



**HAL**  
open science

## Canopée ©, Assistance interactive personnalisée pour l'interprétation

Pierre Beust, Roger Cozien, Serge Mauger

### ► To cite this version:

Pierre Beust, Roger Cozien, Serge Mauger. Canopée ©, Assistance interactive personnalisée pour l'interprétation. Journées de Rochebrune (Rencontres interdisciplinaires sur les systèmes complexes naturels et artificiels) "Echelles et modélisations multi-niveaux", Jan 2011, France. 11 p. hal-01069615

**HAL Id: hal-01069615**

**<https://hal.science/hal-01069615>**

Submitted on 29 Sep 2014

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

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

*Canopée*©  
**Assistance interactive personnalisée  
pour l'interprétation**

Pierre Beust\*, Roger Cozien\*\*, Serge Mauger\*

\* GREYC CNRS UMR 6072 – ISLanD & Pôle ModesCoS de la MRSH  
Université de Caen - Basse Normandie  
14032 Caen Cedex, France  
**{pierre.beust, serge.mauger}@unicaen.fr**

\*\* **EXO MAKINA**  
[www.exomakina.fr](http://www.exomakina.fr)  
[roger.cozien@exomakina.fr](mailto:roger.cozien@exomakina.fr)

Résumé :

Cet article présente le projet en cours de développement intitulé Canopée ©. Il vise à la construction d'un environnement d'assistance à l'interprétation des textes. Le but est de médiatiser l'activité interprétative du sujet et lui proposant des résultats statistiques et des visualisations qui, pour lui, soient porteurs de signification. C'est dans une interaction personne-système que se mettent en place de manière incrémentale les parcours interprétatifs.

Mots clés : Informatique & linguistique, Sens, Sémiotique, Interaction, Interprétation.

## 1. Le contexte de ce travail

Cet article rend compte d'un travail de recherche et de développement au sein de la société **EXO MAKINA** en collaboration avec le laboratoire GREYC CNRS UMR 6072 de l'université de Caen Basse-Normandie. L'objectif est de développer et de commercialiser un outil d'assistance à l'interprétation dédié à l'accès personnalisé aux contenus dans des collections documentaires numériques textuelles, dont celles de grandes tailles.

Dans les tâches numériques quotidiennes de la majeure partie des utilisateurs des services et programmes informatiques, l'appréhension de collections documentaires revêt une importance croissante. Ainsi, l'accès à un dossier ou à un répertoire est déjà une confrontation à une collection potentiellement assez grande. La gestion régulière d'une boîte de courrier électronique est aussi une façon d'appréhender une collection qui par nature est de taille croissante. La réponse même d'une requête à un moteur de recherche est aussi une collection assez vaste (en témoigne les estimations de nombres de réponses fournies) dont l'utilisateur ne regarde dans presque tous les cas que les premiers documents. Enfin, les flux d'informations (dépêches d'actualités, flux RSS, etc.) sont aussi des collections à part entière avec, en plus, la spécificité chronologique. Le projet Canopée cherche à produire une assistance logicielle dans le rapport entre un utilisateur particulier (voir un groupe d'utilisateurs) et une collection, rapport sous-tendu par les capacités interprétatives de l'utilisateur.

Dans une première partie de cet article nous détaillerons notre positionnement épistémologique sur la question de l'interprétation. Nous nous baserons principalement sur un champ théorique linguistique, celui de la sémantique interprétative. Ensuite, nous expliquerons en quoi la sémantique interprétative nous incite à problématiser l'interprétation dans un espace multi-échelles. Alors, nous présenterons la solution en cours de développement comme un environnement de travail implicitement structuré par ces multiples échelles. Enfin, nous ferons tirer quelques pistes pour la poursuite du travail.

## 2. Le positionnement épistémologique sur l'interprétation

Les sciences humaines et les sciences cognitives nous incitent à envisager le document électronique de manière indissociable de l'activité du ou des humains qui les produisent, les recueillent, les indexent et les recherchent. Ainsi, le champ de l'herméneutique nous montre à quel point les rapports intertextuels sont déterminants dans les interprétations des documents. Notre positionnement épistémologique quant à la question de l'interprétation découle de ce champ de l'herméneutique. C'est celui de la Sémantique Interprétative de François Rastier de tradition saussurienne.

En matière d'ingénierie documentaire sur Internet, une tendance actuelle est de renforcer une vision idéaliste du Web comme une vaste base de connaissances, ce qu'évidemment il n'est pas. C'est la démarche considérée dans le projet du Web Sémantique. L'objectif annoncé par Tim Berners-Lee (en 1998), initiateur du Web Sémantique et directeur du W3C, est d'enrichir<sup>1</sup> les documents en ligne à l'aide d'ontologies normalisées, soit automatiquement, soit en assistant leur auteurs. Cet enrichissement a pour but de formaliser la sémantique du document de sorte qu'il

---

<sup>1</sup> Notamment au moyen des technologies développées autour du langage XML.

soit notamment interprétable par des agents logiciels sans la supervision d'une interprétation humaine. Ainsi les agents logiciels et interprétants humains auraient accès à la même signification. Cette position tend à ramener la question de l'interprétation à un « décodage » d'une connaissance que le texte est sensé véhiculer. C'est-à-dire ramener l'interprétation à une forme de compilation.

L'interprétation est une perception sémantique individuelle selon Rastier. En tant que perception du sujet elle fait preuve d'une grande variabilité d'un sujet à un autre, variabilité éventuellement individuelle : il nous est tous arrivé de relire différemment un texte et d'en tirer des interprétations bien différentes. C'est ce qu'explique Théodore Thilivitis dans son concept d'anagnose, tentative d'objectivation des rapports intertextuels des différentes interprétations du passé d'interprétant d'un individu qui conditionnent les interprétations à venir. Il en découle que l'interprétation des textes, comme des images d'ailleurs, est située : elle est liée à l'individu, à ses pratiques et donc, plus généralement, fortement contextualisée et culturellement déterminée.

Le Web Sémantique est dans la lignée de ce que François Rastier appelle la tradition logico-grammaticale<sup>2</sup>. Ceci indique l'ensemble des travaux qui privilégient les signes, la syntaxe et la compositionnalité. On y constate une grande prédominance des modèles et des architectures d'analyse en couches où différents résultats indépendants s'enchaînent relevant successivement des niveaux de la morphologie, de la syntaxe, de la sémantique et finalement de la pragmatique. Dans une perspective de TAL ces modèles, où l'interprétant n'est finalement qu'en bout de chaîne, ne sont pas sans poser des problèmes d'explosion combinatoire car, pris isolément, chaque niveau génère un bon nombre d'ambiguïtés (et donc de faux résultats) que les niveaux suivants ne peuvent pas toujours dissocier des analyses correctes. Le contexte, et donc les conditions d'interprétation, y est aussi relégué en bout de chaîne, dans la phase d'analyse pragmatique. L'importance est d'abord donnée aux analyses co-textuelles (morphologie, syntaxe, sémantique) puis ce qui en sort est contextualisé relativement à une certaine représentation pré-existante du contexte. La contextualisation consiste ici en général à une ultime levée d'ambiguïtés. La culture quant à elle n'intervient jamais.

La question de la contextualisation du sens d'une unité linguistique (un mot, un syntagme ou un paragraphe) est d'une grande importance dans la perspective de la sémantique interprétative. Elle constitue le fondement de l'établissement de parcours interprétatifs. Selon François Rastier, la contextualisation est à la base de l'interprétation d'une unité linguistique (Rastier, 2001, page 92) :

*L'interprétation procède principalement par contextualisation. Elle rapporte le passage considéré, si bref soit-il ? ce peut être un mot : (i) à son voisinage, selon des zones de localité (syntagme, période) de taille croissante ; (ii) à d'autres passages du même texte, convoqués soit pour des tâches d'assimilation, soit de contraste ; (iii) enfin à d'autres passages d'autres textes, choisis (délibérément ou non) dans le corpus de référence, et qui entrent, par ce choix, dans le corpus de travail.*

C'est donc le positionnement en « global » d'un élément local qui permet par contextualisation son interprétation. Ainsi tout rapprochement d'une zone de texte dans les zones de localité du co-texte proche ou de l'intertexte est générateur d'opérations interprétatives (assimilation, dissimilation) analysées en termes d'échanges et de récurrences de sèmes.

En étendant ce principe de contextualisation aux ensembles documentaires et aux environnements numériques de travail, on peut tout à fait imaginer, par exemple, qu'un article de presse présenté dans une source documentaire très sérieuse ne sera pas perçu par l'interprétant de

---

<sup>2</sup> Selon F. Rastier, le cognitivisme constitue l'aboutissement contemporain de cette tradition logique et grammaticale. Elle est à opposer à la tradition rhétorique et/ou herméneutique qui, quant-à-elle prend pour objet les textes et les discours dans leur production et leur interprétation.

la même manière que le même article présenté dans une source plutôt satyrique : le contexte global instauré par les autres articles aiguillant forcément les parcours interprétatifs. La contextualisation émerge d'une activité de rapprochements des textes les uns relativement aux autres sans avoir du en passer par une représentation du contexte.

Dès lors qu'on parle de « vrais » textes, de « vrais » corpus, et pas simplement de phrases d'exemples artificiellement construites en dehors d'un contexte linguistique et pragmatique, la dimension perceptive personnelle fait évidence. Le sens ne repose pas dans le texte, mais dans les conditions d'interprétation. Il en résulte, à notre avis, que le sens ne peut être modélisé à la façon d'un résultat calculatoire qui serait plus ou moins complété ou dégradé d'un interprétant à un autre. Le sens n'est pas de nature symbolique ; c'est un processus sémiotique au centre de l'activité de l'interprétant. Ce processus sémiotique n'est pas un calcul déterministe prédictible, c'est une boucle perception-action. De ce point de vue, et cela tranche par rapport à la tradition logico-grammaticale : interpréter un texte ce n'est pas le transformer en autre chose.

Beaucoup de travaux en sémantique formelle (logique, DRT, SDRT, etc.) ont depuis des années déployé beaucoup d'efforts et de moyens pour obtenir de façon compositionnelle un « calcul du sens » acceptable, cherchant ainsi à remplacer le texte par une forme arbitrairement vue comme plus épurée. Force est de constater qu'un tel résultat reste hors de portée. Il ne s'agit pas ici que d'un problème d'évaluation dont on n'aurait pas encore bien mis en place la méthodologie mais d'un problème beaucoup plus profond : les alternances entre les fonds sémantiques et les formes textuelles sont constitutives de l'espace interprétatif. C'est vrai du point de vue de l'interprétation, comme celui de la parole qui n'est pas préméditée (elle se détermine en cours de production en réaction à elle-même comme l'a montré Jacques Courcil). C'est également le cas de la production écrite, en témoigne les phénomènes de réécriture et de raturage en cours de production écrite et plus généralement, en témoigne aussi les phénomènes de reconstruction et de recomposition de textes antérieurs (Mayaffre 2002) actualisés dans de nouveaux corpus.

Contrairement aux sémantiques formelles cherchant à résumer la donnée textuelle en une expression logique épurée, interpréter un texte, ce n'est donc surtout pas se passer du texte. Au contraire, c'est bien plus y un projeter son anagnose et ainsi démultiplier les possibilités de relations intertextuelles. C'est ce que montrent les opérations interprétatives en terme de dynamique des sèmes dans l'interprétation. Par exemple, considérons le vers :

*La terre est bleue comme une orange<sup>3</sup>*

Une interprétation logique sous forme d'analogie conclurait à un « non-sens logique » (quelque chose que nous pourrions paraphraser de la sorte : *la terre est bleue comme l'orange est bleue* !). De toute évidence l'interprétation est très loin du non-sens et peut notamment s'expliquer linguistiquement comme une actualisation dans l'interprétation des lexies terre et orange d'un sème commun /sphérique/ (cf. [photoblog.ludopics.com/index.php?showimage=14](http://photoblog.ludopics.com/index.php?showimage=14) pour une interprétation plus visuelle).

La théorie de la sémantique interprétative décrit cette opération interprétative d'actualisation comme très fréquente dans le parcours interprétatif porté par le texte. En témoignent les exemples suivants :

- *Women, Fire and Dangerous Things : What Categories Reveal about the Mind* (Titre d'un livre de G. Lakoff, 1987) : Actualisation dans *Women* du sème /dangereux/

---

<sup>3</sup> Premier vers du 7ème poème du premier chapitre "Premièrement" composant le recueil "L'amour la poésie", Paul Eluard, 1929

- *Il y a musique et musique* : Actualisation de 2 sèmes contradictoires (i.e. une dissimilation), l'un /bonne/ dans la première occurrence de musique, l'autre /mauvaise/ dans la seconde occurrence
- *Il l'emparouille et l'endosque contre terre* (extrait de "Le grand combat" de H. Michaux, 1966, p.14). les lexies en question n'ont pas de contenu. Les compréhensions qui peuvent en être faites sont uniquement dues par des actualisations de sèmes afférents (de même pour le fameux langage des *stroumphs*).

Relativement à ce positionnement épistémologique fortement ancré sur la sémantique interprétative de François Rastier, la question de la mise en œuvre d'une assistance logicielle à l'interprétation doit consister en une interaction entre le sujet et les textes dans une approche très fortement centrée utilisateur. Le sens prend forme dans une interaction entre un sujet, un corpus, des ressources lexicales et plus largement une culture. Il est question de médiatiser l'espace interprétatif du sujet en lui permettant, sur ses corpus (textes, emails, etc.) et à l'aide de visualisations ses facultés d'oppositions, d'établir des rapprochements, des regroupements. En somme, de déployer ses facultés interprétatives. La construction de ces visualisations et la suggestion de rapprochements et d'oppositions est rendue possible par une instrumentation des différents niveaux et échelles de l'espace interprétatif. Cet espace est structuré à la rencontre des échelles de textualité et par des zones praxéologiques du sujet. Ce que nous expliquons dans la partie suivante.

### 3. Echelles et interprétation

Le rapport d'un texte à d'autres textes est constitutif de l'espace interprétatif. C'est vrai des textes qui composent avec lui le corpus dont il est extrait, comme c'est également vrai des textes qui sont consultés avant, après ou plus ou moins en même temps que le texte en question. Ce rapport du texte à la collection n'est pas seulement une détermination de la collection à partir des textes qui la composent. La globalité de la collection agit sur les conditions d'interprétation d'un document. C'est une détermination du global sur le local qui est tout aussi importante que la détermination inverse. C'est typiquement un problème qui est évacué dans le positionnement du WEB sémantique où l'accès à un corpus n'est ni plus ni moins considéré que comme un accès itéré à des documents.

En analysant ces rapports de détermination du global sur le local, nous mettons en évidence qu'il existe des niveaux d'échelle qui influent sur l'interprétation des textes.

#### Les échelles de textualité

Les échelles de textualité sont des outils analytiques de la Sémantique Interprétative pour caractériser la dynamique sémique à l'œuvre dans la détermination du local par le globale. François Rastier décrit trois principes relevant trois échelles de textualité :

- le principe de **contextualité** : deux signes ou deux passages d'un même texte mis côte à côte sélectionnent réciproquement des éléments de signification (sèmes). Cette sélection réciproque s'analyse notamment en termes de récurrences de sèmes appelées isotopies. La portée dans le texte de l'isotopie détermine une zone de localité sémantique renforcée par l'étendue dans le discours des relations de prédication et d'anaphore. Cette zone, appelée *Période* ([Rastier & al. 94, p. 116]), définit le champ d'étude de la mésosémantique.
- le principe d'**intertextualité** : deux passages de textes différents, si brefs soient-ils, et fussent-ils réduits à la dimension d'un signe, sélectionnent réciproquement dès qu'ils sont mis côte à côte, des éléments de signification (sèmes). Le principe d'intertextualité généralise la notion d'isotopie au-delà de la linéarité syntagmatique du texte;

- le principe d'**architextualité** : Tout texte placé dans un corpus en reçoit des déterminations sémantiques, et modifie potentiellement le sens de chacun des textes qui le composent.

Ces 3 échelles sont des *principes* bien plus que des *paliers*, c'est pourquoi il ne faut pas entendre qu'il faille absolument partir du « barreau » de dessous pour aller au « barreau de dessus ». Le principe de contextualité est bien à l'œuvre dans les deux autres principes mais il n'est pas autant une étape préalable. L'interprétation d'un document au sein de sa collection produit les trois échelles. Une métaphore en terme d'échelle plus appropriée consisterait en une sorte d'échelle circulaire à deux sens où le dernier barreau permettrait de monter au premier ainsi que de descendre du premier barreau au dernier. Ce qu'illustre cette métaphore c'est que l'appréhension des rapports « global – local » dans l'interprétation procède à tous les niveaux de contextualité, d'intertextualité et d'architextualité d'alternance entre les formes et les fonds. Ainsi une forme n'a pas de fond tant qu'il n'y a pas interprétation et un fond doit nécessairement s'actualiser sur une ou plusieurs formes. L'instrumentation logicielle que nous visons doit donc médiatiser cette alternance en considérant notre échelle circulaire comme un cercle vertueux où se déploient les richesses d'interprétation du lecteur.

### Les niveaux de contexte d'interprétation

Le sens n'est pas de nature symbolique, mais dynamique. Un texte, un signe n'a pas de *sens* en tant que mais, fait sens pour un interprétant. On retrouve ici la triade sémiotique de Pierce. Pour Rastier la dynamique du sens s'inscrit dans une théorie de l'action du sujet dans et par le langage, ce qu'il nomme une praxéologie. Cette théorie de l'action place l'interprétation dans un espace fait de trois zones autour du sujet :

- La zone identitaire : la dimension personnelle et subjective de l'interprétation consciente ou inconsciente. C'est là que s'exprime l'individualité, les points de vue particuliers. C'est la zone de la première personne verbale (*je, nous*) dans l'ici et le maintenant.
- La zone proximale : elle représente l'environnement linguistique immédiatement proche du sujet. C'est là que se trouve la communication avec l'autre. Dans les approches interactionnistes en psychologie (Brassac & al. 1996), le contexte de cette zone se ramène à la notion de terrain commun mis en évidence (Clark & al. 86, Vivier 1992). Le terrain commun permet de borner le contexte, évitant ainsi le recours à une description ontologique du monde dans son ensemble. C'est la zone linguistiquement marquée par la deuxième personne verbale (tu, vous), par le passé proche et futur immédiatement probable.
- La zone distale : elle représente les normes et ancrages culturels du sujet. Les deux premières zones sont celles des expériences propres du sujet. La zone distale est celle de ce qui est étranger mais connu. C'est linguistiquement la zone de la troisième personne verbale, du récit, du non-présent et de l'irréel.
- 

Ces trois zones ou, niveaux praxéologiques des sujets, éclairent les apports de la dimension subjective aux phénomènes interprétatifs. Considérer que l'interprétation ne fait intervenir que des connaissances ontologiques (comme c'est le cas pour le WEB sémantique) c'est ne voir que la troisième zone et passer sous silence ce qui relève des réalités empiriques du sujet, à savoir les deux premières zones. Il est donc fondamental dans un environnement d'assistance à l'interprétation que les trois zones puissent être utilement mises en œuvre. Cela a pour conséquence, que les ressources qui guident les processus calculatoires qui interviennent dans l'interaction (extractions statistiques, visualisations, etc.), doivent être très massivement personnalisables pour donner toute l'importance qu'il convient aux deux premières zones. Au sein de la tâche du sujet, c'est par un travail d'aller retour entre des corpus et des ressources

termino-ontologiques personnelles que l'environnement logiciel peut amener à caractériser des formes sémantiques sous l'angle des moments stabilisés de parcours interprétatifs.

#### 4. Le logiciel Canopée - fondements théoriques

Les utilisateurs de messageries électroniques n'ont jamais été aussi nombreux et il est habituel qu'ils utilisent conjointement plusieurs comptes de messagerie électronique. Du point de vue des usages, les applications de clients de messagerie (Outlook, Thunderbird, Mail, etc.) ainsi que les services WebMails (Gmail, Hotmail, etc.) sont de moins en moins vu comme des applications et services parmi tant d'autres. De même que les périphériques de connexions réseau, ils tendent à devenir des éléments de base incontournables des entreprises au même titre, par exemple, que les gestionnaires de fichiers. Dans bon nombre de cas, on constate même que la différence entre architecture de dossiers/fichiers et boîtes de courriers électroniques est une distinction de plus en plus artificielle relativement aux usages et tâches des utilisateurs.

Les interactions offertes par les clients de messagerie et les architectures WebMail sont relativement pauvres du point de vue des fonctionnalités offertes aux usagers. En dehors de l'architecture hiérarchique des boîtes et des fonctions de recherche, l'utilisateur ne dispose pas vraiment d'outils pour manipuler et appréhender ses courriers électroniques globalement alors que, paradoxalement, la messagerie est à la fois le contexte de travail et l'environnement d'archivage du travail en cours. Or, ce constat est certainement celui qui a conceptuellement le plus de portée et il nous semble que ce détournement d'usage soit durablement installé.

La messagerie électronique est un mode de communication asynchrone dont l'usage premier était de délivrer en une fois une quantité d'information limitée. Dans un deuxième temps, il lui a été donné la possibilité d'embarquer une pièce jointe. Depuis, on ne compte plus les emails sans texte et qui véhiculent uniquement des pièces jointes. Cette pratique s'étant accélérée avec les contraintes actuelles de sécurité et de confidentialité. Le contournement d'usage, total ou partiel, n'est pas le problème. Au contraire, une technologie s'impose et perdure auprès d'un large public si justement, il lui est trouvé de multiples usages détournés. Ainsi, les boîtes emails, personnelles ou professionnelles sont-elles devenues le réceptacle d'une partie importante et critique de la *vie numérique*<sup>4</sup> des usagers. Tout y est conservé pèle mèle avec de nombreuses tentatives de classement dont l'efficacité est inversement proportionnelle à l'âge du message. C'est pour ces raisons que l'on cherche des moyens supplémentaires pour démarquer un email important au milieu de la masse des messages, sans compter les spams.

Ceci est éminemment paradoxal. En effet, la théorie de Shannon, confirmée dans sa dimension thermodynamique par les travaux de Brillouin, nous enseigne qu'il existe fondamentalement deux grands types d'information : l'information libre et l'information liée (que Brillouin nomme *néguentropie*). Dans le cadre d'une communication, principalement asynchrone, les utilisateurs sont demandeurs et émetteurs d'information liée et en général d'une petite (voire très petite) quantité d'information libre. Plus la communication (asynchrone) est ancienne, plus la quantité d'information libre est faible à chaque nouveau message. L'intérêt d'un message réside justement dans la quantité d'information liée qu'il va délivrer<sup>5</sup> à un utilisateur donné. Souvenons-nous que la néguentropie est hautement relative au récepteur/destinataire du message. L'observation de

---

<sup>4</sup> Terme très médiatique mais à combattre sans relâche tant il est vrai qu'il ne signifie presque rien, au mieux une pratique quotidienne et non éclairée des moyens (presque) informatiques. Il n'y a pas de vie numérique, il y a la vie tout simplement.

<sup>5</sup> Très précisément, un message ne *délivre* aucune information, libre ou liée. Il faut à ce sujet revenir à des considérations sémiotiques basiques. Il faut qu'il y ait interaction entre ce message et un récepteur/destinataire donné. L'information qui sera alors produite est hautement liée à la nature même de ce destinataire. Un message est d'abord un événement (au sens de Shannon) qui embarque un message symbolique (au sens de la sémiotique). Tout ceci implique la nécessité d'interactions.



l'événement [*un email vient d'être délivré*] apporte à son destinataire une quantité restreinte d'information libre  $H$  et une quantité virtuellement plus grande d'information liée (néguentropie)  $N$ .

S'il peut arriver que  $H \approx N$ , l'objectif premier des emails est de provoquer  $N \gg H$ . C'est très généralement le cas pour la grande majorité des emails et une bonne façon de définir, à contrario, les spams. De la même façon, sur une durée arbitrairement longue, considérons deux distributions statistiques : la première  $\lambda$  est faite de toutes les quantités (quantum) d'information libre  $H_i$  associées à tous les emails reçus sur la période considérée. La deuxième  $\eta$  est fabriquée de la même façon pour les quantités d'information liée  $N_i$  sur la même période. Nous savons que :  $\bar{\lambda} \ll \bar{\eta}$  et de la même façon :  $\sigma_\lambda \ll \sigma_\eta$ .

Or, l'accumulation des boîtes emails, la multiplication des messages et des thématiques abordées produisent une accélération dans la dynamique naturelle des systèmes de communication et qui les pousse vers leurs états d'équilibre thermodynamique caractérisés par le plus haut niveau d'entropie et où le deuxième principe règne en maître absolu. Lorsque le système est proche de cet état, une mesure sur le système apportera une faible quantité d'information liée. Il s'agit d'états où il y a convergence entre  $H$  et  $N$ . Par conséquent les moyennes et écarts types ne sont plus séparées par plusieurs ordres de grandeur. Au contraire, elles sont très proches en valeur.

Nous observons le même phénomène dans l'usage moderne de la messagerie électronique. Soulignons au passage que ceci implique que la combinaison des usages décrits auparavant avec un groupe de messageries se comporte comme un système thermodynamique. Dans de tels systèmes, le démon de Maxwell a été exorcisé par Brillouin après de nombreuses années où les physiciens restèrent plongés dans la plus grande perplexité. On sait depuis qu'obtenir une quantité  $\delta h$  d'information libre dans un tel système proche de son équilibre et à fortiori, une quantité d'information  $\delta n$  de néguentropie (quand cela a un sens) requière de dégrader une quantité d'énergie importante ce qui produira une quantité d'entropie  $\delta s$  tel que :  $\delta s \gg \delta h \gg \delta n$ .

L'exorcisme du démon de Maxwell est également valable dans nos systèmes de messagerie électronique ou, plus exactement, dans les systèmes qui naissent de l'usage moderne des messageries électroniques asynchrones. En effet, l'accumulation des messages de toutes natures et thématiques *éloigne* le destinataire de l'information liée pour le limiter à l'information libre. Au mieux, l'utilisateur va-t-il puiser une quantité de néguentropie  $\delta n$  dans le nom de l'expéditeur, éventuellement dans l'objet du message, la date et la taille du mail. Cependant, cette quantité restera, dans tous les cas, très limitée relativement à l'information potentiellement apportée par la lecture du contenu de l'email et qui repose sur la simple question : *de quoi me parle le message ?*

Ainsi, progressivement, l'utilisateur troque-t-il de la légitime néguentropie contre, d'une part, une quantité limitée d'information libre et surtout, contre du temps (plus exactement de l'énergie du genre temps (Cozien 2002)), qu'il va dégrader et par conséquent, accroître l'entropie de son environnement. Ce processus et cette logique nous sont strictement imposés par la thermodynamique, celle de Boltzmann, mais surtout celle de Shannon, Brillouin et Szilard.

Ces considérations prennent toute leur portée lorsque l'on quitte l'instantanéité de la réception des emails pour revenir sur sa base de messages et, par exemple, en rechercher un en particulier. C'est d'autant plus vrai lorsque l'on recherche, plus que le mail lui-même, son contenu, gisement de néguentropie, en utilisant des critères de faible valeur néguentropique (date, taille, objet, etc.). Par exemple : *je recherche l'ensemble des emails parlant de Rochebrune2011*. Si l'on se pose cette question c'est que justement, au moment de leurs réceptions, l'utilisateur n'aura pas fait l'effort de dégrader

de l'énergie du genre temps (en entropie) pour aller chercher la néguentropie qu'ils contenaient. Mais surtout, s'agissant de parcelles conjuguées de néguentropie, sur un même sujet, il n'aura pas fait cette opération extraordinairement couteuse en thermodynamique et qui consisterait à réaliser un document de synthèse où il agrègerait ces parcelles successives de néguentropie. Ce processus nommé *enregistrement* est celui qui maximise N mais qui dans le même temps, provoque une forte augmentation de l'entropie de l'environnement (Cozien 2002).

Le paradoxe est donc là : pour, après coup, reconstruire N, l'opérateur se sert de fragments  $\delta n$ . Cela va potentiellement lui coûter beaucoup d'énergie sans certitude du résultat : par exemple, si un ou plusieurs emails sont effacés, perdus ou mal classés. Brillouin nous dit que pour maximiser N, il faut mettre en place ce qu'il appelle un *laboratoire* constitué d'un observateur (ici l'utilisateur de la messagerie) d'un objet d'étude (le corpus des emails) et d'un instrument de mesure dont le but est de réaliser, certes une mesure mais formellement, un transfert croisé de néguentropie du sujet observé vers l'observateur puis, d'entropie, de lui-même (l'instrument) vers le milieu externe.

Dans cette acception thermodynamique, Canopée est notre instrument de mesure. Grâce à lui, l'utilisateur/observateur va troquer son ignorance (pré-entropie) qu'il a sur l'état de sa base d'emails contre de l'entropie que l'instrument de mesure ne manquera pas de déverser dans l'environnement du laboratoire.

En d'autres termes, Canopée vous permet de ne plus être tributaire du démon de Maxwell qui vit dans votre base d'emails. Par son approche lexico-sémantique, il autorise, au fil de la vie de la base d'emails de maintenir disponible, au profit de l'utilisateur, la connaissance (quasi-)instantanée sur l'état du système. Certes cette entreprise requière une quantité non négligeable d'énergie au fil du temps. Mais, on démontre, que cette quantité d'énergie est inférieure de plusieurs ordres de grandeurs à la quantité qu'il faudrait pour instantanément obtenir la même quantité de néguentropie sans ce travail constant de mémorisation.

A l'usage, le logiciel est architecturé autour de trois principaux composants : des ressources lexico-ontologiques, un corpus de documents (dans notre cas des emails) et des moyens de visualisation. Ces moyens sont de différentes natures. La plus spectaculaire repose dans la possibilité d'explorer le corpus via une carte en trois dimensions où sont représentés les thèmes par ordre d'importance et de présence. Nous réaliserons une démonstration complète de ses possibilités lors de notre présentation à Rochebrune.

## 5. Conclusions et perspectives

Le sens ne connaît pas d'état final « stable ». On peut toujours donner lieu à des nouvelles variations interprétatives. Interpréter un texte, qui plus est au sein d'une collection documentaire, est une activité dont l'aboutissement n'est pas une condition nécessaire. La façon dont un cherche à accéder au contenu, c'est le contenu. Le contenu n'est pas prédictible, car les interprétations ne sont observables qu'a posteriori. De plus le sens d'un passage de document est largement moins le fait de son auteur que celui de son lecteur. C'est donc bien l'activité de lecture et de confrontation à une collection documentaire qu'il faut chercher à instrumenter.

Le projet Canopée cherche à fournir à son utilisateur une médiatisation personnalisée de son espace d'interprétation. En cela, il se rapprocherait plus dans le principe du champ de la réalité augmentée que du traitement automatique du langage. L'objectif de Canopée n'est pas de calculer du sens à la place de l'utilisateur mais de lui renvoyer des informations qui au sein de ses parcours interprétatifs prendront du sens et apporteront des fonctionnalités d'accès au contenu textuel.

Les échelles de textualité et les niveaux de praxéologie sont ici au centre de la problématique du couplage entre le logiciel et le (ou les) utilisateurs. Les conditions de ce couplage sont un domaine de recherche à part entière à l'interface de l'informatique et des sciences cognitives.

Du point de vue de la thermodynamique, le couplage entre un utilisateur et sa base de messages électroniques via Canopée n'est rien d'autre qu'un *laboratoire* au sens de Brillouin. Par conséquent des flux thermodynamiques croisés, d'entropie et de néguentropie se mettent en place afin que l'observateur (ici l'utilisateur de la base d'emails) voit son ignorance décroître sachant que dans le même temps, l'entropie générale du laboratoire augmentera considérablement. Ainsi, l'objectif de Canopée, en tant qu'instrument de mesure, est de prendre à son compte l'ignorance de l'observateur pour la transformer en connaissance sur l'état du système (dynamique) des emails. Canopée réalise des couplages agissants sur différentes échelles, du message seul à la connaissance de l'utilisateur (échelle cognitive).

Le logiciel entrera en production puis sera rapidement commercialisé durant l'année 2011. Il se déclinera en versions appliquées aux grands corpus de documents, aux messageries électroniques et aux bases photographiques.

## 6. Références

- (Berners-Lee 2001) Berners-Lee T, 2001, « The Semantic Web », *Scientific American Magazine*, May 17.
- (Brassac & al. 1996) Brassac, C., Stewart, J., 1996. Le sens dans les processus interlocutoires, un observé ou un co-construit ? Cinquièmes Journées de Rochebrune "Du social au collectif ", 29 janvier - 3 février 1996.
- (Clark & al. 1986) Clark H. H., Wilkes-Gibbs D., 1986, Referring as a Collaborative Process, *Cognition*, n° 22, p. 1-39.
- (Courcil 2000) Coursil J., 2000, La fonction muette du langage, Ibis Rouge Editions, Presses Universitaires Créoles.
- (Dionisi et Labiche 2006) Dionisi D., Labiche J., 2006, Enaction et informatique : les enjeux de l'opérationnalisation technologique d'une théorie de la cognition, in Actes du colloque ARCo 2006, 6 au 8 Décembre 2006 – Bordeaux.
- (Groupe v 2008) Groupe v, (2008). Conception et usages d'un environnement numérique de travail pour une aide à l'interprétation de documents juridiques. Actes du 11<sup>e</sup> Colloque International sur le Document Electronique (CIDE), 28-31 octobre 2008, Rouen.
- (Mayaffre 2002) Mayaffre D., 2002, Les corpus réflexifs : entre architecturalité et hypertextualité, *Corpus*, n°1 novembre 2002, revue en ligne <http://corpus.revues.org>.
- (Michaux 1966) Michaux H., 1966, L'espace du dedans, Paris, Gallimard.
- (Peirce 1978) Peirce C. S., 1978, Écrits sur le signe, rassemblés, traduits et commentés par Gérard Deledalle, Paris, Seuil.
- (Rastier & al. 1994) Rastier F., Cavazza M., Abeillé A., 1994, Sémantique pour l'analyse, Paris, Masson.
- (Rastier 1987) Rastier F., 1987, Sémantique interprétative, Paris, Presses Universitaires de France.
- (Rastier 2001a) Rastier, F., 2001, Arts et sciences du texte. Presses Universitaires de France, Paris.
- (Rastier 2001b) Rastier F., 2001, L'action et le sens pour une sémiotique des cultures, *Journal des anthropologues*, n° 85-86, mai 2001, pp. 183-219.
- (Roy 2007) Roy T., (2007). Visualisations interactives pour l'aide personnalisée à l'interprétation d'ensembles documentaires, Thèse de doctorat en Informatique – Université de Caen Basse Normandie (<http://roythibault.free.fr/these/index.html>)

(Thilivitis 1998) Thilivitis T., 1998, Sémantique Interprétative Intertextuelle : assistance informatique anthropocentrée à la compréhension des textes, Thèse de l'Université de Rennes 1,  
(Vivier 1992) Vivier J., (1992). Faire et dire ce qu'il faut faire pour ... Physique naturelle et discours : Projet d'étude développementale. Diplôme d'habilitation à diriger des recherches, Université de Caen.  
(Cozien 2002) Cozien R., (2002) Premiers éléments de la théorie du calcul singulier. Thèse de doctorat de l'université de Reims.