Remerciements

- Je remercie

  Monsieur Daniel Herman pour avoir bien voulu accepté de présider le jury de cette thèse.

  Monsieur Patrice Enjalbert et Monsieur François Rastier, les deux rapporteurs de cette thèse, pour l'intérêt qu'ils ont tous deux manifesté à l'égard de mes travaux, ainsi que pour la pertinence de leurs remarques et de leurs conseils.

  Monsieur Jean-Pierre Barthélemy pour avoir dirigé ces travaux. Qu'il reçoive ici ma reconnaissance pour la bienveillance et la bonne humeur dont il a agrémenté ces quelques années de recherche.

  Monsieur Jacques Siroux pour m'avoir fait l'honneur de participer au jury,

- Je remercie particulièrement Monsieur Ioannis Kanellos, qui a encadré, guidé et éclairé mon parcours parfois sinueux. J'espère que cette thèse reflétera fidèlement l'humilité et le respect de la difficulté inhérente des problèmes liés à la formalisation de la langue qu'il a su, au cours de longues discussions, et de courtes altercations, m'inculquer.

- Je remercie également Madame Michèle Noailly, et toute l'équipe du séminaire *Sémantique en contexte* de l'Université de Bretagne Occidentale, pour m'avoir fait découvrir les nombreuses facettes de la linguistique, et m'avoir fait comprendre la relativité d'une formalisation face à une discipline qui laisse aussi place à la sensibilité.

- Je remercie chaleureusement Madame Monique Le Coz, et les membres du département Informatique de la Faculté des Lettres de l'Université de Bretagne Occidentale, pour la confiance et la sympathie qu'ils m'ont accordées, et les facilités d'organisation qu'ils m'ont permises pour terminer la rédaction de ce travail.

- Je remercie la joyeuse équipe du Département IASC de Télécom Bretagne, qu'ils soient enseignants, ex-membres, thésards, ex-thésards ou stagiaires. Je doute qu'un travail de thèse puisse se dérouler dans une ambiance plus franche et sympathique que celle qui y règne au quotidien.

- Je remercie mon ami Pierre pour son soutien sans faille, ses créations théâtrales et les dimanches-après-midi-thé.

- Je remercie bien sûr tous mes collègues, amis, et ma famille pour avoir supporté les sautes d'humeur liées à ce travail.

- Je remercie enfin tous ceux qui prennent le temps de lire les remerciements que l'on trouve au début d'une thèse.

Résumé

Nous proposons dans cette thèse une modélisation informatique de la sémantique interprétative de F. Rastier. Après une première critique des traitements classiques du langage naturel sous ses aspects sémantiques, nous concluons à une non-réductibilité du sens d'un texte à une structure calculée à partir de descriptions locales, et à la nécessité de déterminations globales. Nous étudions ensuite la théorie linguistique de la sémantique interprétative, afin d'en retirer les concepts centraux captables par une approche informatique. Nous proposons un modèle formel de description de ceux-ci, ainsi que des mécanismes de manipulation des structures sémantiques descriptives. Enfin, nous proposons un logiciel appliquant cette formalisation, sous la forme d'une assistance à l'analyse des textes. Nous rejoignons une appoche de coopération homme-machine, en distinguant les activités interprétatives propres à l'utilisateur de leur manipulation formelle par la machine. Nous concluons par une mise en relief des avantages suggestifs et descriptifs d'une approche informatique de certains aspects de l'interprétation.

Abstract

This thesis proposes a computer model for F. Rastier's interpretative semantics. After a criticism of classical natural language processing for semantics, we conclude that there is a need for a description that focuses on the impact of global determination on local descriptions. Through an analysis of the linguistic theory, we identify the main descriptive concepts and then propose a formal description. This formalism consists of a number of semantic structures and operators to manipulate them in order to describe an interpretation process. We then propose a software system which applies these concepts, and provides a computer aid for the semantic analysis of texts. According to a human-computer cooperation paradigm, we distinguish between human interpretative processes and their formal description for management by the machine. We conclude by an emphasis on the advantages of computer assistance for both suggestion and representation purposes.

Avant Propos

1 Environnement

Cette thèse a été réalisée au sein du Laboratoire d'Intelligence Artificielle et Sciences Cognitives de l'École Nationale Supérieure des Télécommunications de Bretagne, sous la direction de Jean-Pierre Barthélemy et l'encadrement de Ioannis Kanellos. Elle a bénéficié d'un financement sous la forme d'une allocation de recherche du Ministère de la Recherche.

2 Cadre scientifique

Le travail présenté dans ce mémoire pourrait s'inscrire confortablement dans ce domaine interdisciplinaire qu'est le traitement de la langue naturelle, également appelé informatique linguistique (mais dont l'idéal de référence serait celui d'une linguistique informatique). Il rejoindrait ainsi nombre de travaux plus ou moins spécialisés tentant de modéliser le sens d'un énoncé en langue naturelle. Cependant, nous désirons revendiquer une approche différente de cette problématique. Le problème du sens, sans doute le plus vertigineux de ceux rencontrés par l'homme, a souvent, à notre avis, été sous-estimé par les informaticiens et autres artificiers de l'intelligence, selon le terme consacré.

C'est dans ce cadre général de réflexion, concrétisé déjà par un petit ensemble de travaux, que nous nous situons. À partir d'une réflexion menée sur les concepts fondateurs de la représentation de connaissances en IA, et sur les notion d'identité dans un cadre formel général, nous proposons d'aborder ici, sous cet angle, le problème de la formalisation du sens dans le traitement de la langue naturelle. Le problème central restera celui de l'objectivité illusoire du sens, telle qu'elle est affirmée ou sous-entendue par les approches formelles de l'informatique appliquée au langage. La subjectivité, elle, déclarée détestable car ennemie d'une simplification opératoire, n'a jamais, ou presque, été abordée directement, respectée voire revendiquée.

Ces considérations générales, heureusement, entrent en résonance avec d'autres visions plus centrées sur le langage. La revendication de la centralité de l'individu dans les rapports à la langue transparaît dans certaines théories linguistiques, dont nous avons sélectionné la sémantique interprétative de F. Rastier.

Cette théorie nous a séduit par plusieurs aspects. Au-delà de forts louables déclarations d'intention, elle constitue en effet une avancée majeure dans la rationalisation des phénomènes de sens, sans pour autant considérer l'immanence de celui-ci. La notion d'interprétation, pilier central de cette approche, définit en effet le sens comme influencé par un ensemble de considérations « extérieures » dont on ne peut guère décrire que l'incidence dans l'ordre linguistique. En ce sens, elle se rapproche des travaux de traitement automatique de la langue naturelle, où certains problèmes centraux du langage sont résolus *ad hoc* par l'invocation subite d'un nébuleux « contexte ». Il s'agira ici de l'intention première, en reconnaissant que le sens se construit, et que l'environnement (au sens large) de l'acte d'interprétation prime.

Ces grands principes sont cependant coûteux pour une opérationnalité immédiate. Malgré un appel à des notions structuralistes rigoureuses, ne pourra être décrite qu'une possible interprétation d'un texte, et non une compréhension que certains voient idéalement comme absolue.

3 Principes de notre approche

Nous prenons en quelque sorte le relais de l'avancée théorique linguistique. Notre premier travail sera de préciser le rôle de la machine dans la prise en considération de ces principes généraux. Nous nous rapprocherons ainsi du paradigme récent de la coopération homme-machine anthropocentrée, que l'on peut voir comme une alternative opératoire au vu des échecs des automatisations complètes issues du cognitivisme dogmatique. Il s'agira, parallèlement de cerner la zone d'immodélisabilité des activités humaines, et de la respecter. L'ordinateur, en abandonnant son autonomie illusoire, devient alors un assistant, doté de qualités qu'il faudra aussi reconnaître : sa capacité de calcul, qui ne doit pas prépondérer, sa faculté de représentation et sa possibilité de gérer certains ordres logiques, et de veiller sur certains aspects de sa cohérence.

Dans un second temps, nous devrons analyser plus en détail les aspects conceptuels de la théorie linguistique, en traçant, ici aussi, des limites entre le formalisable et l'inaccessible pour la machine. Nous nous permettrons dans certains cas de prendre parti, et de modifier en toute conscience certains aspects, mais respectant la cohérence de l'ensemble de la vision interprétative.

Il sera alors temps de réfléchir aux modalités de cette fameuse coopération homme / machine dans cette situation précise d'un utilisateur interprétant un texte. Le rôle principal de la machine sera alors celui d'une motivation à la rationalisation. Sans réfuter une traduction par trop formelle des concepts linguistiques dégagés, nous les revendiquerons comme garants d'une cohérence globale de l'interprétation. Cherchant à obtenir une identité formelle des unités sémantiques, le logiciel que nous proposerons permettra ainsi un « dialogue » avec l'interprète, il l'aidera dans la construction d'une nouvelle forme de discursivité mise au profit de l'analyse. L'interprète se verra sollicité tout au long de son acte d'attribution de sens.

Nous proposons ainsi un prototype d'interface homme / machine destiné à une assistance au cours d'une analyse de texte. Nous l'avons baptisé PASTEL (approximativement pour Programme d'Aide à l'Analyse de TExtes, même Littéraires).

Par moments fastidieuse, par d'autres sources de suggestions, cette nouvelle relation entre l'homme et la machine possède deux aspects techniques intéressants.

Le premier est l'absence d'utilisation de données linguistiques générales : la seule connaissance de la langue mise en place dans l'utilisation du logiciel proviendra de l'utilisateur. Celui-ci pourra donc en toute liberté utiliser les descripteurs sémantiques (également exprimés en langage « naturel ») pour traduire son interprétation, et ne sera pas contraint par des formes figées de la langue, ni limité dans son application à un domaine sémantique particulier. D'un autre côté, il pourra effectivement « dire n'importe quoi », mais le second point l'en dissuadera.

Effectivement, en contrepartie de cette souplesse d'expression, le logiciel mettra en place une structure de données, s'appuyant aveuglément sur les données descriptives que l'utilisateur lui communique, pour organiser formellement le résultat de l'interprétation. Un ensemble de contraintes formelles à ce sujet seront autant de requêtes d'éclaircissement qu'il transmettra à l'utilisateur. Ce qui mettra à rude épreuve une vision fantaisiste d'un texte, qu'il faudra justifier plus loin que la simple énonciation de thèmes généraux. En quelque sorte, la vision première de l'utilisateur / interprète sera explorée plus profondément peut-être qu'il ne l'aurait fait seul, et peut mener à des découvertes d'un nouvel aspect sémantique du texte analysé.

4 Plan du mémoire

Notre présentation se déroulera en cinq chapitres, dont le contenu général sera le suivant :

1. Premier chapitre, à forte teneur introductive, dans lequel nous plaçons notre approche au sein du traitement de la langue naturelle, en y dégageant quelques principes directeurs. À travers une critique des systèmes existants, pour lesquels la syntaxe joue un rôle central, nous précisons les besoins d'un affranchissement des considérations que celle-ci véhicule (principe de compositionnalité, limite de la phrase, etc.). Dans un deuxième temps, toujours en prenant appui sur des travaux en TALN, nous précisons les limites de la considération du sens comme information, et de la sémantique référentielle, pour justifier le recours au paradigme différentiel, en développant la notion de situation du sens, et la pertinence de la notion d'interprétation. Enfin, nous discutons des approches récentes dans le domaine de la coopération homme/machine, précisant l'orientation de notre approche vers l'assistance et non l'automatisation, en redonnant à la machine un rôle de suggestion et non de décision.

2. Second chapitre, consacré à l'étude de la sémantique interprétative comme fondement théorique du modèle présenté par la suite. Après une description des grands principes de la théorie, nous passons en revue les aspects plus formels de celle-ci (notions de sème, sémème, taxème, isotopie, etc.), en précisant leur utilisation, leur remaniement ou leur rejet, en fonction de l'application informatique.

3. Chapitre plus formel, où nous décrivons la structure de données décrivant le résultat d'une interprétation d'un texte, en reprenant les principes et considérations dégagés au chapitre précédent. Cette structure est ici décrite de façon statique, comme finalité de l'acte d'interprétation, et comporte un ensemble de contraintes quant à l'organisation structurelle des différentes entités sémantiques

4. Dans le même esprit que le chapitre précédent, nous y décrivons la structure sous son aspect dynamique, en établissant notamment un « protocole » d'interprétation, ou mise en place de la structure. Nous prenons soin d'y détacher les étapes automatiques des nécessaires (et nombreux) recours à la compétence humaine. Nous y insistons également sur l'aspect positif des contraintes précédemment évoquées

comme source de suggestion et de créativité. Y sont également proposées quelques transformations générales de la structure, afin de permettre la cohésion de celle-ci avec des contraintes externes, ou comme source possible de l'approfondissement d'une interprétation.

5. Dernier chapitre, décrivant les aspects informatiques de la thèse, notamment la transcription de la structure formelle, l'organisation générale du logiciel proposé, et de l'interface graphique qui le rend utilisable et convivial. Le principe général de cet outil informatique est d'accompagner l'utilisateur dans la description de son interprétation d'un texte par repérage de sèmes. Il n'est pas fait appel à des bases de données stables, l'utilisateur ayant toute liberté (et donc nécessité) d'attribuer les sèmes qui lui semblent pertinents aux éléments du texte qu'il désire prendre en compte. Les contraintes précédemment citées ont donc deux rôles : garantir autant que faire se peut la cohérence de l'interprétation, en garantissant l'identité sémantique de chaque signifié, mais également, inciter l'utilisateur a expliciter au maximum son interprétation, par un jeu de questions-réponses, une réponse étant ici le repérage d'un sème et son attribution a un signifié.

Le travail présenté ici a donné lieu à quelques publications : [26], [27], [59].

Chapitre 1

Problématique au sein du Traitement Automatique de la Langue Naturelle

1 Introduction

Dans ce premier chapitre, nous tenterons de définir nos objectifs, ainsi que les principes qui les sous-tendent.

Nous nous intéresserons particulièrement au traitement de la sémantique du langage naturel. Les principes que nous revendiquons sont les suivants :

- Tout d'abord, nous reconnaîtrons la primauté de la sémantique dans le langage, dont le traitement doit ainsi se dégager des spécificités des méthodes appliquées à la syntaxe. À ce sujet, nous envisagerons les distinctions entre syntaxe formelle et syntaxe linguistique. Nous récuserons ainsi les tentatives de manipulation / définition du sens à travers une analyse, si fine soit-elle, des phénomènes de surface du langage, qui correspond à une vision empruntée aux systèmes formels de l'articulation sémantique/syntaxe. Le principe de compositionnalité qui y est supposé devra également être explicité et rejeté dans le cadre de la sémantique. De plus, nous accorderons un statut premier au texte, et non à la phrase, dont le régime, du point de vue logique, est celui de la limite supérieure de l'application des méthodes syntaxiques (ce qui n'a plus lieu d'être si l'on aborde la sémantique, où au contraire c'est le texte (et le contexte) qui prime(nt)).

- Ensuite, nous nous placerons dans une position interprétative. Tout d'abord, nous accuserons les assimilations entre compétences interprétative et générative, en explicitant leur origine dans une vision restrictive du langage comme vecteur d'information. Voulant aborder le sens d'un énoncé linguistique comme éminemment *situé*, et donc dépendant des conditions de communication et des intentions de son destinataire, nous récuserons la symétrie précédente et nous nous dirigerons ainsi vers une vision différentielle de la sémantique. Cette prise de position dans le structuralisme nous permettra de nous interroger sur l'identité des unités sémantiques, qui devra être construite et structurée et non simplement vue comme le résultat d'un décodage.

- Enfin, nous expliciterons plus en détail le rôle que nous assignons à l'outil informatique dans le cadre du traitement de la langue naturelle. Au vu des précédentes constatations, nous préciserons notre passage d'une automatisation impensable, puisque seul l'humain peut supporter les notions de l'interprétation, à une assistance et une source de suggestions. Après avoir explicité les conditions de cette coopération entre l'homme et la machine, nous discuterons également des types d'applications envisageables (et envisagées), tant du point de vue de la méthode que du domaine linguistique.

2 Place de la syntaxe dans le TALN[1]

À partir de la prédominance constatée de la syntaxe dans les applications du TALN, alors que tout le monde reconnaît le sens comme objectif, il faut bien se poser la question de sa nature et de son rôle dans l'approche du sens. De plus, la syntaxe telle qu'elle est définie pour les systèmes de traitement de la langue (systèmes informatisés donc formels) est-elle bien celle dont on parle dans des considérations plus centralement linguistiques? Nous tenterons donc de dégager les tenants et les aboutissants de la profusion d'applications centralement syntaxiques dans le TALN, et comparer ceux-ci avec nos objectifs et prises de position.

2.1 Prédominance de la syntaxe dans le TALN

Revenons quelques temps sur les tentatives initiales en traduction automatique, telles qu'elles fleurirent après la dernière guerre, motivées par la volonté américaine de traduire automatiquement les messages russes. Les productions en ce sens, avant d'être déclarées inutilisables par la comité ALPAC[3] en 1965, s'appuyaient essentiellement sur l'utilisation de corres-

1. TALN est un AQL[2] pour Traitement Automatique du Langage Naturel.
2. Acronyme de Quatre Lettres.
3. Automatic Language Processing Advisory Committee, dont l'initiative a vu le jour en 1952.

pondances mot-à-mot. Les traditions sous-jacentes à l'époque provenaient plus de la cryptographie (dont le succès dans son traitement informatique était validé) que de la linguistique véritable.

À partir de là, deux programmes majeurs furent envisagés au vu des échecs : affiner l'analyse syntaxique ou apporter des connaissances générales au système. On peut citer par exemple Weaver, reconnaissant que le «décodage du russe vers l'anglais» [66] est plus complexe qu'une simple correspondance, et Bar-Hillel [3], qui reconnaît le besoin de connaissances sur le monde, et non plus sur la langue (lexique et grammaire) pour parvenir à une traduction acceptable. S'affirment donc deux programmes : le premier vise l'affinement syntaxique (comme affinement/complexification de la structure) ; le second concerne le mode de coordination du thème de la connaissance avec l'objet linguistique.

Pour l'instant, nous étudierons la première possibilité : développement des méthodes d'analyse morphologique (dictionnaire de formes simples et règles de composition des mots, cela dès les premières tentatives de traduction, pour diminuer la place en mémoire des dictionnaires), et surtout syntaxique, afin de déterminer la structure de la phrase, et atteindre ainsi une meilleure compréhension, donc traduction. Dans le cas de la traduction, cela suppose toujours une correspondance, non plus au niveau du mot, mais au moins au niveau du syntagme et de la proposition.

Ceci se traduit également par le développement de formalismes syntaxiques originaux, à partir des simples grammaires syntagmatiques, en cherchant à atteindre un taux de couverture maximal des phrases correctes analysées. À ce propos nous ne pouvons bien entendu pas nous affranchir de citer les travaux de Chomsky [11, 12], et leur abondante tradition critique qui a nourri presque toute la recherche en TALN, qui s'inscrivent pleinement dans cette vision du sens (structure profonde) accessible par la forme (structure de surface). Nous ne nous lancerons pas dans une critique maintes fois proposée, mais reconnaîtrons simplement que la théorie de la grammaire transformationnelle constitue l'approfondissement le plus notable de cette vision générale.

Ces formalismes ne furent pas utilisés exclusivement à des fins de traducteurs automatiques. Une grande majorité d'outils de TALN utilisent un analyseur morpho-syntaxique ou *parser*. Que ces outils servent à l'interrogation de bases de données en langage naturel, à la production de résumé ou l'indexation de documents, en bref des outils qui visent à une certaine forme de compréhension de la langue, ils passent presque toujours par une phase purement syntaxique. C. Fuchs [19] reconnaît à la syntaxe un statut de nécessité pour la généralité de l'outil, dans le sens de la variété des énoncés que celui-ci traite. À travers des domaines variés, de la météorologie à la classification des bateaux [24], en passant par la simulation de dialogues psychanalytiques [68] (de nombreux ouvrages discutent de ces différents outils, notamment [4], [15], [55]), la constante serait donc la structure syntaxique, sorte de substrat originel de la langue. D'ailleurs cette primauté doit être remise en question, dans la mesure où elle contredit des données établies en psychologie expérimentale. On peut en effet établir facilement l'accès de l'enfant à la langue par la sémantique, et l'apparition tardive d'une compétence syntaxique (voir à ce propos les travaux de J. Piaget [45]). Parle-t-on alors de la même syntaxe dans les deux cas?

2.2 Syntaxe linguistique et syntaxe formelle

Nous nous interrogerons ici sur la séparation syntaxe / sémantique telle qu'elle est définie dans les systèmes formels. Cette distinction, entièrement justifiable dans un rôle formalisateur très général en logique, semble impliquer une projection directe dans la langue, renforcée par d'autres analogies (morphologique et pragmatique), comme indiqué par la figure 1.1, alors que la syntaxe de la langue n'est pas indépendante. En témoignent déjà les travaux de Benvéniste [6, 7] et sa reconnaissance du fait que le sens de la phrase influe sur le sens des mots. Pour résumer son raisonnement, disons que ce sont les rapports forme / sens qui traduisent ces notions de composition. La forme d'une entité est sa composition en entités inférieures et son sens est fonction de sa place dans une entité supérieure. Ainsi, la forme de la phrase ne peut être la suite des mots qui la composent, car le sens de ces mots dépend de celui de la phrase, et elle n'a pas de sens puisqu'elle ne rentre pas dans la composition logique (propositionnelle) d'une autre entité, étant elle-même une proposition.

Ceci récuserait donc le principe de compositionnalité du sens, qui posait déjà des problèmes à Frege (cf. son article *Sens et dénotation* dans [18]), et a motivé l'arrivée des logiques intensionnelles. De ces problèmes que pose la compositionnalité, nous retiendrons également la centrale question de l'identité des unités sémantiques, dont nous parlerons plus loin, en discutant de l'approche différentielle. La possibilité ou impossibilité de substituer des unités linguistiques d'un contexte à l'autre présuppose en effet une identité de ces unités au travers de leurs diverses utilisations: il s'agit d'une identité de forme, mais que dire alors de la polysémie? La notion d'identité sémantique devra donc être considérée avec attention, et devra sans doute être construite et non présupposée. Nous verrons plus loin la nécessité de la considérer comme construite par le sujet qui traite l'énoncé, en y intégrant

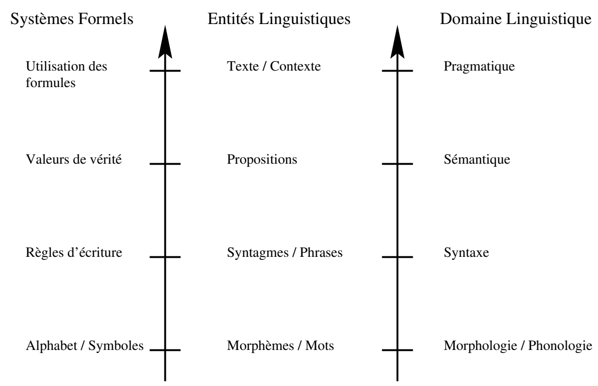


FIG. 1.1 – *Systèmes formels et niveaux linguistiques traditionnels*

des données externes, relatives à la situation de ce traitement.

Même dans l'hypothèse d'une influence mutuelle des différents sens des mots dans la phrase, on serait conduit à une complexité exponentielle (au sens technique du terme, il s'agit d'un problème NP-complet), qui de toute façon supposerait une connaissance absolue des significations et emplois possibles pour chaque élément.

Si la syntaxe possède cette vertu compositionnelle, si toute phrase isolée peut être analysée de façon grammaticale, alors le sens ne peut s'y réduire, il fait entrer en jeu des considérations qui dépassent au moins la limite de la phrase : d'où la nécessité d'un recours au contexte, et ce contexte est plus qu'une phrase, voire quelques phrases.

À ce propos, la notion de contexte a récemment connu une poussée productive. Mais les récentes interrogations formalistes sur le contexte (voir par exemple [10] et [71]) ne traduisent pas autre chose. La constante mise en situation ou contexte de l'unité linguistique (constante quelle que soit la nature de cette unité) ne peut qu'infléchir les développements dans la recherche d'une description formelle de ce contexte. Cependant, le contexte est plus que cela, il est plus qu'une fonction de choix des sens possibles, plus que la trace d'une heuristique sémantique (comme il est souvent invoqué dans certaines approches du TALN, comme [8]). C'est un moment essentiel dans la constitution de l'objet linguistique : il ne vient ni après ni avant mais en même temps que lui.

2.3 La syntaxe pour la sémantique

L'articulation entre syntaxe et sémantique, projetée dans le monde du TALN, se traduit de différentes façons. Il y a tout d'abord le schéma souvent invoqué d'un traitement séquentiel de l'information langagière, à l'aide de différents modules spécialisés, dans l'ordre : module morphologique (reconnaissance des mots), module syntaxique (identification des syntagmes et des structures de phrases), module sémantique (modélisation du contenu des mots et de la phrase), et module pragmatique (utilisation des données sémantiques, en fonction du type d'application : traduction, réponse à une question, inférences, etc.) Un système répondant à ce schéma central est par exemple le célèbre SHRDLU de Winograd [69] qui manipule un robot à partir d'ordres énoncés en anglais, ou le LUNAR de Woods [70] qui traduit les questions en requêtes pour une base de données.

Même si la syntaxe s'analyse correctement, ne laissant subsister que quelques ambiguïtés, le travail du traitement sémantique sensé la suivre n'est jamais résolu correctement (en tout cas, sûrement pas dans le cas d'une généralité d'énoncé que l'analyse syntaxique est sensée mettre à sa portée). Ceci orienta donc la recherche vers le développement de formalismes «mixtes», comme la LFG (*Lexical Functional Grammar*) [1], où les notions sémantiques, traduites par des formules logiques, sont intégrées directement dans le traitement syntaxique. On voit également des tentatives d'ajouter des considérations sémantiques à des formalismes d'analyse syntaxique, comme le propose A. Abeillé pour les TAGs[4] dans [52].

Mais il faut aussi noter le développement de formalismes purement sémantiques, pour attaquer à nouveau le problème. Ces formalismes rejoignent ainsi des considérations plus générales sur la représentation de connaissances, mais on y retrouve toujours une prédominance des principes fondamentaux de la syntaxe, à savoir la notion de calcul et de compositionnalité. Cette fois, des notions sémantiques sont calculées, mais toujours à partir de données locales attribuées aux mots, même si elles sont bien plus riches que de simples catégories grammaticales. Notons également, mais nous y reviendrons par la suite, les présuppositions que traduit la pluralité d'utilisation de ces formalismes. Leur classicisme en IA, et dans le domaine de la représentation de connaissances induit une prise de position sur le langage, dont il faut prendre conscience. Il s'agit bien de formalismes de représentation de connaissances, pas de représentation de sens. Leur utilisation est compréhensible lors de l'application à un système classique de l'IA, comme la résolution de problèmes ou la réponse à des questions sur une base de

4. *Tree Adjunction Grammar*

données (ou base de connaissances), mais dans un cas général de représentation du sens, la prudence est de mise, d'autant plus que la notion même de connaissance reste opaque. Dans un cadre applicatif, comme celui de l'IA, la notion de connaissance est liée à celle de représentation, et une telle assimilation des deux réalités pose donc le problème de la représentativité du sens, dont nous discuterons par la suite.

Il faut également noter la (ré-)apparition de méthodes d'analyses de corpus s'inspirant du paradigme distributionnaliste [23], mettant en œuvre une approche sans aprioris de la sémantique basée sur des considérations de surface, mais purement empiriques et statistiques. Il faut y voir ici un autre moyen de signifier, voire d'affirmer l'accessibilité du sens par l'analyse de la surface. La seule notion extérieure à la restriction empirique des données traitées vient ici de la reconnaissance de l'unité du corpus (comptes-rendus d'actes médicaux comme dans [22]) : le principe de compositionnalité, s'il n'est pas directement exploité, n'en reste pas moins présent. Il est cependant important de noter la diminution de la complexité des relations syntaxiques (au sens large) utilisées : ne subsistent plus que des considérations d'identité de forme, de différence, et de proximité. Nul besoin en effet pour classifier des formes linguistiques, de les articuler autour de notions grammaticales complexes (et d'ailleurs inapplicables dans le cas de corpus importants) : la complexité ici provient justement du nombre. Nous nous permettrons d'y voir ici un argument pour la reconnaissance de la sémantique, par une diminution de la complexité de la modélisation syntaxique. De plus, et nous y reviendrons à la fin de ce chapitre, ces méthodes ne produisent que des propositions quant à la nature du sémantique, en ce sens qu'elle laissent, pour l'affirmation de phénomènes liés aux corpus analysés, la conclusion finale à l'homme qui va interpréter ces données statistiques.

2.4 Premier principe

Le problème du sens demande à être abordé directement et prioritairement. La rupture épistémologique doit être ici de la même taille que celle proposée par le «tout syntaxique» de Chomsky. Ce qui veut dire qu'il s'agit de négliger la syntaxe dans un premier temps, et de reconnaître la primauté de la sémantique. Cela permet au moins de s'affranchir des principes inhérents à la syntaxe, dont nous avons discuté l'inappropriation précédemment.

Renier le principe de compositionnalité revient également à rejoindre Benvéniste (*Les niveaux de l'analyse linguistique*, [6], p. 119s) pour considérer la phrase comme limite, mais non plus de la linguistique, mais de la syntaxe. Dès lors l'objet de la sémantique devient au moins un objet supérieur à la phrase, sans refuser les paliers inférieurs. Il doit être un tout, pour supporter des considérations globales, qui influenceront ses éléments, et non l'inverse. Cet objet primordial est le texte, seule entité également susceptible de situation (*cf* infra).

Deux questions se posent alors : tout d'abord, existe-t-il une possibilité de représenter formellement suivant ce principe des données sémantiques ? Et si oui, quels seront les rapports d'un tel traitement avec la réalité informatique ?

3 De la nature du formalisme sémantique

Nous discuterons ici plus centralement de la sémantique. À partir d'une constatation de la possibilité de confusion entre les deux pôles de manipulation du sens dans le TALN (génération et compréhension), nous discuterons des restrictions qu'entraîne une vision référentielle et/ou informationnelle de la langue. À la suite de quoi nous plaiderons en faveur d'un paradigme différentiel.

3.1 Langage et information

On peut dans un premier temps repérer deux faces générales intéressant le sens : la génération et la compréhension. Leur comparaison offre dès lors la possibilité d'en envisager la symétrie. Si cette symétrie est reconnue, quelles hypothèses implique-t-elle sur la nature de la langue, et quelles influences sur son traitement ? Si ces deux aspects transparaissent dans les plus ambitieuses des applications envisageables, traduction et dialogue homme/machine, cette situation doit-elle impliquer une symétrie ?

Si nous reprenons un schéma simple d'une situation de communication, tel qu'il est utilisé pour représenter un transfert d'information, on retrouve trois entités principales : le destinateur, producteur du message, le destinataire, et le message proprement dit. Les rapports entretenus par les deux participants avec le message sont alors de même nature : un passage du cognitif (ou conceptuel, bref du non-linguistique) au linguistique, et vice-versa. Pour reprendre un pied volontaire dans la théorie de l'information, ce sont des opérations d'encodage et de décodage, l'outil linguistique, manifesté par le message, est donc un simple vecteur de transmission de cette information. Lorsque Jakobson [25] envisage les fonctions du langage, il reprend comme base ce schéma simple et général, pour le compléter comme nous verrons par la suite.

Si ces supposées fonctions d'encodage / décodage ne sont pas réversibles, elles n'en gardent pas moins une nature similaire, reposant sur l'opposition entre linguistique et non-linguistique. Le sens premier d'un énoncé est en effet vu comme une représentation mentale initiale, devant être reconstituée par le destinataire lors de l'acte de communication, et introduit par ce biais une qualification de la communication, comme transfert accompli ou échoué de ce contenu informationnel, en supposant une comparaison des deux structures de représentation. La reconstitution en question est, en effet, une reconstruction ; on recherche ainsi toujours la même chose : l'identité entre le modèle et sa reconstitution. Tout le comprendre devient une sorte de quête d'identité, et toute la génération un problème dont la solution repose sur la possibilité de sa réversibilité.

Mais surtout, cette vision simpliste, mais sans doute insidieuse subordonne la linguistique à une description de ces structures (que nous pouvons appeler cognitives) encodées. Outre le fait que de telles structures ne sont pour l'instant pas traitées ni modélisées de façon satisfaisante, cela n'améliore pas la situation de la linguistique, dont Saussure [56] énonçait comme premier objectif de se définir elle-même. P. Siblot [57] rappelle ceci à juste titre, en considérant les rapports de la linguistique avec ces disciplines reconnues comme connexes, et parfois hiérarchisées que sont la psychologie, la philosophie, et parfois même la physique.

Certes, une vision ainsi simplifiée à l'extrême ne trouve plus beaucoup d'adeptes, surtout après une telle caricature, mais nous verrons que certaines assertions plus complexes peuvent trouver un écho dans cette constatation, et nous enfermer dans une vision restrictive de la notion de sens d'un énoncé linguistique.

C'est d'ailleurs en suivant cette vision que les formalismes classiques de représentation de connaissances en IA ont trouvé leur place dans le TALN. Moults réseaux sémantiques [9] [13], Frames [41] et graphes conceptuels [58] servirent de support aux fameux modules sémantiques des systèmes de TALN à l'architecture classique. L'activation de ces structures, initialement envisagées pour modéliser la mémoire, et justifiés (ou anéantis) par des expérimentations en psychologie, résulte alors d'un appauvrissement considérable du contenu linguistique, afin de n'en retenir que quelques concepts pré-définis et articulés par des relations pour le moins ambiguës (la plus célèbre étant le *is-a* des réseaux sémantiques). Ce point de rencontre, s'il rejoint la prise de position de Bar-Hillel, à laquelle s'opposait en quelque sorte le programme chomskien, n'a pas non plus arrangé les affaires d'identité de la science du langage par rapport à la psychologie [5].

5. Et l'inverse non plus, les modèles hiérarchiques de représentation de connaissances, et autres modèles de la typicalité [32] s'étant vu accusés de n'utiliser comme validation

3.2 Sens situé et interprétation

Reprenons alors les fonctions de Jakobson dans leur totalité. Les trois premières concernaient les deux protagonistes de l'acte communicatoire, et le «monde», par la fonction référentielle. Il y ajouta ainsi des fonctions mettant en jeu des relations sociales entre les communicants, et des considérations sur le message lui-même ou le «code» employé. Autant de considérations qui laissent transparaître que l'acte de communication est en partie défini par un entourage régi par des systématicités variées, et que le seul message isolé de celui-ci, considéré comme il peut l'être dans les technologies de télécommunications, ne suffit plus.

Dès lors, si l'on considère comme importante cette notion de *situation*, la dualité linguistique / conceptuelle se complexifie d'autant plus, et la belle symétrie initiale se ternit. Ainsi, la notion de compréhension doit s'abstenir d'un simple décodage qui peut ou ne peut pas être «efficace», au vu de ces données complexes. La notion d'efficacité est assurément préjudiciable à l'intérieur de la compréhension. Outre le fait qu'elle traduit une volonté de quantifier le comprendre, elle témoigne d'un attachement au test de Turing [62] qui tente de définir les approches sous le régime du succès. Or, déjà, la notion même de succès est une inférence interprétative, et peut supporter différentes définitions. Elle ramène par ailleurs le vieux mirage béhavioriste à un moment fondamental de la constitution de l'objet linguistique.

La sélection des données «externes» et «internes» au message, et leur intégration dans la signification sous-entend un rôle actif de la part du récepteur. Comme le développe F. Rastier [53], une grande part de créativité lui revient, puisque l'acte de comprendre peut également faire intervenir des données qui lui sont propres ; on parle alors d'interprétation. Son résultat n'est pas moins quantifiable (réussite / échec) que la compréhension, mais sa description peut plus aisément intégrer une notion de cohérence, plus pertinente sans doute que celle de réussite. La notion de performance ne nous informe pas plus sur une prétendue compétence que sur ses propres critères définitoires. Mais surtout, la notion d'interprétation permet d'expliquer la multiplicité des «codes» applicables au langage : à chacun correspond une nature d'éléments extérieurs à la réalité du texte, ou à sa propre structuration. De la notion de norme linguistique aux présupposés et objectifs de l'interprétation, en passant par les aspects historiques des interprétations «traditionnelles» du texte, un grand nombre de facteurs interviennent, ou du moins sont susceptibles d'être pris en compte. D'une théorie de l'interprétation on ne peut bien sûr pas exiger une explicitation exhaustive de

que des résultats provenant d'expérimentations à trop forte teneur langagière.

ces aspects, mais du moins on peut en espérer une porte ouverte vers leur respect, qu'une vision référentielle exclusive rejetterait par principe.

3.3 L'autre sémantique

La constatation précédente implique donc un choix épistémologique sur le type de sémantique à prendre en compte. La notion classique de contenu d'un énoncé linguistique répond au paradigme de la sémantique référentielle (ou dénotationnelle), résultant bien souvent de l'isolement du signe. Une alternative est donc la sémantique différentielle, principe structuraliste intégrant le signe dans un système, et lui conférant une valeur en fonction de ses relations avec d'autres signes. La notion de contenu référentiel est donc subordonnée à ces relations, et peut même, pour des raisons d'assainissement théorique, être éludée. La vision différentielle est l'assise d'un méta-langage de description de l'objet linguistique (le texte) : elle n'annonce pas un retour à un structuralisme désuet, mais la mise en relief du caractère central des relations d'identité et de différence dans la description de l'ordre sémantique. Ce thème critique de la textualité ne peut en effet trouver d'appui à l'intérieur d'une vision structuraliste standard.

Ainsi, sur la base de ces relations formalisables entre signes (en étudiant les rapports entre les signifiants et entre les signifiés), la notion d'interprétation peut être développée, comme méthode de *construction* de l'identité sémantique. Le rôle d'une entité sémantique est en effet à définir pour chaque texte et chaque situation. La théorie de l'interprétation que nous exploiterons par la suite et répondant, comme on s'en doute, à ces critères structuralistes, ne nous informera pas sur la formation du sens à proprement parler, mais se voudra une rationalisation d'un effet différenciateur d'une entité sémantique. Elle nous permettra d'expliciter ce qui fait *un* sens pour *un* texte et *un* interprète dans *un* contexte....

Certes, la part de description formelle ne sera pas éludée de notre approche, pour la simple raison du besoin d'une opérationnalité minimale. Simplement, en nous ramenant à des outils descriptifs simplifiés, l'impact du choix de la méthode formelle est réduit : si c'est bien une syntaxe dont il s'agira, elle ne nous conduira pas à l'apparition de structures présupposées classiques. Sur le substrat de ces fonctionnalités simples, le concept de créativité pourra se développer avec moins de contraintes.

3.4 Deuxième principe

Grâce à ce paradigme différentiel, capable de décrire par des relations simples (et donc formalisables et implémentables) une identité sémantique, il est possible de redonner à l'interprète sa spécificité situationnelle. La volonté de ne pas se limiter à un contenu informationnel du sens suppose en effet une souplesse dans les outils de description, d'où une nécessité de se ramener à des concepts fondamentaux dans l'organisation des unités locales.

De plus, à l'aide d'une description non préalablement figée des unités linguistiques, il est envisageable de permettre l'activation de descripteurs locaux (attribués à des entités réduites, comme les mots et expressions) en fonction de données globales. La base d'apriori, présente habituellement dans la description classique des unités locales, serait alors située au niveau du texte et de son environnement, tel que ceux-ci sont appréhendés par l'interprète. La notion de calcul, fondatrice de l'outil informatique devra donc être repensée dans l'impact de ces données textuelles sur les entités locales.

4 Le texte, l'homme et la machine

Comment concilier les principes précédemment définis avec la notion d'automatisation, ou du moins d'informatisation ? Cela semble impossible : on exige ici une trop grande place accordée a l'homme pour envisager un rapport «classique». Il faut donc redéfinir ces relations, et surtout trouver une nouvelle place pour la notion de calcul de la machine. Cette dernière notion de calcul, la seule supportée par l'ordinateur, implique donc un mode de fonctionnement strictement compositionnel, exclusivement adapté à la formalité logique. On est donc devant une apparente contradiction : il faut utiliser une plate-forme compositionnelle pour mettre en application une théorie qui récuse la compositionnalité. La conciliation ne peut donc se faire que par le déplacement de cette compositionnalité, dans les rapports de l'homme à sa machine.

Le problème existe pour d'autres disciplines que le TALN. Nous tâcherons d'en tirer un enseignement et de proposer un rapport entre la machine et l'acte interprétatif. Nous envisagerons à la suite de cela les portes applicatives qui se ferment, et celles qui s'ouvrent, en proposant de nouveaux types d'outils.

4.1 Coopération Homme / Machine

Avec la chute du paradigme cognitiviste premier, dont l'apogée pragmatique (ou plutôt applicative) fut atteinte avec les systèmes experts, et l'insuffisance de leur puissance inférentielle pour la modélisation d'un comportement cognitif simple, d'autres aspects de l'outil informatique furent pris en compte. La modélisation complète et fidèle de la cognition devrait en effet être abandonnée, au profit d'une véritable collaboration, en reconnaissant l'irréductibilité des activités mentales à un calcul symbolique (ni même subsymbolique). C'est en ce sens que Y.-M. Visetti [65] commente le passage des systèmes experts aux systèmes à base de connaissances, première amorce de ce mouvement vers la coopération. Se situant à un niveau plus général que le nôtre (du point de vue des domaines d'application), il reconnaît les nuisances d'une préconception logique majoritaire, dont les contraintes dans les systèmes dits cognitifs bloquent la communication entre la machine et son utilisateur. Ainsi, en restreignant la place de structures logiques complexes et par trop figées, et en accordant une priorité à la coopération plutôt qu'à une automatisation utopique, il trace les grandes lignes d'un nouveau type d'outil. Ce dernier, qu'il nomme «système interactif d'aide à la modélisation et à la validation des raisonnements» doit abandonner l'exigence de l'autonomie. En réduisant la complexité et l'importance de la formalisation logique, il limite l'application de celle-ci à certaines parties restreintes d'un processus de raisonnement. Pour cet aspect, il faut entendre simplement une économie cognitive de l'utilisateur, exigeant par là même de celui-ci une connaissance suffisante du domaine et des limites même de ce type de coopération.

Mais d'un autre côté, en appliquant une rigueur logique (qui d'après l'auteur ne peut guère provenir de logiques non-standard, d'où une position affirmée dans la logique du premier ordre), il est envisageable de prendre en compte certaines suggestions de la part de l'automatisation comme Y.-M. Visetti l'énonce dans l'élégante analogie suivante :

> Ils ont donc le statut ambigu d'oracles : qu'on prenne ou non le parti de les suivre, ce n'est pas d'eux que l'on attend véritablement un développement de la rationalité des organisations [...]

Concrètement, ces approches nécessitent une moins grande complexité technique (dans l'appareil logique), et une plus grande souplesse aussi, ainsi que la possibilité de rejeter l'adéquation de la représentation formelle avec une réalité intangible des structures cognitives supposées.

Schématiquement, et donc idéalement, un tel outil comporterait deux couches principales, auxquelles s'ajouterait bien entendu l'utilisateur si central. Une première couche logique, support des représentations internes, d'une complexité réduite, et véhicule de rigueur et de robustesse. Une couche sémiotique, sorte d'intermédiaire, où se projettent les actions et conceptions humaines, qui doit donc être porteuse de grande souplesse, et d'interaction forte avec l'utilisateur. Une proposition de schéma est visible dans la figure 1.2.

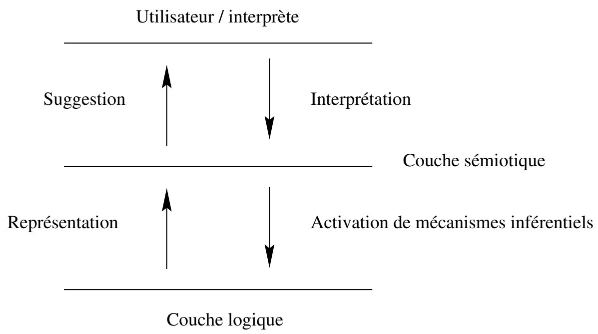


FIG. 1.2 – *Organisation d'un système interactif*

Mais cette structure des couches logiques et sémiotiques ne peut se concevoir que sous une thématisation de la composante humaine. Elle doit être recherchée à tous les niveaux et à toutes les périodes de la constitution d'un tel outil. La couche sémiotique n'est qu'une abstraction, son véritable statut est celui que lui confère l'utilisateur.

Pour terminer, la notion même de subjectivité, et de reconnaissance de la centralité humaine trouve des échos, puisqu'au moins des intentions transparaissent de reconnaître, et non plus de craindre, la variabilité inter-individuelle, mais également les variations temporelles intra-individuelles [67]. Les travaux en ergonomie dans le développement des interfaces homme-machine en témoignent également.

4.2 Informatique et interprétation

Assurément, les applications de ces idées générales couvrent un éventail plus large que celui que nous visons. Il à ce point opportun de revisiter les principes précédemment évoqués pour guider notre approche, en mettant en relief leur adéquation et en envisageant dans les grandes lignes leur mise en place.

Pour ce qui est de la couche logique, nous avons déjà discuté de l'intérêt d'une description formelle simple, se basant sur des concepts fondamentaux, tels que ceux qui sous-tendent la vision différentielle. C'est à ce stade que nous entendons satisfaire le souci de rigueur. Cependant, cette notion de rigueur doit être étendue. Elle ne concerne pas que le régime du calcul - pour lequel elle est presque définitoire. Elle va jusqu'aux régimes multiples d'ajustement et d'adaptation mutuels entre calculatoire et sémiotique.

Pour la couche sémiotique, différentes considérations sont à prendre en compte.

Pour la souplesse, nous la capterons par le principe de description méta-linguistique. Quoi de plus souple en effet que la langue naturelle, à plus forte raison lorsque le discours qu'elle manipule s'adresse de façon privilégiée à son créateur ? Ainsi, sur la base des unités formelles citées plus haut (et développées bien plus bas), l'utilisateur calquera donc ses propres descriptions langagières, avec la liberté que lui laissera leur incompréhensibilité par la machine. Nous tâcherons donc de faire *d'un vice une vertu*, puisque la machine n'influencera pas la valeur fondamentale des ces descripteurs, respectant par là même la prédominance de l'humain dans cette relation. D'un point de vue interprétatif, l'utilisation de la langue comme descripteur rejoint la notion d'appropriation que F. Rastier énonce [53] pour traduire la prise en compte du contexte interprétatif. C'est ici par son propre langage que l'utilisateur décrira son interprétation.

Enfin, par le biais des relations mises en place au niveau formel, «traduites» par le méta-langage naturel de description, la porte est ouverte, sinon aux suggestions de source inférentielle, du moins aux facilités de représentation et de synthèse que peut proposer un outil logiciel manipulant un texte.

Quant à la volonté de construire l'identité sémantique des unités mises en œuvre dans la description interprétative, elle se traduira par un ensemble de contraintes formelles dans la définition de ces unités, qui, projetées à l'attention de l'utilisateur, seront autant de suggestions à la créativité et au repérage des nouvelles descriptions. Restons prudent toutefois sur l'origine de cette identité : sa construction n'est en fait qu'une conséquence de sa présupposition par l'utilisateur. Cette étincelle créatrice initiale entraîne par contre, d'un point de vue formel, des exigences de la couche logique.

Nous verrons au travers de la théorie linguistique motivant notre approche, la traduction de ces principes sous un aspect herméneutique, justifiant tant la circularité méta-linguistique que l'infinité de ce type de processus.

4.3 Quelles applications ?

Ce point de vue, centré sur la subjectivité dans la langue, si attrayant qu'il soit, limite tout de même, outre les objectifs concrets d'une application, ses domaines d'applications. Reprenons un argument de F. Rastier à propos des exemples littéraires qui illustrent ses travaux : «... Outre que ces textes [littéraires] sont des plus agréables à étudier, on peut prétendre que leur complexité met à rude épreuve les instruments descriptifs utilisés. Mais au contraire, choisir des textes littéraires, ne serait-ce pas une solution de facilité ? Leur cohésion guide le scholastique.» [50].

Pour reprendre la première réponse que l'auteur apporte ici à une critique, nous ne pourrons que la renforcer en y ajoutant le statut socialement stabilisé des méthodes exhaustives d'analyse quand elles s'appliquent aux textes. Nombre de travaux littéraires mettent en exergue la rigueur nécessaire à une étude dans ce domaine, et une transversalité des études quant au niveau des entités linguistiques manipulées. Il devient donc plus acceptable de solliciter un utilisateur qui, lui, n'a pas besoin de la machine pour comprendre le texte, pour un poème ou un roman que pour une recette de cuisine ou pour une notice de montage d'un moteur d'avion. Pour ce qui est du deuxième argument, ne nous leurrons pas, l'application à de tels textes garantit également des résultats non triviaux, simplement parce que le style (au sens large) est important. Cependant, des méthodes similaires semblent trouver un écho appliqué à des textes techniques, lorsque les considérations pragmatiques sont l'extraction de connaissance, couplée aux systèmes de représentations plus classiques [2].

Bien sûr, nous ne pouvons prétendre en tout cas à une complète originalité des applications du TALN à la littérature. Au contraire, nous rejoignons en ce sens un mouvement assez riche de collaboration entre l'informatique et les sciences humaines. En effet, nombre de travaux en lexicométrie sont informatiquement appliqués à des études littéraires, pour lesquels l'outil automatique est parfaitement justifié. Un exemple d'une telle mise en œuvre est le programme TACT (*Text Analysis Computer Tools*) du *Center for Computing in the Humanities* de l'Université de Toronto[6]. Constitué d'une grande famille d'outils d'analyse statistique des fréquences de mots, il permet d'extraire des structures stylistiques de textes, basées essentiellement sur la syntaxe (en fait, sur la seule forme des chaînes de caractères), et sur quelques traits sémantiques essentiellement définis comme des classes abstraites d'occurrences syntaxiques. Le principal reproche que nous pourrions

6. Vivons avec notre temps : le logiciel et sa documentation sont disponibles à l'adresse suivante : http://www.chass.utoronto.ca:8080/cch/tact.html

porter à ces approches est la reprise de ce fameux principe issu des modèles cryptographiques, de ne regarder qu'une petite «fenêtre» pour expliciter un terme dans un corpus. La théorie initiale cherchait à définir le nombre de mots nécessaires dont il faut tenir compte avant et après l'occurrence considérée pour la cerner complètement. Il s'agit du fameux KWIC (*Key-Word In Context*) des anglo-saxons, que la coutume a fixé à une taille de 11 mots (5 avant et 5 après). De telles considérations semblent donc peu respectueuses de l'intégrité du texte analysé.

Malgré le caractère essentiellement formel de ces analyses, nombreuses sont les applications de cet outil par des chercheurs en littérature. Il y a donc un besoin (ou du moins un besoin à créer) de l'autre côté de la barrière sciences dures / sciences humaines, de tels outils scientifiques. Ce que nous proposerons ici sera en comparaison un accompagnement de la machine au long d'un acte d'interprétation. Nous avons déjà énoncé un intérêt d'une telle situation dans la gestion de la cohérence de cette interprétation, et dans un pouvoir limité de suggestivité. Mais nous envisagerons également d'autres aspects, liés aux outils précédemment cités, dans la capacité de mémoire et de présentation de la machine.

Mais pas de précipitation dans ce domaine, dont les fondements mêmes sont remis en cause, comme l'explique Mark Olsen dans [43]. Dans cet article, il constate tout d'abord le manque d'innovation et de résultats probants dans le domaine de l'analyse littéraire informatisée, malgré les progrès technologiques et le nombre de travaux croissant. En résumé, rien n'est découvert via ces méthodes technologiques qui ne le serait par une lecture attentive d'une œuvre. Ses conclusions sont doubles : tout d'abord il prône une utilisation de l'ordinateur pour ce qu'il sait faire de mieux que l'homme : manipuler de grandes quantités de données, donc de nombreux textes. Mais surtout, il insiste sur le besoin de corrélation entre théories d'analyse et développements informatiques, indiquant par là un mouvement dans lequel nous estimons nous situer. Nous expliciterons en effet, lors du prochain chapitre les inspirations théoriques qui ont guidé ce travail.

5 Conclusion

En achevant notre première étape, le but que nous nous étions fixé était de placer notre approche dans son contexte au sens large. Sans avoir détaillé exhaustivement les travaux connexes en TALN, nous espérons nous être situé par rapport aux grands axes qui l'organisent.

Nous pouvons résumer ainsi les principes qui ont guidé nos travaux :

- Une prise de position envers les aspects interprétatifs, en rejetant tout d'abord la génération que nous n'envisageons pas comme le décalque de la compréhension.

- Dans cette même voie, un refus de la caractérisation des processus de compréhension comme décodage d'une réalité absolue ou même cognitive. D'où un choix de la notion d'interprétation, plus souple et plus ambitieuse à la fois, donnant une place centrale à la subjectivité dans la condition de l'objet linguistique.

- Une orientation vers la sémantique, au détriment de la syntaxe au sens linguistique. Il s'agira cependant d'une vision structurale de la sémantique, faisant ainsi appel à une notion de syntaxe purement formelle, mais dont les principes plus élémentaires ne soient pas une gêne dans les implications au niveau du sens.

- La reconnaissance du texte comme objet privilégié (et exclusif) d'une vision linguistique du sens en situation.

- Et enfin, pour concilier ces positions, une application résolument tournée vers la coopération entre l'homme et la machine, laissant la notion première de la signification à l'humain, limitant ainsi la machine à un rôle de suggestion et non de décision. Ce qui, de plus, nous pousse à envisager des domaines d'application plutôt littéraires.

Pour terminer, il nous semble intéressant, pour clore ces questions, de nous permettre une petite comparaison, en plaçant notre approche dans la déjà longue histoire du TALN, avec l'historique et la prospective que F. Varela propose pour les sciences cognitives.

Dans son ouvrage «Connaître : les sciences cognitives, tendances et perspectives» [63], F. Varela décrit avec clarté les grandes évolutions des sciences cognitives, de leur genèse à leur avenir supposé. Sans vouloir émettre d'hypothèses fortes quant aux relations réelles entre ces sciences et le TALN, il semble pourtant légitime d'appliquer son schéma à un rapide historique de le discipline qui nous intéresse. Nous reprendrons donc les étapes que l'auteur repère dans le passé, ainsi que la forme didactique qu'il emploie pour synthétiser sa critique, et comparerons enfin notre prise de position à la sienne, en ce qui concerne l'évolution supposée.

L'évolution ainsi présentée est résumée en quatre étapes : depuis les périodes logiques et cybernétiques, en passant par le cognitivisme classique et le subsymbolisme vers la nébuleuse notion d'enaction.

Pour la première époque, où régnait l'optimiste cybernétique, cela se traduisit par la considération de la langue comme simple code. Les premières tentatives de traduction automatique dont nous avons parlé traduisaient bien cette absence de prise en compte de la réalité interne des phénomènes du langage, traité comme de simple chaînes de caractères à mettre en correspondance par un mécanisme simple. Le paradigme de la boîte noire est ici projeté, et cédera sa place aux tentatives d'explicitation par le formalisme logique.

L'étape cognitiviste voit ensuite la notion de symbolisme et de représentation placée sur le devant de la scène. La cognition devient représentation et manipulation de symboles par des règles logiques. Quant au langage, il devient effectivement représentation, captable informatiquement par des logiques complexes, attribuant un sens décrit à des entités qui se combinent, et le sens se construit de façon déterministe. La plupart des formalismes classiques du TALN datent de cette époque, où les systèmes experts faisaient fureur.

Nous reprendrons ici (en les transformant) les questions que pose F. Varela, dont le texte initial était le suivant : Qu'est-ce que la cognition ? Comment cela fonctionne-t-il ? Comment savoir qu'un système cognitif fonctionne de manière appropriée ?

- *Question 1 :* Qu'est ce que la sémantique dans la langue ?
  Réponse : Un mode de représentation des structures conceptuelles destinées à être transmises, de l'information.

- *Question 2 :* Comment cela fonctionne-t-il ?
  Réponse : En traduisant les concepts par des entités linguistiques, dont les relations entre elles (syntaxiques) traduisent également la structuration des concepts.

- *Question 3:* Comment savoir qu'un système de TALN fonctionne de manière appropriée ?
  Réponse : Si l'information décodée et représentée par un formalisme logique permet d'aboutir à des inférences face à un objectif ou des représentations correctes face à une base de connaissances.

L'étape suivante, qui ne suit pas forcément une chronologie bien nette est l'apparition de la notion de subsymbolisme, avec comme plus célèbre manifestation le connexionnisme. Inspiré par les découvertes en neuro-sciences, le principe en comparaison au cognitivisme classique est d'atteindre un résultat non par des règles de haut niveau s'appliquant à des entités bien définies et stables, mais de le faire apparaître ou plutôt émerger par la collaboration d'un grand nombre d'entités minimales, très locales et sans contenu propre. Sans proposer une nouvelle approche du calcul, il ne constitue pas moins un paradigme de représentation.

Nous nous risquerons donc à rapprocher de ce paradigme les travaux sus-cités sur les analyses distributionnelles. En effet, le principe y est aussi de voir «émerger» des considérations sémantiques sur un corpus à partir de la prise en compte d'un grand nombre d'éléments locaux, sans qu'à ceux-ci soit attribué un véritable contenu (décrit de manière formelle). Il s'agit donc de se dégager d'une explicitation de la notion de sens, et de ne rechercher que sa manifestation, comme un principe organisateur sous-jacent.

– *Question 1 :* Qu'est ce que la sémantique dans la langue?
  Réponse : Un système trop complexe et ouvert pour être décrit directement, dont la seule manifestation permet de découvrir l'organisation générale.

– *Question 2 :* Comment cela fonctionne-t-il?
  Réponse : En analysant sans hypothèses fortes la distribution des entités linguistiques, en utilisant des relations simples entre ces unités (identité, co-occurrence, etc.) pour les caractériser à un niveau global.

– *Question 3:* Comment savoir qu'un système de TALN fonctionne de manière appropriée?
  Réponse : Si les classifications issues de ces traitements distributionnels sont interprétables comme des classes sémantiquement valides.

Enfin, F. Varela conclut par la perspective de l'*enaction*, ou le *faire-émerger*. Sans vouloir trop nous avancer sur cette notion opaque, nous retiendrons quelques aspects que nous pensons intéressants, et proches de nos préoccupations et de nos prises de position. Tout d'abord, les liens avec l'étape précédente sont importants : il y a ici refus d'une explicitation trop forte et figée d'un mécanisme général, et donc d'une description exclusivement logique du sens. L'aspect central de la création que l'auteur accorde à ce paradigme trouve également des échos dans la notion d'interprétation. Il remplace également la notion de validité d'un modèle par une viabilité, que nous aimerions concevoir comme cohérence par rapport à l'interprète.

Vouloir y trouver d'autres liens serait hasardeux et sans doute inutile, mais appliquons le jeu des trois questions à notre approche, dans un souci de résumé.

- *Question 1 :* Qu'est ce que la sémantique dans la langue?
  Réponse : Une structure qui se construit, intégrant à partir de données linguistiques les intentions et connaissances de l'interprète.
- *Question 2 :* Comment cela fonctionne-t-il?
  Réponse : Par projection de ces données générales sur la structuration linguistique de l'énoncé, sur les entités locales, et permet de voir apparaître des données globales structurées.
- *Question 3:* Comment savoir qu'un système de TALN fonctionne de manière appropriée?

Cette dernière question n'a pas de réponse pour l'instant, et nous préférons d'ailleurs refuser l'intérêt de celle-ci. Comme nous l'avons déjà évoqué, la notion de performance n'est sans doute pas adaptée au langage. Dans le TALN, bien souvent, la notion d'application directe sert de bouclier face à des remarques théoriques sur une conception sous-jacente de la langue souvent trop pauvre et simpliste. Nul doute qu'ici nous pêcherons par excès inverse...

À notre connaissance, au sein de cette discipline, il ne se trouve pas d'approche susceptible de capter les aspects de la textualité tels que nous les avons énoncés. Une première explication peut se trouver dans la faiblesse du nombre de théories linguistiques qui traite de ces problèmes d'une manière au moins minimalement formelle. Au cours du prochain chapitre, nous nous pencherons donc de plus près sur les travaux de F. Rastier sur la sémantique interprétative, en envisageant sa mise en place formelle et informatique au vu des grandes lignes que nous avons tracées ici.

Chapitre 2

Une approche linguistique du TALN : La Sémantique Interprétative

Dans ce chapitre, nous tenterons, au vu des lacunes et nécessités du TALN exposées précédemment, de justifier l'intérêt d'une théorie linguistique avant tout, mais surtout d'une théorie s'intéressant à l'interprétation. Nous allons donc exposer les principales idées de la Sémantique Interprétative de François Rastier [49], [50], et la façon dont elle peut être appliquée au traitement automatique de la langue tel que nous l'avons envisagé [52].

Dans une première partie, nous expliciterons les origines de la théorie, et les diverses influences qu'elle revendique, puis nous tenterons de mettre en évidence les intérêts majeurs qu'elle présente pour aborder formellement le problème de l'interprétation. Nous rentrerons plus en détail dans cette théorie en en répertoriant les concepts et les outils que ceux-ci peuvent mettre à la disposition du traitement informatique. Enfin, nous parlerons des possibilités concrètes de son utilisation, et bien entendu des limites de celle-ci. Nous nous inspirerons pour ce rapide exposé des enseignements linguistiques des ouvrages [17], [39], [40], [6] entre autres.

1 Approche structuraliste et herméneutique

Dans cette partie nous allons approfondir les diverses influences de la théorie de la sémantique interprétative, à savoir la linguistique structurale et l'herméneutique. La première annonce une volonté de s'ancrer de plain-pied dans le domaine de la linguistique, au détriment des disciplines connexes telles que la psychologie, l'intelligence artificielle et sa représentation de connaissances, et la philosophie du langage. Mais surtout, elle propose une approche différentielle de la sémantique, et un ensemble de concepts sains pour définir l'objet linguistique. La seconde enrichit les apports de la linguistique par une reconsidération de la notion de sens, et inculque quelques méthodes afin de l'appréhender. En effet, les approches herméneutiques permettent d'envisager globalement les problèmes du sens, de sa construction, et plus généralement des phénomènes, sinon de compréhension, du moins d'interprétation. L'herméneutique vise les conditions de signifiance de l'objet linguistique *in situ* (pour un homme, occupant une place dans un société, dotée d'un histoire, etc.).

1.1 La lignée structuraliste, le problème de l'identité

À la base du structuralisme se trouve la recherche de points de comparaison entre langues, issue des grammaires comparées. Que ces langues soient séparées par l'espace, le temps ou les deux, la méthode consiste en la recherche de stabilités entre figures d'expressions et idées véhiculées (par exemple le proto-morphème *pe-* dans les diverses expressions de «pied» et de ses dérivés dans les langues indo-européennes).

La grande idée de Saussure [56] fut donc de considérer dans ce cas la nature de l'élément de base de la comparaison. Reconnaissant la langue comme système, il exprime que c'est la présomption de connaissance de ce dernier qui permet d'aboutir à l'identification de ses constituants (et donc de commencer véritablement la description).

Même si l'on isole le signe dans un schéma de relations entre ses parties (signifié et signifiant), il faut toujours considérer que l'élément est indissociable du tout systémique auquel il appartient et qu'il constitue. L'importance n'est donc pas seulement portée sur la nature et la signification des relations entre les parties du signe, mais surtout sur les relations entre les signes.

Un autre exemple de cette difficulté réside dans la notion même de paradigme. Toujours dans le cadre des comparaisons entre langues, on doit identifier des classes de comparaison pour les termes. Par exemple, pour identifier le «sens» du préfixe *dé-*, il faut considérer un ensemble de termes comme *défaire*, *démonter*, *démolir*, etc. Bien des classes de comparaison sont pourtant disponibles pour chacun de ces termes, comme *faire, refaire, défaire*, mais c'est avant tout la présomption de l'objet à mettre en évidence, et du rôle qu'il joue dans le système de la langue qui détermine les effets locaux utilisés pour mettre en évidence sa nature et même son existence.

Quand nous énonçons le terme de présomption, il faut en fait avouer que l'on s'intéresse à l'*identité* de l'élément. Plus qu'une évidence, cette identité doit en fait être construite. Les relations paradigmatiques qu'entretient une unité signifiante se basent sur son identité, mais en fait elles la forment par là-même. Un mot (comme signification), par exemple, va pouvoir être classé dans diverses classes sémantiques (et supporter des relations au sein de celles-ci) parce qu'il a été identifié, mais cependant ce sont les classes admises et les classes injustifiées (plus qu'impossibles, puisque l'on peut toujours s'amuser à regrouper un ensemble aléatoire de termes et les unifier par un contexte ou même un texte) qui vont construire ou peut-être simplement traduire cette identité. À un autre niveau, si l'on décompose cette entité à classer, c'est sur la base du partage d'un de ses composants avec d'autres entités que son appartenance à une classe sera justifiée. Le morphème 'faire' dans l'exemple précédent est «le même» dans tous les verbes énoncés, alors que d'autres composants y entretiennent des relations de différence ('re' et 'dé' par exemple). Ainsi, dans une approche formelle, un point de départ est nécessaire dans la description d'une unité sémantique (un signifié). Celle-ci peut n'être que partielle, comme l'identification d'une classe qui contient cette unité. À partir d'une spécification minimale, l'ensemble de la notion peut ainsi se construire, cette construction faisant partie d'un processus interprétatif. Ainsi, la fameuse notion de généricité - spécificité n'est donc qu'une conséquence de ces deux principes fondamentaux (identité et différence). L'organisation paradigmatique repose sur la notion de substitution, et suppose donc une identité de l'unité transposée à travers les contextes.

Cet apparent paradoxe n'a, en fait, pas tant d'hermétisme, si l'on considère la double identité du «mot». La première est une identité de forme (une simple chaîne de caractères), la seconde une identité de contenu, du moins pour ce qui est des classes et des relations sémantiques.

Le sauvetage du TALN - C'est ainsi que nous pouvons espérer manipuler le sens via l'outil informatique. La machine manipulera de purs signifiants (suites de caractères), auxquels sont naturellement attribuées des relations inhérentes d'identité et de différence formelles. Le signifié ne sera donc capté qu'au prix de la mise en place de relations entre ces signifiants, en fait de paradigmes. Nous pourrons tout de même nous permettre de «confondre», du point de vue de la machine, ces deux notions, puisqu'une seule d'entre elles sera réellement manipulée. La notion de sémème (*cf. infra*) ne posera donc plus de problème lors de sa formalisation : elle sera considérée dans son autonomie comme étant le signifiant (permettant un pointage vers sa place dans la chaîne syntagmatique), mais supportera la lourde notion de signifié quand elle sera intégrée dans une structure sémantique. De toute façon, la signification ne viendra pas de la machine, nous exigerons simplement que les manipulations qu'elle aura à effectuer soient en cohérence avec la vision de son utilisateur. L'acte interprétatif de mise en relation du signifiant avec le signifié ne relève que de l'humain.

1.2 Le problème du global et du local. L'alternative herméneutique

Reprenons tout de même un instant le paradoxe évoqué précédemment, et qui couvre bien plus de domaines (scientifiques ou non) qu'il n'y paraît. Pour le résumer à nouveau, disons qu'un élément signifiant fait partie d'un système qui lui donne sa signification, basée sur les relations qu'il entretient avec les autres entités, ces relations étant contraintes par la nature même de cet élément... on tourne en rond.

En quelque sorte, pour échapper à ce cercle vicieux (ou vertueux?), force nous est de prendre parti pour une des deux considérations, du global ou du local. Quelques premiers indices sont que l'on s'occupe de langue naturelle, que l'on veut traiter des textes, et que l'on a trop de respect pour la notion de sens pour attribuer une seule de ses occurrences à un texte donné, c'est-à-dire que l'on souhaite prendre en compte des données extérieures dans l'attribution du sens : «contexte» [1], connaissances et intentions de l'interprète, etc.

Il n'y a donc plus de raison de faire durer le suspense : on donnera naturellement la primauté à la notion de globalité, et on couvrira même par la primauté la notion de détermination.

Mais pourquoi ? Naïvement, quand on approche un texte, on n'y verra qu'une suite de symboles, qui, seulement une fois interprétés, permettront d'atteindre une compréhension de l'ensemble. Cette vision est en fait par trop empreinte de la théorie de l'information, et du langage comme simple code à analyser. Le principal problème est en fait la difficulté d'attribuer une signification stable à un élément isolé. Les nombreux travaux sur la polysémie qui ravagent le TALN en sont une bonne preuve. Des systèmes d'une complexité effroyable sont développés pour résoudre des problèmes que ne rencontre aucun interprétant humain, puisque ce dernier ne considère pas (et ne **peut** d'ailleurs pas considérer, la plupart du temps) le terme polysémique de façon absolue (voir par exemple [64]). Mais ne sombrons pas dans une

1. Nous verrons plus loin la raison de ces guillemets précautionneux

critique trop facile. Les objectifs des outils de traitement qui se heurtent de plein fouet à la polysémie ne sont pas les nôtres, puisqu'ils tentent de parvenir à une compréhension automatisée, alors que nous nous contenterons largement d'une interprétation assistée. L'approche formalisatrice est dès lors très différente, la polysémie étant rencontrée sur le chemin du calcul du sens, et la notion d'interprétation ne comporte pas de calcul.

Par contre, et comme nous l'avons vu au premier chapitre, il s'agit bien de l'objet texte qui se trouve en situation : avant d'être perçu comme une composition d'unités inférieures, le texte se voit déjà attribuer un ensemble de considérations. Sa situation sociale et historique, l'intention de l'individu qui l'aborde, sont attribuables prioritairement à la totalité du texte (et dans certains cas, avant même que celui-ci soit analysé). Par contre, ces considérations globales sur le texte peuvent être projetées sur ses éléments, pour s'y trouver justifiées, modifiées ou servir de base dans la découverte d'autres aspects.

D'un point de vue plus formel, la prise en compte de ces rapports de la globalité vers la localité nécessite des outils de description particuliers. La relation est en effet bien plus qu'une simple projection, puisque des phénomènes locaux peuvent également avoir des répercussions sur l'ensemble. Il nous faudra dès lors envisager des rapports bilatéraux, sans nous enfermer dans un dogmatisme purement globaliste. La première tâche sera alors de disposer d'un moyen d'expression compatible avec les deux niveaux. De plus, se limiter à deux niveaux (de façon grossière le mot et le texte) n'est pas non plus suffisant. La notion de localité, toute relative, peut s'étendre du morphème jusqu'au syntagme. Des travaux comme ceux de J. Kristeva [33, 34] n'hésitent pas à descendre jusqu'aux qualifications sémantiques des phonèmes pour appuyer des thématiques globales chères à la psychanalyse. Mais ces problèmes pourront être résolus par une souplesse accordée à l'identité formelle des unités locales.

Un deuxième aspect à considérer dans notre approche a bien sûr trait à la simple et incontournable opérationnalité de tout système informatisé. Les principaux paradigmes de description sémantique acceptables dans un environnement structuraliste se limitaient à l'expression de dépendances fonctionnelles entre entités lexicales, dépendances dont on regrette la stricte et rigide localité. L'alternative que nous sélectionnerons sera le principe de la microsémantique, ou description de la signification par des éléments sémantiques inférieurs à l'unité lexicale. Cette description par définition locale nous semble en effet assez souple pour supporter le poids du global dont nous désirons la charger, et de plus essentiellement pratique car manipulable par la machine. Le principe qui servira de liaison entre localité et globalité sera la justification des entités locales par leur rôle dans des structures dont la zone d'influence n'est rien de moins que le texte. Le principe d'une interprétation sera donc la supposition de l'existence de telles structures, et leur actualisation par des unités locales. D'un point de vue informatique, ceci n'a rien que de très connu, puisqu'il se fait une différence simple entre structure déclarée et instanciée. Aborder un texte quelconque sous la lumière d'une symbolique particulière, par exemple, commencera par la mise en place de structures générales, avec de grands thèmes entretenant entre eux des relations également stabilisées. Ces connaissances a priori permettent une mise en place de telles données globales avec un minimum de relations directes avec le texte (repérage de quelques mots ou expressions). Dès lors, un début de justification apparaît, non pas en ce qui concerne l'attribution ex-nihilo d'un trait à une unité linguistique, mais pour l'utilisation d'un tel trait en fonction des objectifs de l'interprétation. Par la suite, nous verrons comment ces attributions minimales permettent d'en repérer d'autres, et d'enrichir les structures du texte, et d'en repérer de nouvelles. Une interprétation sera donc un continuel aller-retour entre ces deux niveaux.

Dès lors, la question se pose quant à la nature de l'outil descriptif local, capable de supporter ces notions qualifiant le texte.

1.3 La sémantique du local

Le palier local, que l'on peut dans un premier temps associer à la notion de mot ou de morphème, sous-entend généralement une description de l'organisation sémantique du signifié associé. Pour cela, elle met en place des unités sémantiques inférieures, dont la nature et la justification peut varier d'une théorie à l'autre.

1.3.1 Diverses possibilités

Du point de vue du TALN, où domine le souci d'opérationnalité, les questions sur la nature de ces unités inférieures sont rares. Le souci de leur manipulation les ramène classiquement à des unités symboliques manipulées par la logique.

Dès lors que la question se pose, la réponse est souvent recherchée dans le cognitif, lieu privilégié d'explication des phénomènes du langage et de la compréhension comme nous l'avons vu. Les travaux de J.-P. Desclés [16] sur la grammaire applicative et cognitive, cherche ainsi une justification des archétypes descriptifs, la grammaire cognitive de Langacker également [35, 36].

D'autre modèles proposent un point de vue encore plus absolu, comme les travaux mathématiques de R. Thom [61], répercutés dans la langue par B. Pottier [46, 47]. On peut dans ce cas parler de noèmes, par l'universalisme qu'oblige une mathématisation forte. Bien que séduisants, ces modèles perdent malheureusement en opérationnalité, notamment par la perte de la centrale notion de continuité qu'entraîne le passage à la machine.

Une dernière possibilité est de rester dans le linguistique, et de revendiquer son autonomie en utilisant des descripteurs de forme langagière. Cette alternative permet d'échapper aux lourdes présuppositions d'un cognitivisme ou d'un universalisme, tout en restant manipulable. Sa réalisation passe par l'utilisation de marqueurs sémantiques, ou @*sèmes*, et l'on parle alors de *microsémantique*.

1.3.2 Exemples d'utilisation de la microsémantique

Les plus célèbres noms associés à une mise en pratique de ce paradigme sont les célèbres Katz et Fodor [29]. Auparavant, et également après eux, la notion de marqueur sémantique fut utilisée pour guider l'analyse syntaxique (par exemple pour exprimer la nécessité d'avoir un sujet animé pour certains verbes). Mais Katz et Fodor vont plus loin, car ils tentent de repérer toutes les caractéristiques sémantiques d'un mot par l'énumération de ces marqueurs. Leur plus célèbre exemple traduit leur intérêt comme centré autour du problème de la polysémie. En effet, ils utilisent les marqueurs sémantiques pour distinguer les différents sens (ou acceptions ?) de *bachelor*, qui couvrent des domaines divers, du diplômé à l'otarie mâle. Les marqueurs sémantiques forment donc un arbre de signification, dont la racine est le mot lui-même, et chaque branche supporte un marqueur dans le but de distinguer les différents sens (/humain/, /animal/, /mâle/, etc.) ou de décrire le nœud terminal d'un sens (comme /privé de femelle à la saison des amours/[2]).

Comme on peut le voir à la lueur de cet exemple, il s'agit bien d'une polysémie forte qui est traitée ici (voire une homonymie). Cependant, l'intention est louable, puisqu'elle donne aux marqueurs sémantiques (du moins à certains d'entre eux) un rôle différenciateur. Le principal problème de cette méthode est que les sens qui doivent être distingués n'ont pas leurs relations validées sur des critères proprement sémantiques, puisqu'il s'agit d'une identité de forme.

Une autre mise en place d'un système de marqueurs microsémantiques est celle de Paul Guiraud et de ses sèmes [21]. La tâche est plus audacieuse,

2. Il s'agit de l'otarie bien sûr

puisqu'il s'agit de la constitution d'un système de primitives pouvant couvrir toutes les significations. Cette démarche prend un caractère gênant quand l'auteur entreprend de compter les sèmes, et d'extraire une loi numérique liant le nombre de sèmes attribués à une entrée lexicale au nombre de syllabes que celle-ci contient, démontrant ainsi la loi de Zipf (les mots les plus courts sont les plus polysémiques).

Ainsi, répondant à un principe structuraliste classique, les sèmes sont bien utilisés pour distinguer différentes unités sémantiques, mais exclusivement pour résoudre les problèmes de polysémie. C'est-à-dire que les unités que l'on souhaite distinguer n'ont pas de proximité interprétative, où la prépondérance du global empêche justement le rapprochements de significations disparates dont le seul point commun est celui de la forme.

De plus, dans l'hypothèse d'une effective description par ce moyen de l'ensemble des significations des unités lexicales, leur assemblage et mise en cohésion au niveau d'un énoncé atteint des complexités opératoires exponentielles.

La sémantique interprétative, qui elle aussi met en place un système de description reposant sur des sèmes, permet avant tout de préciser le rôle de ceux-ci dans un objectif interprétatif. Elle s'ancre bien dans un paradigme différentiel, mais encore doit-elle bien construire cette notion de différence.

1.3.3 La microsémantique différentielle

Retournons donc nous réfugier au sein du paradigme structuraliste et avant de nous intéresser à la nature de ces sèmes, exprimons clairement les rôles qu'il convient de leur attribuer, et les fonctions qu'ils devront remplir.

- Les sèmes devront traduire les distinctions et similarités entre les sens des unités lexicales, mais sur une base sémantique. Ceci accordera une *justification* à leur nature.
- Les sèmes ne devront pas être utilisés pour capter *le* sens d'une entité lexicale, mais traduire une interprétation de celle-ci. Il s'agira donc d'une *relativisation*.
- Les sèmes devront être la projection sur une entité lexicale de considérations sémantiques globales. Les sèmes devront donc avoir un rôle organisateur, et concrétiseront l'interprétation. Ceci leur accordera donc, dans leur ensemble, et pour un texte, une *cohérence*.

Malgré ces intentions exposées, le danger rôde toujours autour des sèmes, car de telles caractéristiques (surtout la deuxième) ne peuvent empêcher une profusion de leur énumération. Essentiellement, nous verrons que c'est la troisième caractéristique qui nous sauvera du naufrage.

Le paradoxe que nous évoquions sera toujours présent dans une description sémantique à base de sèmes. Ces entités inférieures au «mot» traduiront l'appartenance à une classe pour que se mettent en place les oppositions entre signifiés, et c'est leur présence qui traduira la donnée globale dont elles dépendent, le cercle reste donc toujours aussi fermé... Et pour en sortir, nul artifice, si ce n'est le recours à l'humain, ne gardant de cette description par des sèmes que son aspect formalisateur.

1.3.4 Les sèmes de F. Rastier

Nous allons rapidement résumer les assertions de François Rastier [49] à propos de la nature de ces sèmes.

Forme : L'expression formelle d'un sème sera toute paraphrase métalinguistique, du simple /animé/ au sophistiqué /sert à découper du fromage/.

Nombre : Indéfini : ni trop ni trop peu. Puisque chaque sème sera justifié comme instaurant une relation entre des signifiés, le nombre de sèmes utilisés dépendra de l'organisation sémantique de ces signifiés. Le problème du nombre de sèmes utilisés pour une interprétation sera donc réglé par l'utilisation de classes sémantiques. Quant à la détermination du nombre total de sèmes (dans l'absolu de la langue et non dans la localité d'une interprétation), la question ne peut se poser. Au travers des textes et des interprètes, on peut s'avancer à les déclarer illimités (ce qui règle le problème).

Contenu : Grande question, qui nous mènerait vers des débats philosophiques sans rapport avec notre application. Retenons que les sèmes ne font que découper l'univers sémantique, et ne «contiennent» donc rien de précis. Ce ne sont que des marqueurs, de simples formes. Une autre question couverte par ce point pourrait être l'inclusion d'autres sèmes, ce qui nous ramène à la notion de minimalité si chère aux systèmes de primitives. Mais la justification d'un sème peut ne plus être suffisante à un stade interprétatif donné, et donner lieu à des précisions. Des notions comme celle d'implication logique peuvent être traduites par une inclusion, comme par exemple /a des pieds/ et /a deux pieds/. Mais dans ce cas les deux sèmes ont des statuts et des rôles descriptifs différents, et mettre à jour une notion d'inclusion ne se fera que très localement, comme nous le verrons en abordant les classes sémantiques.

Référence : Le problème de la référence a soigneusement été évacué, mais il reste fort acceptable de pouvoir attribuer un référent à certains sèmes ou groupes de sèmes. Quoiqu'il en soit, la référence ne serait pas une motivation mais un constat *a posteriori*, la *Bedeutung* ne venant qu'après le *Sinn* [18].

Zone de validité : Les sèmes sont-ils universaux ? Autre grande question, qui doit cependant déboucher sur la non-pertinence d'une universalité dans un projet de décrire une interprétation donnée. De toute façon, la perspective interprétative locale induit une liberté de désignation incompatible avec une quelconque universalité.

1.4 Sèmes et informatique

Reprenons ici les assertions de F. Rastier sur la nature et le statut des sèmes, et envisageons les implications informatiques de celles-ci.

- **Forme :** Un sème sera donc repéré par une simple chaîne de caractères, sans aucun besoin d'une analyse de celle-ci, du moins de façon circulairement linguistique. Même sur le plan informatique, il n'y a aucun problème pour l'instant, les simples identité et différence de deux chaînes nous suffiront. Nous échapperons aux critiques de la circularité métalinguistique (ou comment espérer décrire un élément linguistique si les éléments de description sont également de cette nature) par un refus de considérer une analyse des sèmes. La notion informatique de type sera ici fondamentale.

- **Contenu :** De la même façon, en l'absence de tout traitement de type inférentiel sur la base de la description sémique, les notions de référence et de contenu ne sont pas prises en compte.

- **Nombre :** Là non plus rien de grave, le problème de l'infinité étant écarté. L'indéfinition de ce nombre se résoudra par une simple approche incrémentale ; à chaque étape du processus descriptif, cette cardinalité sera fixée. Nous verrons dans notre approche formelle quelques contraintes afin d'évaluer ce nombre en fonction des données générales d'une interprétation. De façon plus pratique, il y aura un certain nombre de sèmes nécessaires, servant à structurer une analyse, et des sèmes supplémentaires, enrichissant cette structure minimale. Nous ne ferons toutefois pas de distinction formelle systématique entre ces deux types. L'important est ici de disposer d'une classe ouverte de sèmes.

- **Zone de validité :** Cette localité de la définition d'un sème est une première barrière à la généralité de tout outil proposé en ce sens. La restriction du repérage d'un sème rend en effet difficile la mise en relation de sèmes provenant d'origines diverses, là où justement on voudrait rechercher l'uniformité. Cette dernière ne devra donc pas être recherchée sur la base de la forme du sème, mais sur son rôle. De façon plus pratique, puisqu'un sème est utilisé par un interprète pour un texte, cela ne garantit pas que la même dénomination se retrouve chez un autre interprète, ni pour un autre texte. Nous devrons donc disposer d'une autre forme d'identité pour ces unités sémantiques.

2 Vers une formalisation de l'interprétation

Nous allons ici préciser les objectifs de l'approche de la sémantique interprétative ainsi que les nôtres, plus proches d'une réalité informatique que d'un souci linguistique. Cependant, nous essaierons de préserver l'articulation de ces deux préoccupations.

2.1 Les phénomènes à prendre en compte

La théorie de François Rastier cherche à décrire et rationnaliser le sens comme interprétation. Les deux notions sont liées, et l'on ne peut donc se contenter de la deuxième. En effet, parler du sens c'est invoquer bien souvent l'immanence de celui-ci, alors qu'une interprétation rattache ce dernier à sa construction, en le relativisant par là-même. C'est ainsi que nous nous intéresserons à un couple structure / processus, c'est-à-dire la structure de représentation du sens et le processus de construction de ce sens qu'est l'interprétation.

Notre objectif n'est pas de proposer un système autonome qui prend un texte en entrée et construit une structure sémantique. Ce serait donner une autonomie illusoire au texte traité, ou un pouvoir immense à la machine. Au contraire, nous nous contenterons de proposer un ensemble de processus élémentaires qui s'appliquent aux intuitions d'un lecteur, naïvement représentées, et qui vont tendre vers une organisation parallèle des connaissances du lecteur et de la structure sémantique du texte. C'est donc une sorte de rationalisation du sens (d'un sens, en fait).

Ces connaissances intuitives initiales représentent les aprioris du lecteur face au texte pris dans son ensemble ; elles couvrent également ses objectifs interprétatifs. Ces derniers relèvent généralement de la pragmatique : on lit une recette de cuisine ou le mode d'emploi d'un grille-pain dans le but d'obtenir une série d'actions à réaliser, un roman policier pour découvrir le coupable avant les dernières pages (quoique !), mais que dire d'un article humoristique ou d'un recueil de poèmes ? Dans ces cas, systématiquement évités par toute approche du TALN à notre connaissance, nous proposerons comme objectif d'interprétation la mise en évidence des structures sémantiques du texte, et du repérage de l'unité textuelle.

2.2 Le processus interprétatif

Dans [50] (p. 18), F. Rastier écrit :

> « Une fois éludé le moment de la compréhension comme intuition globale, le lecteur cesse de maîtriser d'emblée le sens textuel : il n'est qu'un des facteurs de sa constitution. Plutôt que le réceptacle dépositaire d'un sens plus ou moins profond, le texte apparaît comme une série de contraintes qui dessinent des parcours interprétatifs. »

À la base de ce que nous nommerons *processus interprétatif* se trouvent d'ores et déjà les résultats d'une interprétation. Ce point de départ sera donc l'impression thématique laissée par une lecture, et supportée par quelques éléments déclarés pertinents du texte. Par exemple, tel poème d'inspiration romantique peut être déclaré traduire le thème de la mort par des expressions comme 'silence', 'noirs', 'soir', 'Croise tes bras sur ton sein'[3], etc. L'interprétation, ou plutôt la pré-interprétation a donc comme résultats l'identification du thème, et la sélection d'un certain nombre d'unités pertinentes, dont on peut remarquer la disparité formelle (du nom à la proposition). Si, en un certain sens, on pourrait se «contenter» d'un tel résultat, la méthode proposée par F. Rastier dans Sémantique Interprétative permet d'aller bien plus loin puisqu'elle permet de *justifier* un tel état de fait, et nous y rajoutons le fait qu'elle peut servir de tremplin interprétatif vers d'autres thématiques. Nous verrons que cette recherche de la justification peut en effet guider l'interprétant vers des découvertes sémantiques, et au prix d'un travail de systématisation, augmenter la portée du texte interprété. À un autre degré, également, nous verrons qu'une formalisation de ces intuitions permettra de mettre en place une structure textuelle d'autant plus riche dans des textes poétiques ou littéraires, par exemple en se basant sur les relations sémantiques ou syntaxiques entre les différentes thématiques repérées.

Enfin, une sorte de «jeu interprétatif» peut être défini, consistant en la projection de thèmes *a priori* décorrélés des résultats des analyses classiques

3. Ces exemples proviennent de 'En sourdine', d'un certain Verlaine, voir en annexe

du texte, et en la justification de ces projections. Ici peut-être nous nous rapprocherons des travaux de l'OULIPO [44], ou, moins sérieusement, des travaux de certains écoles de psychanalyse, s'efforçant de calquer des thèmes-archétypes (sexuels en général) sur n'importe quel texte, et à tout autre niveau d'analyse.

Plus théoriquement, nous tâcherons de mettre en évidence la possibilité ainsi offerte de récupérer ce fameux axiome de la détermination du local par le global. Ces intuitions premières (que nous nommerons par la suite des pré-isotopies) possèdent déjà dans leur forme cette dualité orientée, puisqu'elles sont constituées d'un thème (global) et d'une liste de supports syntagmatiques (local). Mais l'approfondissement de ces intuitions (la transformation de pré-isotopies en isotopies) va permettre d'étendre la portée de la localité, en justifiant les impressions thématiques, et en exprimant les rapports locaux entre ces thématiques et entre les termes qui supportent un thème donné.

2.3 La notion d'interprétant et le problème de la norme

Nous avons déjà établi l'objectif central de cette théorie qui est de décrire les opérations régissant l'attribution d'un sens à un texte par un interprète donné. F. Rastier introduit à ce stade une notion permettant de qualifier les opérations interprétatives élémentaires : celle d'interprétant. Un interprétant peut être vu comme une entité sémiotique à l'origine d'une attribution d'un trait sémantique à une entité signifiante. En quelque sorte, un interprétant motive l'attribution d'un sème. Et comme nous verrons qu'il y a différents types d'attributions de sèmes, il y a donc différents types d'interprétants. Nous verrons plus tard les différents types de sèmes, mais revenons pour l'instant sur quelques grandes familles d'interprétants que sont les normes. Nous avons vu que l'interprétation était un acte éminemment personnel, mais il se doit tout de même d'être contraint, ne serait-ce que pour confirmer le rôle de communication que remplit avec un succès inégalé le langage. Grossièrement, si l'on peut voir un grand nombre de sens dans un texte, certains sont plus «saillants» que d'autres. Autrement dit, certains interprétants sont, sinon plus prioritaires, du moins plus stables. Ce qui nous conduit aisément à considérer les normes dans le langage comme des formes de stabilité sémantique (du moins nous ne nous intéresserons ici qu'à leur aspect sémantique).

Il convient aussi d'envisager plusieurs niveaux de normes, en fonction de leur stabilité, ou du moins de leur zone de validité [49], p. 39 :

- La localité textuelle, donc les structures relationelles du texte et sa physionomie locale.
- L'idiolecte, ou les conventions personnelles de son auteur, qui se différencient légèrement des premières dans le cadre d'une étude diachronique d'un écrivain par exemple.
- Le sociolecte, ou norme socioculturelle, s'apparente peut-être au genre du discours, et aux pratiques auxquelles il est destiné.
- Le dialecte, ou l'abstraction stabilisée d'une langue donnée.

Concrètement, ces différentes normes permettent de repérer différents sèmes, dont la nature est identique dans l'absolu, mais dont le type d'attribution à une entité linguistique varie.

Idéalement, à chacune de ces normes correspond un ensemble d'associations de sèmes à des signifiés. En contexte, de telles normes peuvent bien sûr être bouleversées, et un certain nombre de priorités seraient envisageables, avec un avantage accordé aux effets locaux.

Ainsi, puisque la notion de langue elle-même n'est qu'une abstraction, il convient de voir en ces normes «supérieures» (sociolecte et dialecte) une stabilisation des sens attribués aux unités signifiantes. La conscience de ces normes est également indispensable en vue de toute notion de créativité, selon le principe que le mouvement ne peut être perçu que par rapport à l'immobile (ou au considéré comme tel).

D'un point de vue formel, nous ne pouvons donc espérer capter ces stabilités si étendues à travers l'espace, le temps et les catégories socioculturelles. Nous pouvons par contre dans bien des cas évaluer la validité des interprétations, c'est-à-dire identifier l'*interprétant* d'un sème. Les interprétants relevant de normes supérieures ne pouvant en règle générale pas être explicités. Par contre, pour les normes locales (mais sont-ce vraiment des normes ?), cela est possible : par exemple expliciter la notion de diffusion d'un caractère sémantique lors d'une association syntaxiquement marquée, comme dans le cas des énumérations.

La sémantique interprétative utilise également les normes afin de qualifier les variations des sens attribués à un même signifiant. Il s'agit ici de polysémie au sens faible du mot, sans tomber dans les extrêmes des sens qui n'apparaissent jamais deux fois identiques en contexte, comme le concluait, empreint de pessimisme, Daniel Kayser [30]. Dans la lignée de ce que proposait Georges Kleiber en réponse au précédent article [31], on peut effectivement espérer retrouver un noyau sémantique pour un terme, mais à un niveau de norme élevé. Sans vouloir pour autant entrer dans le débat acharné qui se jouait dans ces articles, précisons que l'interprétation de *livre* comme « *indexation d'un objet-livre par un documentaliste* » dans « Encore un livre et j'aurai fini cette pile » ne se fait au mieux que via une norme sociolectale, par enrichissement de la notion exprimée en dialecte. Ainsi, sont proposés ([49] p 70) trois paliers de description des « sens »[4] des occurrences, définissant les termes de sens (palier dialectal), acception (palier sociolectal) et emploi (palier idiolectal). Nous verrons par la suite comment l'on peut expliciter techniquement les nuances de ces différents cas, et comment aborder pratiquement les problèmes de polysémie ou de pseudo-polysémie. Pour un même signifiant, les relations entre signifiés dépendent ainsi des normes dont relèvent les composants qui les distinguent. La véritable polysémie (sens différents) suppose ainsi des distinctions par des traits dialectaux. Les variations de *livre* précédemment envisagées, si elles sont situées à des paliers locaux supposent de toute façon une identité suivant un palier plus élevé (et non seulement sur la forme, comme ce serait le cas d'une comparaison avec la monnaie britannique par exemple).

Un corollaire de ce ceci, que nous développerons dans les prochaines chapitres, sera la reconnaissance *a priori* d'une différence de signifié quand des signifiants identiques sont présents dans un même texte. Les rapports d'emplois ou d'acceptions entre ces différents sens n'interviendront qu'une fois l'identité propre (ou commune) de ces occurrences définie.

Nous verrons également par la suite que si l'on ne peut justifier ni identifier automatiquement toutes les normes à l'œuvre dans un processus interprétatif, elles n'en restent pas moins importantes quand il s'agit de comparer différents sèmes ou structures de sèmes. Nous serons donc, une fois de plus, obligés de nous en remettre à la compétence de l'interprète humain pour la qualification même des résultats partiels de son interprétation. Un point positif, toutefois, apparaît lorsque F. Rastier explicite les rapports, au sein de sa théorie, entre linguistique et herméneutique, dans [51] (p. 23) : «La sémantique tente de retracer les parcours interprétatifs. Mais c'est à l'herméneutique critique qu'il revient de problématiser leurs conditions et de hiérarchiser leurs résultats en définissant des degrés de plausibilité.» Ainsi, dans notre approche par trop centrée sur l'interprète, nous nous abstiendrons donc de qualifier le *pourquoi* d'une interprétation, nous concentrant sur un

4. Un terme générique nous fait actuellement défaut

comment descriptif, en lui proposant des outils d'expression, en laissant les impacts des normes implicites.

Nous allons donc dans la suite expliciter les moyens de représentation des informations sémantiques, via un certain nombre de concepts. Dans un premier temps nous résumerons de façon critique les différentes propositions de F. Rastier, puis nous proposerons un formalisme qui les unifiera.

3 Concepts et outils de base

Pour la présentation des différents concepts de la sémantique interprétative, nous avons choisi un ordre d'inspiration constructiviste, donc ascendant. Nous sommes conscients des implications de ce choix, notamment pour le statut de l'isotopie, qui pourrait perdre ainsi son rôle fondateur des autres concepts. En partant en effet des unités locales pour atteindre les structures globales, nous ne faisons pas une publicité très attractive pour la détermination du local par le global. Disons alors que nous préservons le meilleur pour la fin, en établissant les besoins essentiels d'une régulation globale sur ces unités «inférieures».

3.1 Les sémèmes

Nous avons déjà discuté des sèmes, ou éléments distinctifs traduisant des relations entre les signifiés. Nous allons donc continuer en explicitant la nature de ces signifiés, et les rapports qu'ils entretiennent avec les sèmes.

3.1.1 Leur forme : la lexie

Avant de parler du statut de signifié, intéressons-nous à la nature du signifiant. Dans Sémantique Interprétative, F. Rastier proposait le morphème, principalement à cause du fait que des unités strictement minimales peuvent supporter des informations sémantiques. Il ne faut pas confondre cet état de fait avec l'attribution d'un pseudo-sème /pluriel/ au morphème-suffixe '-s' ; ces considérations grammaticales n'ont qu'un rapport indirect avec la sémantique comme nous l'entendons. Par contre, le même morphème peut supporter, sur la même idée, des notions de /multitude/ ou de /nombre/ en un sens purement sémantique.

Mais, hors de ces cas extrêmes, les informations sémantiques reposent généralement sur des composés de morphèmes. Sans pour autant tomber dans le piège du mot, impossible à définir dans une linguistique autre que de surface (par des repères graphiques, et dans certains cas prosodiques) et qui d'ailleurs n'a pas de couverture suffisante, F. Rastier reprend la notion de *lexie* dans «Sémantique pour l'analyse» [52]. Ce terme, proposé par Pottier [47], est assez flou pour couvrir la disparité constatée. Il s'agit en effet d'un groupement stable de morphèmes, répondant à quelques critères syntaxiques, mais dans notre cas les critères seront purement fonctionnels : une lexie sera tout signifiant repéré par l'interprète comme pertinent pour son analyse. Ainsi, la notion de stabilité s'efface pour des considérations plus locales.

3.1.2 Signifiant ou signifié ?

Comme nous l'avons déjà évoqué, dans une approche de traitement automatique, la distinction ne peut se faire dans l'absolu, et de plus la machine ne manipule que des signifiants, et encore. Pour elle ce ne sont que des étiquettes, entités restreintes de la couche symbolique logique, ne devenant signifiantes que par l'intermédiaire de la couche sémiotique et la bonne volonté de l'interprète. Notre outil ne s'occupera donc que de chaînes de caractères, et la notion de signifié ne sera véritablement atteinte que par une structure de données qui établit des relations étiquetées entre ces chaînes.

Reprenant la terminologie classique, on nomme *sémème* le signifié correspondant à un morphème, et *sémie* celui qui correspond à une lexie. La distinction n'a d'intérêt que pour étudier les effets de sens de la composition morphématique dans les unités supérieures. Nous contournerons le problème en considérant tous les signifiés comme des sémèmes, qu'ils aient des morphèmes ou des lexies comme signifiants [5]. Nous préférerons en effet gagner en souplesse descriptive sans poser le problème de la composition des morphèmes en unités plus complexes. Toujours dans un objectif applicatif, nous ne rendrons pas nécessaire la description de toute unité signifiante du texte interprété : la sélection des unités pertinentes fait partie de l'acte interprétatif, et l'on doit juste permettre une grande variété de supports.

De même, par extension terminologique, nous nommerons sémème tout élément terminal de la structure sémantique, même si cette structure, au stade de la description concerné, n'est pas suffisamment élaborée pour justifier un statut pleinement sémantique.

Nous allons maintenant expliciter les types de relations mises en place entre les sémèmes.

5. Avouons que les habitudes prises durant l'étude intensive de Sémantique Interprétative sont difficiles à perdre.

3.2 Limiter les éléments distinctifs : classes et oppositions

Répétons un truisme structuraliste : dans la langue il n'y a que des différences et des identités. Mais dans une microsémantique, ces différences entre sémèmes s'expriment par des différences de sèmes, et les identités de sèmes impliquent des équivalences de sémèmes. Si l'on utilise un sème pour distinguer chaque couple de sémèmes, on se heurte à un problème de complexité combinatoire. Pour diminuer ces identifications, il suffit dans un premier temps de classer les sémèmes, donc de grouper les oppositions. C'est bien entendu une justification opératoire, qui ne nous contente qu'à un certain point.

À l'autre extrémité de l'échelle de justification se trouve le principe aristotélicien que l'on ne peut rien dire de ce qui n'est pas inscrit dans un genre. Il s'agit en fait d'une conséquence du principe de la détermination du local par le global. Ceci nous aide également à limiter les éléments distinctifs, puisque ceux-ci nécessitent avant tout une zone de validité, *i.e.* d'opérationnalité.

Quoiqu'il en soit, ce principe introduit un premier partage dans les types de sèmes envisageables. Une première catégorie (celle des sèmes dits *génériques*) jouera un rôle classificateur, et traduira l'appartenance d'un sémème à une classe sémantique. Une seconde (celle des sèmes dits *spécifiques*) se chargera d'exprimer les différences entre les éléments d'une telle classe.

Intéressons nous maintenant à la nature de ces classes, tant au point de vue formel qu'à celui de leur statut sémantique.

3.3 Taxèmes : classes minimales

La sémantique interprétative identifie trois types de classes sémantiques, dont une seule est obligatoire, car c'est d'elle que dépendent tous les repérages de sèmes. Il s'agit de la classe minimale, ou *taxème*, dont F. Rastier accepte la définition de Coseriu [14] : «structure paradigmatique constituée par des unités lexicales se partageant une zone commune de signification et se trouvant en opposition immédiate les unes avec les autres».

Première remarque : l'identification d'un taxème ne se fait que par la reconnaissance *a priori* d'une zone de localité sémantique. Il y a donc une autre globalité qui détermine le taxème.

Deuxième remarque : le principe de l'opposition des sémèmes au sein d'un taxème sera par contre le but visé par l'identification d'un taxème. Nous l'expliciterons en abordant les sèmes spécifiques par la suite.

Revenons donc au problème initial de la genèse d'un taxème. Ou, en formulant autrement, quelle est la justification de cette zone de localité sémantique? Une réponse vient aussitôt : la norme. Mais laquelle? La question est loin d'être vaine, car du taxème, en tant que fondation de la structure sémantique, nous devons chasser toute imprécision, et préparer sa formalisation.

3.3.1 Justifications des taxèmes

En résumé, deux propositions s'offrent à nous : soit les taxèmes n'ont de validité que locale et contextuelle, c'est-à-dire qu'ils servent en quelque sorte de focalisation de l'interprétation sur des effets de sens estimés centraux par l'interprète, soit ils correspondent à la stabilisation d'une norme supérieure. Que cette dernière soit sociolectale ou dialectale n'est pas pertinent pour l'instant, le problème de leur identification sera résolu (ou déclaré insoluble) de la même manière.

Or, il se trouve que la sémantique interprétative propose différents exemples de taxèmes qui semblent couvrir les deux possibilités. Citons le taxème des couverts, comprenant 'fourchette', 'couteau' et 'cuiller', donc *a priori* sociolectal voire dialectal, et plus loin est envisagé le cas d'un taxème *local*, relevant d'une norme idiolectale. L'exemple cité est tiré d'un texte de Giono, et présente en une énumération un ensemble d'aliments relevant de différentes paradigmes dialectaux. Mais en fait, n'est-ce pas le seul *contenu* de ce taxème qui correspond à une norme locale? La notion de gourmandise est certes socialement stabilisée, et un ensemble de sémèmes «typiques» peut y être aisément associé. La nature «en langue» du taxème correspond bien à une norme globale, pouvant être transgressée localement par le rapprochement de notions disparates.

Il semble donc y avoir une opposition langue / discours incontournable. Nous devons cependant trancher, et appeler la réalisabilité à notre secours. S'en remettre à des normes indéfinissables précisément est hasardeux, alors que justifier toute identification par un objectif local n'engagera que la compétence de l'interprète.

Cependant, l'identification et/ou la formation de classes contextuelles ne peut s'opérer simplement dans tous les cas. Imaginons l'exploration d'un texte hermétique à l'interprète, ne repérant lors d'une première lecture aucun effet classificateur autre que relevant d'une généralité stérile (animé, concret, etc.) L'échappatoire est donc de replacer chaque sémème dans un taxème sans tenir compte de son rôle dans le texte étudié (bien que cela soit impossible de façon absolue), puis d'explorer plutôt le rôle joué par le sémème au sein du taxème : les relations découvertes peuvent donner des indices vers l'unité du texte.

Mais on peut préciser que le taxème est une «classe de sémèmes minimale en langue» (p. 276). De quelle notion de minimalité est-il ici question? Nous avons laissé présager que d'autres types de classes seront mis en place par la suite, mais que seul le taxème était nécessaire. La minimalité concerne donc *a priori* la non inclusion de toute autre classe dans un taxème, y compris un autre taxème. Rendre effective cette notion de minimalité suppose donc la connaissance du découpage des signifiés en classes. De plus, il faut le préciser d'avantage : c'est une classe «en langue», ce qui rejette la possibilité de concevoir des taxèmes purement contextuels, alors que c'est principalement grâce à eux que l'on peut espérer retrouver les phénomènes sémantiques intéressants d'un texte.

3.3.2 Le taxème comme focalisation de l'interprétation

En nous appuyant sur ce principe que le taxème n'est justifiable que par les relations qu'il va permettre d'expliciter entre ses éléments (les sémèmes), il en résulte que le taxème n'est donc qu'une zone relative de focalisation sémantique. Dès lors, il est dépendant des objectifs interprétatifs. D'un point de vue formel, des conclusions fonctionnelles peuvent être tirées.

Premièrement, aucun sémème, considéré lors d'une étape stabilisée d'une interprétation, ne peut appartenir à deux taxèmes distincts.

Deuxièmement : puisque le taxème est une classe minimale, il ne peut contenir d'autres taxèmes.

Il est par contre envisageable de déplacer la zone de localité qu'il représente en la spécifiant. Par exemple, en partant d'un taxème de forte généralité comme l'alimentation, où se regroupent des termes liés à la confection culinaire, aux arts de la table et à l'outillage nécessaire à l'absorption de nourriture. Si la zone de signification recouverte par cet ensemble est présente dans l'intégralité du texte (par exemple si ce texte est une recette de cuisine), il se trouvera une multitude d'éléments à opposer en son sein. Ainsi, même sur des critères purement quantitatifs, un découpage de cet ensemble est justifiable et souhaitable, pour les mêmes raisons qui nous faisaient critiquer les marqueurs sémantiques de Katz et Fodor. De plus, la présence simultanée d'un sémème dans deux taxèmes traduit la co-occurrence de deux interprétations de la même lexie. Si l'on admet que l'identité d'un *sémème* n'est pas liée à sa forme (la lexie), mais à son sens (ses sèmes), vouloir placer le même signifiant dans deux classes sémantiques distinctes traduit simplement le fait que ce signifiant supporte deux signifiés différents. Un exemple d'une telle polysémie liée à une seule occurrence lexicale ne nous apparaît pas pour l'instant, et nous verrons lors de l'explicitation du processus interprétatif pourquoi un tel cas est impossible. Il existe par contre des moyens dans la sémantique interprétative de préciser des rapports sémantiques entre différentes classes (taxèmes).

3.3.3 Taxème et sème micro-générique

Maintenant que nous avons en partie défini la nature du taxème, revenons aux sèmes à proprement parler. Avant même de permettre d'identifier des sèmes exprimant des oppositions entre les éléments de la classe, le taxème identifie déjà un sème attribué aux sémèmes qu'il contient. Tous les sémèmes d'un taxème auront donc un sème commun, et ce sème aura pour eux le statut de *microgénérique*. Nous verrons par la suite que ce principe de généricité est valable pour les autres classes sémantiques plus élevées, d'où l'appellation, ici distinctive, de *micro*. Il convient donc de différencier ces deux notions : le taxème est en quelque sorte l'extension du principe de localité sémantique, et le sème microgénérique associé en est l'intension. Moins par convention et par économie que par un principe de dualité, nous attribuerons la même étiquette aux deux entités, en prenant simplement soin de les distinguer typographiquement. Ainsi, le taxème //couvert// implique la présence du sème microgénérique /couvert/ *dans* les sémèmes qu'il contient. Ce *dans* doit être abordé avec précaution : d'un certain point de vue, les sémèmes peuvent être considérés comme des ensembles de sèmes, si on les voit localement. Mais un sème est avant tout une relation entre sémèmes, dont la projection sur un des sémèmes de la relation est en quelque sorte une occurrence de sème, comme l'exprime F. Rastier lorsqu'il définit le sème comme «élément d'un sémème, défini comme l'extrémité d'une relation fonctionnelle binaire entre deux sémèmes». Le meilleur moyen à nos yeux de distinguer ces deux aspects est d'utiliser les notions d'antériorité logique : le sème ne peut être vu comme élément d'un sémème que parce qu'il exprime une relation [27].

Formellement, la notion de taxème sera donc une structure à deux facettes : un contenu, exprimable par un ensemble fini de sémèmes, et une qualification par un sème. Cette structure, comme d'autres que nous verrons par la suite, servira en quelque sorte de transition organisatrice entre sèmes et sémèmes.

La relation exprimée par un sème microgénérique est une relation d'équivalence entre les sémèmes du taxème, et le taxème en constitue précisément la classe. Vus de l'extérieur du taxème, les sémèmes sont en fait indistinguables, tant que les oppositions qu'ils entretiennent ne sont pas exprimées via les sèmes spécifiques.

Toujours conséquence de cette antériorité logique, plusieurs sémèmes contenant un sème commun ne forment pas pour autant un taxème ; encore faut-il pour cela que le sème commun soit microgénérique pour tous ces sémèmes. Autrement, il ne forme qu'(!)une isotopie, ce qui maintient la notion d'identité des unités formelle au seul niveau de l'interprète-utilisateur, et non comme résultat d'un calcul.

Mais voir le sémème comme un ensemble, dès lors qu'il ne nous guide pas vers une fausse route en considérant les sèmes comme donnés *a priori*, et à assembler en sémèmes (avec un principe de compositionnalité mal placé), permet tout de même de préciser un peu plus sa structure. F. Rastier distingue en effet deux parties d'un sémème ainsi considéré, le *classème* et le *sémantème*. Le premier n'est autre que l'ensemble des sèmes génériques supportés par le sémème, alors que le second est son correspondant pour les sèmes spécifiques. Cette terminologie permet de retracer le chemin à l'envers jusqu'à Greimas, pour qui le classème est un sème contextuel traduisant la cohérence d'une partie de discours ([20] p. 53), donc un sème générique correspondant à une classe contextuellement définie. Greimas ne fait donc pas l'opposition classème / sémantème, ce dernier terme n'étant pas repris dans sa théorie. Il est donc intéressant de noter que notre vision du taxème comme classe contextuellement justifiée, sans doute plus par souci de pragmatisme que par volonté de respecter l'intégrité de la sémantique interprétative, nous rapproche en quelque sorte de la sémantique greimassienne. Nous verrons cependant lors de l'explicitation du processus interprétatif, pour ce qui est de la «création» des taxèmes, que les rapprochements sont encore plus flagrants.

3.4 Les sèmes spécifiques : de la différence à l'opposition

Maintenant que nous avons circonscrit la zone de localité sémantique qu'est le taxème, dans le but de traduire les différences de signifiés en oppositions, précisons la nature de ces relations spécifiques. Cette étape, se situant pleinement dans ce que nous avons identifié comme la couche logique de notre application, doit rejeter toute imprécision dans son développement. La simplicité générale de cette dernière doit être équilibrée par une grande rigueur formelle, et nous verrons que la notion de sème spécifique et celle d'opposition qu'elle véhicule est centrale.

Si nous reprenons l'exemple obsessionnel du taxème //couverts//, contenant 'fourchette', 'couteau' et autre 'cuiller', il va nous falloir exprimer via des sèmes les oppositions entre ces unités.

Or, on trouve ([49] p. 52) : « ... la relation entre deux sèmes spécifiques « permettant d'opposer deux sémèmes voisins » est une relation d'incompatibilité ; elle induit entre les sémèmes qui les incluent une relation de disjonction exclusive ».

L'opposition (entre les sémèmes) est donc bien cette relation de disjonction exclusive dont parle la sémantique interprétative. Par contre, la notion d'incompatibilité doit être étudiée plus attentivement. En logique, l'incompatibilité est une relation très forte, exprimée par le fait que deux entités incompatibles ne peuvent *jamais* coexister dans une entité supérieure. Prenons l'exemple que donne F. Rastier pour les couverts : il attribue /pour prendre/ à fourchette et /pour couper/ à couteau, et se base sur l'incompatibilité de ces deux sèmes (donc *a priori* une relation absolue entre entités d'un niveau d'abstraction élevé) pour induire l'opposition entre ces deux sémèmes. Le schéma 2.1 illustre cet état de fait.

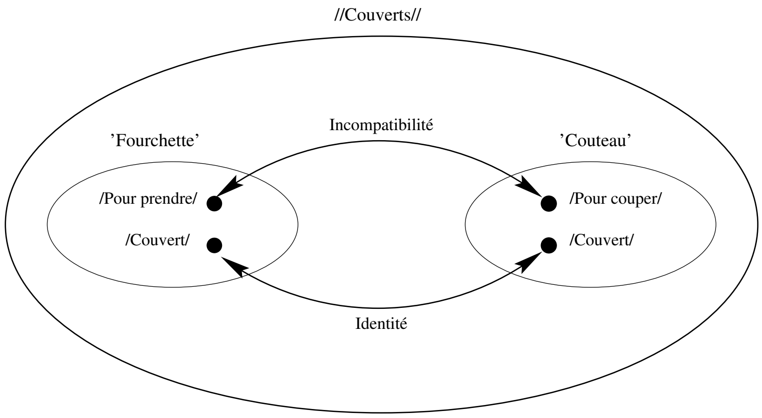


FIG. 2.1 – *Incompatibilité des sèmes spécifiques et opposition des sémèmes*

Considérer l'incompatibilité dans un sens logiquement fort implique donc que l'on ne peut trouver aucun sémème auquel soient attribués ces deux sèmes. Cela suppose également que la langue nous fournit un réseau complet de sèmes incompatibles, comme les éléments de définition d'un système logique. Ou bien la notion d'incompatibilité logique est trop forte, et nous pouvons nous contenter de relations locales (par exemple limitées à la zone du taxème), auquel cas ce terme d'incompatibilité devrait être évité.

Un autre point intéressant est le suivant : un seul sème spécifique défini de la façon précédente ne peut suffire à exprimer une relation binaire entre sémèmes : l'incompatibilité doit être établie entre *deux* sèmes pour induire l'opposition entre sémèmes. Par conséquent, la notion de taxème à un seul élément est inintelligible.

Toutes ces considérations vont donc nous conduire à remanier quelque peu la «forme» des sèmes spécifiques, avec en tête les objectifs suivants :

- Éviter le recours à des notions globales d'incompatibilité.
- Rendre au sème spécifique sa fonction de relation binaire.
- Préserver le principe de «positivité» de l'opposition traduite par les sèmes spécifiques.

3.4.1 Forme des sèmes spécifiques

Nous avons, en abordant la notion de taxème, dû distinguer entre la forme et le contenu de ces classes minimales. L'opposition taxème / sème microgénérique nous a d'ailleurs paru fondamentale, à tel point que nous sommes obligés de projeter cette dualité sur le pendant spécifique des sèmes génériques. Nous allons donc envisager différentes formes d'oppositions binaires et discuter de leur utilisation pour les sèmes spécifiques.

Incompatibilité unaire explicite - Il s'agit en fait de la notion, classique dans l'approche structuraliste, que nous avons rappelée précédemment. Chaque sème spécifique attribué à un sémème exprime une opposition par la présence d'un sème incompatible dans un autre sémème. Les principaux désavantages sont les connaissances globales nécessaires sur les sèmes (une sorte de tableau universel des sèmes incompatibles).

Incompatibilité présence / absence - Si l'on se base pour notre petite étude sur la notion d'incompatibilité, nous pouvons nous rapprocher de celle, incontournable, qui existe entre les deux notions de présence et d'absence. Une entité quelconque ne peut être à la fois présente et absente dans une autre entité. Un sème (spécifique) ne peut donc être présent et absent dans un sémème. Donc, si l'on considère le sémème 'couteau' auquel on attribue le sème /pour couper/, il est en opposition avec tous les sémèmes du taxème qui ne contiennent pas le sème /pour couper/, par exemple 'fourchette'.

Les désavantages de ce choix sont :

- L'universalité de l'opposition ainsi traduite : en quelque sorte, 'couteau' se trouvera opposé à tout sémème, du même taxème ou non, qui ne comporte pas le sème /pour couper/.

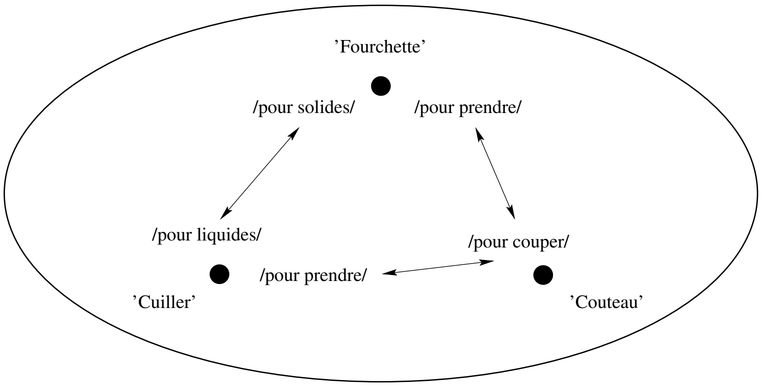


FIG. 2.2 – *Incompatibilité unaire explicite*

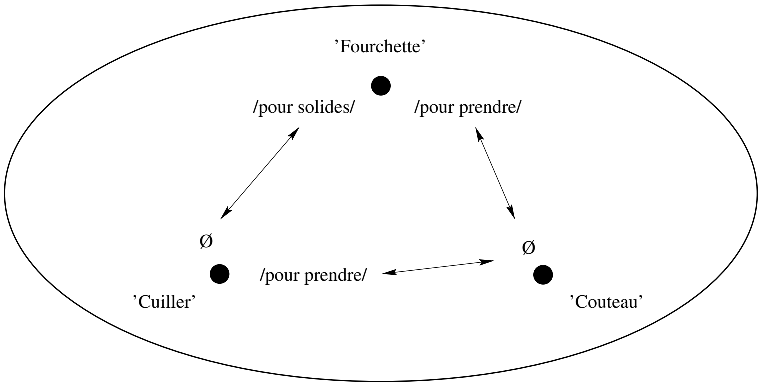


FIG. 2.3 – *Incompatibilité par présence / absence*

– La nécessité de disposer d'un inventaire complet des sèmes attribués aux sémèmes d'un taxème avant de trancher (!) sur les oppositions.

– L'absence de caractérisation positive de l'information véhiculée par l'absence (un sémème sans sème spécifique s'oppose à tous les sémèmes qui en contiennent au moins un).

– L'impossibilité de traduire des différences graduelles du style /grand/ (par rapport à ...).

Incompatibilité A / non-A - À mi-chemin entre les deux possibilités précédentes, l'incompatibilité logique entre un trait A et sa négation non-A pourrait être appliquée. Par exemple, on pourrait attribuer /pour couper/ à 'couteau' et /non - pour couper/ ou même non-/pour couper/ à 'fourchette'. Nous ne nous attarderons pas plus longuement sur cette possibilité, car il est clair qu'elle cumule les deux séries de désavantages que nous avons précédemment énoncées.

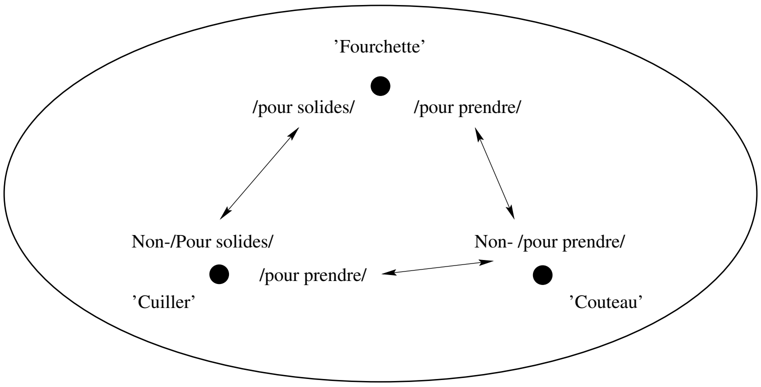


FIG. 2.4 – *Incompatibilité A / non-A*

Incompatibilité binaire explicite - Nous allons, comme il va de soi, expliciter en dernier la possibilité que nous retiendrons par la suite. Elle s'inspire de la forme que donne F. Rastier aux sèmes spécifiques, avec les nuances suivantes :

L'incompatibilité est rejetée comme notion globale : l'attribution d'un sème spécifique à un sémème s'accompagne de la distinction explicite de ce sémème d'un autre sémème du taxème.

Ainsi, le problème de l'inventaire des sèmes spécifiques est *restreint au taxème considéré*.

La forme d'un tel sème spécifique est donc un couple (sémème_1 , sémème_2). Le sème est en fait attribué à sémème_1, pour le distinguer de sémème_2. Ainsi, /pour couper/ ne sera pas attribué à 'couteau' simplement, mais en y ajoutant l'information que ce sème sert à distinguer 'couteau' de 'fourchette', sans avoir besoin dans l'immédiat d'expliciter un sème incompatible pour 'fourchette'. En quelque sorte, seul le premier sémème du couple bénéficie d'une assertion sémantique. Du même coup, un sème spécifique ne peut servir à la distinction d'avec un autre sémème quelconque qui aurait reçu un sème incompatible.

Le problème à ce stade est bien entendu l'absence de symétrie dans la caractérisation positive : de 'fourchette' on ne peut rien dire, si ce n'est qu'on ne peut lui attribuer le sème /pour couper/ afin de le distinguer de 'couteau' (cela annulerait effectivement le rôle distinctif du premier sème spécifique). Mais toutes les autres possibilités d'attribution sont encore ouvertes, y compris l'attribution de /pour couper/ à 'fourchette' par rapport à un troisième sémème. Changeons donc de thème pour exprimer ceci, et revenons nous rafraîchir dans les cours d'eau (taxème approfondi par M. Noailly dans [42] pour l'étude de *fleuve*). On peut attribuer /grand/ à 'fleuve' par rapport à 'rivière', et /grand/ à 'rivière' par rapport à 'ruisseau'. Nous avons ainsi des caractérisations positives des deux premiers sémèmes.

Une implication non négligeable de cette forme est, entre autre, que le sème spécifique devient une relation non symétrique, au contraire de ce que propose la sémantique interprétative. De plus, cette solution évacue radicalement les problèmes de relations globales entre les sèmes : il n'est plus besoin de les manipuler par paires, ni d'avoir recours au méta-langage de description pour trouver l'«opposé» d'un sème (du type /pour couper/ - /pour prendre/). Cette possibilité est cependant toujours ouverte, puisque dans le taxème des couverts nous pouvons attribuer le sème /pour prendre/ au couple ('fourchette', 'couteau').

Nous explorerons plus systématiquement et de façon plus rigoureuse les implications de ce choix lorsque nous aurons introduit un cadre plus formel dans le troisième chapitre. Ce que nous pouvons conclure pour l'instant est que la proposition que nous venons de faire ne résout bien sûr pas tous les délicats problèmes des sèmes spécifiques, surtout celui de l'inventaire des sèmes nécessaires à la «cohésion» d'un taxème. Néanmoins nous sommes restés dans un structuralisme fort, en explicitant directement le rôle relationnel du sème spécifique.

Mais surtout, cette forme possède une vertu dans le principe de la co-

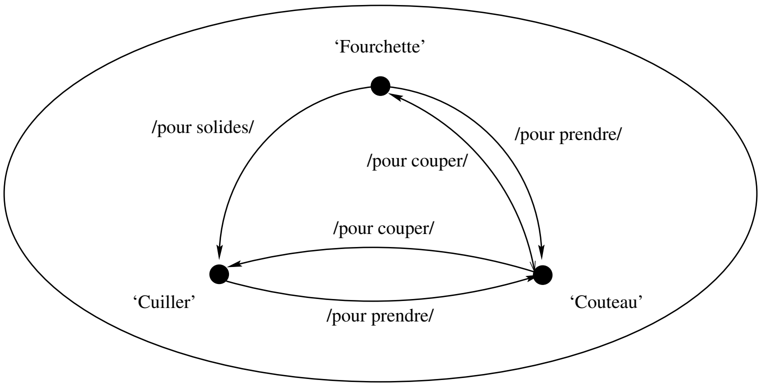


FIG. 2.5 – *Incompatibilité binaire explicite*

opération homme / machine que nous avons présenté qu premier chapitre. Un sème spécifique, par son extension d'une telle forme, peut être repéré simplement par l'utilisateur à qui l'on ne présente que trois éléments : le premier sémème, source de l'opposition et à qui sera attribué le sème, le second sémème duquel le premier se distingue, et le « contexte » de cette opposition, à savoir le taxème. D'un autre côté, l'information nécessaire est simplement le sème, sans qu'il soit besoin d'y ajouter un ensemble de sèmes incompatibles (comme il serait le cas pour une incompatibilité unaire explicite) ou un sous-ensemble de sémèmes du taxème qui ne possèdent pas ce sème (pour l'incompatibilité par absence). La simple adéquation de la structure formelle pour le protocole de communication accorde donc à cette alternative un avantage majeur.

Remarque : Ce que nous venons de discuter ici est bien la «forme» du sème spécifique, tout comme, en quelque sorte, le taxème était la forme (extensionnelle) du sème micro-générique. Nous verrons plus clairement dans le formalisme que nous gardons intacte la notion de sème spécifique, mais que nous modifions (ou plus précisément proposons) ses caractéristiques formelles, que nous ne pourrons plus nous passer de nommer. Nous avons choisi pour compléter analogiquement les notions de F. Rastier le néologisme *spécème* comme étant la composante extensionnelle du sème spécifique. Tout comme le taxème pouvait être vu comme un ensemble de sémèmes, séparable de la notion de sème générique, le spécème sera un couple de sémèmes, apte à supporter un (ou plusieurs) sème(s) spécifique(s). La nature typologique du sème sera d'ailleurs fixée par l'entité formelle qu'il qualifie ; ainsi, un même sème pourra être envisagé comme tantôt générique, tantôt spécifique, suivant les sémèmes et l'interprétation considérés.

3.5 Le taxème comme typologie des sèmes : l'inhérence et l'afférence

Le problème de la norme intervient à plusieurs autres endroits dans la théorie. Nous avons déjà vu deux types de sèmes, les génériques et les spécifiques. Les deux exprimaient des relations entre des sémèmes appartenant à une même classe : le taxème. Autrement dit, il s'établit une sorte de ruption entre deux sémèmes de taxèmes différents : on ne peut rien en dire. Mais ce n'est pas toujours le cas, la langue étant le siège de phénomènes bien plus complexes, et de relations bien plus riches que celles pouvant se mettre en place au sein d'une taxonomie comme celle exprimée par les classes sémantiques.

Deux mots à propos du principe de la taxonomie en linguistique. Ce principe a surtout trouvé une oreille attentive depuis les premières tentatives d'informatisation de la langue, sans doute plus motivées par un souci d'économie de description que d'une véritable volonté de voir ce phénomène à l'œuvre dans le système langagier. Quoiqu'il en soit, le besoin ressenti de classer les unités lexicales au moyen de classes arborescentes, articulées selon l'apparemment simple mais véritablement vertigineuse notion du *est-un*. Outrepassant le fait, comme le souligne F. Rastier, que ce type de description ne s'applique naturellement qu'à une petite sous-partie de la langue (la zoologie, la botanique...), les formalismes en sont venus à utiliser des classes d'un niveau d'abstraction tel (*abstrait, concret, animé, inanimé*, et surtout la trop célèbre *chose*, ou l'*objet*) qu'elles en sont devenues injustifiables linguistiquement. Quand parfois le principe d'économie de cette relation d'héritage n'était pas le moteur de son utilisation, la justification des classes était liée à la forme, comme chez Katz et Fodor, mais nous en avons déjà parlé.

Revenons à nos moutons taxémiques : il n'est que trop clair qu'un découpage rigide en classes sémantiques comme l'établit la notion de taxème ne peut suffire. D'une part cela supposerait la prédominance d'une norme (quelle qu'elle soit), assortie de l'hypothèse qu'une norme locale ne peut remettre en cause les pertinences sémantiques d'une norme plus générale. En fait, le taxème ne fait qu'établir la structure d'une norme, à laquelle on donne une simple priorité, et non une dominance. La réflexion que nous proposons s'émancipe à notre avis de toute prise de décision quant à la véritable norme traduite par un taxème (dialecte ou plus local).

Dans la sémantique interprétative, les taxèmes dépendent du système fonctionnel, mais ne doivent pas empêcher le repérage de sèmes motivés socioculturellement par exemple. Cette possibilité recouvre toutes les notions de connotation, de symbolisme et autres topoï. L'attribution de la violence à la couleur rouge, attribuable sans hésitation à une norme sociale fortement stabilisée, n'est cependant pas admissible en langue, où le rouge n'est qu'une couleur, comparable au vert, jaune, bleu, et la violence une notion abstraite comparable au calme, à la paix, à l'espoir par exemple. Comment donc traduire la proximité sémantique de deux sémèmes appartenant à des taxèmes différents (et que l'on peut par anticipation supposer très distants, ne se trouvant pas eux-mêmes dans une même classe supérieure) ? Il suffit simplement d'assumer que des normes différentes, sans pour autant instaurer un nouvel arrangement parallèle des classes sémantiques, établissent des liens hors de ces classes. C'est donc ce qu'on appelle dans la sémantique interprétative les sèmes *afférents*, à la différence des sèmes tels que nous les avons présentés jusqu'ici, qui instauraient des relations entre des éléments d'un même taxème, appelés sèmes *inhérents*. La distinction est donc la suivante (p. 46) : «Un sème inhérent est une relation entre sémèmes au sein d'un même taxème, alors qu'un sème afférent est une relation d'un sémème avec un autre sémème qui n'appartient pas à son ensemble strict de définition : c'est donc une relation d'un ensemble de sémèmes vers un autre». Il s'agit bien sûr d'une caractérisation formelle, et non pas causale, ni même une définition.

L'auteur met ensuite l'accent sur le caractère non-symétrique de la relation d'afférence. Les sèmes inhérents, tels qu'il les a définis, sont en effet des relations symétriques (identité ou incompatibilité), alors que pour un sème afférent il va falloir distinguer le sémème-but du sémème-source. Le premier est le sémème «connoté» par le sémème source.

Par exemple, pour exprimer le symbolisme violent de la couleur rouge, la relation d'afférence ira de 'rouge' vers 'violence', traversant donc les taxèmes. Au résultat, nous aurons un nouveau sème dans le sémème «rouge», qui n'aura d'autre forme que celle du sémème-but lui-même, à savoir /violence/. Cette translation de forme est bien entendu autorisée par l'absence de contraintes liées à l'expression des sèmes.

Avant de nous intéresser aux problèmes de compatibilité entre cette nouvelle dichotomie et celle exprimée précédemment (générique / spécifique), revenons une fois de plus sur le problème de la norme. Théoriquement, l'afférence provient de normes plus locales que celles établissant les taxèmes. Peut-il en être autrement ? Autrement dit, si l'on considère que les taxèmes dépendent de normes locales, une relation d'afférence dépendant du système fonctionnel de la langue a-t-elle un sens ? Le principe est que ces relations extra-classes n'ont aucun rôle organisateur global, qu'elles ne sont pas suffisantes à une organisation des unités lexicales. En effet, plusieurs normes sociolectales d'origines différentes peuvent être à l'œuvre dans une interprétation, mais elles ne couvrent pas à elles seules toutes les zones du lexique. De plus, l'origine de ces normes est en partie l'existence d'une stabilisation préalable sur lesquelles elles peuvent effectuer des modifications : les langues évoluent, mais elles ne se créent pas.

Au sein d'un texte, cependant, il est important de remarquer l'impact d'une norme sur l'interprétation que l'on en fait, puisque le texte lui-même en est une, et que différentes normes, même si elles aboutissent à des résultats similaires, sont elles-mêmes interprétables. Ainsi, dans notre approche informatique, nous ne pourrons nous abstraire de la reconnaissance des normes qui régissent les interprétations élémentaires effectuées par l'utilisateur à assister.

3.5.1 Actualisation et virtualisation

Ces notions de normes différentes entrant en jeu dans une interprétation peuvent aisément induire des problèmes d'incohérence. Divers niveaux de systématicité peuvent en effet avoir des effets sémantiques incompatibles, qui, sans aller jusqu'aux quiproquos, doivent tout de même être gérés. F. Rastier cite à ce sujet le topos (relation d'afférence socialement normée) qui attribue une notion de faiblesse à la femme. Le sème /faiblesse/ est en effet présent dans de nombreux contextes (classiques, certes), surtout lorsque la femme est comparée à l'homme. Mais dans certains cas cette afférence est tout à fait inacceptable, notamment si le contexte la contredit explicitement, comme dans «Jeanne d'Arc était assurément une forte femme». En fait, dans ce cas, les normes les plus hautes s'effacent devant les plus locales. Que deviennent alors ces sèmes «sous-jacents» ; si l'application locale du topos n'est plus valide, le topos n'en reste pas moins valable. Ces phénomènes se retrouvent à d'autres niveaux, en partant du système fonctionnel. Dans [49], on retrouve l'exemple de 'flic', portant le trait /vulgarité/, mais non-pertinent lorsqu'un ministre de l'intérieur déclare être « le premier flic de France ». Le principe mis en place est donc que les assertions des normes « hautes » peuvent disparaître en contexte. On dit alors que le sème est virtualisé : on reconnaît sa pertinence en langue, mais son inutilité actuelle. À l'inverse, si la pertinence est valable, le trait est conservé, ou *actualisé*. Ces deux opérations complémentaires permettent ainsi une grande autonomie du contexte par rapport à la norme, puisqu'une des thèse de la sémantique interprétative est que tout sème inhérent peut être virtualisé, et qu'aucun sème n'est actualisé en tout contexte.

Le problème de cette grande flexibilité est double : d'une part, elle nécessite pour être mise en œuvre une grande systématicité définitionnelle en langue, mais surtout elle remet en cause la notion même d'identité du sémème. Si un sémème en langue et son utilisation en contexte peuvent varier si fortement de par leurs sèmes, est-ce encore le *même* sémème ? F. Rastier a déjà répondu à cette question en distinguant les emplois, sens et acceptions, qualifiant rigoureusement les rapports entre sémèmes suivant les partages et différences sémiques.

Ainsi, si nous voulons dans notre approche nous abstraire d'un inventaire sémique exhaustif des normes supérieures, avons-nous encore besoin de ces deux notions ? Il semblerait effectivement que certaines opérations purement locales les fassent intervenir. Reprenons l'exemple de Jeanne d'Arc : on justifiera la non-actualisation de /faiblesse/ puisque dans le contexte figure explicitement un sémème, 'forte', contenant un sème incompatible avec /faiblesse/. Ce que nous pouvons discerner derrière cette opération, peut-être de mauvaise foi, est le transfert du sème /force/ de 'forte' vers 'Jeanne d'Arc'. Cette sorte de condensation sémantique, si on la systématise, revient en fait, en se basant sur des considérations syntaxiques, à détruire la structure syntagmatique, ou plus pratiquement à résumer le discours. La justification d'un tel ensemble d'opérations peut se trouver ponctuellement, si par exemple on cherche à définir la structure sémantique de l'«acteur» 'Jeanne d'Arc' : on résumera ainsi, via des sèmes, les attributions effectuées par le texte. Ou bien, si l'on raisonne en résolution de problème, ce transfert peut se contenter du prétexte de blocage d'afférence locale. Il est donc impératif de délimiter les conditions de ces afférences locales, au risque de les voir se répandre au sein de la théorie et de détruire la textualité même. Pire encore, la récurrence des sèmes, pivot de la notion d'isotopie si centrale, serait galvaudé, si les sèmes se propagent ainsi, les isotopies deviendraient légion et trivialité.

Pour des raisons de clarté, nous préférons suspendre ici ce débat, pour n'y revenir que lorsque l'intégralité des notions de la théorie auront été étudiées, surtout cette fameuse isotopie.

3.5.2 Relations entre les différents types de sèmes

Nous allons ici nous contenter d'un simple schéma afin de regrouper les différents types de sèmes que nous avons répertoriés. L'exemple traduit une interprétation reconnue du titre du roman de Stendhal, «Le rouge et le noir», comme métaphore colorée des aspirations carriéristes du héros. La description nécessite donc deux taxèmes, //couleur// et //carrière//, contenant chacun deux sémèmes, entre lesquels se mettent en place des relations de spécificité, généricité, inhérence et afférence (voir figure 2.6, inspirée de [49] p. 54, pour laquelle le signe | représente la relation d'incompatibilité entre sèmes, et = celle d'identité).

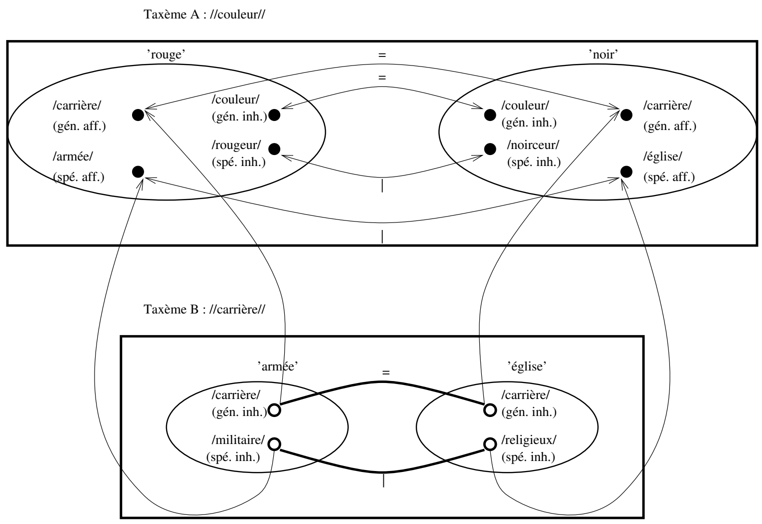


FIG. 2.6 – *Exemple des différents types de sèmes*

Le problème se pose en fait quand on cherche à attribuer un statut spécifique ou générique aux sèmes afférents. Dans l'exemple précédent, /carrière/ est générique afférent pour 'rouge', alors que /armée/ est spécifique afférent. Faut-il en conclure que le statut des sèmes afférents dépend directement du statut que possède ce sème dans ses occurrences inhérentes ? Sans doute, mais alors, y a-t-il seulement une justification pratique de ces statuts rapportés ? Dans le cas cité, /carrière/ apparaît effectivement dans les deux sémèmes du taxème des couleurs, mais si l'on étend aisément ce dernier, en y rajoutant les sémèmes 'jaune' et 'violet', il est peu probable que l'on repère une afférence carriériste sur ces deux notions. Donc, le sème afférent ne joue guère de rôle générique *stricto sensu* dans ce taxème, puisqu'il n'a pas de couverture suffisante pour cela. Cette constatation, ajoutée au fait que la notion d'afférence ne permet pas toujours d'expliciter clairement son origine, nous conduira dans notre formalisation à différencier simplement les sèmes spécifiques et génériques (inhérents par défaut), des sèmes afférents.

De plus, puisque nous avons vu que la forme du sème afférent était en fait, du moins dans le cas des afférences de forte stabilité, celle de la lexie du sémème-source, nous pourrons dans ces cas précis nous passer de l'expression directe de la relation d'afférence elle-même. Exprimer la métaphore carriériste du titre de Stendhal peut se faire par la simple expression des trois sèmes afférents, sans avoir à exprimer directement le taxème qui les possède comme inhérents.

Enfin, nous rappellerons ici une autre origine des sèmes afférents, de teneur purement locale. Les cas concernés sont entre autres les énumérations ou les coordinations syntaxiques. Si nous reprenons le taxème des gourmandises cité précédemment, il est utilisé dans le cadre d'une énumération : «...du café, du chocolat, du sucre, du poivre en grains, du sel fin, de la confiture, un sac de cassonnade, cinq morues sèches, de l'estoquefiche». La notion de gourmandise couvre à un niveau dialectal une sous-partie seulement des sémèmes énumérés. Cependant, l'unité syntaxique qu'est l'énumération semble induire une unité sémantique traduisible par des sèmes communs (ce que nous ne pourrons bientôt plus nous priver de nommer une isotopie), d'où la solution pratique d'attribuer le sème /gourmandise/ à 'morue', avec le statut de sème afférent. L'interprétant de cette afférence est donc syntaxique, et la relation va du taxème repéré via l'énumération vers un sémème de cette même énumération. C'est donc une afférence relevant de ce que nous pouvons appeler une norme locale, ou textuelle. L'expression formelle de la relation est donc la suivante : d'un taxème vers un sémème. L'expression du sème afférent est donc le sème microgénérique initial, et non un sémème. Ainsi, la multiplicité des formes *initiales* des sèmes afférents nous conduira à ne pas rechercher une trop grande systématicité dans l'expression des relations d'afférence.

3.5.3 Pour en finir avec l'afférence

Au bout du compte, résumons les propriétés du concept de l'afférence :

- Elle correspond à une norme plus locale, du moins plus locale que celle du taxème

- Elle n'a donc aucun rôle organisationnel (elle n'est pas «nécessaire»)

- Elle permet d'établir sur cette base des rapports entre des sémèmes qui auraient été séparés par des normes plus hautes

- Son statut importe surtout pour gérer des processus d'actualisation concurrentes de sèmes.

– Si elle se résume comme l'attribution d'un sème à un sémème, ce sème provient d'une attribution plus stable, dans une autre zone de la langue, où il peut prendre des statuts variés, mais inhérents.

Discutons de ces cinq assertions.

Les trois premières sont d'une applicabilité croissante. Si la notion de norme a été, à regret, déclarée insoluble directement, elle se ramène ici à un figement, une stabilisation, qui peut tout de même être transgressée par cette notion d'afférence.

La quatrième, elle, pose un problème applicatif et théorique plus profond : la sémantique interprétative n'a pas pour vocation d'expliciter tous les mécanismes et les raisons d'une attribution de sème. Elle en donne des indices et des moyens d'expression, mais elle ne propose pas une explication de la genèse du système de la langue : pour cette fois la circularité méta-linguistique constitue une barrière. Ainsi, les aspects d'actualisation et de virtualisation, comme phénomènes dynamiques, se rapprocheraient, s'ils étaient systématisés, de considérations psycho-linguistique. De notre point de vue plus applicatif, les processus concurrents menant à l'actualisation / virtualisation seraient hors de propos. Si une automatisation de ceux-ci est proposée localement dans [52], elle s'éloigne à notre goût de l'esprit interprétatif, par l'utilisation d'une forme de règle d'inférence sémantique. Nous nous limiterons ici à une description de l'organisation du *résultat* d'une interprétation. Nous verrons que la rigueur descriptive de l'outil informatique introduit tout de même des aspects dynamiques en amenant l'utilisateur vers des questions qu'il ne se serait peut-être pas posées.

Que va-t-il donc nous rester ? Peu de choses, si ce n'est cette propriété structurelle de l'afférence de s'affranchir d'un découpage rigide, que ce dernier relève d'une norme forte ou d'une volonté formalisante. Pour nous, l'afférence sera donc un moyen d'exprimer des relations sémantiques entre des sémèmes éloignés par une classification unique.

En d'autres mots, il s'agira d'une « soupape conceptuelle » face à un ensemble de contraintes organisationnelles. Nous avons effectivement jusqu'ici dégagé des principes formels qui, comme toute forme de structuration, sont exprimables comme contraintes. La liberté qu'elles limitent est ici d'associer *ad libidum* des sèmes à des sémèmes, au risque d'énoncer ainsi une chose et son contraire. Par exemple, la notion de taxème, en plus d'être un outil d'attribution de sème, sert également à restreindre les sèmes spécifiques. Toute relation sémantique identifiée par l'interprète et non-justifiable au sens de notre formalisation sera donc traduite comme afférente.

La dernière assertion nous permettre d'ajouter une contrainte de justification externe : si un sème est utilisé comme afférent, il devra également être explicité par une voie classique, c'est-à-dire inhérente. L'interprète devra par exemple expliciter le contenu et l'organisation d'un taxème qui possède le sème ici afférent comme sème générique, ou bien des sémèmes qui s'opposent par ce sème.

Ainsi, vouloir outrepasser une contrainte sera une source d'enrichissement thématique pour la structure sémantique produite durant l'interprétation. Ce principe est en quelque sorte un contrepoids à la souplesse précédente : si toute relation peut être déclarée par l'interprète comme afférente, elle se fera au prix d'une explicitation « classique », au sens du respect des contraintes précédemment citées, d'une zone de la langue où ce sème est utilisé comme inhérent. Cette opération prendra donc son appui en-dehors du texte, utilisant s'il la faut des sémèmes qui ne sont pas présents dans l'objet de l'analyse, d'où un enrichissement de la structure sémantique descriptive.

Ce dernier point traduit également le fait qu'une afférence est le but d'une relation, qu'il s'agit d'un glissement de sens qui provient bien d'une autre zone sémantique.

Dès lors, nous ne revendiquerons pas une adéquation entre notre concept d'afférence et celui de F. Rastier. Nous préserverons tout de même l'appellation, par optimisme et par intuition que certains effets de glissement de sens sont justement apparus sous la condition d'une contrainte. De plus, la modélisation de l'afférence dépasse à notre avis l'ordre informatique, puisqu'elle traduit des mécanismes d'une finesse et d'une complexité relevant d'une sensibilité de la langue.

Enfin, le fait de définir différemment la notion d'afférence induit également la possibilité de glissement typologique des sèmes que nous formaliserons. Ainsi, un sème pertinemment afférent (selon F. Rastier) pourra très bien être représenté par une relation spécifique, pour peu que l'interprète qui en est à l'origine lui attribue bien un rôle organisationnel.

3.6 Domaines : vers la référence

Nous avons précédemment laissé entendre l'existence de classes sémantiques supérieures au taxème. Nous allons donc rappeler ce que l'approche interprétative propose comme classe immédiatement supérieure : le *domaine*. «Le domaine est un groupe de taxèmes, tel que dans un domaine il n'existe pas de polysémie. ... La composition et l'inventaire des domaines sont liés à des normes sociales ...» (pp. 49-50). Il appellera également *mésogénérique* le sème correspondant à un domaine, et qui est donc attribué à tous les sémèmes contenus dans ce premier.

Il s'agit donc avant tout d'un *groupe* de taxèmes, donc une classe qui respecte le découpage primordial. Un taxème peut-il appartenir à plusieurs domaines? Cette question n'est pas directement envisagée dans la théorie, mais on peut trouver quelques indices ([49] p. 59) : «En somme, dans le cas où un sémème comporte plusieurs sèmes génériques, son sème microgénérique implique *son* sème mésogénérique ...» (souligné par nos soins). Comme de plus on insistera sur la possibilité qu'auront les représentants de la troisième catégorie de classes sémantiques de rompre les frontières établies par les taxèmes (et les domaines), il apparaît clairement que, tout comme les taxèmes établissaient une partition des sémèmes, les domaines établiront une partition des taxèmes (à moins que l'on n'autorise des taxèmes sans domaines).

Revenons donc au problème de la norme : les domaines dépendent de normes sociales, et leur identification se situe donc au palier sociolectal. Il est intéressant de montrer que, si les taxèmes dépendent du système fonctionnel de la langue, les afférences (sociales ou plus locales) donnaient déjà une possibilité de s'affranchir de leur découpage. Les domaines, qui organisent les taxèmes, dépendent également de normes sociales, nous ne pouvons donc nous contenter de considérer l'impact des différentes normes comme régi par une hiérarchie qui irait naïvement du plus général au plus particulier, chaque norme de niveau *n* ne pouvant que spécifier les organisations sémantiques d'un niveau *n+1*. Tout ceci ne nous éclaircit décidément pas sur ces notions envahissantes... Il semble cependant que la stabilité des normes liées aux domaines soit très forte, puisque l'auteur indique que certains domaines sont attestés dans les dictionnaires, comme //alimentation//, //militaire//, etc. Il est donc envisageable de construire *a priori* un ensemble stable de sèmes mésogénériques, à des fins d'automatisation de leur repérage, mais la question qui se pose alors est celle de la pertinence dans une analyse. Une simple suggestion d'utilisation de ces sèmes est plus acceptable.

Le principe de la non-polysémie à l'intérieur d'un domaine semble plus un constat qu'une assertion identitaire. Les cas cités sont classiques : 'chinois', placé dans le domaine de l'alimentation, n'offre plus la confusion avec une nationalité asiatique.

Par contre, un aspect intéressant des domaines, que nous retrouverons dans l'isotopie, est la notion d'*impression référentielle*. La stabilité socioculturelle du découpage de l'univers référentiel par ces classes sémantiques induit des effets de référence. Ces effets sont atteints, via la répétition le long de la chaîne syntagmatique, de sémèmes relevant d'un même domaine, comme le cas de la phrase «L'amiral fit carguer la voile». L'établissement du domaine //navigation// se fait aisément par la spécificité des trois termes principaux, et permet aisément de traduire une réalité extra-linguistique. On retrouve en quelque sorte le principe de Sapir-Whorf sur le découpage du réel par la réalité linguistique. Ce principe attribuable à toute notion de classe sémantique est d'autant plus fort que la norme sociolectale est prépondérante.

3.7 Dimensions : une porte vers les effets métaphoriques

La sémantique interprétative envisage également un troisième type de classe, d'un statut formel différent des deux premières. La *dimension*, à laquelle est associé un sème *macrogénérique*. Elle est définie ainsi ([49], p. 50) : «Une dimension est une classe de généralité supérieure. (...) À la différence des taxèmes ou des domaines, des dimensions peuvent être articulées entre elles par des relations de disjonction exclusive (cf. //animé// vs //inanimé//)», et plus loin : «... la dimension, classe de généralité supérieure, en intersection avec tous les domaines, et incluant certains taxèmes». Il apparaît ainsi que la dimension n'a pas le rôle organisateur des deux premiers types de classes, qui étaient compatibles. Des notions comme //animé// apparaissent effectivement assez universelles pour les retrouver dans tous les domaines, et dans la plupart des taxèmes.

De plus, si l'on considère le découpage que font ces dichotomies dimensionnelles des taxèmes, ces sèmes macrogénériques rentrent-ils en conflit avec les sèmes spécifiques ? On retrouve bien dans la disjonction exclusive qui articule //animé// et //inanimé// la notion d'incompatibilité propre aux sèmes spécifiques. Les sèmes macrogénériques seraient alors des «sèmes spécifiques de grande généralité». Quel peut être leur rôle explicatif dans une interprétation ? Justement leur statut indépendant des classes : tout comme les sèmes afférents, ils permettent, au niveau du système fonctionnel de la langue, de rapprocher des éléments de classes incomparables. Deux sémèmes aussi éloignés que 'Grille-pain' et 'Église' sont au moins comparables via cette dimension de l'inanimation. Certes, cette identification n'a rien que de très banal, dans ce cas précis, mais de telles trivialités sont en fait d'excellents indices d'établissement de connections plus fines et pertinentes. Elles constituent en quelque sorte une «degré zéro» de l'interprétation, une sorte de garantie d'isotopie dans toute analyse. Leur automatisation est également envisageable techniquement, mais se pose le même problème de pertinence que pour les domaines.

Ce qui nous ramène à un aspect intéressant des sèmes spécifiques, à savoir leur potentiel à effectuer des relations inter-classes, qualifiables de métaphoriques. Si un même sème est utilisé de façon spécifique dans deux taxèmes distincts, les deux sémèmes qui le supportent entrent en relation d'équivalence par l'identité de ce sème. C'est le cas des schémas métaphoriques «classiques», comme par exemple l'opposition exprimée par /bâtiment/ entre 'église' et 'prêtre' dans un taxème religieux et entre 'mairie' et 'maire' dans l'administration locale. Au sein d'un parcours interprétatif complexe, en cas de «tâtonnement», la trace de telles relations peut très bien être une articulation dimensionnelle du type /animé/ - /inanimé/. Nous proposons donc de voir dans les dimensions de tels guides pour repérer des relations transversales plus pertinentes entre des classes, et nous la cantonnerons d'ailleurs à ce rôle. Nous ne définirons donc pas formellement de classe supérieure au domaine, puisque, comme nous l'expliciterons bientôt, une dimension sera traitée comme une isotopie spécifique.

Mais pour éclaircir ces phénomènes, et rendre à la théorie de F. Rastier son intégrité remarquable, nous allons aborder dans la suite le concept central et incontournable qu'est l'isotopie.

4 Globaliser l'interprétation : les isotopies et leurs présomptions

4.1 Deux façons de voir l'isotopie

Revenons sur les objectifs que nous avions fixés à l'analyse sémique : nous avions cité trois principes, la justification, la relativisation et la cohérence. En exprimant les différentes typologies de sèmes proposées, nous avons partiellement rempli l'objectif de justification, puisque tout sème identifié au cours d'une interprétation possède une validité comme expression de relations entre signifiés. Le principe de relativisation est très général, mais il concernait entre autres la contextualité des taxèmes. Quoiqu'il en soit, nous n'avons pas encore parlé de la cohérence, ni même exprimé l'utilité directe de tous ces sèmes. Il est grand temps de se préoccuper de la place des sèmes dans un texte, et non plus seulement dans la langue ou dans toute autre abstraction paradigmatique.

Nous allons donc présenter et analyser une notion si fondamentale qu'elle apparaissait en filigrane dans les discussions précédentes, qui ne seront d'ailleurs entièrement validées que par elle. Il s'agit de la notion d'*isotopie*. Il existe bien des façons d'aborder cette notion, mais puisque nous avons pris le parti de l'exposer après les notions de sème, taxème etc., nous la définirons tout d'abord à partir de ces dernières.

4.1.1 L'isotopie comme constat

L'isotopie dans sa considération statique n'est donc rien d'autre que la récurrence d'un *même* sème dans des sémèmes entretenant des relations syntagmatiques. Ceci exprime simplement le fait que les signifiants alignés dans un texte ont des sens «proches», d'une proximité de caractère purement sémantique. Ainsi la simple phrase «Mange ton gâteau avec ta fourchette !» est le siège d'au moins une isotopie, puisque le sème /alimentation/ peut être repéré dans les descriptions de 'mange', 'gâteau' et 'fourchette'. Ainsi, cette simple notion de répétition permet d'ores et déjà de valider le sème récurrent comme description d'une «thématique» du discours. Cette description n'est bien entendu pas exhaustive, mais n'en est pas moins valide. Si l'on considère les différents sémèmes de cette phrase analysés indépendamment, la reconnaissance d'un sème commun établit une sorte de cohérence du discours, et de l'analyse qui en a été faite. L'isotopie constitue ainsi une sorte de «condensation sémantique» du texte analysé. Une isotopie est constatée dès lors que deux sémèmes d'un texte possèdent un sème en commun. De plus, le sème récurrent n'a pas besoin d'avoir le même statut pour toutes ses occurrences, que celui-ci soit générique, spécifique, inhérent ou afférent. Nous verrons d'ailleurs plus loin comment établir sur cette base une caractérisation des isotopies. Les sémèmes qui supportent une isotopie ne sont pas nécessairement aussi proches sur le plan syntaxique que dans notre petit exemple précédent : F. Rastier propose ainsi une isotopie du sème /intensité/ tout au long de *L'assommoir* ([49], p.115).

Au vu de toutes ces possibilités, il apparaît donc qu'une isotopie est chose fort courante si on la considère *a posteriori*, et qu'en tout cas elle ne fait que résumer le résultat de l'interprétation : c'est en effet en chasser tout aspect constructif.

4.1.2 L'isotopie comme processus : de la présomption à la validation

Ce que nous voulons exprimer ici est qu'une isotopie est également, et surtout, un moyen efficace de guider l'interprétation d'un texte, et un outil formel pour capter les objectifs et présuppositions interprétatifs de l'interprète. Interpréter peut, à un certain niveau déjà suffisant dans une approche informatisée, être vu comme la construction et l'explicitation d'isotopies. Comme le dit F. Rastier : «En général, on considère l'isotopie comme une forme remarquable de combinatoire sémique, un effet de la combinaison des sèmes. Ici au contraire, où l'on procède paradoxalement à partir du texte pour aller vers ses éléments, l'isotopie apparaît comme un principe régulateur fondamental. Ce n'est pas la récurrence de sèmes déjà donnés qui constitue l'isotopie, mais à l'inverse la présomption d'isotopie qui permet d'actualiser des sèmes, voire *les* sèmes.»[49](p. 12).

Encore plus loin, la volonté de mettre en place des isotopies lors de l'interprétation d'un texte permet de réduire la multiplicité des identifications de sèmes. Le fait d'avoir déjà utilisé /couvert/ comme sème pour 'fourchette' peut guider l'utilisateur à interpréter 'assiette' comme ustensile alimentaire et non comme synonyme d'équilibre. C'est là un grand principe, qu'en interprétant un texte on s'attend à sa cohérence, donc à pouvoir y repérer des isotopies ; c'est la fameuse «présomption d'isotopie» de F. Rastier ([49] p 12). Ainsi, en abordant une recette de cuisine, il est clair que le domaine de l'//alimentation// y sera utilisé, et que l'isotopie correspondant à ce sème sera extensionnellement importante dans un tel texte. Plus que cela, tous les repérages de sèmes, même de ceux qui seront «distants» du culinaire, seront tout de même influencés par cette présomption. La présomption d'isotopie peut donc se contenter d'être formellement pauvre, mais se concrétisera en une véritable et explicite isotopie le long du processus interprétatif.

Et c'est ainsi que déjà se profile la zone d'assistance informatique que nous proposerons par la suite : elle concernera la mise en forme et la construction enrichissante de données brutes.

4.2 L'isotopie approfondie

Revenons donc plus attentivement sur ce qui se cache réellement derrière cette notion. Elle fut tout d'abord proposée par A.J. Greimas, puis reprise tous azimuts dans la communauté linguistique ; il est vrai que la simplicité de son énonciation permet d'y placer un grand nombre de notions, même de non sémantiques. C'est en fait à la récurrence que l'on peut s'attacher facilement. Pour A.J. Greimas, l'isotopie consistait principalement en la répétition du même classème (au sens particulier que nous avons évoqué plus haut). F. Rastier a donc largement étendu cette notion, puisqu'une isotopie, selon lui, peut s'exprimer par la récurrence de n'importe quel type de sème. Cette proposition donne en fait une véritable souplesse au rapport entre les ordres syntagmatique et paradigmatique. En effet, le statut des sèmes dépend de l'organisation des signifiés en classes sémantiques, donc de l'interprétation qui les repère ; et il en est de même de leur forme. Ainsi, cette notion de récurrence est une abstraction par rapport à ces considérations, et tend à unifier une interprétation *en soi*. Mais nous verrons plus tard que la nature des sèmes récurrents, en ce qu'elle traduit un niveau de systématicité, peut également avoir des conséquences non négligeables sur le rôle de ces isotopies.

4.2.1 Syntagmatique et paradigmatique

Revenons donc sur cette dualité de l'isotopie, à laquelle F. Rastier s'intéresse d'ailleurs, pour y accepter la récupération d'un principe de Jakobson [25], celui de la «projection du principe d'équivalence de l'axe paradigmatique sur l'axe syntagmatique». Ceci est parfaitement visible en ce qui concerne les isotopies dont le sème est générique dans ses occurrences. Par exemple, «Mange ton gâteau avec ta fourchette» est le siège d'une isotopie *mésogénérique*, les trois sémèmes qui la supportent étant des éléments du même domaine. Réunis dans la structure sémantique, il entretiennent des relations du même ordre en étant réunis dans un même discours. On passe bien de la langue au discours, en ne préservant de la structuration complexe de la première qu'une partie, mais en la confrontant à la réalité du second, où elle sera confrontée à d'autres formes de complexité.

Mais nous verrons que ces isotopies n'ont qu'un intérêt limité dans le cadre de l'analyse d'un texte à vocation littéraire, du moins si l'on considère ces classes sémantiques comme stabilisées à un haut niveau. Nombre d'effets stylistiques se distinguent justement par leur transversalité par rapport à ces classes «typiques».

4.2.2 Isotopie et niveaux sémantiques

La notion d'isotopie a un rôle d'unification sémantique que nous nous devons de mettre en valeur à ce stade. Face à nos préoccupations «globalistes», nous avons en quelque sorte fait des concessions opérationnelles en choisissant une approche par traits sémantiques. Mais c'est l'isotopie qui nous sauve, par cette remontée vers des phénomènes macrosémantiques, grâce à leur statut de structures textuelles.

Les travaux plus récents de F. Rastier se sont portés justement vers des notions plus globales, en grande partie basées sur cette notion d'isotopie qui devient alors le matériau de base d'un autre niveau d'analyse des textes. Même si les méthodes et les préoccupations sont identiques, l'appareil conceptuel mis en place dans ces approches nous a semblé à ce stade insuffisant pour supporter une informatisation même partielle. Nous nous concentrerons donc dans nos travaux sur cette notion d'isotopie, en espérant que les assertions formelles que nous ferons à son encontre lui accorderont une souplesse suffisante pour être utilisée avantageusement dans le prochain épisode de cette aventure. Nous donnerons tout de même plus loin quelques propositions sur quelques aspects de la macrosémantique qui peuvent être atteints via une légère refonte de l'isotopie.

4.2.3 Caractérisation d'une isotopie

Intéressons-nous avant cela aux différents types d'isotopie, et aux propriétés qui les caractérisent. Nous disposons, sur la base du statut des sèmes, de deux dichotomies permettant de caractériser plusieurs types d'isotopies :

- **Isotopies génériques** : il y en a trois, types, autant que de sèmes génériques.
  - Les isotopies microgénériques traduisent une cohésion très forte du discours qui les supportent. Par exemple : «Le couteau se place à droite, et la fourchette à gauche de l'assiette» s'inscrit fortement dans ce taxème //couverts// qui nous accompagne depuis le début. Dans le cas d'une définition en langue du taxème, une telle isotopie traduit des rapports extrêmement précis entre ses sémèmes, et une grande spécificité du texte. Dans le cas d'une justification contextuelle du taxème, une telle isotopie est en fait *fondatrice* d'un taxème local : parmi l'ensemble des possibilités du discours, son repérage indiquera l'identité du taxème.
  - Les isotopies mésogénériques apportent ce que F. Rastier appelle une impression référentielle, par la confirmation qu'elles opèrent du respect d'un univers et de pratiques sociales. Par exemple «L'amiral fit carguer la voile» produit un tel effet. Cette distinction d'avec d'éventuels effets référentiels produits par une isotopie microgénérique nous échappe cependant.
  - Les isotopies macrogénériques peuvent par contre s'abstraire à un certain point du respect des classes sémantiques classiques. Par exemple «Le chat du commissaire aime regarder les oiseaux» est le siège d'un telle isotopie via le sème /animé/, même si la pertinence d'une telle récurrence n'est guère évidente (si ce n'est par son extension aux sémèmes 'aime' et 'regarder', qui serait sans doute laissée pour compte dans une approche IA classique). Bien des isotopies macrogénériques sont dans ce cas. Par contre, pour aborder des phrases comme «Les avions à réaction sont les grille-pains des anges», une telle isotopie du sème /concret/ ou /matériel/, une des plus aisées à mettre en place face à un énoncé aussi disparate, peut être un premier pas organisateur.

- **Isotopies spécifiques** : elles présentent l'intérêt majeur de présenter des liens sémantiques entre sémèmes de classes différentes. Elles permettent ainsi, face à des classes sémantiques fortement stabilisées de traduire des effets poétiques et/ou métaphoriques. F. Rastier cite l'exemple du vers «L'aube allume la source» de Éluard comme siège de la récurrence du sème /inchoativité/, dont le rôle spécifique apparaît dans au moins trois taxèmes différents. Cet aspect de l'isotopie donne également un rôle organisateur transversal aux sèmes spécifiques, qui est d'ailleurs quelque peu gênant conceptuellement, et sur lequel nous reviendrons par la suite. De telles isotopies ne peuvent être établies qu'une fois les classes sémantiques mises en place, avec l'expression de leur organisation interne.

- **Isotopies mixtes** (selon la généricité): ce sont les isotopies qui ne sont ni purement spécifiques ni purement génériques. Il faut noter que la généricité est obligatoirement homogène, même dans une isotopie mixte : un même sème ne peut servir à dénommer un taxème et un domaine. Par contre, un sème microgénérique peut avoir un rôle spécifique dans un autre taxème que celui qui lui est propre. Par exemple, 'Dimanche', dans le taxème des //jours de la semaine// peut être distingué des autres par une notion de /religion chrétienne/, cette même notion étant tout à fait acceptable pour dénommer un taxème contenant 'église', 'pape' et autres 'carême'. Le repérage d'une telle isotopie se fait *a priori* à partir de sa partie générique et par extension, sans pour autant faire appartenir les autres sémèmes qui la supportent à la classe originelle.

- **Isotopies inhérentes** : c'est principalement de celles-ci que nous avons parlé précédemment. Elles correspondent à des interprétations «classiques» respectant des normes stables.

- **Isotopies afférentes** : le problème de l'afférence n'étant toujours pas résolu, nous nous contenterons de reprendre les remarques de F. Rastier à leur propos. Une isotopie afférente relève d'une interprétation par des normes plus locales et relatives. L'isotopie de /carrière/ dans *Le rouge et le noir* est ainsi typiquement afférente. De telles isotopies peuvent également résulter de ces interprétations qui calquent sur un texte *a priori* décorrélé des thèmes psychanalytiques et/ou religieux. Le repérage d'une telle isotopie ne peut se faire qu'*après* la mise en place d'isotopies inhérentes dont les schémas structurels peuvent inspirer une translation thématique.

- **Isotopies mixtes** (selon l'inhérence) : encore une fois la souplesse est de mise, et certaines isotopies peuvent posséder des parties inhérentes et afférentes. Il s'agirait, comme dans le cas précédent, de l'extension directe d'une isotopie inhérente vers d'autres zones d'un texte, toujours par un principe de présomption d'isotopie, présomption qui alors porte surtout sur la taille extensionnelle de l'isotopie, et de sa couverture syntagmatique.

D'autres critères, syntagmatiques ceux-là, peuvent également intervenir dans la caractérisation d'une isotopie. En jouant sur la dualité de la notion, une isotopie peut être vue comme une classe de repérages de signifiants dans la chaîne ordonnée du discours. Dès lors, nombre de critères numériques interviennent, dont nous ne pouvons nous empêcher de nous réjouir, au vu de nos préoccupations informatiques. Citons-les sans ordre précis, puisque nous les évaluerons en détail une fois la notion d'isotopie formellement éclaircie.

- Poids : simplement le nombre de sémèmes couverts par une isotopie, soit un nombre brut, soit un taux de couverture de l'ensemble du texte.

- Volume : étendue du texte couverte entre le premier et le dernier (dans l'ordre de la chaîne du texte) sémème de l'isotopie.

- Densité : comme le nom l'indique, le rapport poids / volume, qui traduit simplement la force d'une isotopie dans sa zone. Une énumération thématique, par exemple, aura une très forte densité.

- Taux de systématicité : de simples caractérisations numériques des taux de sèmes d'un certain type dans une isotopie mixte (aux deux sens de la mixité). Ces taux dépendent bien entendu plus de l'ordre paradigmatique que les précédentes notions.

Ces aspects calculatoires, quoique simples, sont une porte ouverte par la seule informatique vers d'autres aspects, cette fois non modélisables, du même ordre que ceux proposés par les analyses statistiques de corpus.

4.2.4 Caractérisation de plusieurs isotopies

Mais les plus importantes caractérisations d'un texte à partir de ses isotopies proviennent sans doute des relations instituables *entre* ces isotopies. Nous avons vu que la souplesse de la notion permet le repérage d'un grand nombre de récurrences sémiques, la plupart étant liées par les rapports entre sèmes. Mais il existe également des relations quantifiables sur les extensions de ces structures (les schémas correspondants sont présentés dans la figure 2.7) :

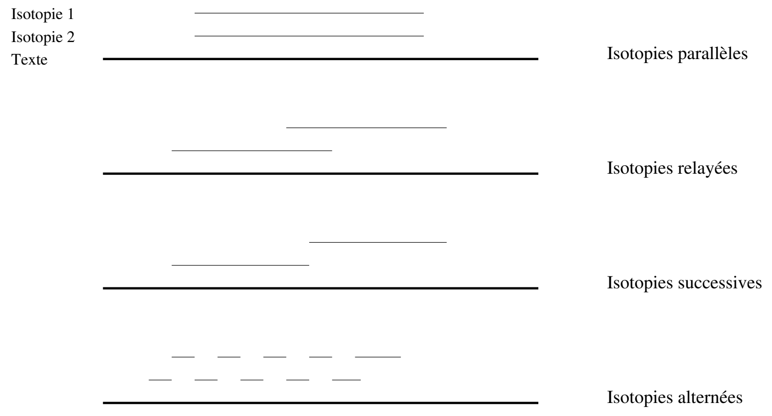


FIG. 2.7 – *Schémas de corrélations d'isotopies*

- Isotopies parallèles : il s'agit d'isotopies possédant des extensions identiques ; plusieurs cas sont alors envisageables. Il peut s'agir d'isotopies génériques superposées, par exemple une isotopie microgénérique «couverte» par l'isotopie mésogénérique associée, par exemple celle du taxème //christianisme// et du domaine //religion//. Il est clair que dans ce cas la plus «pertinente» est la microgénérique, puisqu'à extensions égales elle est plus *focalisée*. Mais il peut s'agir également d'une isotopie inhérente et d'une isotopie afférente, comme /couleur/ et /carrière/ dans «Le rouge et le noir». Ce parallélisme notionnel traduit assurément le transfert thématique dû à l'interprétation, depuis l'isotopie inhérente vers l'afférente.

- Isotopies relayées ou successives : dans le cas où une isotopie s'arrête là où une seconde commence. Deux cas sont alors envisageables, suivant si une partie du texte supporte effectivement les deux isotopies. Ces phénomènes de succession permettent de capter l'évolution thématique d'un discours. Un schéma similaire mais plus complexe peut se mettre en place avec des isotopies spécifiques qui «relient» extensionnellement des isotopies génériques ; ceci entre alors dans le cadre des effets stylistiques sur lesquels nous nous focaliserons plus tard.

- Isotopies alternées : lorsque les deux isotopies démontrent plusieurs phases de succession ou de relais. Deux thèmes sont alors liés dans une zone de texte, sans pour autant concerner les mêmes sémèmes. C'est par exemple le cas des discours comparant deux thèmes, mettant ainsi en place des relations d'un autre ordre entre les différents éléments de ces thèmes. Nous verrons comme exemple de cette alternance un article satirique débattant parallèlement de la politique et de la religion.

4.3 Remaniement de la notion

Nous en avons pour l'instant assez dit sur l'isotopie pour indiquer quelques directions vers notre formalisation du phénomène. Notre idée principale à ce propos s'articule sur les différents statuts de la notion d'isotopie : elle supporte deux dualités. La première concerne les ordres syntagmatique et paradigmatique, pour lesquels l'isotopie joue un rôle articulatoire. La seconde concerne le résultat et le processus de l'interprétation. Il convient donc de tendre vers une représentation de l'isotopie qui accepte ces quatre rôles.

4.3.1 Isotopie et ordre syntagmatique

C'est la notion la plus simple à couvrir, puisqu'elle concerne directement la dualité de signifiance / signification des sémèmes. Une isotopie se rapporte aux sémèmes en tant que signifiés, mais ceux-ci sont repérables sur la chaîne du discours. Une vision réduite de l'isotopie peut d'ailleurs s'abstraire complètement de toutes les notions de classes et d'oppositions exposées précédemment, et ne consister qu'en une classe d'équivalence d'unités lexicales. Cette réduction est en œuvre dans les outils informatiques manipulant des traits sémantiques si ceux-ci sont conceptuellement égaux. Dans ce cas cependant, *exeunt* les considérations typologiques liées aux sèmes...

4.3.2 Isotopie et ordre paradigmatique

C'est justement le manque à gagner soulevé par la dernière remarque. L'isotopie se doit de différencier les rôles joués par le sème dont elle traduit la récurrence pour les sémèmes qui supportent cette récurrence. Mais surtout, l'isotopie se retrouve en quelque sorte en langue ; c'est elle qui se cache derrière les notions de classes sémantiques, et même derrière les oppositions au sein de ces classes. En nous risquant à une formule trop facile, un taxème n'est qu'une isotopie potentielle. Dès lors que nous nous intéressons plus au discours qu'à la langue, nos taxèmes ne seront utiles que si leur extension est réellement, du moins en partie, présente dans le texte analysé. Les deux notions sont alors «assimilées», et nous verrons comment nous nous servirons de l'isotopie comme d'un principe paradigmatique touchant non seulement aux sèmes et aux sémèmes, mais aussi aux classes et aux oppositions. L'entité organisatrice du «lexique» sera alors *la même* que celle qui structure le discours. Cette économie formelle que nous nous autorisons se projettera bien entendu sur le respect des contraintes de l'outil informatique.

Concrètement, nous regrouperons la notion d'isotopie et d'attribution de sème : jouant sur la polysémie du mot, nous laisserons à la machine la précision à apporter , par un jeu de projection d'une structure sur un énoncé. Cette unique notion sera également une facilité implémentatoire pour traiter les nombreux allers-retours de la structure sémantique au texte au cours du processus de l'interprétation.

4.3.3 Le cas des isotopies spécifiques

La récurrence de sèmes spécifiques dans une isotopie implique plusieurs phénomènes et une explicitation de la nature du sème spécifique. La notion de trait distinctif dans le paradigme structuraliste s'appuie, entre autres, sur cette notion de *commutabilité*. L'exemple classique, rappelé par R. Barthes [5], est celui de la définition de 'jument' par les traits /cheval/ + /femelle/. La validité (ou la *valeur*) de ces traits tient en ce que d'autres termes (d'autres sémèmes en fait) supportent une définition comprenant ces mêmes traits, et que chacun de ces traits peut être remplacé par d'autres d'une même famille, et créer ainsi la définition d'un autre terme. Par exemple, remplacer /femelle/ par /mâle/ ou /cheval/ par /cochon/. Des traits distinctifs sont donc valables s'ils appartiennent à une classe paradigmatique, de la même façon que les sémèmes. Dans notre cas, où les seules classes valables le sont sur le syntagmatique, cette contrainte devient donc la *réutilisabilité* des traits spécifiques. L'identification d'un sème spécifique sera *intéressante* si elle apparaît également dans un autre taxème. Dans la théorie de F. Rastier ceci est le cas des dimensions, puisque les traits correspondants sont assez généraux pour pouvoir atteindre presque toutes les classes. Nous ne pouvons bien entendu pas espérer réutiliser tous les sèmes spécifiques qu'une interprétation fait apparaître, mais nous pouvons juger de la pertinence d'un tel sème par la possibilité de le voir apparaître avec un rôle différent (une autre classe, un autre genre).

4.3.4 Isotopie et processus

Les considérations précédentes prenaient l'isotopie comme un fait accompli, même dans son aspect d'organisation des structures sémantiques. S'y limiter renierait outrageusement à la fois la réalité de l'acte interprétatif et nos intentions théoriques. Une isotopie se construit avant tout, et ne se manipule qu'ensuite. La simple association sème-sémèmes, si elle peut suffire à une description sommaire du texte, ne constitue en fait qu'une première intuition de sa structure : elle ne fait que tracer les grandes lignes le long desquelles l'interprétation proprement dite prendra son ampleur. Nous nous refusons d'ailleurs à nommer ce premier type d'associations une isotopie, et nous l'affublerons par la suite du néologisme de *pré-isotopie*. Lors de la principale phase du processus interprétatif (qui a commencé dès lors que l'on a identifié des sémèmes), cette pré-isotopie va devoir obtenir son statut d'isotopie, donc de *justifier* l'attribution qu'elle fait d'un sème à des sémèmes. La prise en compte du contexte et des intentions de l'interprète comme justification des classes sémantiques se placera à cet endroit dans le processus : nous considérerons que les premières intuitions du lecteur serviront de cadre d'exploration. Ainsi, les pré-isotopies initialement repérées (et c'est seulement pour ces phénomènes que nous parlerons de pré-isotopie), acquerront donc *a priori* un statut d'isotopie générique (du moins à fort taux de généricité). À partir de ces classes initiales, une des premières opérations du processus interprétatif sera la mise en place des oppositions structurant ces taxèmes, et vont donc conduire au repérage de sèmes spécifiques, donc à des isotopies spécifiques. Le processus interprétatif va donc construire la complexe structure de classes et oppositions que nous avons vue précédemment, et par là-même donner naissance à de nouvelles isotopies, dans un cycle créateur que l'on peut supposer infini, ou du moins asymptotique par rapport à un idéal de stabilité interprétative. Il nous faudra tout de même identifier des étapes dans ce processus durant lesquelles des phases d'évaluation (selon les critères récemment répertoriés) auront lieu. Nous pouvons d'ores et déjà résumer les *premières* étapes d'un processus interprétatif dans le schéma 2.8.

Quant au rôle de la machine, il sera celui d'un arbitre, disposant d'un ensemble de règles à faire respecter, évitant des incohérences et des confusions. Elle sera garante de la préservation de l'identité sémantique des constituants. Les contraintes, pour être respectées, entraîneront bien souvent l'utilisateur à « en dire plus », le motivant dans son explicitation, et pouvant ainsi provoquer le repérage de notions sémantiques jusqu'ici ignorées par l'interprète.

5 Conclusion

Dans ce chapitre nous avons tenté d'exprimer l'intérêt extraordinaire que présente la sémantique interprétative dans l'analyse de textes du point

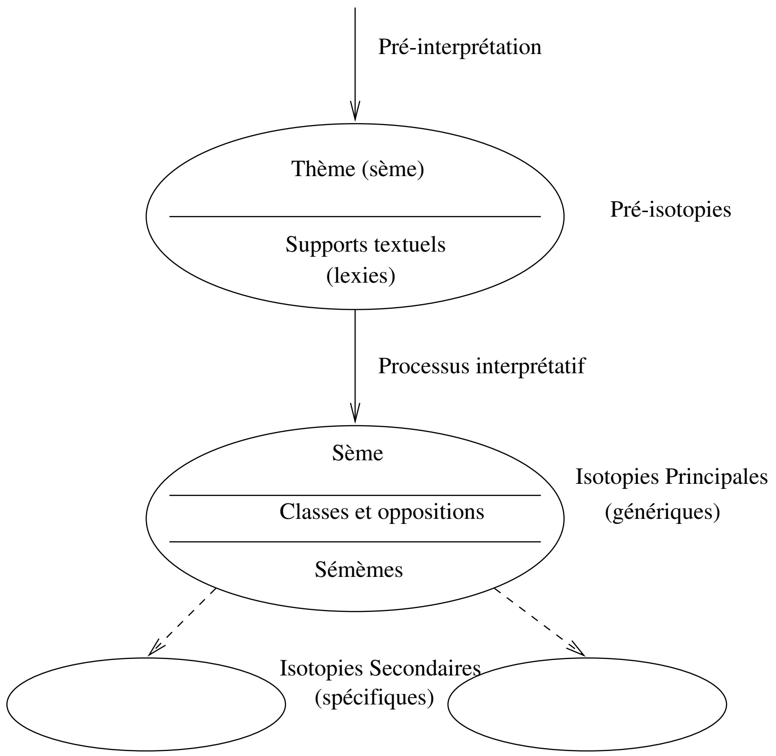


FIG. 2.8 – *Premières étapes du processus interprétatif*

de vue du TALN. Nous avons évoqué, sans toutefois les approfondir suffisamment, les grands principes épistémologiques qu'elle sous-entend, et leur réapparition au fil de la théorie. Si les aspects clairement formalisés par F. Rastier peuvent être tentants pour aborder directement une implantation informatique, nous ne devons pas nous leurrer quant au parcours encore à faire. Et les ombres rôdent toujours autour de notre entreprise ; elles ont pour noms norme, afférence, opposition langue et discours, pour ne citer que les plus visibles. Mais félicitons-nous tout de même des aspects positifs que présente cette théorie : la possibilité de proposer une forme de rationalisation, sinon du sens, du moins de l'interprétation, seul phénomène capable de nous y mener. Nous avons trouvé dans cette approche de la sémantique des moyens élégants pour capter et surtout pour manipuler certaines incidences globales de toute attribution de sens.

Chapitre 3

Description formelle de la structure sémantique

Nous présentons dans ce chapitre la formalisation des concepts retenus lors de l'étude du chapitre précédent, et explicitons les relations que ceux-ci entretiennent. La description que nous proposons ici est celle d'une structure complète et stabilisée. Par contre, nous verrons au prochain chapitre un protocole de mise en place de celle-ci, ainsi que son évolution. Pour cette présentation, l'ordre que nous utiliserons sera, pour des raisons de cohérence, différent de celui du chapitre 2. Nous commençons par préciser les différentes entités, des sémèmes aux isotopies. Certaines exemples dont nous agrémenterons ces considérations formelles sont inspirés d'une analyse du texte « Chirac envisage de se convertir, mais à quoi ? » disponible en annexe.

1 Les sémèmes

Le sémème est l'entité de base de cette axiomatique. Formellement, un sémème est *identifié* par (mais ne se limite pas à) une simple chaîne de caractères, permettant ainsi des relations d'identité et de différenciation. Nous verrons plus loin les lacunes de ces relations, mais nous nous en contenterons pour l'instant.

Définition : S est l'ensemble non vide des sémèmes, ses éléments seront notés s, indicé ou non. L'identité des éléments de S est inhérente à cet ensemble (basée naïvement sur la forme de ces éléments), et notée $=_S$.

La cardinalité de l'ensemble S est susceptible d'évolution au cours du processus interprétatif : elle n'en reste pas moins finie.

Exemple : 'Apôtre', 'Messe', 'La Mecque', 'Colombey-les-deux-églises' sont des sémèmes.

2 Les taxèmes

Les taxèmes sont des ensembles de sémèmes, et leur organisation forme une partition de S. Les taxèmes forment donc une catégorisation des sémèmes permettant par la suite d'opérer des différenciations entre les sémèmes sous la forme d'oppositions entre eux.

Définition : T est l'ensemble des taxèmes, ses éléments seront notés t, indicé ou non. T est un sous-ensemble non-vide strict de l'ensemble des parties de S : $(T \subset \mathcal{P}(S) \setminus \{\emptyset\})$. t est donc un ensemble non vide d'éléments de S, de cardinalité $|t|$ supérieure ou égale à 2.

$$\forall t \in T,\ t \in \mathcal{P}(S),\ |t| \geq 2$$

L'identité de deux taxèmes se base sur les sémèmes qui leur appartiennent : c'est une identité extensionnelle. D'où :

$$\forall t_1, t_2 \in T,\ t_1 \neq_S t_2 \Leftrightarrow (\exists s \in S \text{ t.q. } (s \in t_1 \text{ et } \ s \notin t_2) \text{ ou } (s \in t_2 \text{ et } \ s \notin t_1))$$

Il s'agit donc d'une extension de la relation $=_S$.

Axiome : Tout sémème de S appartient à exactement un taxème :

$$\forall s \in S,\ \exists t_1 \in T \text{ t.q. } s \in t_1 \text{ et } (\forall t_2 \in T \text{ si } s \in t_2 \text{ alors } t_1 =_s t_2)$$

Corollaire : L'intersection de deux taxèmes distincts est vide.

Corollaire : Aucun taxème n'est un sous-ensemble d'un taxème distinct de lui-même.

La cardinalité de T dépend bien entendu de celle de S, mais avec des variations importantes envisageables. Les bornes sont : $1 \leq |T| \leq |S|/2$, en écartant le cas trivial d'un ensemble S possédant moins de deux éléments.

Définition : La relation extensionnelle reliant un taxème à ses éléments sera notée $tax : T \to \mathcal{P}(S)$. La relation inverse tax^{-1} est une fonction de S vers T associant à un sémème le taxème unique qui le contient.

Définition : Cette dernière relation induit naturellement une relation d'équivalence $\equiv_T$ sur S, avec $s_1 \equiv_T s_2 \Leftrightarrow tax^{-1}(s_1) =_S tax^{-1}(s_2)$.

Cette relation exprime, précisément, l'appartenance de deux sémèmes à un même taxème.

Exemple : L'ensemble formé par les sémèmes 'Apôtre', 'Messe', 'Foi' est un taxème. (À condition bien sûr qu'il n'existe pas déjà un taxème qui contienne cet ensemble ou qui soit en intersection non vide avec lui.)

3 Les spécèmes

Rappelons la différence mise en évidence entre les spécèmes et les sèmes spécifiques : les seconds sont supportés par les premiers. Le spécème sera donc un support pour des sèmes, il sert donc à préciser à quel sémème est attribué un sème spécifique, et de quels autres sémèmes du même taxème il sert à le différencier.

Exemple : Dans le taxème //couverts//, /pour prendre/ s'applique à 'fourchette' et 'cuiller' pour les différencier chacun de 'couteau'. Ce sème n'a donc pas de rôle distinctif entre les deux sémèmes auxquels il s'applique.

3.1 Les spécèmes en général

Définition : Un spécème est de la forme (s, s'), avec $s, s' \in S, s \equiv_T s'$ et $s \neq_S s'$.

SP désignera l'ensemble des spécèmes, notés sp (indéxé ou non).

SP est donc un sous-ensemble strict non vide du produit cartésien $S \times S$. Nous écartons comme d'habitude les cas triviaux où S est réduit à moins de deux éléments et où T est vide.

À tout sémème s qui appartient au taxème t correspondent $(|t| - 1)$ spécèmes dont la première composante est s.

Donc, la cardinalité de SP est connue et vaut

$$|SP| = \sum_{1 \leq i \leq |T|} |t_i|(|t_i| - 1) \text{ avec } t_i \in T$$

.

Nous appellerons *sémème opposé* la première projection d'un spécème, et *sémème opposant* la seconde projection.

Définition : Nous noterons ces deux projections σ_1 et σ_2, chacune définie sur $SP \rightarrow S$. Ce sont des fonctions, à l'inverse de leurs réciproques σ_1^{-1} et σ_2^{-1}, définies sur $S \rightarrow \mathcal{P}(SP)$.

Au sein d'un taxème donné, les spécèmes couvrent donc un graphe complet d'oppositions entre les spécèmes. Comme nous le verrons par la suite, il n'est pas nécessaire que chaque spécème corresponde à un sème (ni qu'il se limite à supporter un seul sème, d'ailleurs).

On peut citer quelques relations entre les spécèmes ainsi définis :

- Identité : deux spécèmes seront identiques s'ils ont les mêmes premières et secondes projections. Il s'agit encore ici d'une extension de l'identité $=_S$:

  $$\forall sp_1, sp_2 \in SP,\ sp_1 =_S sp_2 \Leftrightarrow (\sigma_1(sp_1) =_S \sigma_1(sp_2) \text{ et } \sigma_2(sp_1) =_S \sigma_2(sp_2))$$

- Opposition : deux spécèmes seront opposés s'ils sont de la forme $(s,\ s')$ et $(s',\ s)$. On note $sp_1\ opp\ sp_2$. Il est important de noter que cette relation d'opposition n'est pas celle exprimée dans la sémantique interprétative, puisque cette dernière a lieu entre sémèmes. L'opposition de F. Rastier est en fait exprimée par la forme même du spécème : (s, s') exprime une opposition entre les deux sémèmes s et s', avec $s \neq s'$.

- Isosémémie : deux spécèmes seront isosémémiques s'ils ont le même sémème opposé (première projection). Les sèmes spécifiques qu'ils peuvent supporter s'attribueront donc au même sémème. L'identité est un cas particulier de l'isosémémie. On notera $sp_1 =_{is} sp_2$.

- Isotaxémie : deux spécèmes seront isotaxémiques si leurs sémèmes opposés sont deux sémèmes du même taxème (donc leurs sémèmes opposants aussi). L'identité, l'opposition et l'isosémémie sont donc des cas particuliers de l'isotaxémie. On notera $sp_1 =_{it} sp_2$.

Les diverses relations binaires entre spécèmes d'un même taxème sont exemplifiées par la figure 3.1.

3.2 Spécèmes et graphes d'opposition

Au sein d'un taxème donné, les spécèmes doivent couvrir un ensemble de relations de distinction (d'opposition en fait). On obtient de cette façon un graphe orienté dont les nœuds sont les sémèmes d'un même taxème. Ce

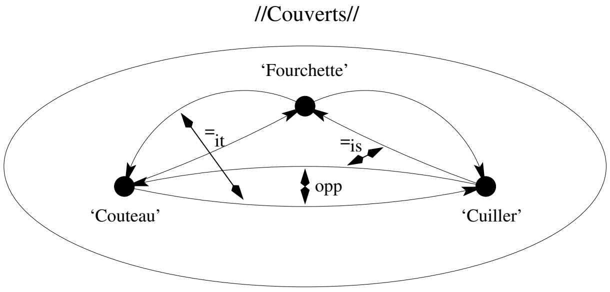


FIG. 3.1 – *Relations formelles entre spécèmes*

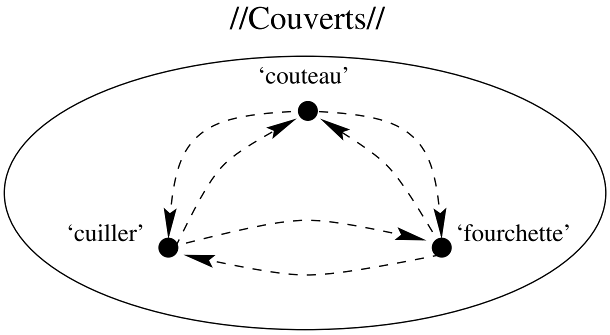


FIG. 3.2 – *Activation du graphe d'opposition d'un taxème - première étape*

graphe est complet, il y a pour chaque couple de sémèmes s, s' d'un taxème t un spécème $sp = (s, s')$ défini.

Par la suite, nous distinguerons les spécèmes *activés, i.e.* ceux qui supportent effectivement un sème, des spécèmes *inertes*, qui ne correspondent qu'à une caractérisation formelle. Ces derniers ne sont en fait que des sèmes spécifiques potentiels. Grâce à cette distinction, nous définissons ainsi un graphe partiel de spécèmes activés. Nous allons ici envisager les contraintes qui doivent organiser ce graphe partiel. Les spécèmes inertes seront en pointillés dans les différentes figures, alors que les spécèmes activés seront dessinés en continu.

Nous approfondirons l'exemple du taxème //couverts// pour exprimer les contraintes s'appliquant aux spécèmes. Sans aucune activation, le taxème est initialisé comme le montre la figure 3.2.

Considérons maintenant le spécème ('couteau', 'fourchette'), dans le taxème //couverts// : si on lui attribue le sème /couper/ (nous verrons com-

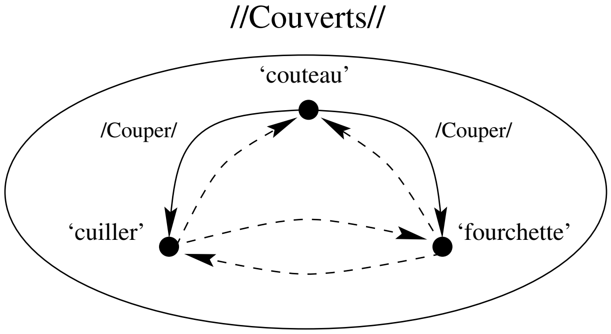


FIG. 3.3 – *Activation du graphe d'opposition d'un taxème - deuxième étape*

ment plus loin), cela signifie simplement que le couteau sert à couper, par *opposition* à la fourchette. Nous avons ainsi une caractérisation positive de 'couteau', mais pas de 'fourchette'; 'couteau' possède donc un sème spécifique, mais pas 'fourchette'. Une première conclusion est l'incompatibilité de cette attribution de sème à celle qui correspondrait à attribuer le même sème /couper/ au spécème ('fourchette', 'couteau') . La relation d'opposition entre les deux sémèmes serait alors symétrique, et perdrait tout son sens. De la même façon, on peut attribuer /couper/ à ('couteau', 'cuiller'), c'est-à-dire rajouter un rôle distinctif au sème spécifique de 'couteau'; mais 'couteau' ne reçoit pas pour autant de nouveau sème spécifique. L'activation de ces deux spécèmes est représentée dans la figure 3.3. Le problème à ce stade est l'absence de caractérisation positive de 'cuiller' et 'fourchette'.

Pour attribuer une caractérisation positive aux deux sémèmes 'cuiller' et 'fourchette', nous pouvons activer les spécèmes ('cuiller', 'couteau') et ('fourchette', 'couteau') par le sème /prendre/. Ces deux sémèmes ont donc un sémantème contenant chacun le sème /prendre/, donc supportent bien une caractérisation positive. À ce stade, représenté par la figure 3.4, nous avons bien la présence pour chaque nœud d'un arc qui y prend son origine, donc une caractérisation positive pour chaque sémème.

Mais 'fourchette' et 'cuiller' possèdent alors exactement les mêmes sèmes spécifiques. Du point de vue de leurs caractérisations sémantiques, la fourchette et la cuiller sont donc à ce stade indistinguables. Si l'on se contente pour palier à cette lacune de sur-activer le spécème ('fourchette', 'couteau'), par exemple en lui attribuant un autre sème comme /piquer/, nous distinguons alors les ensembles de sèmes spécifiques de 'fourchette' et 'cuiller'.

Un petit détour théorique s'impose alors. Rappelons que par les phé-

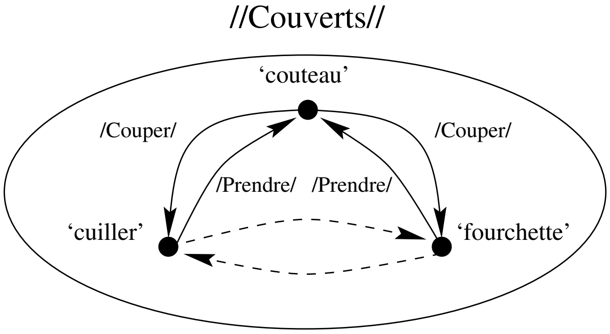


FIG. 3.4 – *Activation du graphe d'opposition d'un taxème - troisième étape*

nomènes et processus que nous décrivons, nous souhaitons, tout comme F. Rastier, proposer la description d'*une* interprétation. Rappelons également qu'une interprétation n'est jamais véritablement achevée, et que la structure que nous décrivons actuellement ne capte qu'un minimum de stabilité nécessaire pour établir cette interprétation. Autrement dit, rien ne nous assure à ce stade de la description du taxème des couverts que le sème /piquer/ ne s'applique pas également au spécème ('cuiller', 'couteau').

C'est-à-dire que l'absence de détermination n'est pas une caractérisation.

La seule certitude que nous puissions avoir est que deux spécèmes opposés ne peuvent supporter le même sème. Par exemple, que l'on ne peut pas, sans bien sûr modifier la description déjà effectuée, attribuer /couper/ à ('cuiller', 'couteau').

Nous basant sur cette constatation, nous pouvons clairement voir le seul moyen de distinguer avec certitude les deux sémèmes récalcitrants : activer un des deux spécèmes qui les relient. C'est ce que nous faisons en attribuant le sème /pour solides/ à ('fourchette', 'cuiller'), comme le montre la figure 3.5. La situation que nous atteignons est donc une complétude non-orientée et la présence d'un arc sortant pour chaque nœud.

Note : Il est important de noter que les problèmes d'identité sémantique que nous avons en partie traités ici sont loin d'être résolus. La notion d'identité des sémèmes au sein d'un taxème prend en compte dans notre précédente discussion la totalité de l'information contenue dans la forme du spécème défini comme un couple de sémèmes. D'autres formes d'identité seront étudiées par la suite, qui nous forceront à nous repencher sur ce problème, sans pour autant modifier les contraintes telles que nous les avons énoncées.

En conclusion, ce graphe doit être complet (sans tenir compte de l'orientation des arcs) et de chaque nœud doit être issu au moins un arc. Ces

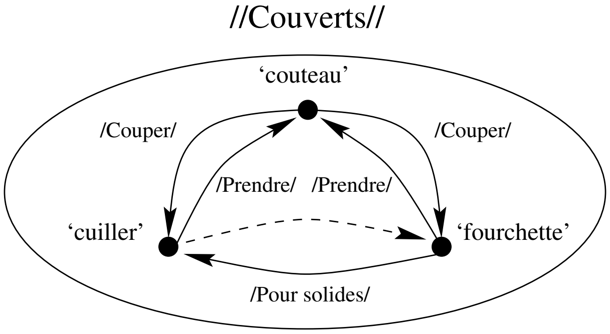
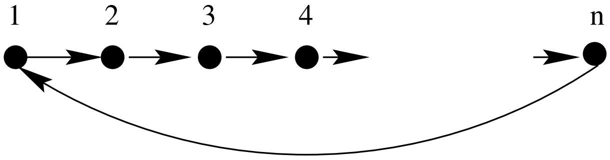


FIG. 3.5 – *Activation du graphe d'opposition d'un taxème - dernière étape*

contraintes seront précisées plus loin, une fois la notion d'isotopie introduite. Nous pouvons ainsi calculer le nombre minimal de spécèmes activés pour un taxème de n sémèmes. Il suffit tout d'abord d'activer n spécèmes afin de mettre en place un cycle entre tous les nœuds (n spécèmes) comme présenté dans la figure 3.2, puis de compléter le graphe pour qu'il soit complet (sans se soucier de l'orientation de ces derniers arcs).



En tenant compte de ces contraintes, pour un taxème à n éléments, un minimum de $\frac{n(n-1)}{2}$ spécèmes doivent être *activés* (sauf pour le cas d'un taxème à deux éléments qui doit, lui, posséder deux spécèmes activés).

4 Les sèmes

Intuitivement, les sèmes sont des « étiquettes » qui seront attribuées aux différentes entités définies jusqu'ici (sémèmes, taxèmes, spécèmes). Le but de ce formalisme étant d'attribuer finalement les sèmes aux sémèmes, mais par des méthodes différentes instaurant autant de types, nous allons maintenant définir une nouvelle et dernière entité, l'isotopie, qui va assumer ce rôle.

Les sèmes sont les éléments d'une classe d'objets SE non vide, notés se indicé ou non. La définition des sèmes est du même type que celle des sémèmes (inhérence de la relation d'identité, notée $=_{SE}$).

Exemple : Au vu des exemples précédents, les sèmes vont nous permettre d'expliciter d'une part les notions que supportent les classes mises en place (les taxèmes), donc le point commun entre les sémèmes que contiennent ces classes, et d'autre part les rapports sémantiques qui existent entre les sémèmes de ces classes. Par exemple, nous pourrons exprimer le fait que les éléments du taxème t_1 sont liés à la notion d'Islam, et que ceux du taxème t_2 sont liés à celle de Christianisme, et également que l'ensemble de ces deux taxèmes est lié à la notion de Religion. /Islam/, /Christianisme/ et /Religion/ sont donc trois exemples de sèmes. Il va nous falloir maintenant expliciter la façon dont on peut les attribuer aux éléments formels mis en place.

5 Isotopies

5.1 Définition

Ce que nous appelons à ce stade isotopie est un concept techniquement plus étendu que celui de la théorie de F. Rastier (nous n'en tirons aucune prétention). La motivation de cet élargissement est la volonté de réunir sous une seule entité toute attribution de sème. L'isotopie est donc le point d'entrée des sèmes dans la structure, tels qu'ils ont été définis. L'idée sous-jacente à la définition de l'isotopie qui va suivre est donc de réunir les sèmes spécifiques inhérents, les sèmes microgénériques inhérents et les sèmes afférents (sans distinction de généricité / spécificité).

La fonction d'isotopie I sera donc la fonction qui attribue un sème aux entités précédemment définies. I est une fonction sur les ensembles suivants :

$$I \ : \ SE \mapsto \mathcal{P}(SP) \times \mathcal{P}(T) \times \mathcal{P}(S)$$

L'image de I, $Im(I)$, permet donc le repérage de tous les types d'entités (sémèmes, taxèmes et spécèmes), participant à la formation de la notion d'isotopie, ou affectées par elle.

Définition : Considérons l'image d'un sème se par la fonction I, qui sera appelé isotopie. On a donc en extension la forme suivante pour une isotopie :

$I(se) = (\{sp_1, \ldots, sp_m\},\ \{t_1, \ldots, t_n\},\ \{s_1, \ldots, s_p\})$

Soit, en se ramenant uniquement sur l'ensemble S, on obtient :

$I \ : \ SE \mapsto \mathcal{P}(\mathcal{S} \times \mathcal{S}) \ \times \ \mathcal{P}(\mathcal{S}) \ \times \ \mathcal{P}(\mathcal{S})$

Avec

$$I(se) = (\{(s_{p_1}, s'_{p_1}), ..., (s_{p_m}, s'_{p_m})\}, \\ \{\{s_1^{t_1}, ..., s_{m_{t1}}^{t_1}\}, ..., \{s_1^{t_n}, ..., s_{m_{tn}}^{t_n}\}\}, \\ \{s_1, ..., s_p\})$$

5.2 Réflexions sur la fonction I

Notre choix d'une telle définition pour aborder la notion d'isotopie permet de relier diverses considérations. De façon primordiale, la forme générique que nous lui attribuons évite la multiplication des notions de généricité, spécificité. En effet, comme on le verra plus clairement dans la suite, elle regroupe la notion d'isotopie selon F. Rastier et le principe d'attribution des sèmes, attribution qui s'accompagne d'une typologie. L'isotopie permet donc d'attribuer des sèmes à des sémèmes, mais par des chemins conceptuels divers, mettant ainsi en place les notions de sème spécifique, générique et afférent.

Plus globalement, le fait de considérer *une* fonction I possède également des avantages conceptuels. Avant tout, l'image de I, comme sous-ensemble du produit cartésien des ensembles $\mathcal{P}(S)$, $\mathcal{P}(T)$ et $\mathcal{P}(SP)$, introduit le rôle de I comme fonction de *choix* sur l'ensemble des possibles attributions de sèmes. Cette simple donnée de la définition de I prépare déjà la grande notion d'interprétation. Pour une interprétation (c'est-à-dire un texte, un utilisateur, un contexte et un ensemble de conditions), il n'existe qu'une seule fonction d'isotopie, qui opère une sélection et une organisation sur des données empiriques du texte (rappelons que celles-ci sont déjà le résultat d'une interprétation, au sens de l'identification de l'ensemble des sémèmes).

Notation : Nous noterons ainsi les trois composantes d'une isotopie : $I_{sp}(se)$ pour les spécèmes, $I_t(se)$ pour les taxèmes et $I_{aff}(se)$ pour les sémèmes (afférences). Ces trois composantes sont donc les trois projections de l'image d'un sème par I.

La présence d'un sémème dans une telle isotopie traduit l'attribution du sème de cette isotopie à ce sémème. Le type d'entité au travers de laquelle ce sémème est présent dans l'isotopie (respectivement spécème, taxème ou directement le sémème) traduit le type de sème pour le sémème considéré (respectivement spécifique, générique ou afférent).

Nous pouvons dès lors définir les différents types d'attribution d'un sème à un sémème :

Définition : *Sème générique inhérent* : se est générique inhérent pour s si le taxème qui contient s est présent dans la deuxième partie de l'isotopie de se. Soit :

se est générique inhérent pour s ssi $\exists t \in T$ t.q. $s \in t$ et $t \in I_t(se)$.

Remarquons au passage que nous confondons à ce stade les statuts de microgénérique et mésogénérique. Nous verrons plus loin comment les distinguer formellement.

Définition : *Sème spécifique inhérent* : se est spécifique inhérent pour s si il existe un spécème dont s est le sémème opposé et qui est présent dans la première partie de l'isotopie de se. Soit : se est spécifique inhérent pour s ssi $\exists sp \in SP$ t.q. $\sigma_1(sp) = s$ et $sp \in I_{sp}(se)$.

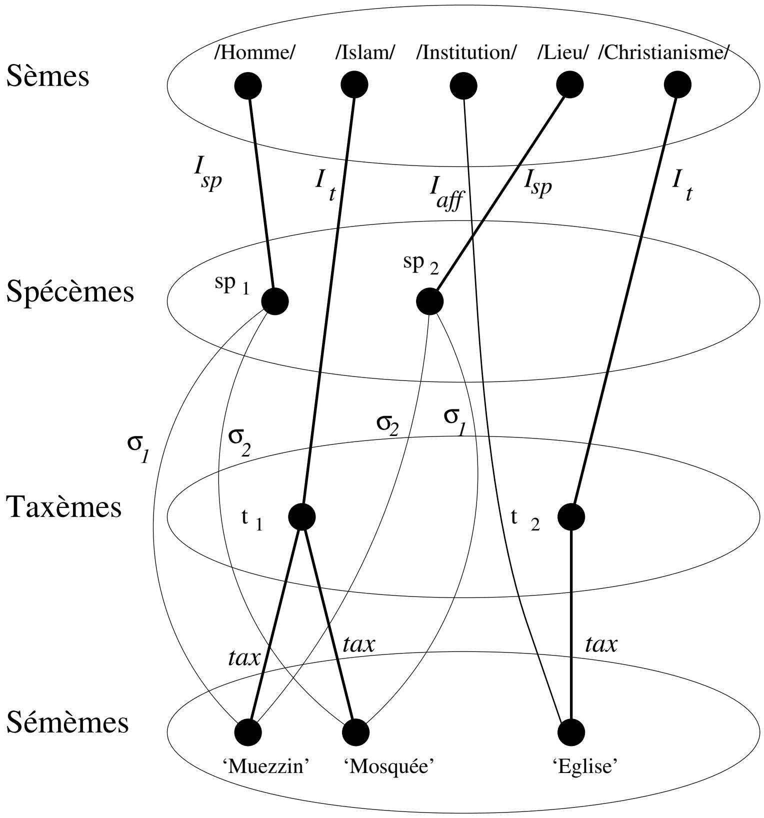
Définition : *Sème afférent* : un sème se est afférent pour s si s est présent dans la troisième partie de l'isotopie de se. Soit : se est afférent pour s ssi $s \in I_{aff}(se)$.

Nous ne distinguerons pas les sèmes spécifiques afférents des sèmes génériques afférents.

Exemple : Voyons par exemple la façon dont est exprimée l'isotopie du sème /Christianisme/, tel qu'il s'applique au taxème t_1 qui contient 'Pèlerinage' et 'Messe'. L'isotopie est de la forme : $(\emptyset, \{t_1\}, \emptyset)$. Si en plus nous rajoutons le sémème 'Dimanche', comme possédant le sème /Christianisme/ de façon afférente (sans que ce sémème appartienne au taxème t_1), nous aurons $(\emptyset, \{t_1\}, \{\text{'Dimanche'}\})$. De la même façon, nous allons utiliser une isotopie pour exprimer les sèmes spécifiques. Pour attribuer le sème spécifique /Voyage/ au sémème 'Pèlerinage', nous allons construire une nouvelle isotopie, à laquelle sera attribué ce sème, qui sera de la forme : $(\{(\text{'Pèlerinage', 'Messe'})\}, \emptyset, \emptyset)$. Bien entendu, par rapport à la définition originelle de la sémantique interprétative, il y a ici une déformation de l'utilisation de l'isotopie, puisque le sème /Voyage/ ne s'applique qu'à un seul sémème (il n'y a pas de récurrence). Il s'agit d'un cas limite de notre reformulation de la notion d'isotopie, mais la rcurrence reste une potentialité par réutilisation de ce sème pour une description ultérieure d'un autre sémème.

Un autre exemple de mise en place de la structure est celui représenté figure 3, avec deux taxèmes : //Islam// et //Christianisme//, deux sèmes spécifiques : /Voyage/ et /Lieu/ et un sème afférent : /Institution/. Par souci de clarté, nous n'avons pas représenté l'intégralité de cette structure pourtant minimale (le second taxème n'a qu'un seul élément, et pas de spécèmes).

FIG. 3.6 – *Exemple de structure sémantique*



En résumé, pour cette figure, nous avons pour chacun des trois sémèmes les sèmes suivants, et leurs statuts :

Sémème	**Sème**	**Statut**
'Muezzin'	/Islam/	générique
	/Homme/	spécifique
'Mosquée'	/Islam/	générique
	/Lieu/	spécifique
'Église'	/Christianisme/	générique
	/Institution/	afférent

5.3 Justification des principes

La forme complète de l'isotopie ainsi définie permet de couvrir toutes les typologies d'isotopie de la théorie de la sémantique interprétative. Bien souvent, cependant, la réalité d'une isotopie sera plus restreinte (un seul type d'élément). De plus, toute attribution de sème devant passer par une isotopie, il peut ne pas y avoir de récurrence représentée par une telle entité (un seul spécème traduisant une seule occurrence comme sème spécifique). L'idée sous-jacente est justement de voir toute attribution de sème comme «réutilisable» et donc, au moins virtuellement, isotopante.

5.4 Contraintes sur les composantes d'une isotopie

5.4.1 Spécèmes - partie spécifique

Reprenons ici la discussion proposée lors de la présentation des spécèmes, pour éclaircir la partie d'une isotopie qui les concerne. Un spécème activé sera dès lors un spécème présent dans au moins une isotopie. Voici donc les contraintes et/ou propriétés associées aux spécèmes dans les isotopies :

- Un même spécème peut appartenir à plusieurs isotopies, donc supporter plusieurs sèmes.
- Deux spécèmes opposés ($(s,\ s')$ et $(s',\ s)$) ne peuvent appartenir à la même isotopie. Soit :

  $\forall sp, sp' \in SP$ t.q. $sp\ opp\ sp'$, $\nexists se \in SE$ t.q. $\{sp, sp'\} \subset I_{sp}(se)$
- Pour tout spécème, lui ou son opposé doit appartenir à au moins une isotopie. Soit :

  $\forall sp, sp' \in SP$, t.q. $sp\ opp\ sp'$, $\exists se \in SE$ t.q. ($sp \in I_{sp}(se)$ ou $sp' \in I_{sp}(se)$)
- Tout sémème doit posséder un spécème activé dont il est la première projection. Soit :

  $\forall s \in S, \exists sp \in SP,$ t.q. $\sigma_1(sp) =_S s$ et $\exists se \in SE$ t.q. $sp \in I_{sp}(se)$

5.4.2 Taxèmes - Partie générique

Il est essentiel qu'à un taxème ne soit associé qu'un seul sème de façon privilégiée : ce sera le sème micro-générique de ce taxème. Un taxème peut également être contenu dans un domaine. Comme nous n'avons pas défini d'entité pour représenter un domaine, nous devrons capter cette notion à partir des seules isotopies. Ainsi, nous autoriserons la présence de plusieurs taxèmes dans un même isotopie, comme nous l'avons vu dans l'exemple plus haut. Par contre, pour définir *le* sème d'un taxème, nous utiliserons une isotopie qui ne contient que le taxème en question.

Ensuite, pour un sémème donné, et pour un sème donné, ce dernier ne peut être attribué au premier que d'une seule façon. Si ce sème est spécifique pour lui, il ne peut être générique pour lui.

Voici donc les contraintes que doit supporter l'isotopie en ce qui concerne sa deuxième projection :

- Tout taxème de T doit appartenir à au moins une isotopie de façon isolée.

- Un taxème ne peut appartenir qu'à deux isotopies (distinctes) au maximum, et une au minimum. Il faut entende par isotopies distinctes des isotopies qui sont les images de deux sèmes distincts.

- Un taxème t ne peut appartenir à deux isotopies distinctes de façon isolée (*i.e.* comme seul taxème des isotopies ; autrement dit, il n'est pas permis de disposer de deux isotopies dont la seconde projection soit le singleton $\{t\}$). Soit, pour résumer les trois assertions précédentes :

  $$\forall t \in T \quad \begin{array}{ll} 1) & \exists ! se \in SE \text{ t.q. } I_t(se) = \{t\} \\ 2) & \forall se_1, se_2 \in SE \text{ t.q. } se_1 \neq se,\ se_2 \neq se, \\ & \text{si } (t \in I_t(se_1) \text{ et } t \in I_t(se_2)) \text{ alors } se_1 =_{SE} se_2 \end{array}$$

- Un taxème ne doit pas contenir la première projection d'un des sp présents dans la même isotopie que lui. Soit :

  $\forall se \in SE, \forall t \in T$ t.q. $t \in I_t(se)$, $\forall s \in tax(t)$, $\forall sp \in SP$ t.q. $\sigma_1(sp) =_S s, sp \notin I_{sp}(se)$.

Nous pouvons maintenant exprimer formellement la différence entre sème microgénérique et sème mésogénérique.

Définition : Un sème se est microgénérique inhérent pour s si le taxème qui contient s est le seul présent dans la partie générique de l'isotopie de se. Soit :

se est microgénérique inhérent pour s ssi $\exists t \in T$ t.q. $s \in tax(t)$ et $I_t(se) = \{t\}$

Définition : Un sème se est mésogénérique inhérent pour s ssi le taxème qui contient s est présent dans, mais ne constitue pas, la partie générique de l'isotopie de se. Soit :

se est mésogénérique inhérent pour s ssi $\exists t_1, t_2 \in T$ t.q. $s \in t_1$, $s \notin t_2$, $t_1 \in I_t(se)$ et $t_2 \in I_t(se)$

5.4.3 Sémèmes - Partie afférente

Rappelons les résultats de notre analyse de la notion d'afférence du chapitre précédent : un sème afférent doit, pour une interprétation, avoir également une attribution comme sème inhérent. La conséquence de notre formalisation est qu'une isotopie quelconque ne pourra être purement afférente. Ensuite, nous devrons ici encore empêcher l'attribution d'un sème à un sémème de deux façons différentes.

- Aucun sémème présent dans la partie afférente d'une isotopie ne doit être la première projection d'un des sp de l'isotopie, ni un élément d'un des taxèmes de cette isotopie. Soit :

  $\forall s \in S, \forall se \in SE$ t.q. $s \in I_{aff}(se)$, $\nexists sp \in SP$ t.q. ($\sigma_1(sp) = s$ et $sp \in I_{sp}(se)$) et $\nexists t \in T$ t.q. ($s \in t$ et $t \in I_t(se)$).

- Il ne peut y avoir de sémèmes dans la partie afférente d'une isotopie si celle-ci ne contient aucun spécème ni aucun taxème (un sème afférent provient obligatoirement d'une occurrence en tant que sème inhérent). Soit :

  $\forall se \in SE$ t.q. $I_{aff}(se) \neq \emptyset$, $\exists t \in T$ t.q. $t \in I_t(se)$ ou $\exists sp \in SP$ t.q. $sp \in I_{sp}(se)$.

6 Les relations entre entités

6.1 Notations

Par extension de la fonction d'isotopie I, on définit par la même notation les sèmes attribués aux entités intermédiaires :

$$I_t^{-1}(t) = \{se \in SE | t \in I_t(se)\}$$

$$I_{sp}^{-1}(sp) = \{se \in SE | sp \in I_{sp}(se)\}$$

$$I_{aff}^{-1}(s) = \{se \in SE | s \in I_{aff}(se)\}$$

En rappelant les relations existant déjà entre les spécèmes, les taxèmes et les sémèmes, nous obtenons le schéma suivant :

$$\begin{tikzcd}
\text{Sème} \arrow[r, "I_t"] \arrow[d, "I_{sp}"'] \arrow[dr, "I_{aff}"] & \text{Taxème} \arrow[d, "tax"] \\
\text{Spécème} \arrow[r, "\sigma_1, \sigma_2"'] & \text{Sémème}
\end{tikzcd}$$

De même, nous pouvons établir des relations directes entre sémèmes et sèmes suivant les types suivants :

- Générique : l'ensemble des sèmes génériques d'un sémème (son classème) est noté : $I_{gen}^{-1}(s) = \{se \in SE \mid \exists t \in T,\ t \in I_t(se)\}$.

  Nous avons donc $I_{gen} = tax \circ I_t$.

- Spécifique : l'ensemble des sèmes spécifiques d'un sémème (son sémantème) est noté : $I_{spe}^{-1}(s) = \{se \in SE \mid \exists sp \in SP,\ sp = (s, s'),\ sp \in I_{sp}(se)\}$.

  Nous avons donc $I_{spe} = \sigma_1 \circ I_{sp}$.

- Afférent : l'ensemble des sèmes afférents d'un sémème est noté : $I_{aff}^{-1}(s) = \{se \in SE \mid s \in I_{aff}(se)\}$

6.2 Identité et différence définitoires

Les deux types d'entités qui constituent la hiérarchie précédemment décrite sont donc les sémèmes s et les sèmes se. A chacune est attribuée une relation d'identité définitoire : $=_S$ et $=_{SE}$.

6.3 Propagation de l'identité des sémèmes

Rappelons ici ce qui a déjà été établi :

- Taxèmes : $t_1 =_S t_2 \ \Leftrightarrow\ tax(t_1) =_S tax(t_2)$
- Spécèmes : $sp_1 =_S sp_2 \ \Leftrightarrow\ \sigma_1(sp_1) =_S \sigma_1(sp_2)$ et $\sigma_2(sp_1) =_S \sigma_2(sp_2)$

- Isotopies et sèmes : deux sèmes seront S-identiques si et seulement si leurs images par I sont strictement S-identiques. $I(se_1)$ et $I(se_2)$ ont donc la même structure et les mêmes éléments. Soit :

  $I(se_1) =_S I(se_2) \Leftrightarrow se_1 =_S se_2 \Leftrightarrow (I_t(se_1) =_S I_t(se_2)$ et $I_{sp}(se_1) =_S I_{sp}(se_2)$ et $I_{aff}(se_1) =_S I_{aff}(se_2))$

Il est également envisageable de considérer une relation plus souple entre deux sèmes qui recouvriraient les mêmes sémèmes mais par des entités de type différent. Pour éclaircir ce point, il nous faut disposer de notions de projections, ce que nous ferons plus loin.

6.4 Propagation de l'identité des sèmes

Le point de départ de cette propagation est donc l'identité inhérente aux sèmes de SE (notée $=_{SE}$). On parlera également de SE-identité ou de SE-équivalence.

- Spécèmes : Deux spécèmes seront SE-identiques s'ils sont présents dans les mêmes isotopies (au sens de $=_{SE}$), donc si les mêmes sèmes leurs sont attribués.

  $sp_1 =_{SE} sp_2 \Leftrightarrow I_{sp}^{-1}(sp_1) =_{SE} I_{sp}^{-1}(sp_2)$.

- Taxèmes : Deux taxèmes seront SE-identiques s'ils appartiennent aux mêmes (ou à la même) isotopie(s).

  $t_1 =_{SE} t_2 \Leftrightarrow I_t^{-1}(t_1) =_{SE} I_t^{-1}(t_2)$.

- Sémèmes (identité forte) : Deux sémèmes seront SE-identiques **fortement** s'ils appartiennent aux mêmes isotopies et que leurs types d'appartenance à ces isotopies sont les mêmes (spécifiques, génériques, afférents) pour chaque isotopie. Soit :

$$\begin{array}{ll} s_1 =_{SE} s_2 \Leftrightarrow & I_{gen}^{-1}(s_1) = I_{gen}^{-1}(s_2) \\ \text{et} & I_{spe}^{-1}(s_1) = I_{spe}^{-1}(s_2) \\ \text{et} & I_{aff}^{-1}(s_1) = I_{aff}^{-1}(s_2) \end{array}$$

- Sémèmes (identité faible) : Deux sémèmes seront SE-identiques **faiblement** s'ils se voient attribuer les mêmes sèmes, mais par des entités différentes par I. Par exemple, un sémème peut posséder le sème de façon générique, l'autre sémème de façon spécifique. Soit :

$$s_1 \equiv_{SE} s_2 \Leftrightarrow \begin{array}{l} I_{gen}^{-1}(s_1) \ \cup \ I_{spe}^{-1}(s_1) \ \cup \ I_{aff}^{-1}(s_1) \ =_{SE} \\ I_{gen}^{-1}(s_2) \ \cup \ I_{spe}^{-1}(s_2) \ \cup \ I_{aff}^{-1}(s_2) \end{array}$$

6.5 Compatibilités entre S-identité et SE-identité

Le grand nombre de contraintes gérant la structure sémantique implique des relations plus fortes entre ces identités et différences. Nous envisagerons les différents cas, afin de statuer sur les distinctions à préserver, et sur celles qui n'ont pas lieu d'être. Nous allons donc présupposer une SE-identité entre deux entités de même nature, et voir si cela entraîne leur S-identité.

6.5.1 Taxèmes

C'est le cas le plus simple : si t_1 et t_2 sont SE-identiques, ils sont présents dans les mêmes isotopies, reliées aux sèmes se_1, se_2, ou seulement se_1.

t_1 doit avoir un seul et unique sème microgénérique, disons se_1. La partie générique de l'isotopie de se_1 est donc :

$I_t(se_1) = \{t_1\}$

Mais t_2 doit également appartenir à cette isotopie, donc $t_2 \in \{t_1\}$, d'où la rapide conclusion que $t_1 =_S t_2$.

Théorème : Deux taxèmes SE-identiques sont S-identiques.

Plus clairement peut-être, si un sème a un statut microgénérique dans la structure, il ne le possède que pour un seul ensemble de sémèmes. Il devient donc légitime de repérer un tel ensemble de sémèmes par son sème microgénérique, autant que par son extension sur S.

6.5.2 Spécèmes

Il est aisé de trouver un contre-exemple à une extension du phénomène démontré pour les taxèmes : considérons ce qui se passe au sein du taxème //couverts//, désormais bien connu. Le spécème ('couteau', 'fourchette') appartient à l'isotopie de /pour couper/, et à celle-là seulement. Il en va de même pour le spécème ('couteau', 'cuiller'). Il n'en reste pas moins que ces deux spécèmes sont distincts au sens de S, puisqu'ils caractérisent tous deux le sémème 'couteau' en l'opposant à deux sémèmes distincts ('fourchette' et 'cuiller', de formes distinctes).

On peut bien évidemment envisager un tel cas réparti sur plusieurs taxèmes, et nous ajouterons même qu'il s'agit là de l'intérêt des isotopies spécifiques.

Ainsi, la différence de statut des identités entre ces deux notions que sont le taxème et le spécème montre bien le statut définitoire du taxème. En effet, on ne peut aborder la notion de spécème que lorsque celle du taxème a été définie, ainsi cette dernière définition ne doit supporter aucune ambiguïté. Il est donc naturel de constater une plus forte cohérence pour les taxèmes que pour les spécèmes.

6.5.3 Isotopies

Rappelons simplement que la fonction isotopante I associe un et un seul triplet extrait du produit $\mathcal{P}(SP) \times \mathcal{P}(T) \times \mathcal{P}(S)$ à un sème. L'identité selon SE entraîne donc, de façon définitoire l'identité selon S. Par contre, il est intéressant de se poser la question inverse. Deux isotopies ayant les mêmes composantes au sens de S correspondent-elles obligatoirement à un même sème ?

Un premier contre-exemple vient à l'esprit grâce aux spécèmes. Un même spécème peut supporter plusieurs sèmes, et donc apparaître dans plusieurs isotopies. On peut donc envisager les deux isotopies spécifiques minimales identiques : $I(se_1) = I(se_2) = (\{sp\}, \emptyset, \emptyset)$. Un exemple d'un tel cas de figure serait l'attribution de /pour couper/ au spécème ('couteau', 'fourchette'), et celle de /se tient de la main droite/ à ce même spécème. Il faut donc supposer également la non-récurrence de ces deux sèmes, afin que les isotopies ne soient pas étendues différemment l'une de l'autre. Rappelons que nous nous situons en contexte, et que même si l'attribution de ce sème à une multitude (voire une infinité) d'autres entités est possible en langue, nous n'en étudions ici que la réalisation locale.

La question se pose par contre de façon plus complexe pour des isotopies au moins en partie génériques, car nous disposons du théorème sur l'identité des taxèmes précédemment exposé. Nous allons donc envisager les différents cas suivant la nature de la partie générique des isotopies.

Premier cas : les isotopies sont de la forme $(X, \{t\}, Y)$ avec $X \subset SP$ et $Y \subset S$.

Ces deux isotopies correspondent donc au sème microgénérique du taxème t (seul cas où un taxème apparaît de façon isolée dans une isotopie). Un taxème n'ayant qu'un seul sème microgénérique, les deux sèmes sont donc identiques (au sens de SE).

Deuxième cas : les isotopies sont de la forme $(X, \{t_1, ..., t_n\}, Y)$.

Ce sont donc des isotopies en partie mésogénériques. Le taxème t_1 ne peut appartenir au maximum qu'à une seule isotopie (au sens de SE) dans laquelle il n'est pas isolé, donc les deux isotopies sont identiques au sens de SE.

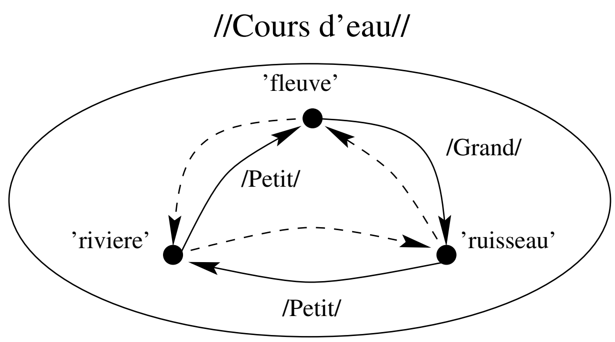
Théorème : Deux isotopies S-identiques possédant une partie générique non vide sont SE-identiques.

6.5.4 Sémèmes

C'est bien ici le problème le plus délicat, puisque nous avons déjà vu qu'il y a au moins deux manières de considérer la SE-identité des sémèmes.

Intéressons-nous pour l'instant à la SE-identité *forte*. Si s_1 et s_2 sont fortement SE-identiques, ils possèdent les mêmes sèmes, et ces sèmes ont pour eux les mêmes types. Par exemple, ils possèdent obligatoirement le même sème microgénérique, et le même sème mésogénérique si celui-ci existe. Ce sont donc, d'ores et déjà, deux sémèmes d'un même taxème. Donc, dans le graphe d'opposition de ce taxème, il existe un spécème activé, soit (s_1, s_2), soit (s_2, s_1), soit les deux. Considérons que ce soit (s_1, s_2), et qu'il lui soit attribué le sème se. Ce sème se est donc spécifique pour s_1. Il doit également être spécifique pour s_2 (hypothèse de travail). Donc, il doit exister un spécème (s_2, s_3) dans $I_sp(se)$. s_3 ne peut être identique à s_1 (principe d'opposition), donc nous avons déjà la contrainte que le taxème de s_1 doit contenir au moins trois éléments. Il apparaît donc que le cas où la S-identité forte n'est pas établie entre s_1 et s_2 est celui d'un cycle formé par les spécèmes de l'isotopie spécifique étudiée. Par exemple, au sein du taxème des //cours d'eau//, les sémèmes 'ruisseau', 'rivière', 'fleuve', peuvent s'articuler de façon minimale comme indiqué par la figure 3.7.

FIG. 3.7 – *Exemple de cycle*



Dans ce cas, les deux sémèmes 'ruisseau' et 'rivière supportent le même sème spécifique /petit/ : leurs sémantèmes sont identiques, même si les contraintes précisées pour les isotopies spécifiques sont respectées dans ce taxème. En fait, la véritable identité sémantique d'un sémème, centrée sur ses sèmes spécifiques, prend en compte non seulement le sème, mais aussi le sémème auquel ce sème sert à opposer le premier. C'est cette idée d'identité structurelle qui nous a conduit à préciser les contraintes (en fait à les limiter). Ainsi, considérer la simple collection de sèmes spécifiques attribués à un sémème, sans considérer la seconde projection des spécèmes correspondants, ne permet par la distinction de deux sémèmes d'un même taxème.

Cette vision d'un sémème comme ensemble de sèmes, même typés, n'est en fait qu'une projection de la structure sémantique complète à des fins calculatoires et/ou typologiques comme nous le verrons par la suite. La véritable forme de l'identité sémantique serait donc la suivante :

$$id(s) = (se_t, se_{t'}, \{(se_i, s_i)\}, \{se_j\})$$

La fonction id permet donc de résumer toutes les considérations sémiques précisément typées attribuées à un sémème, avec les spécifications suivantes :

- se_t est le sème microgénérique du sémème.
- $se_{t'}$ est le sème mésogénérique du sémème, s'il existe.
- La famille des $\{(se_i, s_i)\}$, avec $se_i \in SE$ et $s_i \in S$ rappelle les sèmes spécifiques du sémème s, en précisant les sémèmes de son taxème qui s'opposent à lui par le sème se_i. En fait, ces couples traduisent l'existence d'un spécème (s, s_i) dans l'isotopie du sème se_i.
- Enfin, les se_j sont les sèmes afférents du sémème s.

Théorème : L'identité sémantique id coïncide avec la S-identité. Soit :

$$\forall s, s' \in S,\ s =_S s' \Leftrightarrow id(s) = id(s')$$

Démonstration : La partie directe de l'équivalence est triviale. Considérons donc deux sémèmes s et s' distincts tels que :

$$id(s) =_{SE} id(s') = (se_t, se'_t, \{(se_i, s_i)\}, \{se_j\})$$

Donc,
$\exists t \in T$ tel que $I_t(se_t) = \{t\}$, $s \in t$ et $s' \in t$.

Comme les deux sémèmes appartiennent au même taxème t, il existe un spécème $sp = (s, s')$ ou un spécème $sp' = (s', s)$ qui soit activé, disons sp. Soit se le sème tel que $sp \in I_{sp}(se)$. Dans ce cas, $se \in I_{gen}^{-1}(s)$, donc dans $id(s)$, nous avons dans la troisième composante le couple (s', se). Idem pour $id(s')$, ce qui traduit le fait que le sémème s' s'oppose au sémème s' par le sème se : c'est impossible, d'après la définition des spécèmes (les sémèmes opposé et opposant doivent être distincts selon $=_S$). s et s' ne peuvent donc être distincts selon $=_S$.

Cette image de s par id est donc unique pour tout sémème de S, mais reste en grande partie inutilisable, et nous verrons par la suite des exemples d'utilisation d'une projection de celle-ci.

7 Vers une forme manipulable de l'isotopie

Rappelons que l'identité formelle que nous avons attribuée à la notion d'isotopie s'éloigne de la définition de F. Rastier comme simple récurrence d'un sème le long d'une portion de texte. Cette vision est cependant indispensable pour «utiliser» l'idée de l'isotopie, notamment dans le cadre plus calculatoire du TALN. Il nous faut donc préciser comment obtenir une telle projection de la structure sur l'ensemble des sémèmes.

7.1 Projection de l'isotopie sur S

Nous allons donc définir une forme projective d'une isotopie associée à un sème se sur S, via la fonction I_S, avec :

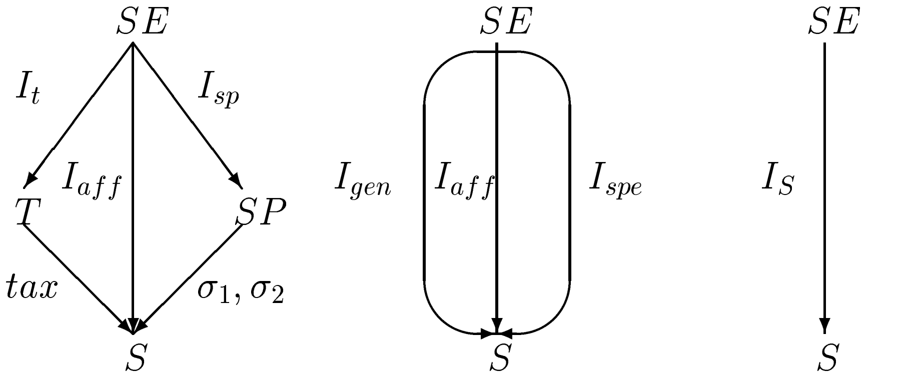
$$I_S : SE \mapsto S$$

avec

$$I_S(se) = I_{gen}(se) \cup I_{spe}(se) \cup I_{aff}(se)$$

Nous retrouvons ici une forme simple de l'extension d'une isotopie, dans laquelle les informations sur les types d'attribution de sèmes ont été perdues. Seules restent les composantes thématiques (quel sème) et tactiques (quels sémèmes) de l'effet de récurrence. Nous verrons plus loin comment utiliser ces notions simplifiées dans la caractérisation des textes par leurs thèmes et leurs structures.

Pour résumer les différents avatars de la fonction I, nous pouvons consulter la figure 3.8. Les différentes modifications proposées précédemment sont



hbt

FIG. 3.8 – *Projections de I sur S*

présentées de gauche à droite, au long d'un processus de simplification et d'effacement progressif des informations.

Mais, maintenant que nous avons tout ramené sur le sol fertile de l'ensemble S (objectif primordial dans l'esprit de la sémantique interprétative), nous allons devoir affiner un peu sa structuration interne, et notamment définir une notion naturelle d'ordre sur ses éléments. Cet ordre ne sera autre que la disposition des unités signifiantes sur l'axe syntagmatique. Avant tout, cependant, nous devons une fois de plus nous intéresser aux délicats mais passionnants problèmes d'identité des sémèmes, et pour cela introduire une nouvelle notion, celle d'épisémème.

7.2 Introduction de la notion d'épisémème

Jusqu'ici, nous n'avons considéré les sémèmes *en soi* que comme des éléments identitairement typés de l'ensemble S. La réalité de la relation d'identité $=_S$ se situe au niveau de la forme (chaîne de caractères) des lexies formant les sémèmes. Les problèmes apparaissent lors de la répétition d'une même chaîne dans un même texte. Autrement dit, si le mot «chat» apparaît 42 fois dans un texte, y a-t-il 42 sémèmes ou un seul répété à 42 positions différentes ? Nous retombons ici en pleine polysémie, phénomène obsessionnel dans le TALN. Nous avons déjà évoqué les problèmes de l'identité sémantique des sémèmes et nous sommes parvenus à la conclusion que deux sémèmes différents se situaient à deux places bien distinctes dans la structure sémantique (même si certaines projections de ces emplacements pouvaient être identiques). Nous devons tout de même être en mesure de traiter le cas de deux chaînes identiques supportant des considérations sémiques distinctes.

Citons comme exemple un cas classique de polysémie forte (que F. Rastier qualifie comme des *sens* distincts) : 'pomme' comme fruit (taxème des //fruits//) ou comme partie d'instrument ('pomme' d'arrosoir). Dans ce cas, aucun sème micro ou mésogénérique ne sera commun entre ces deux sémèmes. Les formalismes classiques de l'IA parleraient de pomme#1 et de pomme#2, ou de deux concepts distincts. Il est cependant possible que ces deux sémèmes aient des sèmes spécifiques ou afférents en commun, par exemple /sphéricité/.

Mais il y a des différences plus subtiles (des *acceptions* selon la sémantique interprétative), comme par exemple entre 'église' comme bâtiment : «la porte de l'église» ou comme institution : «les pères de l'église» (ou «de l'Église» ?). Il est sain de songer que les sèmes génériques (du moins micro- et méso-génériques) de ces deux sémèmes seront identiques (/religion/ et/ou /christianisme/ par exemple). Mais des distinctions peuvent apparaître dans leurs sèmes spécifiques ou afférents : /bâtiment/ ou /institution/ entre autres.

Enfin, le long de la même échelle nous pouvons glisser jusqu'à des cas limites mais néanmoins très intéressants car généralement laissés en friche par le TALN classique. Pour reprendre un exemple déjà en partie explicité lors du précédent chapitre, citons l'adjectif 'rouge' dans «Le rouge et le Noir». À part dans le titre, cette lexie peut apparaître dans le texte sous une occurrence que le contexte ne pousse guère à interpréter comme symbolisant l'armée. Par exemple, quand Julien doit décorer l'église avec du damas rouge. Pourtant, dans les deux cas de figure, il s'agit du même sémème rouge, défini dans le taxème des //couleurs// par le sème spécifique /rougeur/. En fait, seule l'afférence *locale* varie [1]. C'est donc le cas de deux *emplois* distincts.

Dans tous les cas plus ou moins classiques que nous venons de citer, sans pour autant vouloir les justifier, nous devons admettre que la simple forme ne suffit pas à distinguer les simples éléments de S. Il nous faut en plus un critère positionnel afin de respecter l'unicité des occurrences. Cependant, il est tout de même appréciable de pouvoir *assimiler* plusieurs occurrences qui supportent des considérations sémantiques unifiées. La répétition d'un même terme le long d'un texte est de façon typique la source d'isotopies, triviales certes, mais à prendre en compte systématiquement.

Donc, si nous séparons les sémèmes suivant leurs positions, nous devons préserver leur unicité sémantique, sans pour autant avoir à comparer leurs

1. Rappelons une dernière fois que nous nous plaçons ici dans une vision statique de la description : il est clair que cette afférence militaire du rouge se construit au long d'une lecture du roman.

sèmes. De plus, selon les définitions et contraintes que nous avons posées, il est impossible que deux sémèmes distincts suivant $=_S$ supportent des considérations sémantiques strictement identiques (identité forte selon SE).

C'est donc après ces longs détours théoriques que nous allons introduire la notion d'épisémème, afin de tenir compte des positions *physiques* des termes dans un texte. Nous verrons de plus par la suite comment utiliser ces critères positionnels pour qualifier les isotopies.

L'ensemble des positions des termes interprétés d'un texte forme l'ensemble E des *épisémèmes*. L'identité inhérente à E se base *strictement* sur la position. L'identité des chaînes de caractères n'est donc plus qu'un phénomène périphérique et non suffisant. L'ensemble des sémèmes S devient donc un ensemble extensionnel basé sur E : un sémème sera un ensemble d'épisémèmes. Le reste des considérations sur S demeure bien entendu inchangé.

Nous intégrons dans les relations entre E et S les contraintes suivantes :

- Tout épisémème de E doit correspondre à un et à un seul sémème de S.

- Un sémème de S peut ne pas avoir d'extension sur E. C'est le cas de certains sémèmes que nous appellerons «non lexicalisés», c'est-à-dire absents du texte, mais utilisés dans la structure sémantique à des fins définitoires pour d'autres sémèmes (complétion d'un taxème à un seul élément par exemple), ou provenant d'opérations de réécriture.

En fait, la relation liant S à E est une simple relation de *positionnement*. Nous ne nous étendrons pas longuement sur les propriétés de cette relation ni sur celles de E, vu que l'utilisation de celle-ci sera restreinte aux seules isotopies, pour des considérations de *tactique*. Cependant, l'intérêt majeur de ce nouvel ensemble E est qu'il supporte une structure d'ordre total, en partie récupérable par S, comme nous allons le voir ci-dessous.

Définissons tout de même la réciproque de cette relation, l'extension de S sur E : $ext : S \mapsto \mathcal{P}(E)$, qui a un sémème s associe un sous-ensemble de S contenant les épisémèmes qui lui correspondent, si ceux-ci existent ($ext(s)$ peut être l'ensemble vide).

7.3 Définition de l'ordre sur E

La réalité «physique» d'un texte peut aisément être perçu comme une simple suite de termes positionnés le long de l'axe syntagmatique. Cet axe étant orienté, nous pouvons comparer deux épisémèmes quelconques par la relation $\leq$, qui est un ordre total.

$$e_1 \leq e_2 \text{ si la position de } e_1 \text{ précède celle de } e_2 \text{ dans le texte.}$$

Il est important de noter que notre volonté de mettre en place un ordre total nous empêche de considérer des sous-chaînes du type 'église' et 'pères de l'église', c'est-à-dire deux extraits d'une même portion de texte. Certes, la seconde chaîne prend son origine avant la première, mais leurs couvertures (en terme de graphèmes par exemple) sont en intersection. Il n'est donc pas aisé dans ce cas de décider quel est le premier de ces deux épisémèmes. De tels cas seront traités sur le critère de la position du premier graphème repéré[2], avec un choix arbitraire dans le cas où deux chaînes débutent au même point ('bateau ivre' $\leq$ 'bateau', par exemple).

Nous allons maintenant étudier les conséquences de cet ordre sur les autres niveaux de la structure sémantique.

7.4 Ordres induits

7.4.1 Ordre sur S

L'ordre total sur E ne se projette pas directement ni intégralement sur S. Il y a deux raisons à cela.

Tout d'abord, certains sémèmes ne sont pas lexicalisés, comme nous l'avons dit, et ne peuvent donc être comparés par leurs positions, puisqu'ils n'en ont pas. De plus, un même sémème peut avoir plusieurs extensions dans E, et il nous faudra donc affiner certaines notions d'ordres induits pour ordonner S. Mais nous ne nous en donnerons pas la peine, puisque nous pouvons allègrement nous intéresser à l'ordre induit sur SE, *via*, bien entendu la fonction, I.

7.4.2 Ordre des isotopies

Rappelons que nous pouvons projeter le triplet qu'est une isotopie sur S à l'aide de la fonction I_S. Si maintenant nous considérons la composée $ext \circ I_S$, nous obtenons une relation de SE dans E. Nous obtenons ainsi la position d'un sème. Cette dernière est un sous-ensemble (pouvant être vide) de E, donc une famille ordonnée suivant leurs positions. C'est ainsi que nous

2. C'est du moins ainsi que le traite le programme PASTEL.

prétendrons capter en partie ce que F. Rastier appelle la *composante tactique* d'un texte interprété. Dès lors, nous pourrons comparer deux isotopies sur leurs simples positions.

- Concordance : les deux isotopies concernent les mêmes épisémèmes.
- Chevauchement : les deux isotopies possèdent un sous-ensemble d'épisémèmes en commun, et chacune un sous-ensemble propre.
- Inclusion : tous les épisémèmes d'une isotopie sont également présents dans l'autre.
- Indépendance : les deux isotopies ne possèdent pas d'épisémème commun. Dans ce cas, les positions relatives des sémèmes induisent plusieurs possibilités :
  - antériorité / postériorité : si les deux groupes d'épisémèmes ne se croisent pas.
  - alternance : si les épisémèmes de l'une s'alternent avec ceux de l'autre.

Toutes ces considérations sont bien entendu miscibles et graduelles.

D'autre part, d'autres considérations unitaires peuvent s'appliquer aux isotopies :

- Poids : le nombre d'épisémèmes supportant l'isotopie
- Volume : nombre total d'épisémèmes compris entre le premier et le dernier épisémème de l'isotopie
- Densité : rapport Poids / Volume, caractérisant l'importance de l'isotopie dans sa zone de validité.

Nous verrons par la suite les significations stylistiques et typologiques de ces critères quantitatifs, sur des exemple poétiques entre autres. Rappelons que F. Rastier considère les rapports entre thématique et tactique comme fondamentaux dans tout acte typologique, et même dans tout acte interprétatif. Le cas d'un texte réduit à une énumération n'échappe pas à cette généralité.

Nous approfondirons ces considérations lorsque nous exposerons l'outil logiciel PASTEL. En attendant, nous allons voir la mise en place dynamique de cette structure, ainsi que son évolutivité.

Chapitre 4

Évolution de la structure

Nous allons nous intéresser dans ce chapitre aux différents aspects dynamiques de la structure que nous avons décrite durant le chapitre précédent. Étant donnée la complexité intrinsèque des relations entre entités, et surtout le nombre de contraintes régissant ces relations, nous allons dans un premier temps décomposer les opérations de manipulation en opérateurs élémentaires. Nous construirons par la suite, sur la base de ces opérateurs minimaux, des opérations de plus en plus complexes, en nous efforçant à chaque palier d'intégrer la gestion des contraintes. Nous obtiendrons en effet un ensemble de transitions possibles d'une structure sémantique stable (c'est-à-dire vérifiant l'ensemble des contraintes exprimées) à une autre structure de stabilité identique.

Une fois ces opérations décrites, nous les mettrons en œuvre de trois façons différentes :

- La mise en place initiale de la structure, avec la définition formelle du parcours interprétatif.
- L'ensemble des modifications locales applicables sur une structure (ajout ou retrait d'éléments).
- Les modifications globales de la structure, à des fins de comparaison et de fusion entre structures différentes.

Parmi ces opérateurs formels, nous distinguerons ceux qui font directement appel à la compétence de l'interprète. Ils constitueront en quelque sorte les fonctions d'entrées de notre structure.

Cependant, avant de nous lancer dans ces considérations formelles, nous allons préciser ici le déroulement d'une telle interprétation. À chaque étape, nous préciserons les opérateurs dont nous aurons besoin, et qui seront précisés au cours de ce chapitre.

1 Vision générale du protocole

1.1 Première étape : pré-interprétation

Le but de cette étape est de capter les intuitions générales de l'interprète, et d'alimenter la structure de données d'une façon minimale. Cette première étape est effectuée sans aucune contrainte : les éventuels conflits avec les règles d'organisation de la structure seront résolus par la suite.

Le principe général en est donc la déclaration par l'interprète d'un ensemble de traits sémantique (sèmes), d'un ensemble de chaînes extraites du texte analysé (épisémèmes), et d'associations non contraintes entre ceux-ci.

Les opérations formelles nécessaires sont donc celles qui correspondent à l'ajout d'un sème ($+se$), l'ajout d'un épisémème ($+e$), et la déclaration d'une association de ces entités (par définition de la relation PI_E).

1.2 Seconde étape : établissement des classes

Une fois ceci fait, la structure doit commencer à mettre en place les contraintes précédemment définies. Il faut avant tout construire l'ensemble des sémèmes ($+s$), les associer aux épisémèmes (qui eux sont de simples pointeurs sur le texte analysé, avec $+e(s)$), et les associer en classes, notamment en taxèmes ($+t$, $add(s,t)$), et reporter sur ceux-ci les relations sémèmes-sèmes précédemment déclarées par l'interprète (avec activation de la fonction I).

À ce stade seront gérées les contraintes sur l'intersection et l'inclusion des classes, pour lesquelles l'interprète aura à choisir entre plusieurs possibilités, en fonction ses propositions initiales. Cette étape fera donc intervenir un opérateur de choix (*select*, et des opérations liées au déplacement des sémèmes par rapport aux taxèmes ($sub(s,t)$).

1.3 Spécification des classes

Dans cette étape, les taxèmes précédemment créés et remplis devront être spécifiés, en mettant en place l'organisation de leur graphe d'oppositions. Il nous faudra dès lors des opérateurs de création de spécèmes ($+sp$), et d'activation de ceux-ci ($activ$). Ce dernier sollicitera l'interprète dans le repérage d'un sème ($+se$), et l'associera au spécème concerné ($+i(sp,se)$). Les contraintes liées à l'organisation d'un tel graphe d'opposition seront gérées notamment par l'opérateur de complétude ($comp(t)$), qui exigera l'activation d'un nombre minimal de spécèmes pour un taxème donné.

Une fois cette étape effectuée pour tous les taxèmes définis, la structure sémantique de l'interprétation est *stable* : toutes les contraintes définies au chapitre précédent sont satisfaites.

1.4 Modifications ultérieures

À partir d'une structure stable, nous donnerons la possibilité à l'interprète d'apporter quelques modifications à son analyse. Il pourra donc ajouter ou soustraire tout type d'éléments (opérateur d'élimination $-s$, $-se$, etc.), tout en conservant l'intégrité de la structure, ces opérations entraînant parfois un remaniement total de la structure par propagation des contraintes. Nous envisagerons des modifications locales (ajoute ou retrait d'un élément, modifications directes des relations, comme le déplacement d'un sémème d'un taxème à un autre), ou bien des opérateurs globaux. Parmi ceux-ci, nous envisagerons plus centralement le cas d'un opérateur qui crée un taxème à partir d'une isotopie spécifique ($igen$). Cette possibilité est intéressante si l'interprète a découvert, au cours de son parcours, un nouveau thème central du texte. Comme celui-ci ne faisait pas partie de sa vision initiale, il ne constitue pas une classe sémantique. S'il désire donc expliciter plus finement la structuration de ce nouveau thème, sa constitution en taxème en est un bon moyen.

2 Outils formels de manipulation

Nous allons une fois de plus surcharger le lecteur de nouvelles définitions, en décrivant les opérations minimales applicables à une structure. Dans un premier temps, nous ne tiendrons pas compte des contraintes, et nous contenterons de décrire les ajouts et retraits des différentes entités dans la structure.

2.1 Opérateurs d'ajout et de retrait

Ces opérateurs ne nécessitent guère plus d'explicitation que leur simple dénomination. Nous désignerons ces opérateurs d'ajout par $+X$, avec X qui prendra la forme générique d'une entité parmi : s, se, sp, t. De façon similaire, les opérateurs de retrait d'une entité seront exprimés par $-X$.

Nous allons rapidement décrire les particularités de certains de ces opérateurs.

2.1.1 Cas des épisémèmes

Nous n'envisagerons pas le cas des relations entre un épisémème et les types d'entités autres que les sémèmes, puisque ces relations se font intégralement par l'intermédiaire du sémème de l'épisémème.

2.1.2 Ajout et retrait d'un sémème

Lorsqu'un nouveau sémème est repéré, il peut être lexicalisé, et son repérage nécessite alors un paramètre de position. Autrement, cette opération ne s'accompagne que de la donnée d'une chaîne de caractères manifestant le signifiant associé. Dans le cas d'un sémème lexicalisé, ce repérage s'accompagne donc de celui d'un ou plusieurs épisémème(s). Nous ne traiterons cependant pas rigoureusement le cas des épisémèmes, puisque, nous l'avons vu, leur intérêt est restreint à la tactique des isotopies. Quoiqu'il en soit, leur introduction et/ou leur retrait ne présente pas de difficulté conceptuelle.

Lors de l'élimination d'un sémème de la structure, nous considérons également la suppression de tous les épisémèmes qui lui sont associés.

2.1.3 Ajout et retrait d'un spécème

Lors de l'ajout d'un nouveau spécème, nous préciserons pour l'opérateur $+sp$ les deux composantes du couple. Nous noterons ainsi $+sp(s_1, s_2)$ l'ajout du spécème (s_1, s_2). Il en va de même pour la suppression.

2.1.4 Ajout et retrait d'un taxème

La création d'un taxème se fait, à la différence des spécèmes, hors de toute considération portant sur les sémèmes. Autrement dit, un taxème t créé par l'application de $+t$ donne comme résultat un taxème t vide et sans sème associé. De même, nous apportons comme restriction sur $-t$ le fait que l'extension de t dans S doit être vide. Un tel taxème est par définition instable, mais l'opérateur est ici constitutif d'un ensemble plus global d'opérations.

2.1.5 Ajout et retrait d'un sème

De la même façon que pour les taxèmes, nous considérons que l'ajout d'un sème par $+se$ se fait sans aucune attribution de ce sème par la fonction

I. Le retrait d'un sème par $-se$ ne peut également s'effectuer que si ce sème est associé à une isotopie vide ($I(se) = (\emptyset, \emptyset, \emptyset)$).

2.2 Instauration et suppression des relations entre entités

Nous allons maintenant décrire une autre famille d'opérations, portant cette fois sur les relations entre deux entités de types différents. Ces opérateurs ne tiennent toujours pas compte de toutes les contraintes inhérentes à la structure.

2.2.1 Cas des épisémèmes

De même, ici nous n'envisagerons pas le cas des relations entre un épisémème et les types d'entités autres que les sémèmes, puisque ces relations se font intégralement par l'intermédiaire du sémème de l'épisémème. Nous proposons donc deux simples opérateurs d'ajout et de retrait d'un épisémème à un sème : $+e(s)$ et $-e(s)$. Ces deux opérations modifient en fait la relation ext reliant un sémème à ses épisémèmes.

2.2.2 Sémème - taxème

On ajoute un sémème dans un taxème par l'opérateur $add(s, t)$. Le sémème s ne doit appartenir à aucun taxème avant l'application de cet opérateur. De la même façon, on retire un sémème d'un taxème par l'opérateur $sub(s, t)$.

2.2.3 Sémème - spécème

Ce cas est couvert par l'opérateur de création ou de suppression des spécèmes ($+sp(s, s')$ et $-sp(s, s')$), puisqu'un spécème n'est repéré que par ses projections sur S.

2.2.4 Sémème - sème

Nous allons ici envisager le cas de l'ajout ou de la suppression d'un sème comme afférent à un sémème. C'est en effet le seul cas d'un rapport direct entre un sémème et un sème. L'opérateur modifie la fonction I, et est donc de la forme $+i(s, se)$, avec $s \in S$, $se \in SE$. Le résultat de cet opérateur est l'ajout de s dans la troisième composante de l'isotopie $I(se)$ ($I_{aff}(se)$).

Toujours de façon duale, le retrait d'un sème, en tant qu'afférent à un sémème, se fait par $-i(s, se)$, avec suppression de s dans la troisième composante de $I(se)$.

2.2.5 Taxème - sème

On attribue un sème à un taxème via l'opérateur $+i(t, se)$, de façon similaire au cas précédent. Notons que le véritable statut du sème se (microgénérique ou mésogénérique) dépend de la cardinalité de la deuxième composante de l'isotopie, $I_t(se)$. L'opérateur modifie en effet la fonction I de façon à ajouter t dans la seconde composante de $I(se)$. Si celle-ci contient par la suite plus d'un taxème, se sera mésogénérique (du moins pour la partie générique de l'isotopie), sinon il sera microgénérique, et sera alors *le* sème du taxème t.

On peut également soustraire un taxème d'une isotopie par l'opération $-i(t, se)$, si t est présent dans la seconde composante de $I(se)$.

2.2.6 Spécème - sème

On peut trivialement ajouter ou supprimer un spécème dans une isotopie, par les opérateurs $+i(sp, se)$ et $-i(sp, se)$, en modifiant la première composante de $I(se)$.

2.3 Opérateurs évolués

Nous allons ici définir des opérateurs d'un niveau plus élevé, intégrant certaines contraintes jusqu'ici laissées de côté. Nous obtiendrons ainsi une famille de «macro-opérateurs», utilisables par la suite pour les passages d'une structure stable à une autre.

2.3.1 Activation d'un spécème

Cette opération n'a pas pour but d'atteindre le respect d'une contrainte précise, mais nous facilitera plutôt la tâche par la suite, lorsque nous aurons à gérer l'ensemble des contraintes liées aux graphes d'oppositions des taxèmes. Nous allons donc définir l'opérateur $activ(sp)$, destiné à attribuer un sème à un spécème désactivé. Nous partons donc pour cela d'un spécème existant ($sp = (s_1, s_2)$) mais absent de toute isotopie. Supposons également que s_1 et s_2 sont deux sémèmes du taxème t, de sèmes génériques se_t et éventuellement se'_t (son sème mésogénérique s'il existe). L'interprète doit donc à ce stade proposer un sème, afin de l'affecter à sp. Concrètement, il propose une chaîne de caractères, qui peut être déjà un sème.

Nous utilisons donc un appel à une fonction de choix, $spec(s_1, s_2)$, qui nous renvoie la chaîne désirée, ch. Si le sème de forme ch existe déjà, et que ce sème se est différent de se_t et de se'_t, et qu'enfin l'isotopie $I(se)$ ne contient pas le spécème (s_2, s_1) (le spécème opposé de sp), il ne reste qu'à appliquer $+i(sp, se)$.

Si toutefois le sème choisi comportait le sémème s_1 dans sa troisième composante (partie afférente). Cette information n'est plus utile, elle est même redondante. Le sème se étant déclaré spécifique (donc inhérent) pour s_1, il faut donc enlever ce dernier de la partie afférente de $I(se)$ par $-i(s_1, se)$. Il est important de noter que cette transformation est plus riche en information que l'état précédent, puisque l'on sait maintenant par opposition à quel sémème le sème se_i qualifie s.

Si la chaîne proposée ne correspond à aucun sème, il faut créer ce sème, par $+se$, et appliquer sans contrainte $+i(sp, se)$.

2.3.2 Cas particulier pour l'identité des sémèmes

Discutons ici d'un point central de notre approche, qui concerne la notion délicate d'identité des sémèmes. Nous avons vu au précédent chapitre l'aspect fondamental de la relation d'identité sur l'ensemble des sémèmes. Pour prendre en compte les variations de sens par des normes locales, nous avons envisagé le cas de deux sémèmes distincts bien que leur forme (chaîne extraite du texte) soit identique, ou similaire à quelques variations flexionnelles près. La notion qui les distingue est alors celle d'épisémème. Si le même mot apparaît deux fois dans un même texte, il y aura deux épisémèmes distincts, et donc, à un certain moment de l'analyse, deux sémèmes (entités formelles) distincts. Or, il se peut très bien que l'interprète n'envisage pas de distinction de sens entre ces deux occurrences : en ce cas, il sera bien gêné de les distinguer par un sème spécifique. Nous introduisons donc une possibilité de fusionner ces deux sémèmes, qui n'ont aucune distinction sinon de référence dans la chaîne du texte (épisémèmes).

Cette opération s'insère alors dans l'opérateur d'activation des spécèmes, sous une condition : à ce stade de l'interprétation, aucune distinction sémantique ne doit exister entre les deux sémèmes. Ils doivent en effet être identiques au sens de id. À partir de là, si l'interprète le souhaite, il peut choisir de fusionner ces deux sémèmes plutôt que de les distinguer par un sème spécifique. Dans ce cas, les extensions de ces deux sémèmes (en termes d'épisémèmes) sont fusionées, et un des deux sémèmes est éliminé.

Notons que cette opération peut se faire même si les lexies correspondantes dans le texte sont distinctes, il s'agit alors d'un cas de synonymie locale absolue. Par exemple, si dans un même taxème se trouvent un sémème correspondant à 'chat', et un autre correspondant à 'chats', la fusion est possible.

Le cas «par défaut» est présenté dans la figure 4.1, avec quatre épisémèmes, dont deux de chaînes similaires. Par défaut, à chacun de ces épisémèmes est associé un sémème.

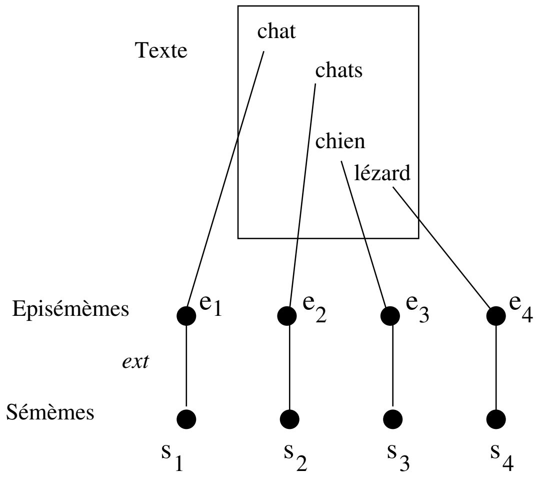


FIG. 4.1 – *Association sémème - épisémème par défaut*

Plus tard, un taxème sera mis en place, contenant les quatre sémèmes précédemment cités, le taxème des //animaux familiers//. Lors de la spécification des oppositions au sein de ce taxème, il sera demandé à l'utilisateur de distinguer, entre autres, 'chat' de 'chats', comme le montre la figure 4.2, par l'opérateur *activ*('chat', 'chats'). On suppose les deux sémèmes concernés égaux par ailleurs (ils n'ont pas de sèmes afférents, ou les ont en commun).

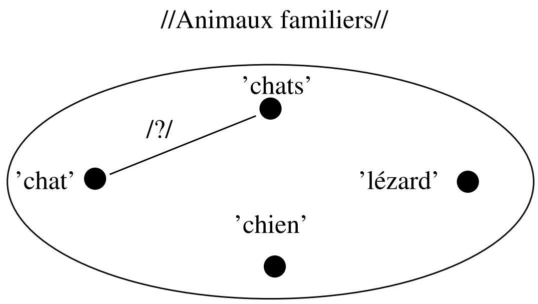


FIG. 4.2 – *Taxème proposant une fusion de sémèmes*

Si aucune distinction n'est apportée par l'utilisateur, et qu'il décide la fusion, le second sémème disparaît, et la structure devient donc elle présentée dans la figure 4.3.

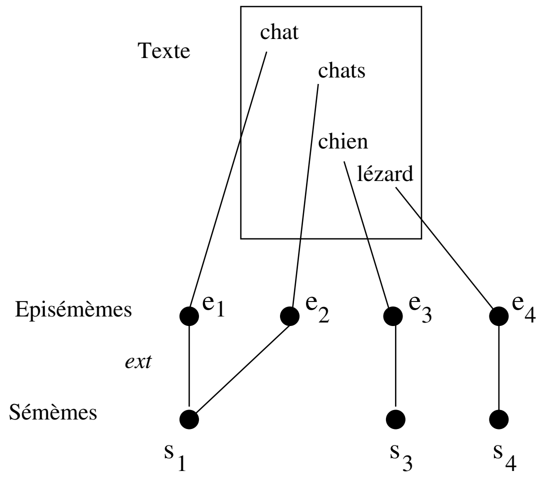


FIG. 4.3 – *Association sémème - épisémème après fusion*

2.3.3 Complétude du graphe d'opposition d'un taxème

Une des contraintes les plus fortes concerne le statut du graphe d'opposition d'un taxème. Rappelons que l'ensemble des spécèmes activés opposant les sémèmes d'un taxème doit former un graphe faiblement complet. La seconde contrainte est l'obligation d'avoir au moins un spécème activé partant de chaque sémème du taxème. Cependant, nous verrons que lors de la création et du «remplissage» d'un taxème, cette contrainte peut être vérifiée aisément. Il ne reste donc qu'à nous intéresser à la faible complétude.

Nous allons donc décrire l'opérateur $comp(t)$, qui va satisfaire cette contrainte, en appliquant l'opérateur précédent, d'activation d'un spécème.

Soit le taxème t, comportant les sémèmes $s_1, ..., s_n$. Pour chaque couple de sémèmes s_i, s_j, $i \neq j$, il faut donc vérifier si (s_i, s_j) ou (s_j, s_i) est activé. Si ce n'est pas le cas, il faut donc activer l'un des deux. Le choix du spécème à activer est arbitraire, disons (s_i, s_j), avec $i < j$. Il suffit donc d'appliquer l'opérateur $activ(s_i, s_j)$.

Une fois tous les couples envisagés, la contrainte est vérifiée.

2.3.4 Contrainte de l'afférence pure

Nous appelons contrainte de l'afférence pure l'axiome qui interdit à une isotopie quelconque de n'avoir que sa troisième composante non vide. Un sème ne peut en effet être qu'afférent. Cependant, un tel état de fait peut apparaître lors de la suppression de certaines entités. Par exemple, si l'isotopie du sème /liquidité/ était réduite à $(\{sp\}, \emptyset, \{s\})$ avec $sp =$ ('pleurs', 'sanglots') et $s =$ 'larmes', et que l'on élimine le spécème sp, l'isotopie se ne satisfait plus cette contrainte.

Comment résoudre ce problème ?

Deux façons sont envisageables : soit la suppression pure et simple de l'isotopie, avec effacement de l'afférence pour le sémème s, soit le «remplacement» de la partie inhérente, par exemple par la création du taxème //liquidité//. La première méthode est la plus simple à gérer formellement, mais appauvrit la structure, alors que la seconde, plus complexe, permet d'enrichir l'analyse du texte considéré. Nous choisirons la seconde possibilité.

Il nous faut donc créer un nouveau taxème, qui n'existe bien entendu pas, puisque le sème n'est pas utilisé de façon inhérente. Nous nous contenterons pour cette opération d'un taxème minimal, à deux éléments.

Nous nommerons la vérification de cette contrainte $pb_{aff}(se)$, qui se décompose en la série d'opérateurs suivante, qui s'applique dans les cas où $I(se) = \{\emptyset, \emptyset, X\}$ avec $X \neq \emptyset$:

$+t$	création du taxème t vide
$+s_1$	création du premier sémème, lexicalisé ou non
$+s_2$	création du deuxième sémème, lexicalisé ou non
$+i(t, se)$	déclaration du sème se comme microgénérique du taxème t
$+sp(s_1, s_2)$	création du premier spécème
$+sp(s_2, s_1)$	création du deuxième spécème
$activ(s_1, s_2)$	activation du premier spécème
$activ(s_2, s_1)$	activation du deuxième spécème

Le taxème t vérifie bien évidemment les deux contraintes de caractérisation positive pour tous ses sémèmes, et de complétude (forte dans le cas d'un taxème à deux éléments).

2.4 Conclusion

Nous avons établi ici quelques opérateurs fort utiles pour la suite de notre présentation. Nous verrons par la suite d'autres opérateurs du même type, mais quoiqu'il en soit ceux-ci nous suffisent pour l'instant pour décrire la phase de mise en place de la structure, ce que nous allons développer maintenant. Nous commencerons par décrire la notion de pré-interprétation, afin d'expliciter les données sur lesquelles nous travaillerons.

3 La pré-interprétation

3.1 Pré-isotopies

La notion de pré-isotopie introduit à ce stade de notre développement le principe d'initialisation de la structure présentée jusqu'ici. Au sein d'un processus d'interprétation, la pré-isotopie nous servira à capter formellement les intuitions qui conditionnent toute approche d'un texte. Bien évidemment, l'identification d'une pré-isotopie elle-même est la résultante d'une première interprétation, mais reste indispensable en tant qu'amorce d'un processus rationnalisant. Il serait de toute façon illusoire de cerner l'origine absolue d'une interprétation. Nous nous situons en un point de départ arbitraire d'un processus infini, qui n'a de justification que comme établissement nécessaire des conditions formelles minimales de notre modélisation.

Formellement, une pré-isotopie sera une association directe non contrainte entre un ensemble d'épisémèmes et un sème. Nous pouvons donc la définir, à l'instar de l'isotopie, par le biais d'une fonction pré-isotopante PI_E. Avec :

$$PI_E \ : \ SE \mapsto \mathcal{P}(E)$$

Une pré-isotopie sera donc l'image d'un sème se par PI_E, $PI_E(se)$, un sous-ensemble de E.

3.2 De l'épisémème au sémème

Si l'on désire rapprocher la notion de pré-isotopie d'une des formes que nous avons envisagées pour l'isotopie, nous devons pour cette dernière considérer la projection sur E. Dans une telle comparaison, la notion de sémème est donc absente. La structure d'une pré-isotopie n'est donc que forme, simple repérage de zones pertinentes dans un texte, avec bien entendu l'attribution d'un sème à des groupes de telles formes. Nous souhaitons en effet repousser l'apparition de la notion de sémème à la phase de mise en place de la structure complète. Nous verrons en détail cette mise en place par la suite. À ce stade donc, l'ensemble E, support de la fonction pré-isotopante PI_E, est décorrélé de S. La forme d'un épisémème peut bien entendu être trouvée par repérage de la chaîne de caractères correspondant aux coordonnées de l'épisémème en question dans le texte.

Nous allons ainsi définir une autre fonction de pré-isotopie, PI qui, elle, associe un ensemble de *sémèmes* à un sème :

$$PI : SE \mapsto \mathcal{P}(S)$$

Nous retrouvons donc, comme lien entre PI_E et PI, une relation précédemment définie : ext. Nous avons donc la relation suivante :

$$PI_E = ext \circ PI$$

Pour ce qui est de la pré-interprétation, la fonction ext est définie de façon minimale, associant un seul et unique épisémème à un sème. Le parcours interprétatif que nous allons maintenant décrire précisera également cette relation entre sémèmes et épisémèmes.

3.3 Parcours interprétatif

On appellera parcours interprétatif une suite d'opérateurs permettant de passer d'un **ensemble** de pré-isotopies à un **ensemble** d'isotopies. Il s'agit donc d'une opération globale. Ce parcours contient donc, entre autres, des transformations unitaires d'**une** pré-isotopie en **une** isotopie. La nature de cette transformation unitaire dépend principalement des relations qu'entretiennent les pré-isotopies initiales entre elles. Citons, entre autres, la conservation d'une pré-isotopie (l'extension de l'isotopie résultante est identique), sa spécification (on a découpé une pré-isotopie en plusieurs autres, et on obtient autant d'isotopies), son extension, et surtout la génération d'un certain nombre d'isotopies nouvelles (auxquelles ne correspondent pas de pré-isotopies).

Une caractéristique principale de ce processus est l'intégration progressive des contraintes gérant la structure sémantique. Nous avons vu le nombre et la force de celles-ci au chapitre précédent, et nous avons également déclaré l'absence de toute contrainte en ce qui concerne la fonction PI_E. Le processus interprétatif, du point de vue de la machine, se traduira donc par la satisfaction de ces règles. Du point de vue de l'interprète, il s'agira d'un approfondissement progressif de ses intuitions initiales qui le poussera à repérer de nouvelles relations sémiques, et à clarifier sa vision de l'organisation sémantique du texte.

4 De la pré-interprétation à l'interprétation

Dans cette partie nous allons exposer les différents processus de mise en place de la structure définie au chapitre précédent.

Voici donc les principales étapes et leurs caractérisations formelles.

4.1 Point de départ

Sont connus, à l'origine, les ensembles et relations suivants :

- E, ensemble des épisémèmes extraits du texte interprété.
- SE, ensemble des sèmes repérés.
- PI_E, fonction pré-isotopante, reliant les éléments de E à ceux de SE. Aucune contrainte du type inclusion ou intersection ne régit cette fonction.
- Les ensembles S, T, SP sont vides et les fonctions PI, ext et I n'ont pas d'extension.

4.2 Identification des sémèmes

Une première étape fondamentale est de créer le niveau de base de notre structure sémantique, à savoir S, ensemble des sémèmes. Nous ne disposons pour l'instant que d'un ensemble E d'épisémèmes, reliés par PI_E à des sèmes. Il est important de noter que tout épisémème présent dans E est présent dans au moins une pré-isotopie.

Nous allons donc initialiser l'ensemble S ainsi que la fonction ext, en associant par défaut un sémème différent (par l'opérateur $+s$) à chaque épisémème : S est donc isomorphe à E et ext est une bijection entre ces deux ensembles. Deux sémèmes de même forme (chaîne extraite du texte à l'aide de la position de l'épisémème) peuvent donc être déjà différenciés. De plus, aucune distinction sémantique (sèmes attribués à ces sémèmes par PI) ne permet de justifier cette différence. Dans le cas où l'interprète souhaite d'ores et déjà déclarer deux épisémèmes comme extensions du *même* sémème, nous allons définir un opérateur d'assimilation : $assim : E \times E \mapsto E$. La seule contrainte d'application est le fait qu'un des deux épisémèmes doive être la seule image par ext du sémème qui lui correspond *i.e.* $\{e_1\} = ext(ext^{-1}(e_1))$ ou idem pour e_2.

Si l'interprète désire assimiler e_1 et e_2, nous appliquons $assim(e_1, e_2)$, qui modifie la structure de la façon suivante :

- Vérification de l'identité sémantique des deux épisémèmes, c'est-à-dire $PI_E^{-1}(e_1)$ et $PI_E^{-1}(e_2)$. S'il y a distinction, l'assimilation est impossible.
- Choix d'un épisémème entre e_1 et e_2. Si un des deux épisémèmes n'est pas la seule et unique extension du sémème qui lui correspond, c'est l'autre épisémème qui est choisi. Considérons pour la suite que c'est e_2 qui est choisi.
- Attribution de cet épisémème au sémème $ext^{-1}(e_1)$ par $+e(ext^{-1}(e_1))$. Le sémème correspondant à e_1 voit donc son extension sur E augmentée d'un élément.
- Suppression du sémème $ext^{-1}(e_2)$ par $-s(ext^{-1}(e_2)$. L'information de sa nature sémantique n'est pas perdue, puisqu'il possédait les mêmes sèmes que le sémème correspondant à e_1, et que les informations positionnelles ont été *sauvegardées* dans le sémème de e_1.

Après une seule assimilation de ce type, la fonction ext perd son caractère bijectif. Quoiqu'il en soit, la définition de S et de ext nous permet de manipuler par la suite la fonction PI, sans nous préoccuper la plupart du temps des épisémèmes.

Toutefois, rappelons que la fusion possible des sémèmes lors de l'activation des spécèmes reprend ces notions, et qu'elle permet également l'assimilation des entités.

Nous allons maintenant explorer la phase suivante de l'activation de la structure, en commençant par le niveau des taxèmes.

4.3 Partie générique

4.3.1 Création des taxèmes et isotopies associées

Pour toute pré-isotopie $PI(se)$ qui n'inclut pas une autre pré-isotopie (au sens de S), on crée un taxème t dont l'extension est celle de la pré-isotopie (toujours au sens de S).

Soit, $\forall se \in SE$ tel que $\nexists se' \in SE$ tel que $PI(se') \subset PI(se)$, on applique donc la suite d'opérateurs :

$+t$	Création du taxème vide
$\forall s_i \in PI(se), add(s_i, t)$	Remplissage du taxème
$+i(t, se)$	Attribution du sème au taxème

À ce stade, toutes les pré-isotopies minimales au sens de S deviennent donc des taxèmes. Le critère de minimalité propre au taxème est donc vérifié de façon simple : le taxème est bien une classe sémantique n'en contenant aucune autre. Les contraintes inhérente à T ne sont pas encore toute respectées, puisque des intersections entre taxèmes sont encore possibles. La fonction I est en partie précisée, et conserve à ce stade l'association sémème-sème présente dans PI (sans pour autant la comprendre intégralement).

Si une pré-isotopie n'est pas minimale et que ses sous-pré-isotopies ne couvrent pas tous ses sémèmes, on crée artificiellement une nouvelle pré-isotopie «générique», dont l'extension est le reliquat de la sur-pré-isotopie. Par exemple, si l'on dispose de la pré-isotopie /religion/, contenant les sémèmes 'église', 'mosquée', 'évêque', 'muezzin' et 'foi', et qu'on la découpe en /christianisme/ ('église' et 'évêque') et en /islam/ ('mosquée' et 'muezzin'), il reste 'foi', que l'on place dans une nouvelle pré-isotopie, baptisée /religion générique/. cette dernière pré-isotopie est donc minimale et donnera également lieu à la création d'un taxème.

Soit, si $\exists se \in SE$ tel que $\forall se_i \in SE$ t.q. $PI(se_i) \subset PI(se)$, $\cup_i PI(se_i) \neq PI(se)$, on applique :

$+t$	création d'un nouveau taxème vide
$\forall s \in PI(se) \setminus \cup_i PI(se_i) : add(s, t)$	remplissage du taxème
$+se'$	création du sème « se générique »
$+i(t, se)$	attribution du sème se au taxème t

Exemple : La figure suivante (4.4) exemplifie la première étape de création des taxèmes à partir d'une pré-interprétation. On peut y voir trois taxèmes : //Christianisme//, //Islam// et //Économie//. Le sème /Religion/ ne correspond pas à une classe terminale et ne donne donc pas lieu à un taxème.

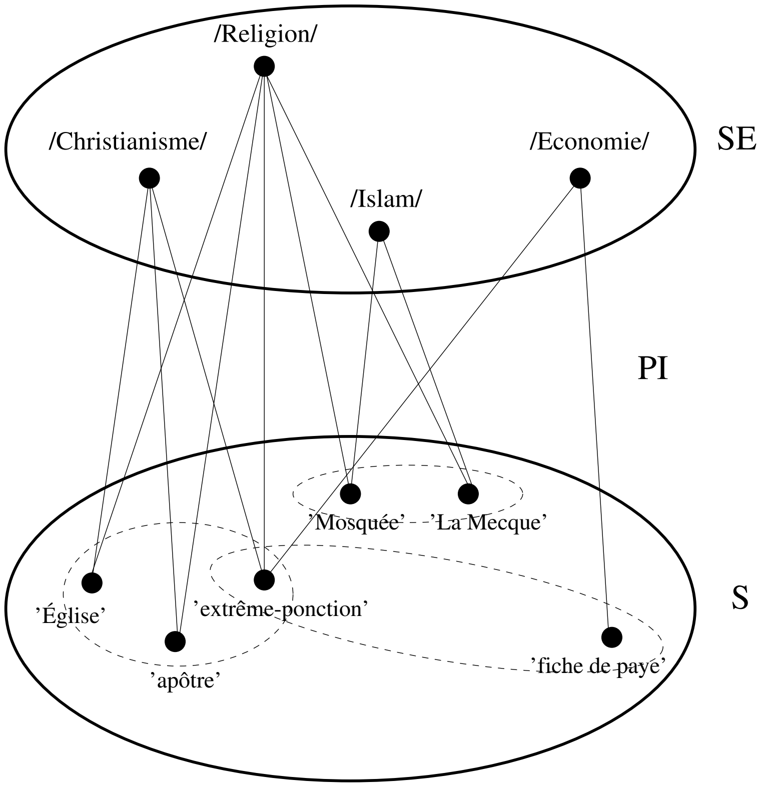


Fig. 4.4 – *Création des taxèmes*

Après cette étape, on dispose de trois taxèmes et de trois isotopies, correspondant aux trois sèmes qualifiant les taxèmes :
$t_1 =$ { 'Église', 'apôtre', 'extrême-ponction' }
$t_2 =$ { 'Mosquée', 'La Mecque' }
$t_3 =$ { 'fiche de paye', 'extrême-ponction' }
et :
I (/Christianisme/) $= (\emptyset, \{t_1\}, \emptyset)$
I (/Islam/) $= (\emptyset, \{t_2\}, \emptyset)$
I (/Économie/) $= (\emptyset, \{t_3\}, \emptyset)$

4.3.2 Création des domaines et isotopies associées

Il faut maintenant traiter le cas des pré-isotopies non terminales, qui contiennent donc les cas précédemment traités. Pour chaque isotopie terminale qu'une telle pré-isotopie «contient» (au sens de l'inclusion de leurs extensions dans S), on crée une isotopie comportant les taxèmes de ces pré-isotopies terminales.

Note : On considère à ce stade qu'il n'y a, dans l'organisation de PI qu'un seul niveau d'inclusion : soit une pré-isotopie est terminale, soit elle ne contient que des pré-isotopies terminales.

Soit, $\forall se \in SE$, tel que $\exists se' \in SE$ tel que $PI(se') \subset PI(se)$, si $t_{se'}$ est le taxème correspondant au sème se', on applique l'opérateur $+i(t_{se'}, se)$.

À ce stade, l'intégralité de la structure organisée par PI est traduite par l'ensemble T et la fonction I. Les ensembles S et SE sont inchangés.

Aucun nouveau taxème n'est créé durant cette étape. Sont générées uniquement des isotopies servant à caractériser les domaines.

Exemple : La suite de notre exemple consiste durant cette étape en la création de l'isotopie /Religion/, portant sur les deux taxèmes //Christianisme// (noté t_1) et //Islam// (t_2) :
I(/Religion/)$= (\emptyset, \{t_1, t_2\}, \emptyset)$

4.3.3 Gestion des contraintes sur T

Comme nous l'avons dit précédemment, toutes les contraintes portant sur l'ensemble T des taxèmes ne sont pas encore respectées. À ce stade la partition de S n'est qu'en partie réalisée : aucun élément de T n'en contient un autre, mais certains sémèmes peuvent encore appartenir à plusieurs taxèmes. C'est donc cet état de fait que l'on doit à présent traiter.

Pour chaque sémème de S, s'il appartient à plus d'un taxème, il faut utiliser une fonction de décision représentant le choix de l'interprète pour n'en choisir qu'un parmi ceux-là. Cette fonction de choix, pour traiter le recouvrement est désignée par F_r, définie sur $S \times \mathcal{P}(T) \rightarrow T$. Si $F_r(s, t_1, ..., t_n) = t_i$, avec $s \in t_1, ..., s \in t_n$, alors s n'appartiendra qu'au taxème t_i. Par contre, le rapport entre le sémème et les sèmes des taxèmes, tel qu'il était exprimé par la fonction PI, doit être conservé dans I, via le phénomène (ici détourné) de l'afférence. Plusieurs opérations en découlent ainsi :

- t_i est conservé tel qu'il l'était après les étapes précédentes

- Par contre, s est retiré des autres taxèmes $t_1, ..., t_{i-1}$ et $t_{i+1}, ..., t_n$ par l'application de $\forall j \in \{1, ..., (i-1), (i+1), ..., n\}$ $sub(s, t_j)$.

- Enfin, s doit être rajouté dans la partie afférente des isotopies contenant les taxèmes dont il a été retiré.
  Soit $\forall j \in \{1, .., (i-1), (i+1), ..., n\}, +i(s, se_j)$ si se_j est le sème microgénérique du taxème t_j. Ces isotopies prennent la forme : $(\emptyset, \{t_j\}, \{s\})$.

À ce stade, l'ensemble des taxèmes T est effectivement une partition de S. Le problème du recouvrement vient d'être réglé, et S n'ayant toujours pas été modifié depuis la pré-interprétation, pas plus que les rapports entre S et SE, tout sémème appartient bien à un taxème. La seule contrainte à vérifier est donc la cardinalité des taxèmes remaniés par l'afférence : en effet, rien ne garantit plus le nombre minimum d'éléments. Il faut donc à ce stade exiger la complétion des taxèmes ne possédant qu'un élément. Les sémèmes exigés par cette opération peuvent être extraits du texte, ou bien au contraire être repérés «en langue», il s'agit alors de sémèmes non lexicalisés. Le rôle de ces derniers est de servir de référence pour expliciter par la suite les rapports (via les sèmes spécifiques) d'un sémème isolé dans le texte avec d'autres éléments d'une de ses classes définitoires.

Soit : $\forall t \in T$, si $|t| = 1$ on applique $+s$ et $add(s, t)$.

Une fois les taxèmes passés en revue, leur cardinalité est bien entendu supérieure ou égale à deux, et les contraintes définies sur l'ensemble T sont toutes vérifiées.

Exemple : Le seul sémème de notre exemple qui appartient à deux taxèmes est 'extrême-ponction', déclaré comme appartenant à //Christianisme// et à //Économie//. La fonction de choix appliquée à cet élément doit donc sélectionner un des deux taxèmes. Étudions les deux cas :

- Choix de //Économie//. Le taxème //Christianisme// ne contient plus dès lors que deux sémèmes ('Église' et 'apôtre'), et //Économie// contient maintenant 'extrême-ponction' en plus de 'fiche de paye'. De plus, le sème /Christianisme/ est déclaré afférent à 'extrême-ponction', et son isotopie devient donc : $I(/\text{Christianisme}/) = (\emptyset, \{t_1\}, \{'\text{extrême-ponction'}\})$

- Choix de //Christianisme//. Le taxème //Christianisme// reste inchangé avec trois sémèmes. Par contre, le taxème //Économie// ne contient plus qu'un seul sémème, et il doit être complété. On peut le faire en y ajoutant le sémème (non lexicalisé) 'salaire'. De plus,

'extrême-ponction' est placé dans la partie afférente de l'isotopie de /Économie/, qui devient :
I(/Économie/)= $(\emptyset, \{t_3\}, \{'\text{extrême-ponction'}\})$
avec $t_3 = \{$'salaire', 'fiche de paye'$\}$.

4.3.4 Résultat du traitement générique de la pré-interprétation

Résumons ici l'état de la structure sémantique telle qu'elle est élaborée à ce stade de l'analyse. Tout ce qu'a fait le lecteur jusqu'ici est de définir un ensemble d'extraits du texte interprété et de classifier ces derniers. Sans vouloir nous aventurer dans les méandres des caractérisations psychologiques ou même cognitives de l'opération de classification, contentons-nous de qualifier ces opérations «pré-interprétatives» de *globales*. Argumentons tout de même : il est clair que ces classes sémantiques ne sont établies et nommées que par corrélation entre les différents éléments repérés. Bien pire que cela, l'établissement d'une classe va inciter le repérage d'autres extraits du texte. Il est donc possible que la globalité affirmée ne soit qu'un masque de la complexité combinatoire des opérations de lecture à ce stade. Le point positif dans notre qualification est tout de même la part d'*aprioris* que l'on trouve dans les classes initialement repérées. Par exemple, pour un poème de Verlaine comme celui que nous présentons en annexe («En sourdine»), les thèmes propres au romantisme se retrouvent comme classes initiales (//mort//, //amour//, //nature//, //temps//, etc.). Mais il s'agit bien entendu, non pas de notions globales, mais de notions *plus* globales. Au vu de la suite des opérations, ou des notions, émergeront de considérations portant sur un petit ensemble d'éléments du texte : les sèmes et isotopies spécifiques, dont nous allons à présent expliciter la mise en place.

4.4 Partie spécifique

Nous devons donc préciser les modes de repérages des isotopies spécifiques, se basant sur les spécèmes, couples de sémèmes définis, durant les précédentes étapes, comme appartenant à un même taxème. Nous allons donc préciser comment ces entités sont mises en place pour un taxème. Nous avons peu de contraintes quant au nombre maximal de spécèmes activés, comme nous l'avons exprimé au chapitre précédent, aussi nous intéresserons-nous dans la suite à l'atteinte de l'activation minimale des spécèmes pour un taxème.

4.4.1 Création de SP

Jusqu'ici, l'ensemble SP des spécèmes était vide. Les propos que nous allons tenir par la suite sur les spécifications des taxèmes concerneront le phénomène d'activation des spécèmes. Encore faut-il que ceux-ci existent. Il suffit simplement, pour chaque taxème, de créer l'ensemble des spécèmes susceptibles d'être activé. Si un taxème t a pour extension $\{s_1, ..., s_n\}$, on crée les spécèmes correspondant à tous les couples possibles de (s_i, s_j) avec $s_i, s_j \in t$, $s_i \neq s_j$ par application de l'opérateur $+sp(s_i, s_j)$.

4.4.2 Traitement des caractérisations positives

Afin de vérifier la première de ces contraintes, nous proposons à l'interprète le protocole suivant :

- Pour tout sémème s du taxème t, celui-ci doit sélectionner un sous-ensemble non vide de $t \setminus \{s\}$. Ceci se fait par un appel à une fonction de choix, sollicitant ainsi l'interprète : $select(t \setminus \{s\})$. Rappelons que la cardinalité de t est supérieure ou égale à 2, ce sous-ensemble n'est donc pas vide.
- Puis, pour chaque élément s_i de ce sous-ensemble, on applique l'opérateur d'activation du spécème (s, s_i), $activ(s, s_i)$.

4.4.3 Traitement de la complétude faible

À ce stade, le graphe n'est pas forcément complet, mais de chaque nœud est issu au moins un arc. Il ne reste donc plus qu'à compléter le graphe (faiblement), par l'application de $comp(t)$, pour tout taxème t de T. À ce stade, bien entendu, le graphe d'opposition est faiblement complet, et de chaque nœud est bien issu un arc.

4.5 Conclusion

Avec ces opérations, toutes les contraintes énumérées au chapitre précédent sont respectées, la structure sémantique globale de l'interprétation est stabilisée.

Il nous reste maintenant à envisager les différentes modifications qui peuvent s'appliquer à une telle structure stabilisée. Nous allons distinguer pour cela deux types de modifications :

- Ajout ou retrait d'éléments, soit des modifications locales de la structure. Nous envisagerons alors les conséquences de telles actions sur l'ensemble des entités.

– Modifications globales, cherchant à réorganiser l'ensemble de la structure. Nous verrons plus loin l'utilité de telles transformations. Contentons-nous d'annoncer que ces transformations concerneront, au départ, la nature des isotopies.

5 Modification locales

Dans cette partie, nous allons encore envisager les opérations d'ajout et de retrait d'éléments dans la structure, mais cette fois-ci, en tenant compte de l'ensemble des contraintes de cette dernière. Nous ne nous intéresserons donc qu'à certains types d'entités, principalement les sémèmes, les sèmes et les taxèmes. En effet, les spécèmes et les isotopies sont des entités plus «intermédiaires» dont la manipulation directe a peu d'intérêt interprétatif.

Nous exprimerons bien entendu ces nouveaux opérateurs «stabilisés» à l'aide des opérateurs minimaux présentés en début de chapitre, ainsi qu'à l'aide des trois opérateurs plus évolués présentés plus haut. Rappelons une dernière fois que toutes les considérations que nous allons prendre en compte se basent sur une structure dont *toutes* les contraintes sont vérifiées.

5.1 Repérage d'un sémème et placement dans un taxème

Cette opération consiste en la définition d'un nouveau sémème, puis en son placement dans un taxème déjà existant. Nous nommerons cette opération $ADD(s, t)$.

On obtient ainsi la suite d'opérateurs suivante :

$+s$	Repérage du nouveau sémème
$add(s, t)$	Placement de ce sémème dans le taxème t
$\forall s_i \in t \setminus \{s\}$, $+sp(s, s_i)$ et $+sp(s_i, s)$	Création des nouveaux spécèmes du taxème t
$select(t \setminus \{s\})$	Choix par l'interprète d'un sous-ensemble des sémèmes de t distincts de s
$\forall s_i \in select(t \setminus \{s\})\ activ(s, s_i)$	Activation des spécèmes sélectionnés précédemment
$comp(t)$	Vérification de la complétude du graphe d'opposition de t

Note : Cette définition satisfait aussi la complétion d'un taxème vide ou ne contenant qu'un seul sémème ; nous verrons plus loin un cas de ce type d'utilisation.

5.2 Élimination d'un sémème

On désire ici éliminer purement et simplement un sémème de l'ensemble de la structure. Nous nommerons cette élimination $elim(s)$. Soit l'opération portant sur le sémème s du taxème t, nous obtenons la suite d'opérateurs suivante :

$\forall s_i \in t \setminus \{s\}$ et $\forall se \in SE$ t.q. $(s, s_i) \in I_{sp}(se), -i(sp, se)$ et $pb_{aff}(se)$	Retrait des spécèmes issus de s de leurs isotopies, et vérification de l'afférence pure de celles-ci
$\forall s_i \in t \setminus \{s\}\ \ -sp(s, s_i)$	Élimination des spécèmes issus de s
$\forall s_i \in t \setminus \{s\}$ et $\forall se \in SE$ t.q. $(s_i, s) \in I_{spe}(se), -i(sp, se)$ et $verif(se)$	
$\forall s_i \in t \setminus \{s\}\ \ -sp(s_i, s)$	Mêmes opérations pour les spécèmes dont s est la seconde projection
$\forall se \in SE$ t.q. $s \in I_{aff}(se),\ -i(s, se)$	Retrait de s des isotopies de ses sèmes afférents
$-s$	Élimination du sémème s
Si $\|t\| = 1$, $ADD(s', t)$	Si le taxème t ne contient plus qu'un seul sémème, identification et ajout d'un nouveau sémème s'.

5.3 Création d'un nouveau taxème

Nous allons répéter ici quelques opérations prenant place à la suite de la pré-interprétation, afin de créer un nouveau taxème t, de sème microgénérique se, et de remplir celui-ci avec deux nouveaux sémèmes. Cette opération consiste en la suite d'opérateurs suivante :

$+t$	Création du nouveau taxème t
$+se$	Création du nouveau sème se (à moins qu'il n'existe déjà comme sème purement spécifique)
$+i(t, se)$	Déclaration de se comme sème microgénérique de t
$ADD(s, t)$	Identification du nouveau sémème s et placement de celui-ci dans t
$ADD(s', t)$	Identification du nouveau sémème s' et placement de celui-ci dans t

Note : C'est dans ce cas précis que nous abusons de la définition de l'opérateur ADD. Sa première utilisation (premier sémème ajouté) ne crée en effet aucun spécème, et n'en active aucun : elle a le même effet que add. La seconde application effectue toutes ces opérations.

5.4 Identification d'un nouveau sème spécifique

Cette opération consiste en la déclaration d'un nouveau sème se comme sème spécifique attribué au sémème $sp = (s, s')$. Nous la nommons $add_{spe}(sp, se)$. La nature de cette opération dépend principalement de l'existence de se.

Si se n'existe pas :

$+se$	Déclaration du nouveau sème
$+i(sp, se)$	Attribution de se à sp

Si se existe déjà, les contraintes à vérifier sont :

- se ne doit pas être un des sèmes génériques des composantes de sp.
- se ne doit pas être attribué au spécème opposé à sp.

Si ces contraintes sont respectées, les opérations sont :

$+i(sp, se)$	Attribution du sème *se* à *sp*
Si $s \in I_{aff}(se),\ -i(s, se)$	Si *se* était afférent à *s*, élimination de cette afférence.

5.5 Déclaration d'une afférence

Cette opération très simple va simplement affecter le sème se comme afférent au sémème s. Les seules conditions sont les éventuelles attributions précédentes de se au taxème de s ou aux spécèmes liés à s. Il faut donc vérifier, si se existe, que :

- se n'est pas générique pour s, donc que t, taxème de s, n'est pas présent dans le seconde composante de $I(se)$.
- se n'est pas spécifique pour s, donc qu'il n'existe pas un sémème s' tel que (s, s') n'est pas présent dans la première composante de $I(se)$.

Sinon, il suffit d'appliquer $+i(s, se)$. Si se n'existe pas, il faut créer une nouvelle isotopie qui lui soit liée, et cette isotopie doit avoir une composante inhérente, et plus spécialement générique. Ce qui revient donc à résoudre la contrainte de l'afférence pure pour se. Soit :

$+se$	Création du sème *se*
$+i(s, se)$	Déclaration de l'afférence de *se* à *s*
$pb_{aff}(se)$	Gestion de l'afférence pure de l'isotopie de *se*

Note : Bien entendu, l'application de l'opérateur de gestion de l'afférence pure entraîne, à coup sûr, la création d'un nouveau taxème.

6 Modifications globales

Bien que nous ayons vu lors de l'étude la propagation des modifications que les sèmes, et donc les isotopies sont modifiées à partir de considérations locales (ajout d'un sémème par exemple), il n'en reste pas moins que les modifications les plus intéressantes se situent au niveau global de la structure sémantique. Plusieurs raisons nous poussent à considérer de telles modifications.

Nous nous intéresserons tout d'abord à la relativité du statut des sèmes. Nous avons déjà vu au second chapitre que les notions de sème générique, spécifique, inhérent et afférent (au sens de F. Rastier) ne sont pas directement captables par une approche locale (*i.e.* sur un ou plusieurs textes). Ces statuts sont en effet liés à la langue et à ses niveaux de systématicité. Concrètement, le fait de définir un taxème comme //absolu//, contenant des sémèmes aussi variés que 'extasiés', 'profond', 'à jamais' (voir l'analyse de «En sourdine» en annexe), n'implique absolument pas la validité de /absolu/ comme sème générique en langue. De même, attribuer le sème /oiseau/ ou /chant/ comme sème spécifique au sémème 'rossignol' au sein du taxème //nature// ne relève que de la localité. Nous n'espérons pas ici capter la norme sous-jacente, ni même décréter que /oiseau chanteur/ est un *vrai* taxème, mais nous devons tout de même nous donner les moyens de modifier *a posteriori* le statut d'une attribution de sème, ne serait-ce qu'à des fins de cohésion inter-textuelle, ou plutôt inter-interprétative.

Encore plus concrètement, nous voudrions pouvoir imposer à la structure de vérifier certains statuts d'origines diverses, et donc de modifier les qualifications isotopiques.

Ces opérations peuvent d'ailleurs rejoindre d'autres considérations, plus applicatives celles-ci. Il peut être envisageable en effet de fondre deux structures sémantiques distinctes, soit par la personne de l'interprète, soit par le texte interprété. Nombre de travaux en littérature s'intéressent par exemple à la description d'un ou plusieurs thèmes récurrents chez un auteur ou durant une période, ou même de façon transversale. Il convient donc d'effectuer une fusion de plusieurs structures sémantiques, pouvant avoir des sèmes (et des sémèmes) en commun. Par exemple, si pour une interprétation d'un texte A un sémème s appartient au taxème repéré par le sème se, et que pour un autre texte B, le même sémème possède le sème se comme spécifique, la fusion devra trancher entre les deux possibilités. Au bout du compte, nous ne voulons qu'une seule structure rendant compte des deux interprétations, *en gardant comme cohérences uniquement celles voulues par l'interprète commun*. Par exemple, si la symbolique de la couleur chez Stendhal est mise en exergue pour *Le rouge et le noir* mais pas pour *La Chartreuse de Parme*, les sémèmes 'rouge' et 'rouge' devront être différenciés par leurs textes d'extraction.

Dans de tels cas de modifications de statuts ou de fusions de structures, notre critère principal devra rester le respect de l'association sémème-sème, indépendamment de tout typage. Ce seront donc les données brutes qui devront être respectées au maximum, et l'organisation globale nous accordera alors les degrés de liberté nécessaires.

Enfin, nous verrons comment un tel remaniement de la structure peut donner des guides pour approfondir une interprétation, et bien rappeler que toute attribution de sens n'est que temporairement stabilisée. Nous allons donc envisager deux grandes familles de transformations applicables à la structure, en les décomposant dans un premier temps suivant les opérations élémentaires que nous venons d'expliciter.

6.1 Généricisation d'une isotopie spécifique : l'opérateur *igen*

Soit un sème se, tel que $I(se) = (\{sp_1, ..., sp_n\}, \emptyset, \{s'_1, ..., s'_m\})$. Nous appliquons à ce sème un transformateur $igen$ qui va créer le taxème t, de sème microgénérique se (par l'application de $+t$ et $+i(t, se)$). Cette transformation va consister en un déplacement des sémèmes correspondant aux premières projections des $sp_1, ..., sp_n$ vers le taxème t. Ainsi, le contenu initial de $I_{spe}(se)$ sera transferré dans $I_{gen}(se)$. Il y a donc conservation d'un des sèmes spécifiques de chacun de ces sémèmes (notons-les $s_1, ..., s_n$), puisqu'il s'agira, après application de $igen$, de leur sème microgénérique. Ce déplacement va donc consister en une suppression de ces sémèmes de leur ancien taxème t_i et en leur ajout dans t : on applique donc les opérateurs $sub(s_i, t_i)$ et $add(s_i, t)$. Cependant, il va nous falloir conserver une trace des autres sèmes de ces sémèmes avant leur suppression. La souplesse de notre définition de l'afférence ne pose aucun problème pour envisager le transfert d'un sémème, mais de nombreux problèmes se posent pour ce qui est des sèmes spécifiques (autres que le sème se, opérande de $igen$). Pour cela, deux cas sont à examiner.

6.1.1 Premier cas : pas d'isotaxémie

Dans le cas où aucun couple de spécèmes sp_i et sp_j n'étaient isotaxémiques (il n'existe pas deux sémèmes parmi les s_i qui appartenaient au même taxème), nous n'avons aucune opposition (s_i, s_j) à conserver. Seules nous intéressent alors les données brutes des sèmes spécifiques supportés par les sémèmes s_i.

On applique ensuite à ces sémèmes l'opérateur $sub(s_i, t_i)$, avec t_i leur ancien taxème. Il faut donc dans un premier temps vérifier la cardinalité des t_i, et la complétude de leurs graphes d'opposition.

On crée maintenant le taxème t de sème microgénérique se par $+t$ puis $+i(t, se)$

On replace les sémèmes s'_i dans t via l'opérateur $add(s_i, t)$, en interceptant par la suite la gestion des nouveaux spécèmes créés. C'est en effet lors de la spécification de ce nouveau taxème que nous tenterons de réinsérer les sèmes précédents. Nous nous trouvons donc dans le cas où, à chaque sémème s'_i de ce nouveau taxème, est associée une famille de sèmes se_i^j.

Dans un premier temps, nous nous occuperons des sèmes précédemment génériques des sémèmes s_i. Une possibilité pour leur traitement est de considérer que leur nature précédemment générique leur donne également un statut d'opposition au sein du nouveau taxème créé par *igen*. Prenons un exemple pour exprimer ce fait. Soit dans un premier temps les deux taxèmes //lieu de culte// et //homme religieux//, contenant respectivement 'mosquée', 'église' et 'muezzin', 'prêtre'. Considérons également que les premiers sémèmes de ces taxèmes supportent le même sème spécifique /islam/ (via des spécèmes dont la seconde composante est le second spécème de ces taxèmes). Nous appliquons donc le transformateur *igen* à /islam/. Le taxème //islam// nouvellement créé contient donc maintenant 'mosquée' et 'muezzin'. Les considérations précédentes sur la généralité des oppositions nous conduisent donc à attribuer automatiquement le sème /lieu de culte/ au spécème ('mosquée', 'muezzin') et /homme religieux/ au spécème opposé.

Un cas de transformation minimale sans isotaxémie est représenté dans la figure 4.5. Les sèmes précédemment génériques des sémèmes transposés deviennent automatiquement spécifiques.

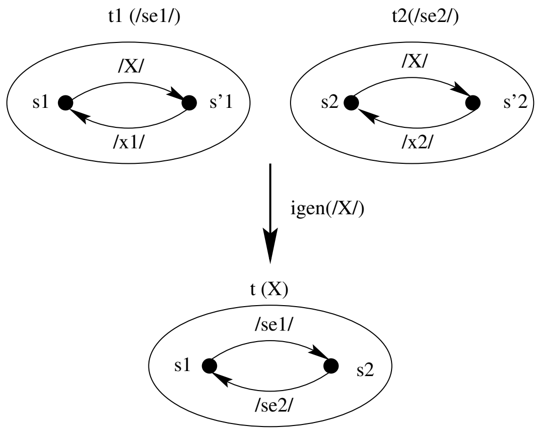


FIG. 4.5 – *igen sans isotaxémie*

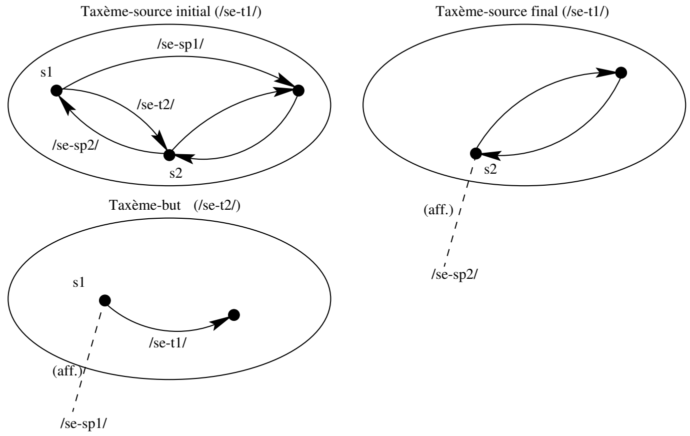


FIG. 4.6 – *Gestion des sèmes spécifiques par igen*

Ensuite, nous devons nous intéresser aux autres sèmes spécifiques initialement supportés par les sémèmes translatés, ainsi qu'aux sèmes spécifiques supportés par les spécèmes dont ils étaient les sémèmes opposants.

Pour les sèmes spécifiques directement supportés par les sémèmes translatés, nous avons toujours comme possibilité de recourir au phénomène de l'afférence, afin de ne pas perdre l'information initiale. Nous perdons simplement la notion d'opposition, mais la liste des sèmes supportés par le sémème n'est pas modifiée. Dans ce cas, nous devons également nous poser le problème de la résolution d'un cas d'afférence pure pour le sème concerné : cela peut donc mener à la création d'un nouveau taxème.

Pour les sèmes spécifiques dont le spécème avait le sémème translaté comme seconde projection, puisque ce spécème devra être détruit durant l'opération de transfert, nous allons voir comment y préserver l'information de ce sème. Deux cas se présentent encore ici : soit le sème considéré était également supporté par un autre spécème qui va, lui, rester intact. Dans ce cas aucune opération supplémentaire n'est à envisager. Soit il s'agissait de la seule occurrence de ce sème spécifique, auquel cas nous nous servirons encore une fois de l'afférence pour le conserver. Les deux cas sont visibles dans la figure 4.6.

Récapitulatif Pour résumer cette opération complexe, voici les différentes étapes et opérateurs associés, présentés dans le tableau de la page suivante.

On applique l'opérateur $igen$ au sème se, qui est initalement attribué aux spécèmes (s_i, s'_i). Les sémèmes s_i appartiennent respectivement aux taxèmes t_i (tous distincts) de sèmes microgénériques se_{ti}.

En conclusion, toutes les informations sont préservées, et il y a eu possibilité de repérer de nouveaux sémèmes et sèmes.

$+t$	Création du nouveau taxème.
$+i(t, se)$	Attribution du sème se au nouveau taxème.
$\forall s_i, add(s_i, t)$	Ajout des sémèmes dans le taxème.
$\forall s_i, s_j \in t, +sp(s_i, s_j)$	Création des spécèmes.
$\forall sp = (s_i, s_j), +i(sp, se_{t_i})$	Attribution des anciens sèmes microgénériques comme sèmes spécifiques.
$\forall s_i, s_i' \in t_i$, $\forall se' \in I_{sp}^{-1}(s_i, s_i')$, $-i((s_i, s_i'), se')$, $-sp(s_i, s_i')$, $+i(s_i, se')$, $pb_{aff}(se')$	Gestion des autres sèmes spécifiques initialement attribués aux sémèmes translatés : ils deviennent afférents à ces sémèmes.
$\forall s_i \in t$, $\forall s_i' \in t_i$, $\forall se' \in I_{sp}^{-1}(s_i', s_i)$, $-i((s_i', s_i), se')$, $-sp(s_i', s_i)$ et si $se' \notin I_{spe}^{-1}(s_i')$ alors $+i(s_i', se')$ et $pb_{aff}(se')$	Traitement des spécèmes dont le sémème translaté était la seconde projection : on supprime les attributions de sème à ces spécèmes avant de détruire les spécèmes eux-mêmes, et on déclare le sème comme afférent si nécessaire.
$\forall s_i \in t, \forall s' \in S, \forall sp$ t.q. $sp = (s', s_i)$, $\forall se$ t.q. $se \in I_{sp}^{-1}(sp)$, $-i(sp, se)$, $pb_{aff}(se)$, $-sp$ et $+i(s', se)$	Elimination des spécèmes initiaux liés aux sémèmes translatés, et suppression des attributions de sèmes sur ces spécèmes (avec traitement du problème de l'afférence pure pour les isotopies concernées, et déclaration d'afférence).
$\forall s_i \in t, sub(s_i, t_i)$	Suppression des sémèmes translatés de leur taxème d'origine.
$\forall t_i$, si $\|t_i\| = 1, +s$ et $ADD(s, t_i)$	Complétion des taxèmes initiaux si ceux-ci n'ont plus qu'un sémème.
$\forall t_i$, $comp(t_i)$	Vérification de la complétude faible pour les taxèmes initiaux.

6.1.2 Deuxième cas : avec isotaxémie

Nous allons maintenant traiter le cas où au moins deux des sémèmes à translater dans un nouveau taxème appartiennent, avant l'opération, à un même taxème. La grande majorité des opérations précédemment définies étant préservées, nous ne nous intéressons qu'aux cas de ces deux sémèmes.

Soient donc s_1 et s_2 appartenant au taxème t_1. Ces deux sémèmes supportent tous deux le même sème spécifique se, auquel nous voulons appliquer l'opérateur de généricisation $igen$. Deux sous-cas vont devoir être étudiés, suivant la nature des spécèmes supportant le sème se :

Pas de lien se entre les deux sémèmes Dans ce premier cas, le sème manipulé par $igen$ ne qualifie ni (s_1, s_2) ni (s_2, s_1). Nous avons donc l'un des deux schémas possibles (voir 4.7).

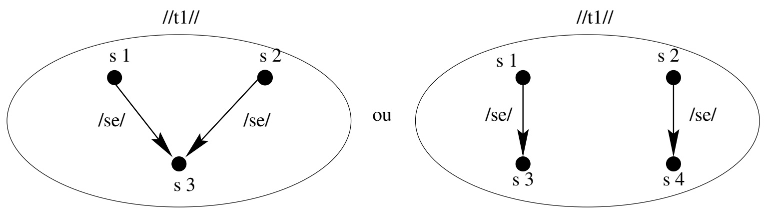


FIG. 4.7 – *Pas de lien se entre les deux sémèmes*

Ces deux schémas seront traités de la même façon. Le principe général est que les oppositions existant déjà entre s_1 et s_2 au sein du taxème t, puisqu'elles sont autres que celles exprimées par se, seront conservées dans le nouveau taxème t', de sème microgénérique se. Les spécèmes (s_1, s_2) et (s_2, s_1), ou le seul d'entre eux qui est activé, seront alors préservés, et leurs isotopies également. Il ne reste plus qu'à qualifier les oppositions entre s_1, s_2 et les autres sémèmes translatés en utilisant le sème microgénérique de t comme pour le cas de non-isotaxémie. Une exception évidente dans ce cas, le sème microgénérique de t ne qualifiera bien entendu pas d'opposition entre s_1 et s_2. Les autres spécèmes de t seront également traités de la même façon. Rappelons qu'un des deux spécèmes (s_1, s_2) ou (s_2, s_1) doit être activé, et que donc les contraintes sur t', nouveau taxème créé par $igen$, seront respectées : la caractérisation positive par le principe de transfert des sèmes microgénériques, et la faible complétude par ce même principe, et par le transfert du ou des spécèmes précités.

Lien *se* entre les deux sémèmes Nous nous trouvons donc dans le cas d'une «opposition transitive» du type représenté dans le schéma 4.8.

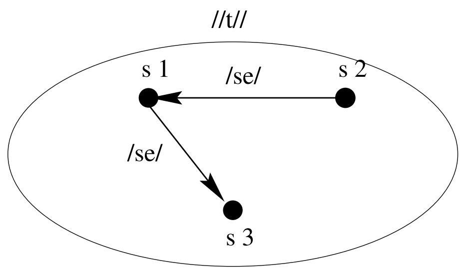


FIG. 4.8 – *Lien se entre les deux sémèmes*

Bien évidemment, ce lien supportant *se* entre s_1 et s_2 ne peut pas être translaté dans le nouveau taxème, puisque ce dernier aura justement *se* comme sème microgénérique (contrainte de la redondance). Par contre, les autres qualifications pouvant exister entre s_1 et s_2 seront translatées. Nous pouvons donc avoir, dans les cas minimaux, une non-complétude dans le taxème final. Un tel cas de figure sera donc traité simplement par application de l'opérateur *comp* sur le taxème final. À part cette opération supplémentaire, les autres étapes sont conservées.

6.1.3 Récapitulatif

Nous reprenons ici le tableau précédemment établi pour l'absence d'isotaxémie, en y ajoutons quelques opérations pour les cas d'isotaxémie. Nous nommerons S_{iso} le sous-ensemble de $I_{spe}(se)$ comportant les sémèmes appartenant initialement à un même taxème.

$+t$	Création du nouveau taxème.
$+i(t, se)$	Attribution du sème *se* au nouveau taxème.
$\forall s_i \in I_{spe}(se), add(s_i, t)$	Ajout des sémèmes dans le taxème.
$\forall s_i, s_j \in t$ t.q. $\nexists t_i \neq t$ t.q. $s_i \in t_i, s_j \in t_i$, $+sp(s_i, s_j)$	Création des spécèmes pour les cas de non-isotaxémie.
$\forall sp \in I_{sp}(se)$, $-i(sp, se)$	Suppression des attributions spécifiques du sème traité.
$\forall s_i, s_j \in t$, tq $s_j \notin t_i \ + i((s_i, s_j), se_{t_i})$	Attribution des anciens sèmes microgénériques comme sèmes spécifiques (pour la non-isotaxémie).
$\forall s_i$, $\forall s'_i \in t_i$ t.q. $si_i \notin S_{iso}$, $\forall se' \in I_{sp}^{-1}(s_i, s'_i)$, $-i((s_i, s'_i), se')$, $-sp(s_i, s'_i)$, $+i(s_i, se')$, $pb_{aff}(se')$	Gestion des autres sèmes spécifiques initialement attribués aux sémèmes translatés (non isotaxémiques) : ils deviennent afférents à ces sémèmes.
$\forall s_i \in t$, $\forall s'_i \in t_i$ tq $s'_i \notin S_{iso}$, $\forall se' \in I_{sp}^{-1}(s'_i, s_i)$, $-i((s'_i, s_i), se')$, $-sp(s'_i, s_i)$ et si $se' \notin I_{spe}^{-1}(s'_i)$ alors $+i(s'_i, se')$ et $pb_{aff}(se')$	Traitement des spécèmes dont le sémème translaté était la seconde projection : on supprime les attributions de sème à ces spécèmes avant de détruire les spécèmes eux-mêmes, et on déclare le sème comme afférent si nécessaire.
$\forall s_i \in t$, $sub(s_i, t_i)$	Suppression des sémèmes translatés de leurs taxèmes d'origine.
$\forall t_i$, si $\vert t_i \vert = 1$, $ADD(s, t_i)$	Complétion des taxèmes initiaux si ceux-ci n'ont plus qu'un sémème.
$\forall t_i$, $comp(t_i)$	Vérification de la complétude faible des taxèmes initiaux.
$comp(t)$	Vérification de la complétude faible du taxème-but.

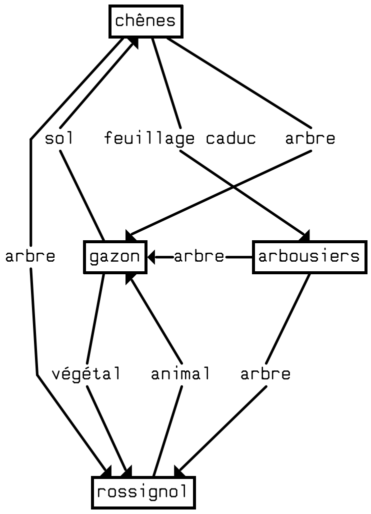


FIG. 4.9 – *Taxème //nature// initial*

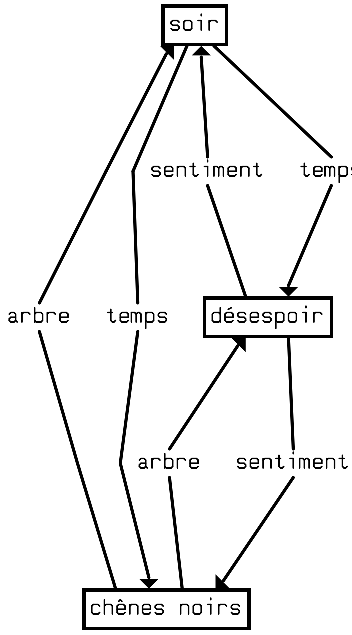


FIG. 4.10 – *Taxème //mort// initial*

6.1.4 Un exemple pour clarifier

Pour résumer cet opérateur, nous allons présenter un cas assez complexe extrait d'une analyse de «En sourdine». Il s'agit donc d'un cas d'isotaxémie, et l'isotopie visée est celle du sème /arbre/. Cette isotopie ne contient pas de partie afférente, seule une partie spécifique, dont les spécèmes sont :
('chênes', 'rossignol') ('chênes', 'gazon') ('arbousiers', 'gazon') et ('arbousiers', 'rossignol') pour le taxème //nature// et
('chênes noirs', 'désespoir') et ('chênes noirs', 'soir') pour le taxème //mort//.

D'autre part, les deux taxèmes concernés ont leurs graphes d'opposition représentés par les figures 4.9 et 4.10 [1]:

Après application de l'opérateur *igen(/arbre/)*, diverses modifications ont eu lieu.

1. Le rendu graphique différent de ces graphes est dû à l'utilisation du logiciel VCG intégré à PASTEL. VCG génère automatiquement les graphes d'oppositions des taxèmes. Ce sera désormais le cas pour tous les exemples «réels» que nous citerons. Le logiciel est disponible à l'adresse suivante : http://www.cs.uni-sb.de/RW/users/sander/html/gsvcg1.html

Tout d'abord, regardons le graphe d'opposition du nouveau taxème //arbre// (4.11).

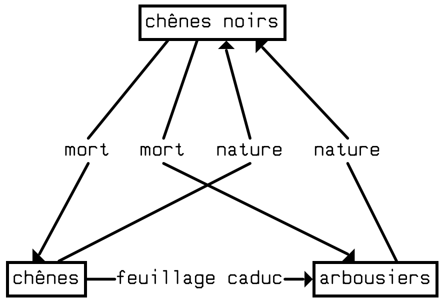


FIG. 4.11 – *Nouveau taxème //arbre//*

Comme il apparaît clairement dans cette figure, le sème spécifique /feuillage caduc/ a été conservé entre 'chênes' et 'arbousiers'; d'autre part les anciens sèmes microgénériques /nature/ et /mort/ ont maintenant des statuts spécifiques.

Regardons maintenant ce que les deux taxèmes initiaux sont devenus (4.12 et 4.13).

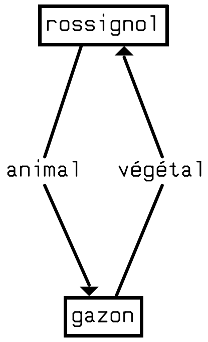


FIG. 4.12 – *Nouveau taxème //nature//*

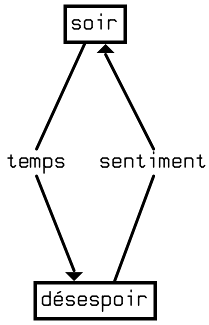


FIG. 4.13 – *Nouveau taxème //mort//*

Dans ces taxèmes modifiés, les spécèmes initiaux ont été conservés. Cependant, nous pouvons remarquer que le sème /sol/, anciennement spécifique à 'gazon' n'apparaît plus dans le graphe de //nature//. Effectivement, la seconde projection du spécème qui le supportait a bien été supprimée du taxème : /sol/ est donc maintenant afférent à 'gazon'. Encore plus loin pour ce sème, puisqu'il n'est plus (pour l'ensemble de la structure) qu'afférent, cette opération va donc créer le nouveau taxème //sol// afin de satisfaire le problème de l'afférence pure. Ce taxème devra donc être rempli par deux nouveaux sémèmes (*a priori* non lexicalisés dans notre cas).

Enfin, aucune complétion des taxèmes précités n'est à effectuer, nous pouvons en effet remarquer que les graphes correspondants sont faiblement complets et que de chacun de leurs nœuds est bien issu au moins un arc.

6.2 Généricisation d'une isotopie quelconque : *igen* étendu

Nous n'avons précédemment traité que le cas où l'isotopie transformée par *igen* n'était que purement spécifique : ses parties génériques et afférentes étaient vides. Puisque cette situation ne couvre pas tous les cas d'attribution d'un sème à un spécème, nous allons maintenant envisager les autres cas.

6.2.1 Cas d'une isotopie déjà générique

Soit se un sème tel que $I(se) = (\{sp_1, ..., sp_n\}, \{t\}, \emptyset)$. La différence principale entre ce cas de figure et le traitement précédemment envisagé est le fait que le taxème-but existe déjà, et qu'il ne faudra donc que le compléter par les sémèmes possédant se comme sème spécifique. L'idée générale restera la même, toutefois, puisque nous tenterons d'utiliser les sèmes microgénériques de ces sémèmes comme spécifiques dans t. Mais il se peut très bien que ces sèmes en question soient déjà présents dans le graphe d'opposition du taxème t initial. Cependant, leur utilisation comme nous l'avons précisée dans *igen* ne violera aucune contrainte sur ce graphe, puisque les nouveaux spécèmes seront activés de la même façon que leurs opposés : plus clairement, le sème précédemment microgénérique ne peut concerner que des spécèmes initialement présents dans t, et donc ne pourra être attribué à deux spécèmes opposés.

Par contre, puisque les sémèmes initialement présents dans t ne sont pas translatés, ils ne peuvent bénéficier d'un sème simple à trouver pour se différencier des nouveaux sémèmes. Mais encore une fois ceci n'est pas gênant, puisque le graphe initial de t vérifiant les contraintes, ils auront bien chacun au moins un spécème activé dont ils sont la première projection. Ainsi, cette contrainte ainsi que la complétude faible du graphe seront respectées. Pour les cas d'isotaxémie des spécèmes de $I(se)$, le problème est donc également identique au cas plus précisément traité dans les paragraphes précédents.

Si maintenant l'isotopie en question est mésogénérique, donc de la forme :

$I(se) = (\{sp_1, ..., sp_n, \{t_1, t_2\}, \emptyset)$, le cas est plus complexe, puisqu'il faudra choisir entre t_1 et t_2 pour la translation des sémèmes. Mais hormis ce nouvel appel à une fonction de choix, nous restons bien dans les cas prévus par *igen* et son extension aux isotopies en partie microgénériques.

6.2.2 Cas d'une isotopie afférente

Si maintenant la troisième partie de l'isotopie $I(se)$ n'est pas vide, nous allons avoir plusieurs cas à envisager :

Isotopie générique et afférente : Nous avons

$$I(se) = (\emptyset, \{t\}, \{s_1, ..., s_m\})$$

La seule opération est donc de translater les sémèmes s_i de leurs taxèmes d'origine vers t, de la même façon que nous avons vue précédemment pour les premières projections des spécèmes de l'isotopie.

Isotopie spécifique et afférente : Nous avons

$$I(se) = (\{sp_1, ..., sp_n, \emptyset, \{s_1, ..., s_m\})$$

avec cet avantage que les première projections des sp_i sont toutes différentes des s_j. Il ne reste donc plus qu'à considérer l'ensemble de tous ces sémèmes d'origines différentes et à appliquer *igen* sur cet ensemble étendu.

Isotopie complète : Dans le cas final, où l'isotopie possède trois parties non vides, nous allons donc nous permettre une fois de plus de préciser l'ensemble des opérations applicables, ce qui servira de résumé final à la notion de généricisation d'une isotopie. Les données initiales sont :

$$I(se) = (\{sp_1, ..., sp_n\}, \{t_1, t_2\}, \{s_1, ..., s_m\})$$

Pour un sémème s_i (qu'il soit une première projection d'un des spécèmes, ou directement un élément de la troisième composant de l'isotopie) t_i est son taxème d'origine.

Pour les isotopies initialement mésogénériques, $c(s_i)$ est celui des taxèmes t_1 ou t_2 qui a été sélectionné par l'interprète pour accueillir le sémème s_i.

Si l'origine d'une famille de sémèmes s_i n'est pas précisée, il s'agit de l'ensemble des premières projections de sp_i, uni à la troisième projection de l'isotopie.

$\forall s_i \in I_{spe}(se), add(s_i, c(s_i))$	Ajout des sémèmes de la partie spécifique dans le taxème choisi.
$\forall s_j \in I_{aff}(se), add(s_j, c(s_j))$	Ajout des sémèmes de la partie afférente dans le taxème choisi.
$\forall t \in \{t_1, t_2\}, \forall s_i, s_j \in t$ t.q. $\nexists t_i \neq t$ t.q. $s_i \in t_i,\ s_j \in t_i,\ +sp(s_i, s_j)$	Création des spécèmes pour les cas de non-isotaxémie.
$\forall sp \in I_{sp}(se),\ -i(sp, se)$	Suppression des attributions spécifiques du sème traité.
$\forall t \in \{t_1, t_2\}, \forall s_i, s_j \in t$, t.q. $s_j \notin t_i\ +i((s_i, s_j), se_{ti})$	Attribution des anciens sèmes microgénériques comme sèmes spécifiques (pour la non-isotaxémie).
$\forall s_i,\ \forall s'_i \in t_i$ tq $si\ s_i \notin S_{iso}$, $\forall se' \in I_{sp}^{-1}(s_i, s'_i),\ -i((s_i, s'_i), se'),$ $-sp(s_i, s'_i),\ +i(s_i, se'),\ pb_{aff}(se')$	Gestion des autres sèmes spécifiques initialement attribués aux sémèmes translatés (non isotaxémiques) : ils deviennent afférents à ces sémèmes.
$\forall t \in \{t_1, t_2\}, \forall s_i \in t,\ \forall s'_i \in t_i$ t.q. $s'_i \notin S_{iso},\ \forall se' \in I_{sp}^{-1}(s'_i, s_i),$ $-i((s'_i, s_i), se'),\ -sp(s'_i, s_i)$ et si $se' \notin I_{spe}^{-1}(s'_i)$ alors $+i(s'_i, se')$ et $pb_{aff}(se')$	Traitement des spécèmes dont le sémème translaté était la seconde projection : on supprime les attributions de sème à ces spécèmes avant de détruire les spécèmes eux-mêmes, et on déclare le sème comme afférent si nécessaire.
$\forall s_i,\ sub(s_i, t_i)$	Suppression des sémèmes translatés de leurs taxèmes d'origine.
$\forall t_i$, si $\lvert t_i \rvert = 1,\ ADD(s, t_i)$	Complétion des taxèmes initiaux si ceux-ci n'ont plus qu'un sémème.
$\forall t_i,\ comp(t_i)$	Vérification de la complétude faible des taxèmes initiaux.
$\forall t \in \{t_1, t_2\}, comp(t)$	Vérification de la complétude faible du taxème-but.

6.3 Généricisation de plusieurs isotopies

Nous avons déjà discuté de la nature globale de l'opérateur *igen*, qui résulte de son application à un sème, sorte de «sommet» de notre structure, et dont les modifications se répercutent, nous l'avons vu par la complexité même de l'opérateur, dans la totalité de la structure. Revenons un instant sur l'objectif de cet opérateur cependant, qui est de donner à une structure dans son ensemble une conformité soit à une autre structure, soit à un ensemble de contraintes d'origines diverses. Dans ce cas, une simple et unique application de *igen* parait insuffisante. Pire que cela, une application séquentielle de cet opérateur n'est pas envisageable telle quelle, puisqu'il peut y avoir des conflits. Prenons le cas d'un taxème minimal de deux sémèmes, opposés l'un à l'autre par deux sèmes spécifiques. Si l'on désire appliquer *igen* à ces deux sèmes spécifiques, et que l'on procède séquentiellement, voici ce qui se passe. Après application du premier *igen*, le taxème initial s'est vu priver de son premier sémème, qui a donc dû être remplacé par un nouveau sémème, *a priori* non lexicalisé, une sorte de sémème-tampon. De plus, le deuxième sème spécifique a disparu du taxème durant l'opération, puisque la seconde projection de son sémème support a été transferrée. Même si nous appliquons tout de même *igen* en forçant le sème spécifique à perdurer, nous nous retrouverons au bout du compte avec un taxème initial toujours présent, mais ne comportant que deux sémèmes non lexicalisés, de pertinence plus que contestable. Le but initial de cette double application pouvait en effet s'exprimer par la volonté de faire disparaître un taxème trop minimal.

Nous nous retrouvons concrètement face à un nouveau problème à résoudre : la validité, contestable, des taxèmes rendus entièrement non lexicalisés par *igen*. Il suffit donc de faire intervenir un test de lexicalisation lorsque l'on complète un taxème. Si le dernier sémème en place est non-lexicalisé, le taxème est éliminé, le sémème aussi. Bien entendu, ce n'est qu'une option.

6.4 Spécification d'une isotopie générique : le non-opérateur *ispec*

Comme nous pouvions nous y attendre, après la transformation d'une isotopie spécifique en isotopie générique, il nous reste à envisager la transformation inverse. Nous nous intéresserons donc au désassemblage d'une classe sémantique, avant tout d'un taxème. Et pourquoi ? Toujours pour des raisons de cohérence de la structure face à une autre structure ou à des contraintes normatives, mais aussi, comme pour *igen*, pour pousser encore un peu plus loin l'étendue d'une analyse.

Le principe de ce nouvel opérateur sera donc de conserver la proximité sémantique des sémèmes d'un même taxème par le biais d'une isotopie spécifique, en translatant les sémèmes vers de nouveaux taxèmes. Au sein de ce nouveau taxème, le sème auparavant générique pour le sémème translaté deviendra donc spécifique, ce qui peut nécessiter l'identification d'un autre sémème pour supporter l'opposition (spécème). Nous l'avons déjà avoué, notre conception formelle du taxème, et surtout son utilisation dans une approche uni-textuelle est assez souple : c'est en effet une classe purement contextuelle. Dans l'océan de la langue, nos sémèmes rassemblés pour un texte dans une même classe ont *a priori* très peu de chance de cohabiter dans une classe définie par une norme générale.

Fort de ces constatations, accordons-nous maintenant sur le fait que $ispec$ ne sera pas exactement l'opérateur dual de $igen$, puisque l'ouverture thématique effectuée par une spécification sera bien plus importante. En effet, nous devrons donc, à partir d'un sème spécifique, demander à l'interprète de retrouver une classe la plus resserrée possible autour de ce sémème.

Ainsi, $ispec$ ne peut véritablement s'envisager comme un opérateur global, puisqu'il ne serait en fait qu'une suite de traitements locaux, ces traitements étant des translations de sémèmes vers des taxèmes *a priori* nouveaux. Le rapport initial entre ces sémèmes, qui était leur présence dans une même isotopie contextuellement définie s'effaçant progressivement.

7 Conclusion

Pour terminer cette discussion sur la formalisation des transformations globales applicables à une structure donnée, revenons maintenant aux objectifs et/ou intérêts que nous nous sommes fixés à propos de ces dernières.

7.1 Relations et transformations intra-structurelles

Tout d'abord, du point de vue d'une seule analyse, le simple fait de rendre génériques certaines isotopies-clés permet d'enrichir substantiellement la structure (nouveaux sémèmes, nouveaux taxèmes, nouveaux sèmes...). Ainsi, à un stade où une analyse s'arrête faute de combativité de la part de l'interprète, une sélection judicieuse de certaines isotopies afin de les rendre génériques peut relancer le moteur interprétatif.

Mais bien évidemment, toutes les isotopies ne peuvent être rendues génériques, par le simple fait qu'une isotopie purement générique, même minimale, nécessite au moins deux isotopies en partie spécifiques (c'est-à-dire deux sèmes spécifiques distincts au moins par taxème). Une première question se pose donc quant aux critères de sélection des isotopies à transformer. Une première réponse se trouve dans les critères quantitatifs chers à notre âme d'informaticien : nous les avons déjà définis et en partie interprétés, et voici donc une utilisation possible de ceux-ci :

Tout d'abord le poids, la plus simple des données. Une isotopie spécifique de poids fort a le plus de chances de nécessiter l'identification de nouvelles entités lors de sa génériciation, puisqu'elle soustraira à de nombreux taxèmes de grands nombres de sémèmes (ou l'un des deux cas). De plus, sa couverture thématique n'est pas négligeable, et mérite un approfondissement par une focalisation en un taxème. Une «belle» isotopie spécifique est d'ailleurs un résultat positif du protocole que nous proposons : il s'agit bien d'un «filon thématique» repéré par l'interprète *après* la phase de pré-interprétation, donc un pas conséquent au-delà de ses intuitions initiales.

Pour ce qui est des autres considérations tactiques, comme le volume et la densité, elles nous semblent moins pertinentes pour les transformations en général. Par contre, une curiosité interprétative qui serait une isotopie de poids fort dans un faible volume, et rendue générique permettrait l'approfondissement des relations morphologiques et syntaxiques entre les sémèmes qui la composerait.

Enfin, le type de systématicité, lui aussi quantifiable, d'une isotopie est un bon guide. Trivialement bien sûr, puisque les opérations qui composent l'opérateur de génériciation dépendent directement de ces données. Nous verrons également par la suite, pour ce qui concerne les liaisons inter-structurelles, un plus grand intérêt de ces données.

7.2 Relations et transformations inter-structurelles

Voilà pour *une* analyse. Envisageons maintenant les possibilités offertes par la prise en considération de plusieurs structures de données. Il y a, avant, tout plusieurs relations de similarité à envisager entre deux structures, suivant le texte et l'identité de l'interprète.

Entre deux analyses du même texte, mais d'origines humaines différentes, nous avons vu que nous pourrions espérer récupérer la problématique des normes langagières, en comparant les types d'attributions des sèmes, si du moins l'on trouve un consensus sur la désignation de ceux-ci. Il est assez aisé, vu la standardisation de la structure que propose PASTEL, de comparer deux analyses, *via* la tactique du moins, et la thématique en partie. Le problème principal est l'accord sur le découpage de la chaîne syntagmatique, où les variations sont infinies. Il est toutefois envisageable, en relâchant les contraintes tactiques, de trouver des formes identiques ou fortement similaires sur deux analyses d'un même texte. À ce propos, et sur les considérations couvertes dans ce chapitre, on consultera avec profit le mémoire de F. Rivière [54]. C'est également dans un tel contexte d'utilisation que notre fameux opérateur *igen* peut s'appliquer, afin d'harmoniser deux structures en prenant comme pivot une ou plusieurs isotopies-clés. Si un même sème apparaît dans deux structures distinctes, rendre les deux isotopies correspondantes purement génériques est une étape qui facilitera bien leur comparaison.

Ce dont il est surtout question dans ce mémoire concerne effectivement les analyses de textes différents. Mais pas si différents que cela, puisque les cas intéressants sont ceux concernés par les études de familles thématiques, ou encore mieux, des œuvres d'une même main. C'est dans de telles études que nous espérons apporter une aide par la mise en place de structures isotopiques simples. L'idéal serait en effet de remarquer des structures propres à un auteur ou à un style, et de qualifier formellement ces dernières.

Par contre, nous pouvons plus aisément envisager une classification / typologie des interprétations plus que des textes. Ce qui n'est guère étonnant vu l'objectif de cette méthode. En effet, les différents critère qualitatifs et quantitatifs attribuables aux isotopies, et leur représentation en diachronie nous donnent de nombreuses pistes pour évaluer le type d'interprétation. Tout d'abord, le nombre et l'ampleur des pré-isotopies : une bonne connaissance d'un texte et de son entour mène directement à l'identification de pré-isotopies étendues, et sans nul doute moins nombreuses que dans le cas d'une interprétation «naïve», où la cohésion thématique n'est pas saillante pour l'interprète. Dans ce dernier cas, en effet, ce sera sans doute le local, donc la considérations de termes «isolés» qui guidera l'interprétation, en multipliant par là-même les classes sémantiques en en diminuant leur cardinalité. Il est cependant difficile de caractériser une pré-interprétation dans l'absolu : il reste plus raisonnable de comparer deux pré-interprétations d'un même texte.

Voilà pour les conditions initiales. Une fois l'interprétation proprement dite effectuée (ou du moins bien avancée et stabilisée), les attributs des isotopies nouvellement créées peuvent indiquer une première notion de succès de l'analyse. La découverte d'isotopies spécifiques étendues est, nous l'avons dit, un point positif. Un autre point est également l'extension des classes initiales (venant des pré-isotopies). Un échec relatif, par contre, serait la multiplication des sèmes spécifiques, et la conservation d'une vision locale de chaque classe, puisque l'explicitation des classes n'a pas apporté de nouveau «concept».

Un dernier point sur lequel nous aimerions insister est l'apport d'une telle exploration des classes. Nous avons évoqué la difficulté qu'éprouverait l'explicitation d'une vision «fantaisiste», puisque la finesse de description que nous exigeons par la satisfaction des contraintes mettrait rapidement à jour des manques dans la vision de l'interprète. D'un autre côté, nous estimons qu'une exploration d'une interprétation justifiée est elle, réconfortante de cette vision. L'hypothèse se trouve subjectivement renforcée, par la mécanique descriptive qui ne rencontre que peu d'obstacles dans son explicitation. Les sèmes spécifiques utilisés pour organiser un thème bien ancré dans un texte jouent alors le rôle d'autant d'arguments en faveur de cette hypothèse.

Chapitre 5

Une application : PASTEL

Au long de ce dernier chapitre, nous allons présenter en détail un exemple d'application du formalisme décrit lors des deux précédents chapitres. PASTEL est un acronyme pour « Programme d'Aide à l'Analyse de TExtes, même Littéraires ». D'un point de vue technique, PASTEL est conçu dans le langage CamlLight, et pour la partie interface graphique, sous le concepteur d'interface CamlWin, conçu par Pierre Saunier [1]. Nous allons dans un premier temps expliciter les choix d'implantations des structures sémantiques désormais bien connues, puis présenter les diverses fonctionnalités du logiciel, agrémentant le tout de copies d'écran du rendu informatique.

Nous tenons ici à remercier les différents utilisateurs de ce programme, qui ont bien voulu se prêter au jeu interprétatif proposé, et dont les remarques et réactions m'ont encouragé à continuer. Les aspects positifs dégagés par ces diverses utilisations par des non littéraires confirmaient en effet nos intuitions.

1 Codage de la structure

Caml étant un langage fonctionnel fortement typé, nous avons choisi de représenter les différents niveaux, correspondant aux diverses entités sémantiques de la structure, par des listes. Le principe général de l'organisation informatique est celle du *double linking*, pour exprimer les relations entre entités de différents niveaux. Nous allons avant tout présenter le codage de chaque type d'entité.

1. Cet outil et sa documentation sont disponibles par *ftp* à l'adresse suivante : ftp://ftp.inria.fr/INRIA/Projects/cristal/caml-light/Usercontribs/camlwin

1.1 Sémèmes

La structure d'un sémème est un enregistrement, comportant les données suivantes :

- Lexie : la chaîne de caractères correspondant au signifiant du sémème.
- Positions : une liste de coordonnées des épisémèmes supportant ce sémème. Une coordonnée est un simple couple d'entiers (numéro de ligne, numéro de colonne) correspondant au premier caractère de la chaîne dans le texte.
- Marqueur de lexicalisation : une valeur booléenne, indiquant simplement si le sémème possède ou non des épisémèmes dans le texte. Cette valeur est donc à *vrai* si la liste des positions n'est pas vide.
- Numéro de taxème : un lien avec le taxème qui contient ce sémème. Cette valeur correspond à la position du taxème dans la liste totale des taxèmes. Elle est donc unique pour chaque sémème.
- Numéros de spécèmes : de la même façon que pour le taxème, une série de liens directs avec les numéros des spécèmes ayant ce sémème comme première projection, correspondant également aux positions de ces spécèmes dans la liste totale des spécèmes.
- Numéros des isotopies d'afférence : une dernière série de liens directs avec les isotopies qui possèdent le sémème dans leur troisième composante (pour traduire un sème afférent au sémème).

Enfin, tous les sémèmes repérés dans un texte sont stockés dans une seule liste, et seront par la suite repérés par leur position dans la liste. Il est important de noter que l'ordre des sémèmes dans la liste est totalement arbitraire, et ne traduit en aucun cas l'ordre syntagmatique.

1.2 Taxèmes

Un taxème est également défini par un enregistrement, comportant simplement deux données :

- Sémèmes : une liste des numéros des sémèmes qui appartiennent à ce taxème (position de ces sémèmes dans la liste).
- Isotopies : une liste des numéros d'isotopies (deux au maximum) qui possèdent le taxème dans leur seconde projection.

Encore une fois, l'ensemble T des taxèmes est traduit dans le programme par une liste de taxèmes, ces derniers étant repérés par les autres entités via leur position dans cette liste.

1.3 Spécèmes

Une dernière fois un simple enregistrement :

- Première composante : le numéro du sémème s si le spécème codé est de la forme (s, s').
- Deuxième composante : le numéro du sémème opposant s'.
- Numéros d'isotopies : la liste des isotopies correspondant aux sèmes supportés par ce spécème. Ce spécème est donc dans la première projection des isotopies repérées dans cette liste.

Il existe également une liste globale de tous les sémèmes pour leur repérage.

1.4 Isotopies

Il s'agit ici d'un quadruplet simulant les différents liens effectués par la fonction I :

- numéro de sème (unique).
- liste des numéros de spécèmes.
- liste des numéros de taxèmes.
- liste des numéros de sémèmes (afférence).

Les isotopies sont également rangées en une liste pour être référencées par leur position dans cette liste.

1.5 Sèmes

Un sème est une simple chaîne de caractères, rangée dans une liste globale pour être repérés dans les isotopies et les pré-isotopies par leur emplacement dans cette liste.

Notons que nous avons séparé la notion de sème de celle d'isotopie pour pouvoir tenir compte aisément du transfert des sèmes de la pré-interprétation (donc des pré-isotopies) à l'interprétation (isotopies).

L'ensemble des entités que nous venons d'expliciter, ainsi que les relations d'indexation qui existent entre elles, agrémentées de leur cardinalité, sont représentées dans le schéma 1.5, dont nous empruntons le style aux méthodes de conception de bases de données relationnelles.

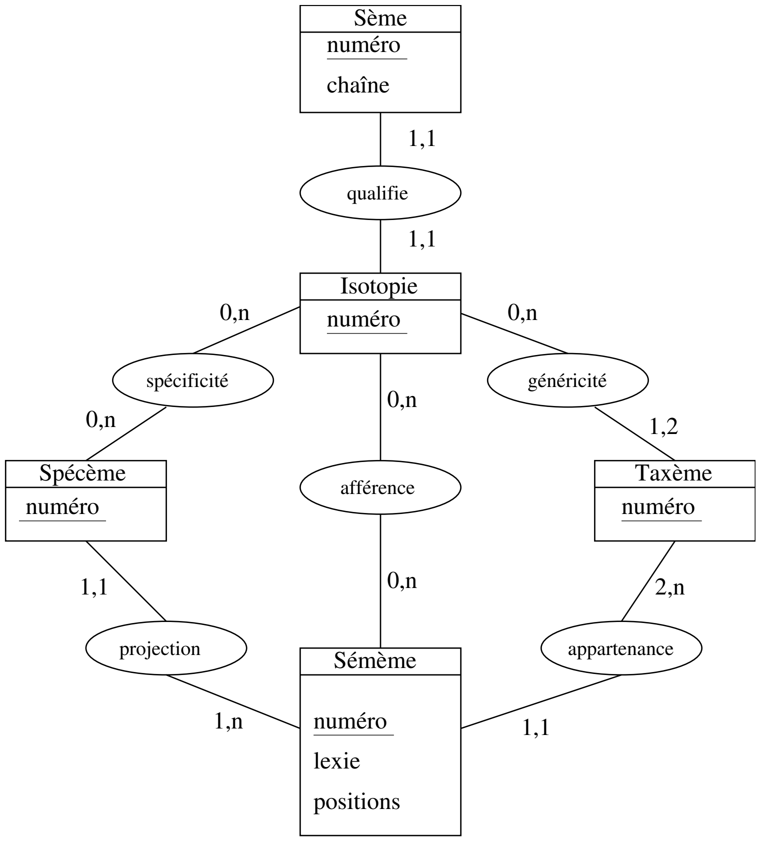


Fig. 5.1 – *Organisation relationnelle*

1.6 Pré-isotopies

Ces entités ne sont utilisées que lors de la première phase d'interprétation. Leur organisation est celle d'un arbre n-aire, afin de capter aisément les notions d'inclusion qui peuvent apparaître lors de leur repérage. Une feuille de cet arbre est donc un couple (sème - liste de numéros de sémèmes), et un nœud est un couple (feuille - liste de nœuds).

2 Organisation de l'interface

La majeure partie de l'outil PASTEL étant visible, nous allons effectivement décrire les différentes composantes de l'interface graphique. Le principe général en est simple : à partir d'une fenêtre principale, comportant une zone de texte, il est possible d'appeler différentes fonctions, ou plutôt différentes familles de fonctionnalités, chacune de ces familles étant liée à un concept de la structure sémantique.

2.1 Fenêtre principale

Un aperçu du rendu effectif de la fenêtre principale est visible dans la figure 5.2.

Outre le texte dans la partie inférieure de la fenêtre, une barre de menu donne accès aux différentes possibilités d'actions concernant les entités : sémèmes, taxèmes, isotopies et pré-isotopies. Les deux derniers menus contiennent les fonctionnalités génériques (sauvegarde et chargement de fichier, résultats de l'analyse, etc.).

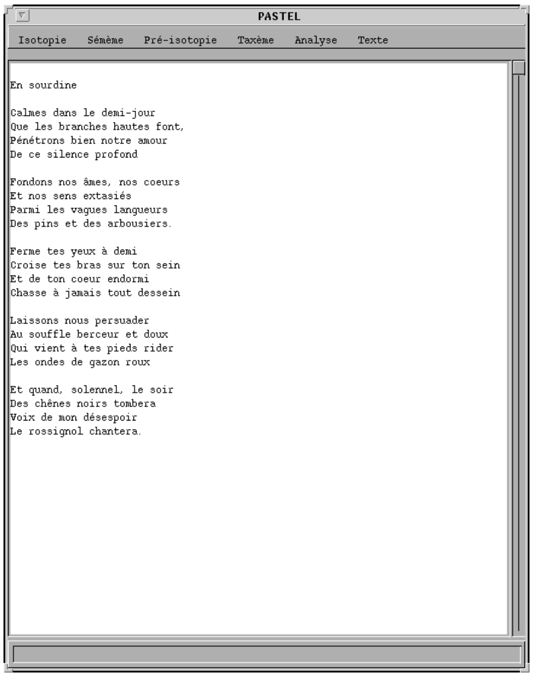


FIG. 5.2 – *Fenêtre principale*

Pour chaque menu, les fonction accessibles directement sont :

- Sémème : affichage de la liste des sémèmes, qui permet par la suite de manipuler ces derniers (cf. plus bas).
- Taxème : Affichage de la liste des taxèmes, qui donnera également accès à d'autres possibilités, et démarrage du protocole de spécification des taxèmes. Cette dernière fonction va appliquer la procédure de spécification (voir plus bas) à chaque taxème non encore spécifié.
- Isotopies : Affichage de la liste des isotopies ainsi que de leur caractéristique, et activation de l'outil de comparaison tactique des isotopies. Ce dernier permet de calculer les corrélations entre deux isotopies de la liste en fonction de leur extension sur l'axe syntagmatique.
- Pré-isotopies : Ce menu n'est accessible que lors de la première phase d'utilisation du logiciel (pré-interprétation). Une fois les pré-isotopies traitées (donc une fois les premiers taxèmes créés), tout accès aux fonctions correspondantes est inhibé. Il contient néanmoins les fonctions suivantes : Affichage de la liste des pré-isotopies (repérées pas leurs sèmes). Ajout d'un sémème dans une pré-isotopie. Cette fonction n'est utilisable que si au moins une pré-isotopie a été repérée (par son sème). Création d'une nouvelle pré-isotopie (par interrogation de l'utilisateur et identification d'un sème).

2.2 Fenêtres des pré-isotopies

Plusieurs fenêtres sont nécessaires afin de mieux gérer la phase de pré-interprétation. Cependant, une fenêtre principale est utilisable pour la plupart des opérations. Elle est représentée dans la figure 5.3. Dans le corps principal de celle-ci apparaissent les sèmes repérés jusqu'ici pour identifier les pré-isotopies. Différentes fonctions sont accessibles par des boutons, à savoir :

- "Liste des sémèmes" ouvre une nouvelle fenêtre indiquant la liste des sémèmes couverts par la pré-isotopie sélectionnée dans la première fenêtre. Cette fenêtre des sémèmes d'une pré-isotopie est visible dans la figure 5.4. Dans cette dernière est donnée la possibilité d'ajouter non pas un nouveau sémème mais un nouvel épisémème, identifié comme étant une occurrence d'un sémème déjà repéré. Cette fonctionnalité évite ainsi les processus ultérieurs d'assimilation des épisémèmes, dont nous parlerons à l'occasion du traitement des sèmes spécifiques.

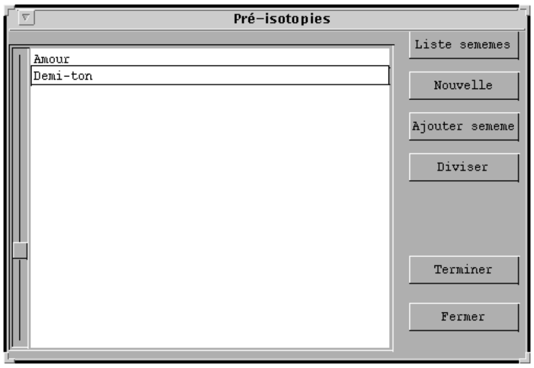


FIG. 5.3 – *Fenêtre des pré-isotopies*

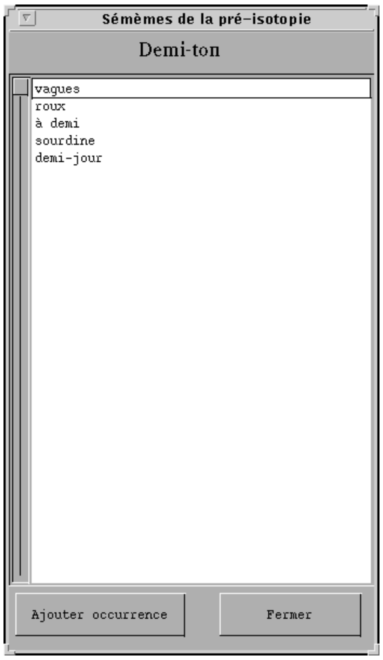


FIG. 5.4 – *Fenêtre des sémèmes d'une pré-isotopie*

- “Nouvelle” permet de créer une nouvelle pré-isotopie, en désignant son sème.

- “Ajouter sémème”, comme son nom l’indique, va ajouter dans la liste des sémèmes de la pré-isotopie sélectionnée le sémème dont la chaîne est celle qui est sélectionnée dans le texte de la fenêtre principale. Cette opération ne créera un nouveau sémème que si cette chaîne n’a pas déjà servi à repérer un sémème. Ainsi, si l’on désire intégrer un même sémème à plusieurs pré-isotopies, ceci sera fait sans ambiguïté par cette même fonctionnalité.

- “Diviser” permet le découpage d’une pré-isotopie existante (celle qui est sélectionnée dans la liste). Cette opération demandera à l’utilisateur un nouveau sème qui servira à identifier une sous-partie de l’isotopie à diviser. Après ceci, la liste des sémèmes contenus dans la pré-isotopie lui sera présentée, afin d’en sélectionner un sous-ensemble pour les intégrer à la nouvelle «sous-pré-isotopie».

- “Terminer” achève la phase de pré-interprétation. Diverses opérations se déroulent alors :

  - Gestion des recouvrements. Un algorithme simple vérifie la présence d’un même sémème dans plusieurs pré-isotopies. Si cela est le cas, et qu’une des pré-isotopies ainsi repérées n’est pas une sous-pré-isotopie d’une autre, l’utilisateur aura à sélectionner une de celles-ci comme pré-isotopie principale. Le sème correspondant sera alors inhérent pour le sémème en question, et les autres pré-isotopies serviront à attribuer un sème afférent.
  - Création des sous-pré-isotopies complémentaires. Lors de la division d’une pré-isotopie, il est possible de laisser des sémèmes non repérés dans un sous-ensemble. Si le cas se présente, une nouvelle sous-pré-isotopie est créée, dont le sème sera /X générique/, si /X/ est le sème de la pré-isotopie divisée.
  - Création des taxèmes et isotopies associées. Une fois les problèmes précédents résolus, le logiciel crée les premières isotopies de la structure, en transformant les pré-isotopies en taxèmes. Ces taxèmes n’ont pour l’instant aucun graphe d’opposition associé, mais les isotopies correspondantes peuvent comporter tout de même une partie afférente provenant des éventuels recouvrements traités précédemment.

Une fois ces opérations effectuées, la phase de pré-interprétation est terminée, et toute notion de pré-isotopie est désormais obsolète. Le menu «pré-isotopies» de la fenêtre principale est désactivé. Par contre, il est maintenant possible de manipuler les isotopies et les taxèmes nouvellement créés, et nous allons justement discuter de cela.

2.3 Fenêtre des taxèmes

Nous allons donc nous intéresser maintenant aux taxèmes. La fenêtre principale les concernant est activée par le menu «Taxèmes» de la fenêtre principale, et active la fenêtre représentée par la figure 5.5.

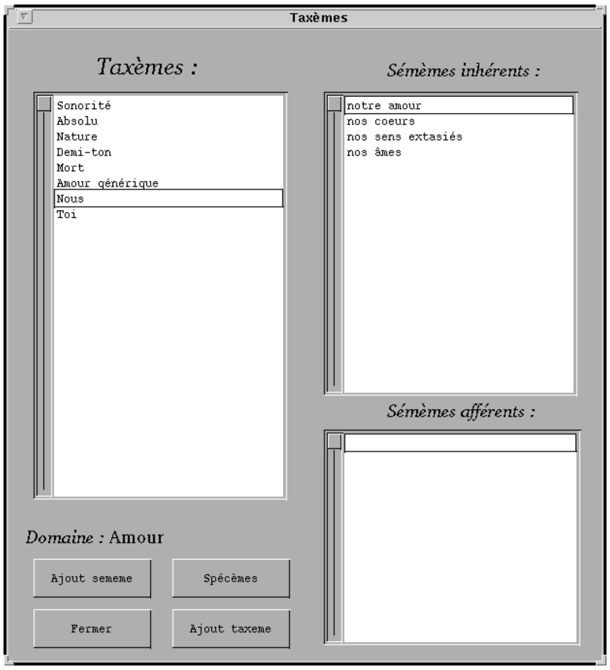


FIG. 5.5 – *Fenêtre des taxèmes*

La partie gauche de cette fenêtre comporte la liste de taxèmes identifiés jusqu'ici. Sous celle-ci est indiqué le nom du domaine comportant ce taxème, s'il existe. En haut à droite se trouve la liste des sémème que contient ce taxème, et en dessous la liste des sémèmes pour lesquels le sème du taxème est afférent. Diverses actions assez simples sont accessibles depuis cette fenêtre :

- Identification d'un nouveau taxème. Il s'agit simplement de repérer un nouveau sème et de le déclarer comme sème microgénérique. Il sera bien entendu demandé à l'utilisateur de remplir ce taxème à l'aide d'au moins deux sémèmes, lexicalisés ou non.

- Ajout d'un sémème dans un taxème existant. Là, non plus, rien de sorcier, si ce n'est que cette opération entraînera un nouveau passage dans la boucle de spécification décrite ci-dessous.

- Spécification du taxème. Si le graphe d'opposition du taxème est inexistant ou incomplet, le bouton marqué «Spécèmes» dans la figure 5.5 indiquera en fait «A spécifier». Il permet ainsi d'accéder à l'algorithme d'activation des spécèmes liés au taxème sélectionné. Nous avons déjà décrit les différentes étapes de cet algorithme au quatrième chapitre, il ne nous reste donc qu'à éclaircir la réalité des étapes qui le composent :

  - Chaque sémème du taxème est présenté tour à tour à l'utilisateur, en rappelant à chaque fois la liste des autres sémèmes du taxème. L'utilisateur doit donc choisir au moins un de ces autres sémèmes, pour lesquels il aura à identifier un sème spécifique. Cette première opération de sélection se fait à l'aide de la fenêtre représentée dans la figure 5.6.

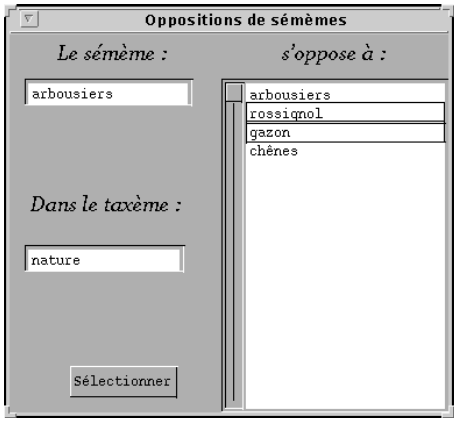


FIG. 5.6 – *Sélection des sémèmes à opposer*

– Dans un deuxième temps, et pour chaque sémème qui aura été sélectionné durant l'étape précédente, l'utilisateur aura à identifier un sème spécifique pour qualifier l'opposition entre deux sémèmes. Ceci se fait à l'aide d'une nouvelle fenêtre, représentée dans la figure 5.7.

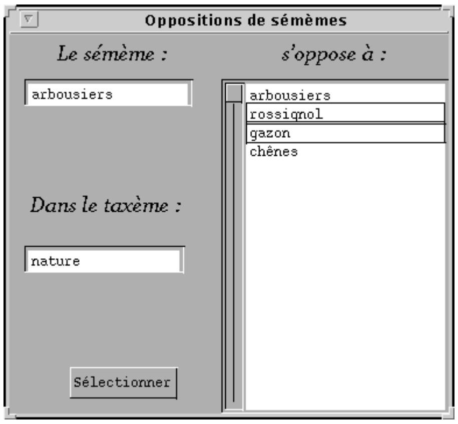


FIG. 5.7 – *Sélection du sème spécifique*

À partir de cette fenêtre, plusieurs actions sont possibles. Avant tout, les deux sémèmes à opposer sont rappelés dans la partie supérieure gauche. Juste en dessous de ces indications l'utilisateur peut rentrer directement la chaîne correspondant au sème qu'il veut utiliser pour qualifier l'opposition. Il peut également utiliser un sème déjà identifié, dans la liste située à droite, qui contient tous les sèmes utilisés dans l'analyse à ce stade. En bas à droite se trouve l'indication de l'éventuel sème spécifique utilisé pour qualifier le spécème opposé à celui qui est considéré actuellement. Suivant le statut du sème entré par l'utilisateur, il y a acceptation par le programme, et retour à la suite de la procédure, ou refus (si le sème proposé est générique pour les sémèmes à opposer, ou s'il qualifie déjà le spécème inverse). Dans ce dernier cas, l'utilisateur est averti et doit proposer un nouveau sème. Enfin, deux autres actions sont possibles : l'assimilation et l'élimination du sémème à opposer. L'assimilation s'opère entre les deux sémèmes à opposer par l'intermédiaire du bouton «Fusionner». Dans ce cas, le deuxième sémème proposé deviendra une simple occurrence du premier, et aucune opposition ne sera nécessaire. Cette action est à utiliser par exemple dans les cas de répétition de lexies dans un taxème. Si par exemple nous avions deux occurrences de 'chêne' et que ces deux occurrences n'avaient pas été déclarées comme telles dans la phase de pré-interprétation (via un ajout d'occurrence dans une pré-isotopie). La deuxième action, la suppression d'un sémème, s'applique dans ce cas au premier sémème à opposer, qui est donc purement et simplement éliminé de la liste. Cette action a été rendue accessible à ce stade puisqu'une erreur d'identification de sémème est incontournable lors de cette phase.

  - Enfin, lorsque chaque sémème a été opposé à au moins un autre sémème du taxème, prend place la phase de vérification de la complétude faible du graphe d'opposition du taxème. Si cette contrainte n'est pas satisfaite, l'utilisateur aura à identifier de nouveaux sèmes pour qualifier les oppositions manquantes, en utilisant les étapes précédemment décrites. Il n'aura cependant pas à sélectionner les sémèmes à opposer, ces derniers lui étant imposés par le programme.

- Enfin, si le taxème est déjà spécifié, donc si son graphe d'opposition satisfait les différentes contraintes, le bouton «À spécifier» s'affiche comme «Spécèmes» et permet de visualiser le graphe d'opposition, ou de rajouter un nouveau sème spécifique dans le graphe.

2.4 Fenêtre des sémèmes

Une autre fenêtre donne accès à la liste des sémèmes repérés durant l'analyse. Elle est représentée dans la figure 5.8.

Dans la partie gauche est représentée la liste de toutes les chaînes correspondant aux sémèmes. Dans la partie droite apparaissent clairement les différents sèmes du sémème sélectionné dans la liste, ainsi que le nombre d'épisémèmes (occurrences) couverts par ce sémème. Enfin, deux actions sont possibles à partir de cette fenêtre : l'élimination pure et simple du sémème sélectionné, qui peut entraîner la violation de certaines contraintes, et donc de nouvelles identifications, et la recherche des occurrences de ce sémème dans le texte, ce qui indiquera la position de celles-ci dans la fenêtre du texte.

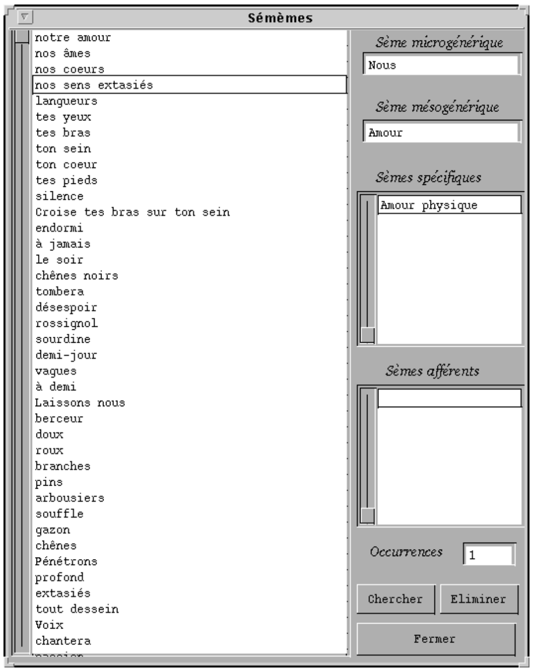


FIG. 5.8 – *Fenêtre des sémèmes*

2.5 Fenêtre des isotopies

Dans cette dernière série de fenêtres sont représentées tous les renseignements intéressants concernant les isotopies produites par l'analyse. La première donne un récapitulatif des renseignements «thématiques» d'une isotopie, dont un aperçu est visible dans la figure 5.9.

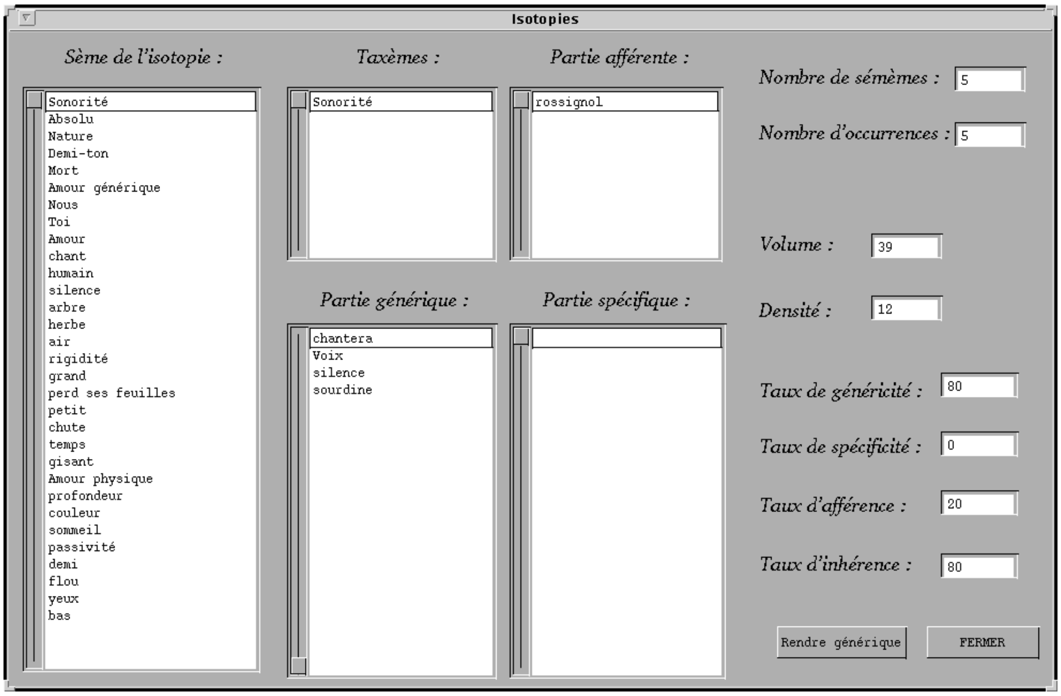


FIG. 5.9 – *Fenêtre des isotopies*

À gauche de cette fenêtre sont donc représentés les sèmes de la structure, et la sélection d'un de ceux-ci donne dans la partie droite les différentes composantes de l'isotopie correspondante : les taxèmes, les sémèmes de la partie générique, spécifique et afférente. Tout à droite sont les critères numériques de l'isotopie : nombre de sémèmes, d'épisémèmes, volume, densité et les différents taux correspondants aux proportions de sémèmes génériques, spécifiques, inhérents et afférents. Il est également possible d'appeler l'opérateur *igen* à partir de cette fenêtre (si l'isotopie n'est pas déjà générique bien entendu).

La deuxième fenêtre dédiée aux isotopies permet de visualiser la «forme» tactique d'une isotopie, et de comparer sur ce critère deux isotopies de la même structure. Cette fenêtre est représentée dans la figure 5.10.

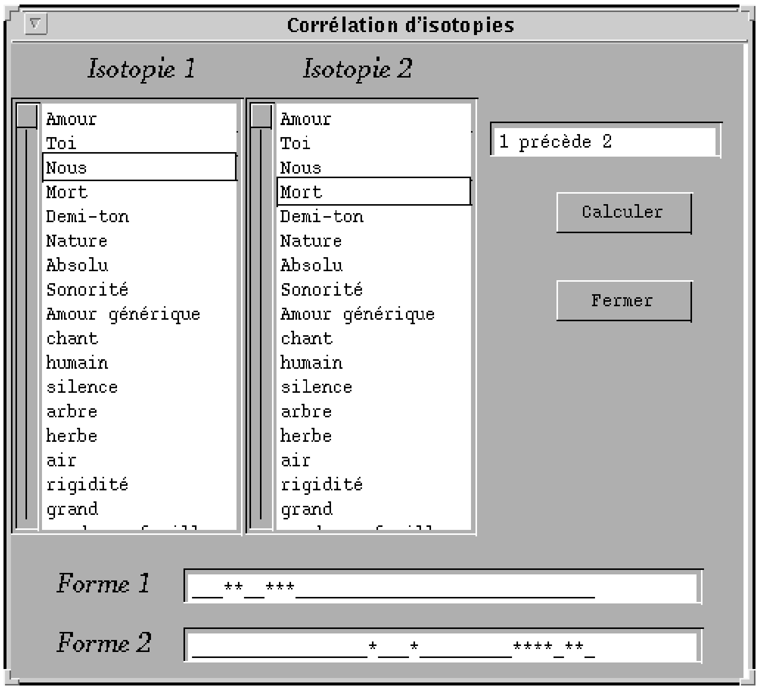


FIG. 5.10 – *Fenêtre des corrélations des isotopies*

Il est possible dans cette fenêtre de sélectionner deux isotopies dans la liste, et de visualiser leurs projections sur l'axe syntagmatique, symbolisées par les deux chaînes de «⋆» et de «−». L'interprétation de ces symboles est simple : la liste ordonnée suivant l'axe syntagmatique des épisémèmes du texte est représentée par une chaîne, le caractère correspondant est «⋆» si l'épisémème possède le sème correspondant (quel que soit son type), et «−» sinon. Un calcul simple sur ces deux formes donne une caractérisation des rapports tactiques entre les deux isotopies sélectionnées, qui apparaît en haut à droite de la fenêtre (précédence, inclusion, etc.).

2.6 Qualification de l'analyse

Une dernière fonctionnalité de l'outil que nous présentons ici n'est accessible qu'une fois l'analyse stabilisée, c'est-à-dire quand toutes les contraintes de la structure sont satisfaites. Nous proposons en effet quelques caractérisations simples du processus interprétatif établi par PASTEL. Ces informations, accessibles par le menu «Analyse» de la fenêtre principale, sont résumées dans une dernière fenêtre. Ces données quantitatives sont :

- Le nombre de classes minimales initiales. Ce nombre est fixé une fois les pré-isotopies repérées, et correspond donc aux taxèmes initiaux. Du point de vue interprétatif, il ne signifie rien dans l'absolu, ce nombre n'est là que pour des raisons de comparaisons avec les autres données.

- Le nombre de classes initiales qui ont été étendues durant l'analyse. Il s'agit en fait des isotopies initiales qui ont été modifiées, et sont devenues en partie spécifique. Ceci traduit généralement une extension des thématiques du départ.

- Le nombre de nouvelles isotopies repérées durant l'analyse. Il s'agit du nombre d'isotopies purement spécifiques. Ici non plus la considération isolée de ce nombre n'est pas significative, mais il sert de comparaison pour les autres données.

- Le nombre d'isotopies spécifiques qui «traversent» plusieurs taxèmes. C'est donc le nombre de «bonnes» isotopies spécifiques, qui ne concernent pas un seul taxème. Leur nombre indique donc la découverte de nouvelles thématiques. Il faut bien entendu considérer essentiellement le rapport entre cette valeur et la précédente pour évaluer en quelque sorte le taux de «découverte» de l'analyse effectuée.

- Le meilleur taux d'inter-taxémie de ces isotopies spécifiques. Parmi ces bonnes isotopies spécifiques, on note ici la plus étendue en terme du nombre de taxèmes traversés. Un nombre important est en fait l'indice d'une thématique forte dans le texte, et découverte durant l'analyse.

- Le taux général d'isotaxémie. C'est la moyenne en pourcentage du nombre de taxèmes traversés par une isotopie spécifique. Là encore l'ordre de subjectivité prime, et le mode de calcul n'est que trop simple. On peut en effet s'estimer satisfait si l'on obtient une seule isotopie spécifique étendue, même si celle-ci se perd dans la masse statistique d'isotopies locales.

3 Conclusion et perspectives sur PASTEL

Loin de vouloir présenter PASTEL comme un outil universel, nous préférons insister sur quelques aspects de la méthode proposée. PASTEL est avant tout une utilisation directe du formalisme que nous avons présenté dans les deux précédents chapitres. Bien que ce logiciel reprenne fidèlement les principes qui composent cette formalisation, il ne propose guère plus qu'une possibilité de sa mise en place. Le tout reprend bien cette vision de l'interaction entre l'homme et la machine que nous avons explicitée dans notre introduction : toute la «créativité» et l'initiative de l'analyse d'un texte relève de l'humain, et la machine ne remplace en fait qu'un support classique du type papier-crayon. Ou plutôt papier-formulaire, formulaire dont les cases sont les moteurs de vérification de contraintes qui constituent la véritable part informatique de notre outil. Il ne pouvait pas en être autrement, au vu de notre revendications initiales de reconnaître le sens comme en dehors de la machine. Que fait alors PASTEL ? Il prend en charge le côté calculatoire de l'interprétation.

Analyser un texte avec PASTEL est donc une sorte de jeu aux règles complexes, dont le but n'est pas clairement défini ; il n'est actuellement pas possible de déclarer une analyse terminée, sinon par abandon du protagoniste. Mais en réalité le problème n'est pas celui d'une interprétation terminée mais achevée. La différence est celle du régime de la satisfaction de l'utilisateur : il trouve sa pleine justification au sentiment d'avoir compris quelque chose, d'être parvenu à exprimer plus profondément ses intuitions initiales, ce qui prend sa forme opérationnelle dans la mise en place des jeux isotopiques.

Pour en justifier l'utilité, le plus simple est de reprendre la devise oulipienne, qui exprime que la création émerge de la contrainte. Vouloir en effet exprimer le grand nombre de trivialités repérables dans un texte, quel que soit le niveau de l'analyse, et en suivant les déviations plus ou moins bien fléchées du protocole entre bien dans ce cadre.

Plus concrètement toutefois, c'est sur le principe structuraliste de l'opposition entre signifiés proches (appartenant à une même classe) que repose l'aide à la créativité. Il est sans doute plus naturel de rapprocher que d'éloigner, d'où la plus grande facilité ressentie durant la phase de pré-interprétation, par rapport à l'activation des graphes d'opposition. Marc Cavazza se posait pour une de ses applications la question de la caractérisation d'un «bon taxème» : Nous proposons, comme ébauche de réponse, la possibilité d'activer le graphe d'opposition correspondant à ce taxème. Ce qui entraîne bien entendu comme critère la cardinalité faible du dit taxème, mais surtout la véritable proximité sémantique des signifiés qui le composent.

Un aspect intéressant de l'étape d'activation du graphe d'opposition à l'aide de PASTEL est la présence de la liste des sèmes déjà repérés lorsque l'utilisateur doit qualifier une opposition. Cette facilité est sans doute plus qu'une simple commodité : à notre avis elle va sauver en partie l'extension des isotopies spécifiques. De façon quelque peu parodique, c'est une sorte de principe d'«économie cognitive» qui est exploité : si l'utilisateur peut qualifier une opposition, même de façon approximative, en sélectionnant simplement sur un élément de la liste, plutôt que de préciser *ex-nihilo* un nouveau sème, que fera-t-il ? Il modifiera en quelque sorte son analyse, de façon satisfaisante, pour se plier aux contraintes de la machine et du protocole. En poussant à outrance ce principe, la structure finale comportera très peu d'isotopies, et les isotopies spécifiques seront très riches. Encore mieux, cela permet de développer l'analyse en deux étapes. De telles macro-isotopies spécifiques, une fois rendues génériques par *igen* nécessiteront à coup sûr l'identification de nouvelles oppositions, puisqu'elles translateront un grand nombre de sémèmes, avec appauvrissement de nombreux taxèmes, etc. L'utilisateur devra donc, d'une manière bien plus souple que pour l'activation d'un graphe d'opposition respectable, approfondir la pensée qu'il n'avait qu'en partie exprimée par économie. Et ce d'autant plus facilement que les oppositions seront nouvelles qualifier, et d'autant plus faciles que certains des sémèmes sur lesquelles elles s'articulent seront non-lexicalisés, donc proposés par l'interprète lui-même.

Une dernière discussion sur ces isotopies spécifiques «généricisables», serait de les associer à la notion proposée par F. Rastier de dimension. Cette classe transversale de grande généralité que nous n'avions pas intégrée dans notre formalisme réapparaît sans doute ici. Nous n'y trouvons peut-être pas les oppositions binaires qui les articulent, mais sans doute leur universalité thématique. Il est clair qu'alors les rendre génériques serait une erreur. Mais il serait peut-être intéressant d'envisager les dimensions comme une autre forme de pré-isotopie, sans qu'elle donne lieu à des classes définitoire.

Conclusion et perspectives

Ici s'achève donc la présentation de nos travaux, laissant en friche bien des aspects de l'approche sémantique du traitement de la langue naturelle, et bien des questions en suspens. Afin d'expliciter les différents aspects de cette conclusion, nous les séparerons en trois parties, depuis la prise en compte des notions et objectifs initiaux aux perspectives applicatives envisageables, en passant par quelques questions sur l'utilisation de la sémantique interprétative.

1 Concessions passées et futures sur les principes initiaux

Du point de vue formel, les principaux objectifs que nous nous étions fixés ont été atteints, par des voies diverses.

La notion de primauté de la sémantique s'est réduite à un abandon de toute méthode d'analyse syntaxique, puisque la seule considération de « surface » dont nous nous servons est la seule disposition des signifiants sur l'axe syntagmatique. De plus, cette simple projection sur l'axe est entièrement soumise au palier de description sémantique, que nous avons placé au centre de nos préoccupations.

Prenant le contre-pied de la compositionnalité, nous avons également réduit la place de mécanismes inférentiels au strict minimum. Les phases de calcul de notre logiciel sont en effet réduites à la vérification de contraintes, dont le repérage n'entraîne qu'une sollicitation de l'utilisateur. Lors de la description des transformateurs globaux, toutefois, nous avons proposé certaines manipulations de la structure sémantique pour lesquelles la phase d'autonomie de la machine était accrue (disons qu'elles constituent un ensemble d'opérations non supervisées). Sans que nous y accordions une place centrale, la simple possibilité de telles phases d'automatisation nous semble importante, puisqu'elle traduit une forme de cohérence et de stabilité de la structure que nous avons proposée, la rendant capable de se plier à un ensemble de contraintes externes.

Toutefois, si une certaine forme de l'automatisation de l'attribution des sèmes devait être appliquée, elle ne serait qu'une simple suggestion. Cette suggestion pourrait aisément provenir d'une base de données du style dictionnaire électronique, modifié de façon à présenter un ensemble de traits sémantiques pour ses entrées. L'utilité de ces sèmes «automatiques» devrait être réduite à certaines zones de l'analyse que l'on pourrait qualifier de périphériques, l'ensemble des contraintes gérées par le système entraînant comme il se doit des requêtes d'information sémantique non nécessaires au vu des objectifs interprétatifs. Mais il s'agit là du prix à payer pour la rigueur de la couche logique dont parlait Y.-M. Visetti [65], comme nous l'avons explicité au premier chapitre.

2 Aspects formels de l'approche

La formalisation que nous avons proposée aux troisième et quatrième chapitres mérite également d'être discutée ici d'un point de vue général. Son inspiration ensembliste ne met finalement en œuvre que des concepts assez simples formellement. Une telle approche aurait peut-être gagné en profondeur par l'utilisation de formalismes généraux plus sophistiqués, aussi nous discuterons ici des raisons de ce choix.

Tout d'abord, la volonté de restriction est originelle : le but était d'éclaircir les assertions formelles déjà présentes dans la sémantique interprétative ; dès lors le formalisme devait se plier à celles-ci. Un formalisme plus évolué aurait, à notre avis, entraîné un certain nombre de biais purement formels qui aurait contraint les principes initiaux. Il aurait introduit, par exemple des opérateurs nécessaires à la cohésion formelle, sans que pour autant ceux-ci aient une projection directe dans la théorie linguistique. Nous n'avons pas, bien sûr, échappé à cela, mais nous estimons que les nouveaux concepts introduits (comme les spécèmes) se situent à un niveau très général, incontournable dans toute approche formelle.

De plus, les métaphores ensemblistes sont présentes à l'origine dans les travaux de F. Rastier. Il y avait cependant quelques pièges à éviter. Si la notion de classe sémantique (taxème, domaine) ne posait que peu de problèmes, la double nature de celle de sémème a dû être élucidée. Considérer un sémème comme un ensemble de sèmes (du moins comme point de vue initial) aurait brisé les rapports entre local et global.

D'autres approches ont toutefois été envisagées, dont il est question dans

[71] et [27]. Déjà, F. Rastier avait rejeté des méthodes classiques en IA [48]. Une possibilité restait envisageable, concernant la trop grande rudesse des relations ensemblistes : celle de la méréologie de Leśniewski [37, 38], mais leur développement formel nous a semblé insuffisant, ou orienté vers d'autres directions.

Enfin, notre choix final nous semblait adapté aux principes de la coopération homme-machine définis précédemment : rapidement mis en place, sans une complexité interne trop contraignante, nous avons pu envisagé, plus centralement, les liens directs avec l'interprétation et la médiation sémiotique.

3 Autres aspects de la sémantique interprétative

Pour ce qui est de la théorie linguistique de la sémantique interprétative qui a guidé nos travaux, là aussi diverses questions restent sans réponse à la fin de ce travail. Nous avons déjà discuté au second chapitre des aspects que nous ne pouvions envisager de capter formellement, comme la notion de norme. Mais la théorie de F. Rastier a également connu certaines évolutions depuis le premier ouvrage. La principale avancée dans ce domaine est à nos yeux la prise en compte des différents niveaux de l'interprétation. Non seulement les différents outils conceptuels de la théorie permettent d'accéder à différents niveaux de granularité du texte, et cela reste cohérent avec la forme initiale de la théorie, mais les mécanismes qui les manipulent dépendent, eux, du niveau d'analyse.

Dans son dernier ouvrage consacré à l'explicitation de ces principes [52], l'auteur envisage trois paliers de description : le niveau microsémantique, centré autour du sémème ou plutôt de la lexie ; le niveau mésosémantique, pour les paliers intermédiaires que sont la phrase ou la période, et enfin le niveau macrosémantique, qui couvre tout le texte.

Notre formalisation, si elle devait se déterminer par rapport à ce découpage, se placerait sans doute autour du niveau mésosémantique, ou peut-être macrosémantique. L'application que nous avons mise en place, puisqu'elle s'applique de façon privilégiée à des textes courts, s'intéresse en effet aux structures plus globales. La description des mécanismes sémantiques locaux sont limités par l'absence de prise en compte des relations syntaxiques, qui constituent la norme majoritaire de ces derniers, et par l'absence de systématicité dans le choix des lexies décrites. Le niveau microsémantique serait donc, au vu de ces deux lacunes principales, le site privilégié d'un retour à une certaine forme d'automatisation. Prenons tout de même une distance devenue instinctive, puisque la reconnaissance de ces trois niveaux n'est pas un découpage absolu : les considérations sémantiques s'entrecroisent entre ceux-ci, et il est hors de question de mettre en place des mécanismes locaux guidés par la seule syntaxe, puisque cette description au niveau des unités minimales doit être entièrement guidée par les considérations globales (méso- et macro-sémantiques).

Indépendamment de ce découpage en niveaux, F. Rastier propose également une classification des descriptions interprétatives en quatre types (dialectique, thématique, tactique et dialogique), dont nous avons déjà discuté au troisième chapitre. Si deux de ces modalités de descriptions sont accessibles depuis notre formalisation (thématique et tactique), les deux autres nécessitent une extension de notre appareil formel, et peut-être le retour vers un mode de description plus proche de la logique[2]. De telles considérations tendent en effet vers une organisation des sèmes (et non plus des sémèmes), en mettant en place des relations casuelles entre ceux-ci. Quoiqu'il en soit, ces modifications seraient, à notre avis, plus à considérer comme une «surcouche» formaliste, se basant en grande partie sur la description sémantique «classique» de la sémantique interprétative, et ne remettent pas en cause les outils descriptifs mis en place.

4 Ouvertures applicatives

Hormis les quelques variations autour du thème que nous avons développées au précédent chapitre, quelques pas plus ambitieux sont envisageables, et certains sont déjà en cours. Les inspirations philologiques de la sémantique interprétative ouvrent une voie vers la description des phénomènes que l'on peut regrouper sous le terme d'inter-textualité. En constituant une avancée de plus vers la compréhension des phénomènes d'attribution de sens, elle projette la globalité non plus sur le texte et un entour reconnu indescriptible de façon satisfaisante, mais sur un ensemble de textes. Cette vision ajoute ainsi un palier de description, et donc de complexité, mais qui permet par contre d'expliciter plus finement l'identité même d'un texte. Des travaux [28, 60] reprennent déjà les principes généraux exprimés ici, en attaquant de front la question de la relation entre textes et ses influences sur les descriptions internes au texte. La place de l'outil informatique dans ce cas est d'autant plus justifiée, par le développement des méthodes d'interfaces homme-machine pour lesquelles la pluralité des représentations, aussi complexes que celles d'un texte décrit comme nous l'avons présenté, n'est plus le principal obstacle.

2. Sans pour autant rejoindre un positivisme logique sans concessions.

Cette remarque rejoint d'ailleurs celle des approches de linguistique statistique, pour reconnaître la supériorité de la machine dans le contexte du nombre de signes traitables, pour peu que celles-ci soient bien décrites [43].

Pourra-t-on un jour prochain envisager des rapprochements plus établis entre les différentes approches du TALN, et enfin proposer un outil véritablement complet dans les travaux de linguistique et littérature ? Il y aura sans doute un grand nombre de concessions à faire, mais ces concessions exigent à notre avis, avant tout, de prendre un parti et de s'y tenir quelque temps... Peut-être même, beaucoup de temps.

Bibliographie

[1] Abeillé (Anne). – *Les nouvelles syntaxes : Grammaires d'unification et analyse du Français.* – Armand Colin, 1993.

[2] Assadi (Houssem). – Interactive semantic analysis for building conceptual models from corpora. *In : Corpus-oriented semantic analysis.* – Proceedings of the ECAI-96 Workshop, 1996.

[3] Bar-Hillel (Y.). – *Language and Information.* – Reading, Mass., Addison-Wesley, 1964.

[4] Barr (A.) et Feigenbaum (E.A.). – *The handbook of artificial intelligence.* – Pitman, 1981.

[5] Barthes (Roland). – *L'aventure sémiologique.* – Seuil, 1991, *Points, collection Essais.*

[6] Benvéniste (Émile). – *Problèmes de linguistique générale.* – Paris, Gallimard, 1966 volume 1.

[7] Benvéniste (Émile). – *Problèmes de linguistique générale.* – Paris, Gallimard, 1974 volume 2.

[8] Béringer (H.). – Disambiguation with purely linguistic knowledge. *In : Actes AFCET 89.*

[9] Brachman (R.J.) et Schmolze (J.G.). – An overview of the KL-ONE knowledge representation system. *Cognitive Science*, vol. 9, 1985, pp. 171–216.

[10] Brézillon (P.). – *Proc. of the IJCAI-93 Workshop on "Using Knowledge in its context".* – Rapport technique n° 13, LAFORIA, Institut Blaise Pascal, 1993.

[11] Chomsky (Noam). – *Structures syntaxiques.* – Paris, Seuil, 1969.

[12] Chomsky (Noam). – *Aspects de la théorie syntaxique.* – Seuil, 1971.

[13] Collins (A.M.) et Quillian (M.R.). – Retrieval from semantic memory. *Journal of verbal learning and verbal behavior*, 1969.

[14] Coseriu (Eugenio). – L'étude fonctionnelle du vocabulaire : précis de lexématique. *Cahiers de lexicologie*, no 29, 1976, pp. 5–23.

[15] Coulon (Daniel) et Kayser (Daniel). – Informatique et langage naturel : présentation générale des méthodes d'interprétation des textes écrits. *Technique et Science Informatiques*, vol. 5, n° 2, 1986, pp. 103–128.

[16] Desclés (Jean-Pierre). – *Langages applicatifs, langues naturelles et cognition.* – Paris, Hermès, 1990.

[17] Ducrot (Oswald). – *Le structuralisme en linguistique.* – Paris, Seuil, 1968, *Qu'est-ce que le structuralisme ?*

[18] Frege (G.). – *Écrits logiques et philosophiques.* – Paris, Seuil, 1971.

[19] Fuchs (Catherine) et al. – *Linguistique et traitements automatiques des langues.* – Paris, Hachette, 1993.

[20] Greimas (A.J.). – *Sémantique structurale : recherche de méthode.* – Paris, PUF, 1986.

[21] Guiraud (Paul). – Les structures élémentaires de la signification. *Bulletin de la société linguistique de Paris*, 1965, pp. 97–114.

[22] Habert (B.), Naulleau (E.), Nazarenko (A.) et Salem (A.). – Construction symbolique puis quantitative des catégories sémantiques d'un corpus. *In : Journées de Sémantique Lexicale Brestoises.* – ENSTBr, 1996.

[23] Harris (Z.). – Distributional structure. *Word*, 1954, pp. 146–162.

[24] Hendrix (G.G.). – Human engineering for applied natural language processing. *In : Proceedings of the fifth IJCAI*, pp. 183–191. – 1977.

[25] Jakobson (R.). – *Essais de linguistique générale.* – Editions de Minuit, 1963.

[26] Kanellos (I.) et Tanguy (L.). – Du sens interne et externe d'un terme. *Cinquièmes journées ERLA-GLAT - Université de Bretagne Occidentale*, 1994.

[27] Kanellos (I.), Tanguy (L.) et Zaldivar-Carillo (V.-H.). – Représentation des connaissances et sémantique interprétative. *In : Actes du Troisième congrès européen de systémique*. – Rome, 1996.

[28] Kanellos (Ioannis) et Thlivitis (Théodore). – Lexical semantics and intertextuality. Linguistic theory and means of computer assistance. *In : 16th International Congress of Linguists.*

[29] Katz (J.J.) et Fodor (J.A.). – The structure of a semantic theory. *Language*, no 39, 1963, pp. 170–210.

[30] Kayser (Daniel). – Une sémantique qui n'a pas de sens. *Langages*, 1987, pp. 33–45.

[31] Kleiber (Georges). – Sur la définition sémantique d'un mot : les sens uniques sont-ils des impasses ? *In : La définition.* Colloque "La définition" du CELEX de l'Université Paris-Nord (18-19 Novembre 1988), pp. 125–147. – Larousse.

[32] Kleiber (Georges). – Prototype et prototypes : encore une une affaire de famille. *In : Sémantique et cognition*, pp. 103–130. – CNRS, 1991.

[33] Kristeva (Julia). – *Histoires d'amour*. – Folio, 1983.

[34] Kristeva (Julia). – *La révolution du langage poétique : l'avant-garde à la fin du XIXème siècle, Lautréamont et Mallarmé.* – Seuil, 1985.

[35] Langacker (Ronald W.). – An introduction to cognitive grammar. *Cognitive Science*, vol. X, 1986, pp. 1–40.

[36] Langacker (R.W.). – *Foundations of cognitive grammar. Vol 1 : theoretical prerequisites.* – Stanford University Press, 1987.

[37] Lejewski (C.). – A single axiom for the mereological notion of proper part. *Notre Dame Journal of formal logic*, vol. VIII-4, 1967, pp. 279–285.

[38] Lejewski (C.). – A note on Leśniewski's Axiom System for the Mereological Notion of Ingredient or Element. *Topoi*, vol. II-1, 1983, pp. 63–71.

[39] Lyons (John). – *Linguistique générale.* – Paris, Larousse, 1977.

[40] Lyons (John). – *Sémantique générale.* – Paris, Larousse, 1980.

[41] Minsky (Marvin). – A framework for representing knowledge. *In : The psychology of computer vision.* – P. Winston, 1975.

[42] Noailly (Michèle). – Dans le sens du fleuve : syntaxe et polysémie. *In : Polysémie et construction du sens*, éd. par Fall (K.), Léard (J.M.) et Siblot (P.), pp. 25–40. – Université Paul Valéry - Montpellier 3, Praxiling, 1996.

[43] Olsen (Mark). – Signs, symbols and discourses : a new direction for computer-aided literature studies. *Computers and the humanities*, vol. 27, n° 5-6, 1993.

[44] OULIPO. – *La littérature potentielle.* – Gallimard, 1973.

[45] Piaget (J.) et Inhelder (B.). – *La psychologie de l'enfant.* – Paris, PUF, 1993.

[46] Pottier (Bernard). – *Théorie et analyse en linguistique.* – Paris, Hachette, 1987.

[47] Pottier (Bernard). – *Sémantique générale.* – Paris, PUF, 1992.

[48] Rastier (François). – Représentation du contenu lexical et formalismes de l'intelligence artificielle. *Langages*, vol. 67, 1987, pp. 77–102.

[49] Rastier (François). – *Sémantique interprétative.* – Paris, PUF, 1987.

[50] Rastier (François). – *Sens et textualité.* – Paris, Hachette, 1989.

[51] Rastier (François). – Le défigement des expressions figées et leur interprétation. *In : Polysémie et construction du sens*, éd. par Fall (K.), Léard (J.M.) et Siblot (P.), pp. 17–24. – Université Paul Valéry - Montpellier 3, Praxiling, 1996.

[52] Rastier (François), Cavazza (Marc) et Abeillé (Anne). – *Sémantique pour l'analyse : de la linguistique à l'informatique.* – Paris, Masson, 1994.

[53] Rastier (François). – Communication ou transmission ? *Césure*, no 8, 1995, pp. 151–195.

[54] Rivière (François). – *Les typologies textuelles et le traitement automatiques des langues naturelles.* – Rapport de dea, ENSTBr / Université de Rennes 1, 1996.

[55] Sabah (Gérard). – *L'IA et le langage.* – Paris, Hermès, 1989.

[56] Saussure (Ferdinand de). – *Cours de linguistique générale.* – Payot, 1972 [1916].

[57] Siblot (Paul). – La linguistique peut-elle traiter de la «représentation des connaissances dans le lexique» ? *Cahiers de praxématique (Praxiling, Université de Montpellier III)*, vol. 21, 1993, pp. 142–161.

[58] Sowa (J.F.). – *Conceptual structures : information processing in mind and machine.* – Reading, MA, Addison-Wesley, 1984.

[59] Tanguy (L.) et Thlivitis (T.). – Pastel : un protocole informatisé d'aide à l'interprétation des textes. *ILN 96*, 1996.

[60] Thlivitis (Théodore) et Kanellos (Ioannis). – Interprétation intertextuelle assistée par ordinateur. *In : Journées Scientifiques et Techniques du Reseau Francophone de l'Ingénierie de la langue.*

[61] Thom (René). – *Modèles mathématiques de la morphogenèse.* – Paris, Christian Bourgois, 1980.

[62] Turing (A.M.). – Computing machinery and intelligence. *Mind*, no 59, 1950, pp. 433–460.

[63] Varela (Francisco). – *Connaître : les sciences cognitives - tendances et perspectives.* – Seuil, 1989.

[64] Victorri (Bernard), Raysz (Jean-Pierre) et Konfe (Amadou). – Un modèle connexionniste de la polysémie. *In : Neuro Nîmes 89*, pp. 97–107.

[65] Visetti (Yves-Marie). – Des systèmes experts aux systèmes à base de connaissances : à la recherche d'un nouveau schéma régulateur. *Intellectica*, vol. 12, 1991, pp. 221–280.

[66] Weaver (W.). – Translation. *In : Machine translation of languages*, éd. par Locke (W.N.) et Booth (A.D.), pp. 15–23. – Technology Press of MIT and Wiley, 1955.

[67] Weill-Fassina (Annie). – Dynamique des représentations et gestion des actions. *In : Représentations pour l'action*, éd. par Weill-Fassina (A.), Rabardet (P.) et Dubois (D.). – Octares, 1993.

[68] Weizenbaum. – Eliza : A computer program for the study of natural language communication between man an machine. *CACM*, no9, 1966, pp. 36–45.

[69] Winograd (T.). – *Understanding natural language.* – New-York Academic Press, 1972.

[70] Woods (W.A.). – Progress in natural language understanding : An application to lunar geology. *In : AFIPS Conference Proceedings.* pp. 441–450. – Montvale, N.J., 1963.

[71] Zaldivar-Carrillo (Victor-Hugo). – Contributions à la formalisation de la notion de contexte : le concept de "théorie" dans la représentation des connaissances. *Thèse de Doctorat - Université de Montpellier 2*, 1995.

Table des matières

Table des figures

Index

Annexe A

Glossaire des principales notions

- **Afférence :** relation non définitoire entre deux sémèmes. L'afférence est à la source de l'attribution d'un sème à un sémème, et se situe au sein d'une norme «locale». Le sème ainsi attribué est donc afférent, par opposition à inhérent, ou définitoire. Dans notre formalisation, l'afférence est une attribution libre et directe d'un sème à un sémème, nous permettant de contourner localement les contraintes qui régissent la structure sémantique.
- **Dimension :** classe de sémèmes de plus haute généralité, correspondant à des sèmes comme /animé/, /inanimé/, etc. Nous considérons ici la dimension comme une isotopie spécifique étendue, et non comme une classe sémantique.
- **Domaine :** classe de sémèmes intermédiaire, entre le taxème et la dimension. Un domaine est traité ici comme un ensemble de taxèmes, auquel est associé un sème unique qui prend le statut de méso-générique.
- **Épisémème :** occurrence d'un sémème. L'identité d'un épisémème se base sur la localisation de la chaîne de caractères qui le manifeste. Un épisémème est associé à exactement un sémème. Un sémème peut posséder plusieurs, ou aucun épisémème, et ceci indépendamment des chaînes de caractères associées à ces épisémèmes.
- **Généricisation :** transformation d'une isotopie afin de la rendre générique. Elle entraîne l'extension ou la création d'un taxème, et diverses modifications de la structure sémantique.
- **Inhérence :** qualification d'une attribution de sème à un sémème. Un sème inhérent traduit l'appartenance d'un sémème à une classe sémantique, ou son opposition à un autre sémème. Cette relation est définitoire pour le sémème.
- **Isosémémie :** relation d'équivalence entre deux spécèmes, signifiant que ces deux spécèmes ont le même première projection, donc qu'ils caractérisent le même sémème (en l'opposant toutefois à des sémèmes différents).
- **Isotaxémie :** relation d'équivalence entre deux spécèmes, signifiant que ces deux spécèmes opposent des sémèmes d'un même taxème.
- **Isotopie :** effet de la récurrence d'un sème dans un texte. Une isotopie est ici plus généralement l'ensemble des unités sémantiques (sémèmes, taxèmes, spécèmes) auxquelles un sème est attribué.

Fonction d'isotopie : fonction responsable de l'attribution d'un sème à un ensemble d'unités sémantiques (sémèmes, taxèmes, spécèmes).

- **Pré-interprétation :** première étape du processus d'interprétation, consistant en la définition de l'ensemble des sémèmes et épisémèmes, et en la déclaration des pré-isotopies. La pré-interprétation se distingue principalement de l'interprétation proprement dite par l'absence de toute contrainte dans son déroulement.

- **Sème :** marqueur sémantique qualifiant une interprétation d'un signifiant. Un sème est exprimé par une chaîne de caractères, et est attribué à un nombre non déterminé d'unités sémantiques (sémèmes, classes, ou spécèmes). Le type d'attribution d'un sème particulier à un sémème particulier entraîne une qualification du sème parmi :

  Sème afférent : si la relation sème-sémème est directe et libre, correspondant à la notion d'afférence.

  Sème générique : si le sème est attribué à une classe sémantique qui contient le sémème considéré. Suivant le type de classe, ou distingue les sèmes **micro-génériques** (taxèmes), **méso-génériques** (domaines) et **macro-génériques** (dimensions, non traitées dans notre formalisme).

  Sème spécifique : si le sème est attribué à un spécème dont le sémème considéré est la première projection. Ce sème caractérise donc un sémème donné en l'opposant à un plusieurs sémèmes du même taxème.

  Sème inhérent : si le sème est définitoire pour le sémème considéré. Dans notre formalisation, un sème inhérent est donc un sème générique ou spécifique.

- **Sémème :** signifié minimal, correspondant à une lexie. Un sémème est décrit, au cours d'une interprétation, par un certain nombre de sèmes de types différents. Un sémème peut donc être vu comme un ensemble de sèmes typés.

- **Spécème :** support formel d'un sème spécifique. Un spécème est un couple de sémèmes appartenant à un même taxème, indiquant une opposition sémantique entre ceux-ci.

  Activation d'un spécème : attribution d'un sème à un spécème, *via* la fonction d'isotopie.

Graphe d'opposition : ensemble des spécèmes activés opposant les sémèmes d'un taxème.

- **Taxème :** classe minimale de sémèmes, au sein de laquelle se mettent en place les oppositions, dont la forme est définie par des spécèmes. Un sémème appartient à exactement un taxème. Le sème unique attribué à un taxème est déclaré micro-générique

Annexe B

Notations utilisées

1 Notations de la sémantique interprétative

- Les sémèmes sont indiqués entre apostrophes : 'exemple'
- Les sèmes sont indiqués entre barres obliques simples : /exemple/
- Les classes sémantiques sont indiquées entre barres obliques doubles : //exemple//
- La relation d'incompatibilité entre sèmes est notée : |

2 Notations du formalisme

2.1 Notations de description

S	Ensemble des sémèmes
SE	Ensemble des sèmes
T	Ensemble des taxèmes
SP	Ensemble des spécèmes
E	Ensemble des épisémèmes
$=_X$	Relation d'identité sur l'ensemble X, prenant ses valeurs parmi ceux cités ci-dessus
tax	Relation liant un taxème aux sémèmes qu'il contient
$\equiv_T$	Relation d'équivalence entre deux sémèmes appartenant au même taxème
(s, s')	Définition en extension d'un spécème, s et s' étant des sémèmes
σ_1, σ_2	Première et seconde projection d'un spécème
opp	Relation d'opposition entre deux spécèmes (s, s') et (s', s)
$=_{is}$	Relation d'isosémémie entre deux spécèmes
$=_{it}$	Relation d'isotaxémie entre deux spécèmes
I	Fonction d'isotopie
I_{sp}, I_t, I_{aff}	Composantes de la fonction d'isotopie
I_{spe}, I_{gen}	Projections des deux premières composantes de I sur S
I_S	Projection de I sur S
id	Relation d'identité sémantique entre sémèmes
ext	Extension d'un sémème vers ses épisémèmes
I_E	Extension de I sur E
$\leq$	Relation d'ordre sur E
PI_E	Fonction de pré-isotopie définie par les épisémèmes
I_S	Fonction de pré-isotopie définie par les sémèmes

2.2 Notations de modification

$+x$	Opération d'ajout élémentaire d'une entité parmi s, t, se, sp, e
$-x$	Opération de retrait parmi s, t, se, sp, e
$add(s,t)$	Ajout du sémème s dans le taxème t
$sub(s,t)$	Suppression du sémème s du taxème t
$+i(x,se)$	Ajout de l'entité x (parmi s, sp,, t) dans l'isotopie du sème se
$-i(x,se)$	Suppression de l'entité x (parmi s, sp,, t) de l'isotopie du sème se
$activ(sp)$	Opération d'activation du spécème sp
$compt(t)$	Vérification de la contrainte de complétude du graphe d'opposition du taxème t
$pb_{aff}(se)$	Vérification de la contrainte de l'afférence pure pour l'isotopie du sème se
$ADD(s,t)$	Ajout d'un sémème à un taxème pour une structure stable
$elim(s)$	Élimination d'un sémème d'une structure stable
$add_{spe}(sp,se)$	Attribution du sème se au spécème sp dans une structure stable
$igen(se)$	Transformation d'une isotopie quelconque pour la rendre générique

Annexe C

Textes

1 En sourdine

Calmes dans le demi-jour
Que les branches hautes font,
Pénétrons bien notre amour
De ce silence profond

Fondons nos âmes, nos cœurs
Et nos sens extasiés
Parmi les vagues langueurs
Des pins et des arbousiers.

Ferme tes yeux à demi
Croise tes bras sur ton sein
Et de ton coeur endormi
Chasse à jamais tout dessein

Laissons nous persuader
Au souffle berceur et doux
Qui vient à tes pieds rider
Les ondes de gazon roux

Et quand, solennel, le soir
Des chênes noirs tombera
Voix de mon désespoir
Le rossignol chantera.

Paul Verlaine

2 Chirac envisage de se convertir, mais à quoi ?

Frédéric Pagès - *Le Canard enchaîné* - 11/10/95

Donc Chirac doit se convertir à l'Islam, selon les injonctions du GIA (ou de ceux qui se cachent derrière). Certains prennent cette hypothèse à la rigolade. Ils ont tort. Car Chirac est un caractère éminemment convertible. C'est de notoriété publique : ce type a plusieurs fois changé de religion au cours des derniers mois.

Après avoir prêché la baisse des impôts directs et indirects pendant la campagne présidentielle, il professe aujourd'hui le contraire, administrant l'extrême-ponction aux contribuables agonisants. Après avoir prêché que «la fiche de paie n'est pas l'ennemie de l'emploi», il propose aux fonctionnaires de faire carême. Après avoir anathémisé le dogme de l'équilibre budgétaire, et joué les fils prodigues, il se repent bruyamment, et, touché par la grâce, proclame sa foi dans la rigueur salariale. Il voulait écraser le démon de la technostructure et interdire les orgies bureaucratiques. Vaines prophéties ! Cet apôtre du franc fort va-t-il se convertir brutalement à la baisse des taux d'intérêt ? Pour son prochain pèlerinage, le Président préférera-t-il La Mecque à Colombey-Les-Deux-Eglises ? On tremble.

Les plus optimistes se rassurent en remarquant que cet été, à Brégançon, on découvrit un Chirac fort confit en dévotion catholique. A la messe, tous les dimanches ! Une ferveur qu'on ne lui connaissait pas et qui serait due, selon les experts chiracologues, à l'influence spirituelle et édifiante de sa femme Bernadette, qui devient ainsi - qu'on se le dise - l'ultime rempart de l'Occident chrétien.

Bilan : notre Président a deux conseillers familiaux : son épouse pour la religion, et sa fille Claude pour l'image. C'est bien là le problème. Après avoir vu cet été Chirac déambuler en bermuda clair et chaussettes noires, on se dit que cet homme hérétique jusqu'au bout des pieds est plus facilement convertible à l'islam qu'à l'élégance.