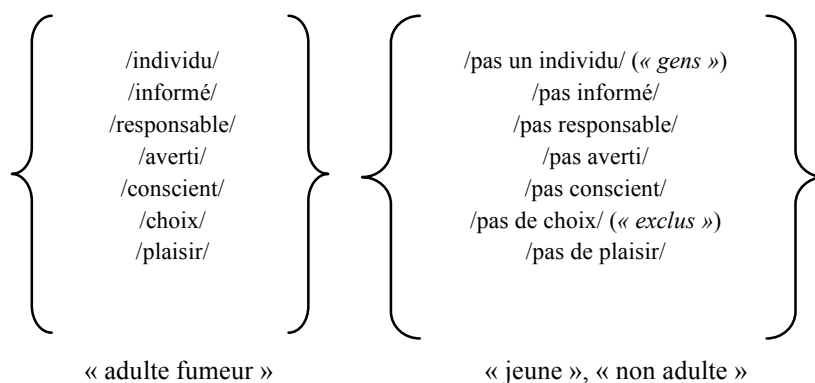


Fig. 4 – Formes sémantiques possibles d'« adulte » et de « non adulte » dans les textes de l'industrie du tabac

La production de nouveaux sèmes pour le concept existant « adulte » participe au succès du discours promouvant le tabac chez les jeunes et les échecs répétés de la lutte contre le tabac : les messages de l'industrie du tabac sont massivement positifs et euphoriques quand ceux des organismes de prévention sont négatifs, dysphoriques⁷.

4. Discussion : ajuster les signifiés et les formes sémantiques

4.1. Les déterminations sémantiques du texte

Le sens d'une unité lexicale dépend autant du contexte dans lequel il apparaît que de sa définition première. En complément de l'opposition sèmes génériques vs sèmes spécifiques (qui permet d'introduire la notion de classe sémantique), la sémantique interprétative opère une distinction entre des sèmes inhérents et des sèmes afférents. Les sèmes inhérents peuvent être considérés comme plus « définitoires » (par exemple, pour « chien », on aura : /canidé/, /quadrupède/, etc.) que les sèmes afférents

⁷ C'est du moins un des résultats d'analyse de l'équipe du projet C-MANTIC auquel nous empruntons ce matériau.

qui relèvent de l'usage qui est fait du mot dans les textes. Les sèmes afférents sont donc issus des contextes d'énonciation. C'est le cas dans l'exemple de la figure 4 où le signifié d'*adulte* n'est représenté ici qu'avec des sèmes afférents. Il s'agit de sèmes hérités d'un ensemble de contextes et susceptibles d'être recontextualisés à l'identique.

De la même manière, le sème /pauvre/ est fréquemment actualisé dans « *population* » lorsqu'il est cooccurrent du lexème « *banlieue* ». Mais il ne s'agit pas d'un sème inhérent : rien dans la banlieue ne la prédispose à accueillir de façon privilégiée une population pauvre. De même, l'expression « *jeunes des banlieues* », dans le discours journalistique (qui est aujourd'hui largement prescriptif) ou dans le discours politique (qui lui ressemble) ne signifie pas tous les jeunes résidant en banlieue, mais certains jeunes, défavorisés, résidant en banlieue, et même, fréquemment, des jeunes gens « *issus de l'immigration* ». Lorsque le quotidien *Le Monde* titre le 30 septembre 2001 : « Les jeunes des banlieues craignent l'amalgame entre musulmans et terroristes », il enrichit implicitement le signifié de l'unité « jeune de banlieue » d'un sème /musulman/. Ces sèmes, /pauvre/ ou /musulman/ sont des sèmes afférents, « subjectifs » ou « socialement normés », c'est-à-dire circonscrits d'un point de vue historique, géographique et socioculturel. Aux Etats-Unis, les pauvres vivent en centre-ville. En revanche, lorsque « banlieue » s'intègre dans certaines lexies composées telles que « *banlieue de l'ouest parisien* », le sème /pauvreté/ est inhibé par le sème /bourgeois/, afférent à « *ouest parisien* », quand même la banlieue ouest de Paris est tout aussi hétérogène que la banlieue dans son ensemble.

Les sèmes contenus dans un signifié ne sont pas tous égaux. Leur nature et leur qualité varient en synchronie (tous les sèmes n'ont pas la même valeur dans le signifié) comme en diachronie (un même sème peut évoluer dans le temps, disparaître, etc.). Les sèmes oscillent entre stabilité et instabilité. Le signifié est constitué de sèmes résistants à la variation, sinon permanents et de sèmes instables. Cette variabilité résulte de l'enrichissement ou de l'appauvrissement du signifié à mesure que le mot est actualisé dans les textes. D'une certaine façon, chaque actualisation d'un mot l'enrichit de son contexte d'actualisation, la fréquence de sa participation à un groupement transversal modifie son signifié. Nous dirons, pastichant ainsi une formule de Rastier, que tout mot placé dans un texte en reçoit des déterminations sémantiques, et modifie potentiellement le signifié de chacun des mots qui le composent⁸.

Ainsi, le sens résulte d'ajustements entre des signifiés et des formes sémantiques. De la même façon que le cortex visuel traite moins d'informations issues du nerf optique que d'information stockée en mémoire, l'interprétation résulte autant – sinon

⁸ La proposition initiale de (Rastier 2001, 92) est que « Tout texte placé dans un corpus en reçoit des déterminations sémantiques, et modifie potentiellement le sens de chacun des textes qui le composent ».

davantage – d’une activité sémique intense (reconfiguration du signifié, ajustement, convocation des afférences possibles en mémoire, etc.) que du texte lui-même⁹.

4.2. La forme sémantique comme *protosémie*

L’élaboration d’un signifié subit des contraintes textuelles et intertextuelles. Les contraintes intertextuelles sont notamment liées aux discours et aux genres, les contraintes (intra)textuelles sont liées à l’économie ou l’organisation sémique du texte¹⁰. Dans ce contexte, et compte tenu de ce que nous avons présenté dans le paragraphe précédent, nous proposons de considérer la forme sémantique comme le signifié potentiel (ou le sémème potentiel) d’un signe sans signifiant synthétique attiré. Soit l’équivalence hypothétique présentée dans la figure 5.

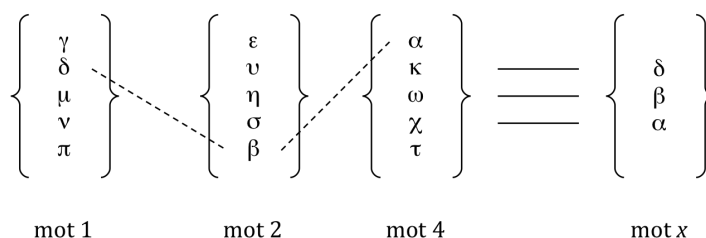


Fig. 5 – La forme sémantique comme protosémie

En effet, signifié et forme sémantique sont tous deux identifiables à un groupement sémique de compacité variable, associé à un signifiant stable et synthétique dans le cas du signifié, discontinu et sans lexicalisation privilégiée dans le cas de la forme sémantique. En bref, nous faisons l’hypothèse que certaines formes sémantiques sont potentiellement des signifiés en devenir, ou les signifiés de protosignes sans signifiant stabilisé ni synthétique attiré. En d’autres termes, certaines formes sémantiques sont des *protosémies*.

Dans le contexte d’une recherche appliquée, en veille lexicale par exemple, le concept de protosémie a pour incidence la possibilité d’identifier le signifié en cours d’élaboration d’un signe en analysant le complexe sémique dans lequel il s’insère. Pour cela, nous empruntons à (Rastier 2006) les concepts de *diffusion* et *sommation* qui rendent compte des échanges sémiques entre les fonds et les formes sémantiques. La *diffusion* relève de propagation des sèmes des formes sémantiques vers le fond sémantique. La *sommation* ressortit à la propagation du fond sémantique vers des formes sémantiques. Si le signifié peut être considéré à la manière d’une forme sémantique, il est susceptible de se *diffuser* dans le fond sémantique. Dans ce cas, il nous est possible de restituer une protosémie, c’est-à-dire un signifié-forme sémantique en postulant la *sommation* de ladite forme à partir du fond sémantique que

⁹ (Rastier & Valette 2009) en présentent quelques exemples dans le cadre de l’évolution sémantique d’une unité lexicale existante (ou *néosémie*).

¹⁰ On lira Valette (sous presse) pour un développement.

nous identifions par le biais d'*isotopies locales*. C'est ce que nous étudierons dans le paragraphe suivant en relatant une étude récente qui illustre notre propos.

4.3. Naissance d'un concept

(Reutenauer *et al.* 2009) évaluent les déformations subies par une forme sémantique dont l'élément stable est constitué de l'unité lexicale « *économie réelle* » dans un corpus de presse constitué de 1587 articles, tirés du *Figaro* et de *l'Humanité* entre septembre 2008 et février 2009. Le thème du corpus est la crise économique et financière. Il se présente sous forme de deux versions parallèles : la version lexicale, d'un million d'occurrences de formes, et une image sémique de ce même corpus, composé de 23 millions d'occurrences de candidats-sèmes. A partir d'un calcul de spécificités (implémentation Lexico3, Salem *et al.* 2003), les auteurs mesurent la sensibilité des informations au contexte éditorial des deux quotidiens du corpus. Ainsi, au voisinage d' « *économie réelle* », *l'Humanité* active les sèmes /bien/ (substantif), /revenu/, /consommation/ et /dépense/, tandis que /ressource/ et /économie/ sont actualisés par *le Figaro*. Les auteurs avancent l'hypothèse aisément corroborable d'une perception plus macroéconomique de l'économie dans *le Figaro* et plus localisante dans *l'Humanité*.

S'appuyant sur une méthodologie similaire, (Reutenauer *et al.* 2010) effectuent une analyse de l'environnement sémique d'un signifié en cours d'élaboration, celui d'« *Outreau* ». Le corpus porte donc sur l'affaire judiciaire dont « *Outreau* », la ville, est l'éponyme. Divisé en cinq périodes, il est constitué d'articles de presse publiés entre novembre 2001 et avril 2006 comprenant au moins une occurrence du mot étudié. (Lecolle 2007), à laquelle est emprunté le corpus, observe que le sens d'« *Outreau* » évolue du toponyme à « l'erreur judiciaire par excellence ». Comme dans l'étude précédente, le corpus se présente sous deux versions parallèles : la version lexicale de 400 000 occurrences de formes et une image sémique de 10 millions de candidats dont a été extraite une sous-image constituée des seuls candidats correspondant à des formes rendues saillantes par un calcul de spécificités effectué là encore avec Lexico3. Ainsi, les auteurs ont à manipuler une sélection de quelques dizaines de candidats-sèmes seulement.

L'une des analyses effectuées par les auteurs nous intéresse particulièrement. A partir de listes de candidats-sèmes spécifiques à chaque période, des regroupements sémantiques sont réalisés manuellement. Selon nous, ces regroupements peuvent être assimilés à des isotopies. Par exemple, le regroupement de candidats tels que /police/, /procureur/, /écrouer/ s'apparente à une isotopie domaniale //judiciaire// particulièrement présente dans les périodes précoces du corpus. En son sein, on peut observer les traces de différents champs génériques : l'//arrestation// apparaît à travers l'ensemble de candidats-sèmes {/écrouer/, /police/, /arrestation/, /incarcération/, /incarcérer/, /prévenu/}. L'isotopie domaniale //politique// préfigurant le sens d'« erreur judiciaire par excellence », apparaît dans la quatrième période et se renforce ensuite. Inversement, l'isotopie taxémique des //dénominations de crimes//

(comportant des candidats tels que /pédophilie/, /meurtre, /viol/) décroît en importance en quatrième période puis disparaît. Les isotopies taxémiques locatives (/lieu d'habitation//, //lieu géographique//) ne sont représentées significativement qu'à la première période. L'isotopie //fiasco// (constitué de /nauffrage/, /drame/, /faillite/, /faute/) est représentative de la cinquième période.

Si nous considérons que les faisceaux d'isotopies locales observables ici sont le résultat d'une diffusion de la forme sémantique d'« *Outreau* » vers le fond, nous pourrions, par *simulation sommatrice* inverse, restituer la forme sémantique, c'est-à-dire le signifié en cours d'élaboration d'« *Outreau* », lequel est minimalement définissable par un changement de sème générique : le sème domanial //judiciaire// se substitue au sème inhérent toponymique /ville/, avant d'être inhibé pour céder sa place au sème domanial //politique//.

Ainsi, le signe « *Outreau* » qui désignait une ville du Nord, se voit investi, par le contexte, de plusieurs valeurs sémantiques qui le conduisent à devenir un concept, celui de scandale politico-judiciaire.

5. Conclusion

Nous avons souhaité exposer ici à grands traits un ensemble de propositions théoriques destiné à mettre en évidence d'une part l'importance du texte dans la formation des unités lexicales et des concepts correspondants, même dans les situations où l'on préjuge d'une stabilité ontologique *a priori* forte (par exemple le *chien*, l'*adulte*, une entité nommée tel qu'*Outreau*) ; d'autre part les promesses d'une description linguistique et sémantique de ces phénomènes de conceptualisation au prisme des textes. En cela, l'outillage théorique de la sémantique des textes (formes sémantiques, complexe sémique, protosémie, etc.) s'avère, selon nous, pertinent. La mise en œuvre d'une telle méthode permettrait de repérer les variations des concepts dans des textes, en synchronie comme en diachronie, et d'en organiser les différentes facettes.

Toutefois, les études signalées dans cet article relèvent de la linguistique de corpus et ont une visée explicitement exploratoire. Elles ressortissent à une sémantique *enrichie* qui nécessite des temps de traitement et d'analyse des données incompatibles avec des exigences industrielles. Leur finalité est de mieux comprendre les mécanismes nécessaires à l'élaboration, la production et l'interprétation des concepts en situation réelle à partir de textes et d'en tirer partie, à terme, pour des applications ciblées. Il reste évidemment à concevoir les outils et les procédures nécessaires à une opérationnalisation.

Références

- BACHELARD, Gaston, (1938) *La formation de l'esprit scientifique*, Paris : Vrin.
- BLUMENTHAL, Peter / HAUSMANN, Franz Josef, éd. (2006) « Collocations, corpus, dictionnaires », in : *Langue française*, Paris, Larousse.
- FOUCAULT, Michel (1969) *L'archéologie du savoir*, Paris, Gallimard.
- KOCH, Peter (1997): « Diskurstraditionen: zu ihrem sprachtheoretischen Status und ihrer Dynamik », in : Barbara FRANK *et al.* (éd.), *Gattungen mittelalterlicher Schriftlichkeit. Tübingen*, p. 43-79.
- LECOLLE, Michelle (2007) « Polysignifiante du toponyme, historicité du sens et interprétation en corpus. Le cas de *Outreau* », *Corpus* n°6, p. 101-125.
- RASTIER, F. (2001) *Arts et sciences du texte*, PUF, Paris.
- RASTIER, François (2006) « Formes sémantiques et textualité », in : *Langages*, n°163, p. 99-114.
- RASTIER, François / VALETTE, Mathieu (2009) « De la polysémie à la néosémie », *Le français moderne*, Salah MEJRI, éd., *La problématique du mot*, n°77, p. 97-116.
- REUTENAUER, Coralie / VALETTE, Mathieu / JACQUEY, Evelyne (2009) « De l'annotation sémique globale à l'interprétation locale : environnement et image sémiques d' "économie réelle" dans un corpus sur la crise financière », *Colloque ARCo'09, Interprétation et problématiques du sens*, (9-11 novembre 2009), Rouen.
- REUTENAUER, Coralie / LECOLLE, Michelle / JACQUEY, Evelyne / VALETTE, Mathieu (2010), « Sémème au microscope : genèse et variation sémiques d'une unité lexicale », in : *Actes JADT'2010* (9-11 Juin 2010), Rome.
- SALEM, André / LAMALLE, Cédric / MARTINEZ, William / FLEURY, Serge / FRACCHIOLLA, Béatrice / KUNCOVA, André / MAISONDIEU, Aude (2003) « Lexico3 – Outils de statistique textuelle. Manuel d'utilisation. », Syled-CLA2T, Université de la Sorbonne nouvelle – Paris 3.
- SAUSSURE, Ferdinand de (2002) *Ecrits de linguistique générale*, Paris, Gallimard.
- SLODZIAN, Monique (2000) « L'émergence d'une terminologie textuelle », *Le sens en terminologie*, Ph. THOIRON et H. BEJOINT, éd., Lyon, Presses universitaires de Lyon (Travaux du C.R.T.T.), p. 61-85.
- VALETTE, Mathieu (sous presse) « Méthodes pour la veille lexicale », in : *Actes de la journée d'étude Le dictionnaire électronique. Quelles perspectives pour les sciences humaines et sociales ?*, Leila Messaoudi (éd.), Publication du laboratoire Langage et société, Université Ibn Tofail Kénitra (disponible sur <http://hal.archives-ouvertes.fr/>).