

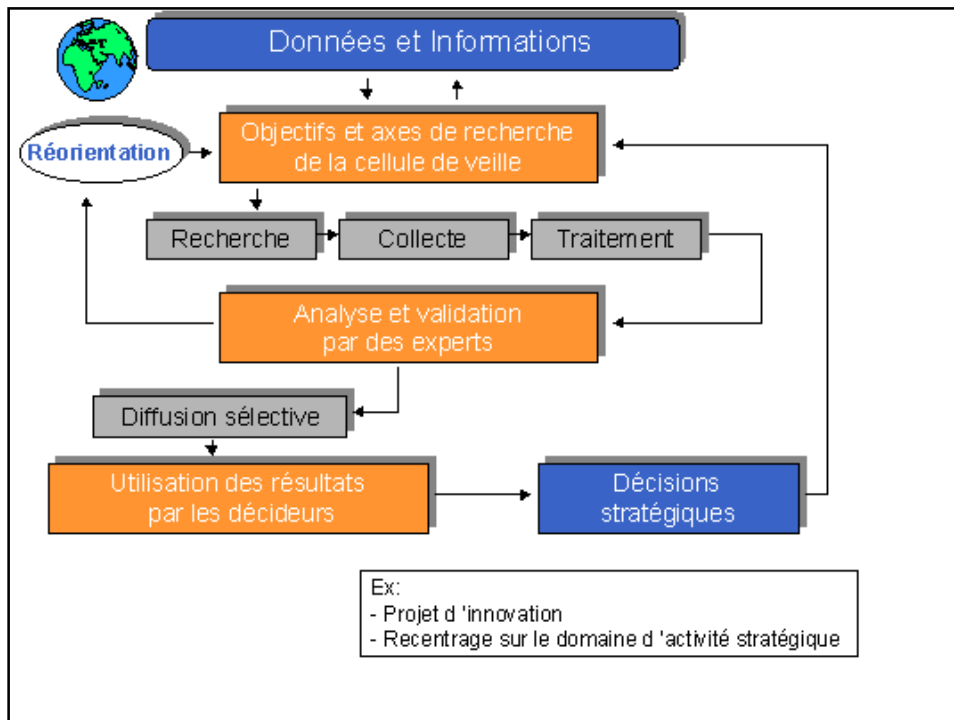
IHEC, Tunis, 17 avril 2007

Veille et Intelligence stratégique

Jean-Pierre Desclés
LaLIC
Université de Paris-Sorbonne

Définition et enjeux de la veille technologique

- **La veille stratégique est l'activité qui comprend la collecte, l'analyse, la mise en forme et la diffusion de l'information et cela dans le but de :**
 - **Détecter des signes de changements.**
 - **Fournir les moyens de décider.**
- **C'est un système d'aide à la décision qui observe et analyse l'environnement de l'organisme pour en déduire les menaces et les opportunités de développement.**



Surveillance, Veille, Intelligence

- 1. Scanning (balayage) (F.J. Aguilar, 1967)
- 2. Problème de la détection des signaux faibles : I. Ansoff (1975)
- 3. Veille Technologique (1970)
- 4. Emergence des systèmes d'intelligence (1950, 1958, 1967 ...)

Terminologie

- Veille et Intelligence stratégique
- Surveillance, « Scanning » (balayage),
- Vigilance, « Competitive Intelligence »,
- Veille stratégique, économique
- Intelligence économique
- Intelligence économique et stratégique
- « Intelligence »

« Intelligence »

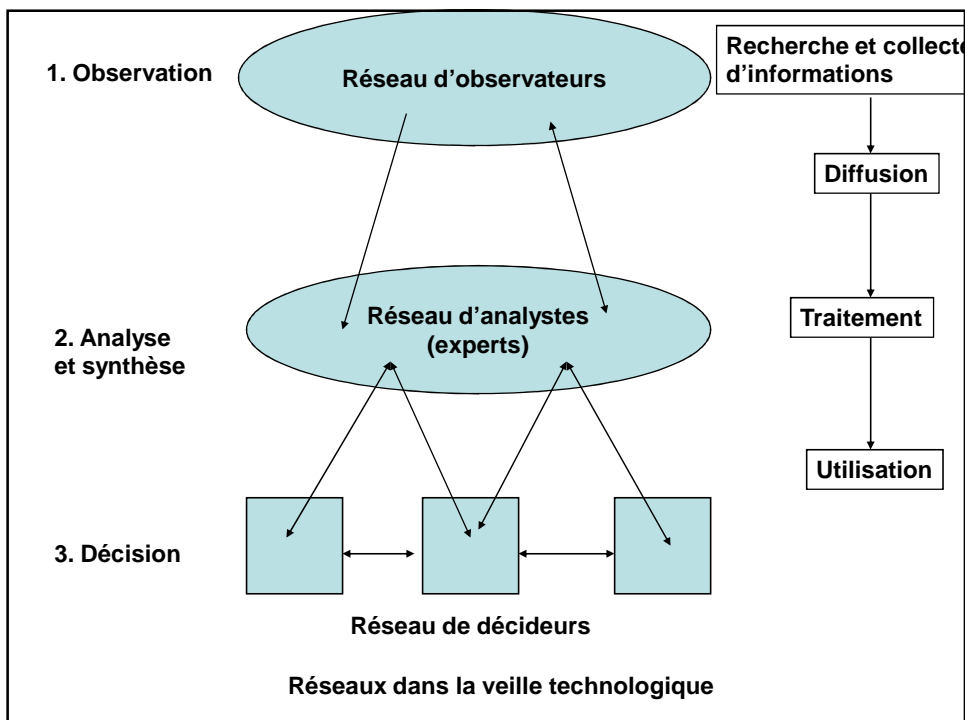
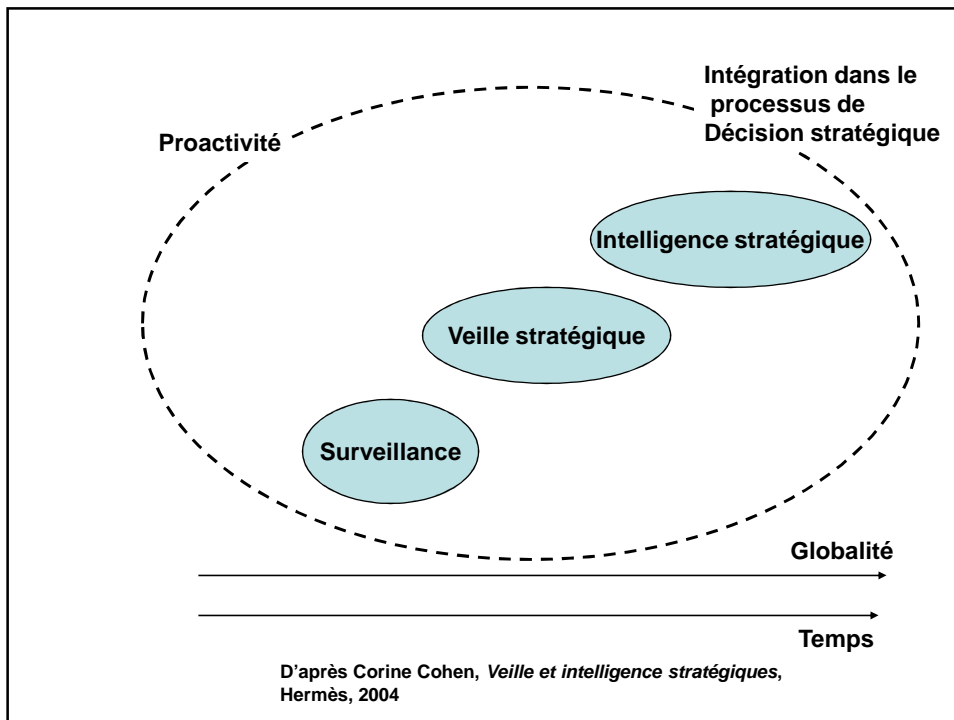
- Intelligence = capacité à appréhender les interrelations entre les faits disponibles de manière à guider l'action vers un but désiré.
- Relier des événements pour leur donner du sens
- Luhn (1958) (chez IBM) : *Business Intelligence System*
- *Competitive Intelligence* (Michel Porter, 1980)
- => Surveillance, Veille, Intelligence

Intelligence stratégique

- Intelligence stratégique (IS) est définie comme un processus formalisé de recherche, collecte et traitement d'informations, de diffusions des connaissances utiles au management stratégique.
- Elle a pour mission d'anticiper les menaces et les opportunités de l'environnement (*fonction anticipative*), de proposer et de mener des actions (*fonction proactive*), dans le but d'aider la prise de décision stratégique et d'améliorer la compétitivité et la performance de l'organisation.
- Elle nécessite une structure organisationnelle en réseaux, des moyens humains, techniques et financiers.

Veille

- La veille doit aller jusqu'à signaler les impacts de tel ou tel événement détecté.
- Elle devient intelligente dès lors qu'elle produit des recommandations, elle préconise, elle fait des prescriptions à l'utilisateur destinataire, a fortiori lorsqu'elle les met en œuvre.



Surveillance, Veille



1. Détermination des besoins en information
2. Recherche et collecte des informations
3. Traitement de l'information (analyse, synthèse, mise en forme, visualisation)
4. Stockage de l'information
5. Diffusion de l'information
6. Utilisation de l'information

Veille sur internet

- Découvrir de nouveaux acteurs potentiels sur le marché.
 - Mieux connaître son environnement technologique.
- => Grâce à cette démarche, l'organisme pourra désormais :
- Mieux se positionner dans son environnement.
 - Disposer de scénarios de repositionnement en fonction des tendances repérées.

Informations ?

Trois types d'informations

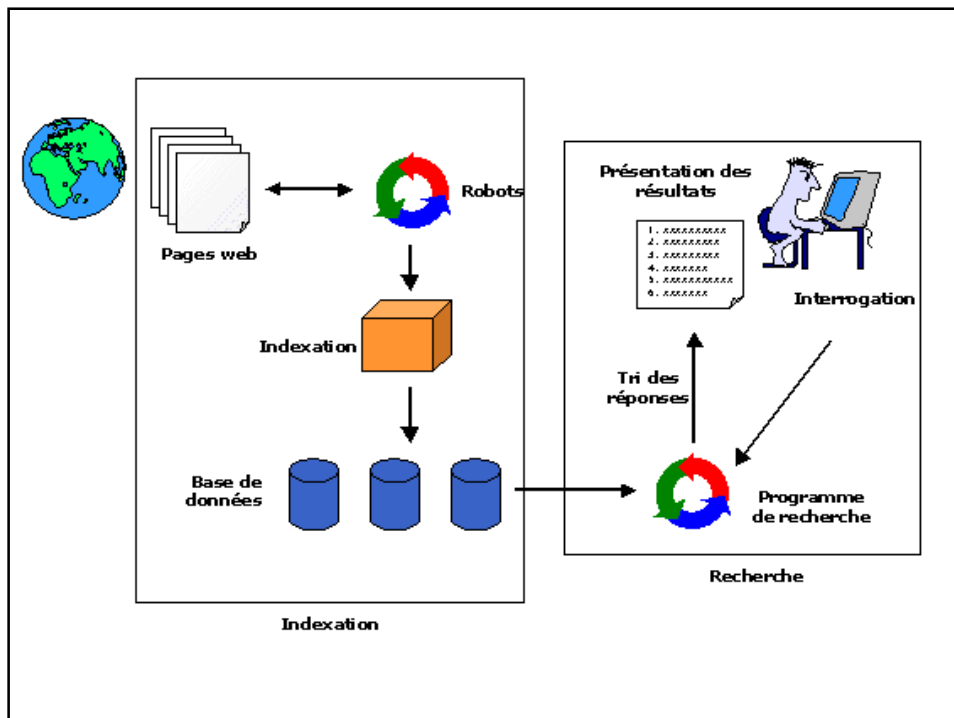
- **L'information blanche** : Publique et accessible, ne fait l'objet d'aucune sécurisation particulière
=> Recherche "classique" avec les outils grand public.
- **L'information grise** : Ne fait pas l'objet de publicité, mais on peut la trouver de manière indirecte ou détournée ; Information sensible d'accès légal
=> Techniques avancées de recherche et de traitement de l'information, groupe de discussion, liste de diffusion
- **L'information noire** : Fait l'objet d'une haute sécurisation
=> Relève de l'espionnage industriel-

Désinformation

- En fait, pour le néophyte, sur Internet, il est quasiment impossible de bien distinguer le bon du mauvais.
- Outre la désinformation volontaire pratiquée par certains sites envers les technologies concurrentes, il est fréquent de trouver des sites Web dont l'information n'est pas actualisée, voire des liens hypertextes non valides.
- Certaines informations sont lancées comme des leurres (tromper les concurrents).

- Informations sur une société : sur sa structure, sur ces produits, etc.
- Thèmes de recherche d'un laboratoire universitaire.
- Précisions sur les normes et le cadre législatif d'un pays.
- Renseignements sur une personne ou un groupe de personne.
- Données statistiques.
- Communiqués de presse.
- Rapports annuels.
- Informations promotionnelles.
- Photographies de produits et équipements.
- Dictionnaires ou lexiques, encyclopédies.
- Catalogues d'éditeurs ou de librairies.
- Des informations financières sur un pays, une société, etc.
- Des articles de presses. ...

Recherche d'informations par indexation et moteurs de recherche



Qualités des informations

- 1. Sources et crédibilité : Identification des auteurs, comité éditorial, cible du site, qualité de la langue, publicité, partenariat
- 2. Contenu : Citations des sources originales, structuration des informations, distinction nette entre le contenu et la publicité, exhaustivité, mise à jour
- 3. Liens hypertextes: Pertinence des liens, sélection, validité
- 4. Design et navigation : Lisibilité du texte, facilité de navigation, nécessité de logiciels spécifiques pour visualiser le contenu, rapidité de chargement des illustrations
- 5. Accessibilité : Présence dans les principaux répertoires et moteurs, adresse intuitive
- 6. Confidentialité et interactivité : Protection des données personnelles, rétroactions

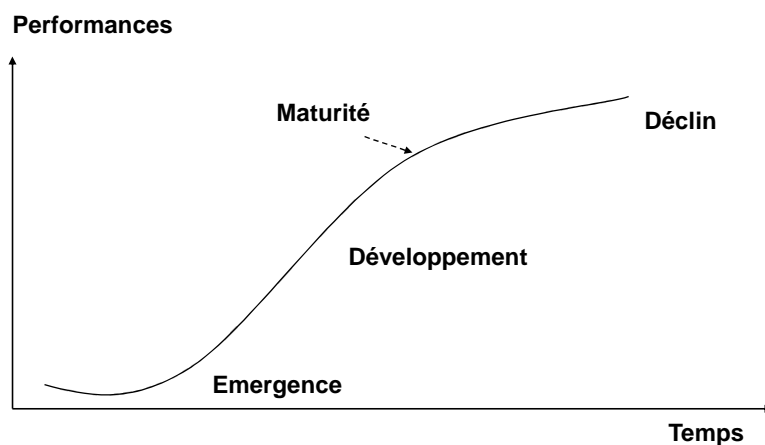
Sources d'informations

- Presse
- Ouvrages
- Médias
- CD ROM
- Brevets
- Etudes privées, publiques
- Sources légales
- Internet
- Rapports d'analystes financiers
- Rapports annuels
- Intranet

Valeur des informations collectées

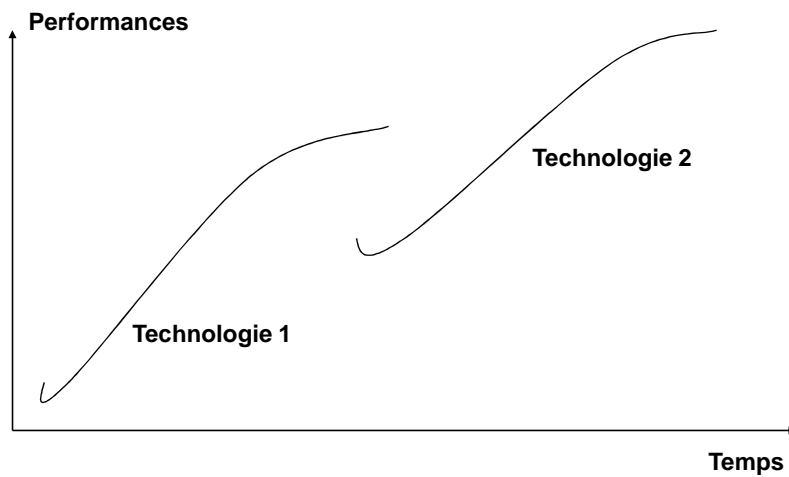
<u>Valeur de la source</u>	Suspecte	peu sûre	digne de foi risque d'erreur subjective	digne de foi
<u>Valeur de l'information</u>				
Inutile	-----			
Utile à l'occasion				
Intéressante				
Prioritaire et importante				++++

Analyse de l'information



Courbe en S (Foster, 1986)

Rupture technologique



**Traitement pour l'analyse et
la synthèse**

Outils de traitement

1. Méthodes fondées sur les statistiques :
 - Visualisation graphique des informations
 - Techniques d'indexation
 - Moteurs de recherche avec requêtes
2. Méthodes « classiques » fondées sur la linguistique
 - Morphologie + syntaxe + sémantique => grosses ressources => coût
3. Méthodes d'analyse sémantique (sémantique des mots, réseaux sémantiques, synonymie)
4. Méthodes appuyées par des ontologies des domaines
5. Méthodes fondées sur des relations discursives, par exploration contextuelle avec des moteurs de recherche;

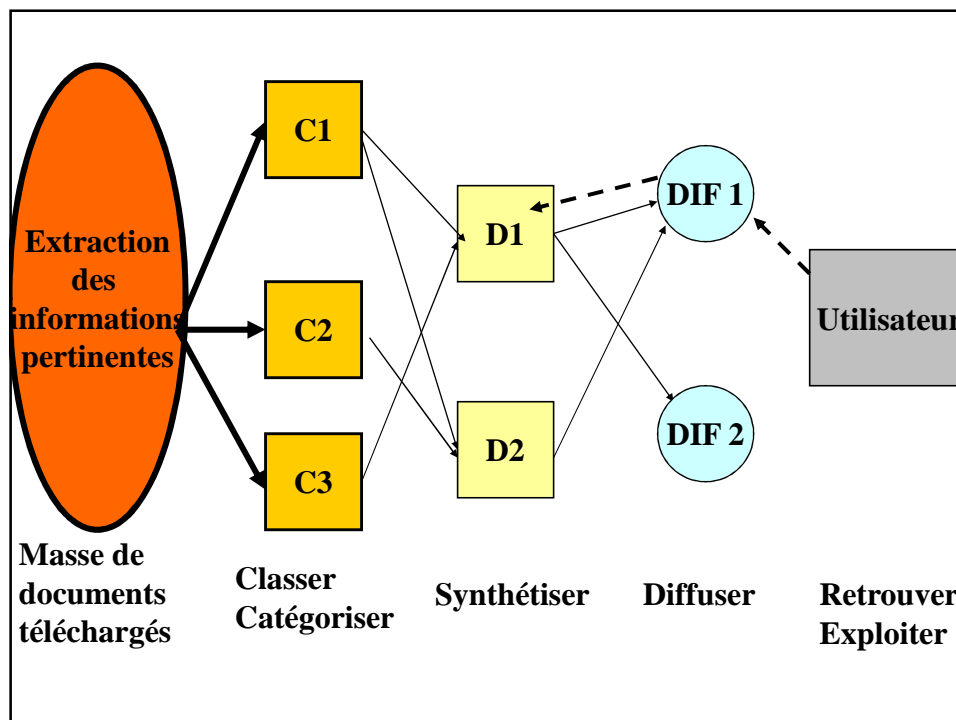
Evaluation des logiciels

Logiciels d'analyse textuelle (moteurs de recherche et outils de visualisation).

- La **formulation de requêtes** dans une démarche de veille ne semble pas pertinente puisque le veilleur ne connaît pas a priori l'information utile (pour lui) qui est noyée dans un ensemble de textes.
- Les **approches statistiques** qui produisent des cartes sémantiques ne conviennent pas car l'information stratégique n'est pas toujours une information fréquente.

Extraire des informations

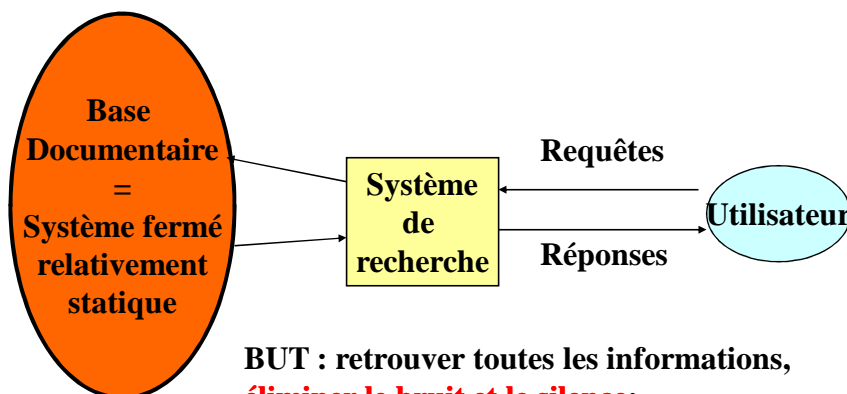
- les chercher, les trouver et les extraire de la masse des documents produits et accessibles (téléchargés)
- les classer et les catégoriser,
- les synthétiser,
- les diffuser,
- les retrouver pour leur exploitation....



Méthodes « classiques » vs « nouvelles »

- Les **méthodes « classiques »** de recherche d'informations **dans des bases documentaires** restent assez mal adaptées à une recherche dans un réseau « ouvert » comme Internet.
- Un **réseau d'informations** (Internet ou des intranets) n'est pas un système documentaire statique et fermé (systèmes documentaires, bases de données), il est **dynamique, ouvert**; il contient en général un très grand nombre de documents dans des formats différents.
- La recherche doit être plus **interactive et orientée** vers les **besoins multiples des utilisateurs**.

Méthodes « classiques » dans des systèmes documentaires



BUT : retrouver toutes les informations,
éliminer le bruit et le silence;
critères de validation : précision / rappel
Méthodes : indexation, mots clés ...

La fouille et la synthèse doivent être orientées vers des destinataires

On ne résume pas de la même façon un rapport technique portant sur l'identification d'une nouvelle molécule, selon que l'on destine ce résumé :

- à la direction générale pour justifier des crédits dépensés
- à service financier sollicité pour des développement ;
- à un laboratoire destiné à exploiter cette molécule ;
- aux juristes chargés de protéger la découverte par un brevet ;
- à des journalistes susceptibles de diffuser l'information ;
- aux étudiants de première année
que l'on souhaite orienter vers un nouveau secteur...

Synthèse des informations pertinentes

- Disposer de presque tous les documents sur un sujet donné serait de bien peu d'utilité dans la mesure où l'on serait, parallèlement, incapable d'en synthétiser rapidement les contenus.
- Diffuser les mêmes informations dans toutes les directions, sans sélection, amène à une véritable pollution informative qui risque de laisser finalement les destinataires relativement peu informés.
- **« trop d'informations = information nulle ».**
- **« un trop grand bruit tue l'information utile... ».**
- Pour être bien informé, il faut donc être capable de sélectionner, en temps utile, les « bonnes informations ».

« Pertinence » d'une information ?

Une information est « pertinente » lorsque

- ou elle vient confirmer d'autres informations déjà connues ;
- ou elle est « nouvelle » dans un domaine déjà exploré ...;
- et elle est diffusée à un destinataire adéquat ...

THESE (linguistique et communication) :

« **Une information pertinente dans un document textuel est indiquée par des marqueurs linguistiques identifiables par un destinataire** »

La « fouille sémantique » de textes

La prise en compte, dès la modélisation, des utilisateurs a conduit progressivement à une conception élargie du résumé synthétique.

Il s'agit maintenant de proposer des systèmes informatiques

- **capables de fouiller les textes, avec des critères sémantiques,**
- avec des points de vue différents,
- les contenus des textes.

=> Constituer automatiquement des fiches de synthèse

Participation interactive de l'utilisateur

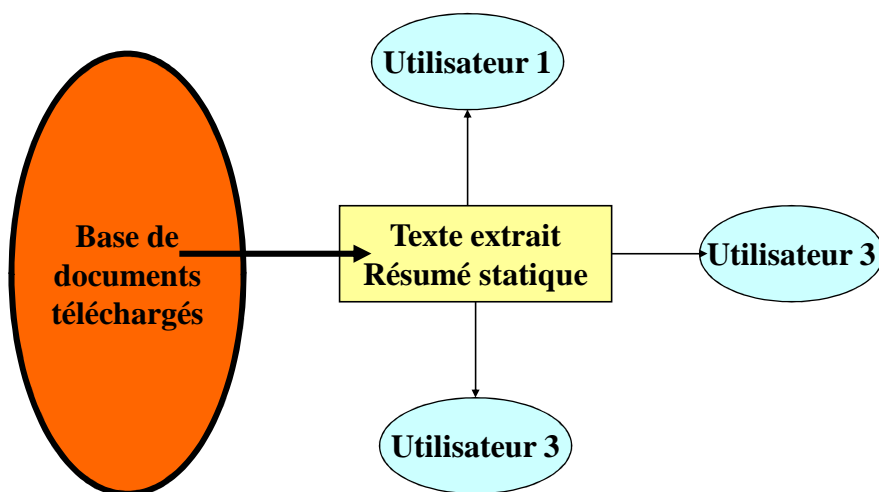
Si le résumé statique sur support papier est identique pour tous les utilisateurs,

la fouille sémantique de textes réclame une participation interactive de l'utilisateur,

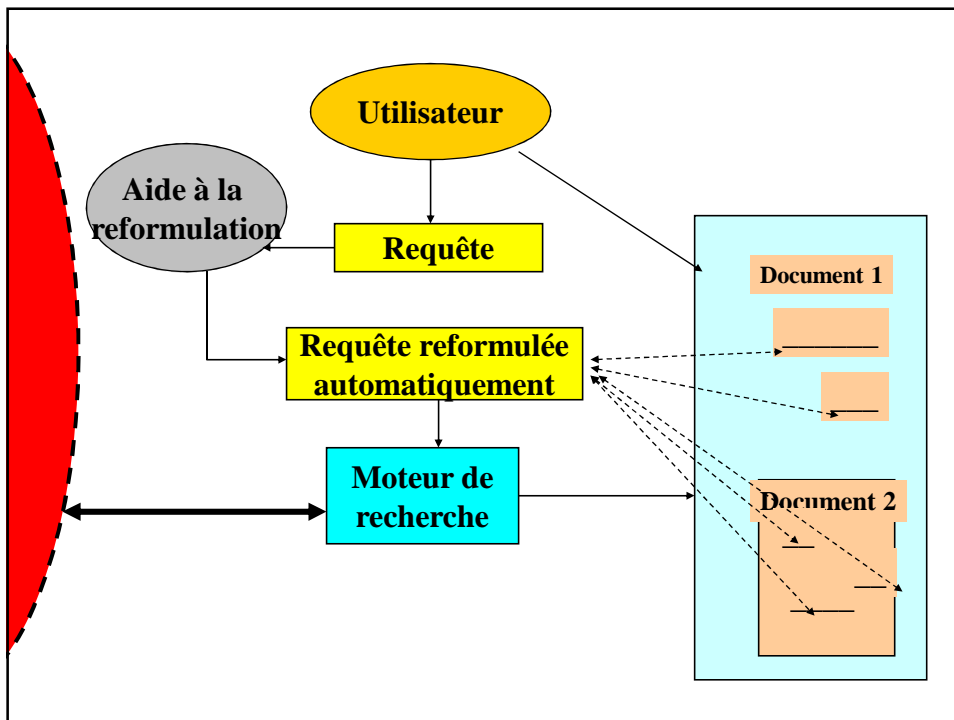
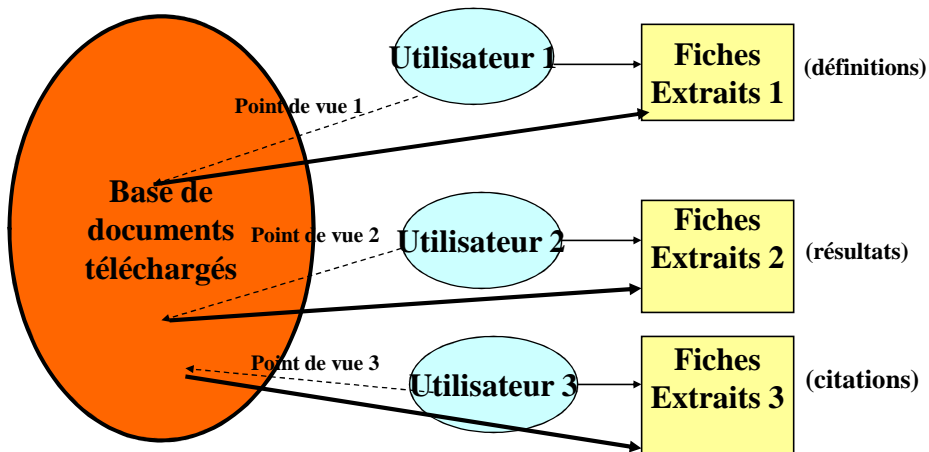
en lui donnant un accès à une plate-forme informatique susceptible de l'aider à la construction de sa propre synthèse;

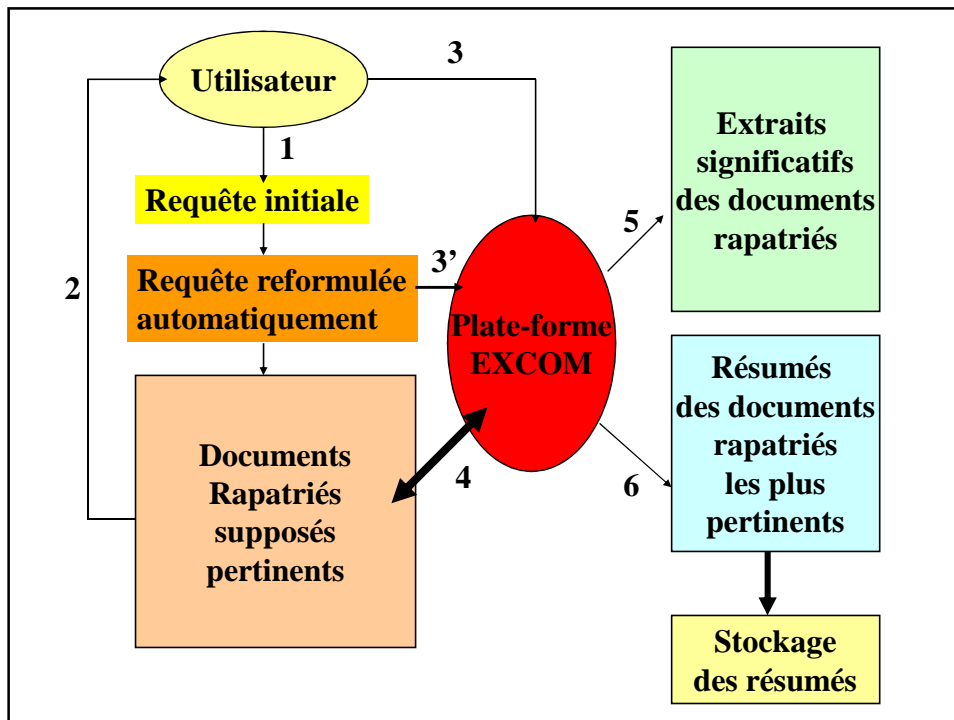
l'utilisateur fera alors appel à un ou plusieurs « agents » qui réaliseront différentes tâches d'extraction guidées par les objectifs.

Résumé statique = identique pour tous les utilisateurs



Synthèses dynamiques : chaque utilisateur construit « son » résumé en fonction des « ses intentions »





Exemples de recherches d'informations

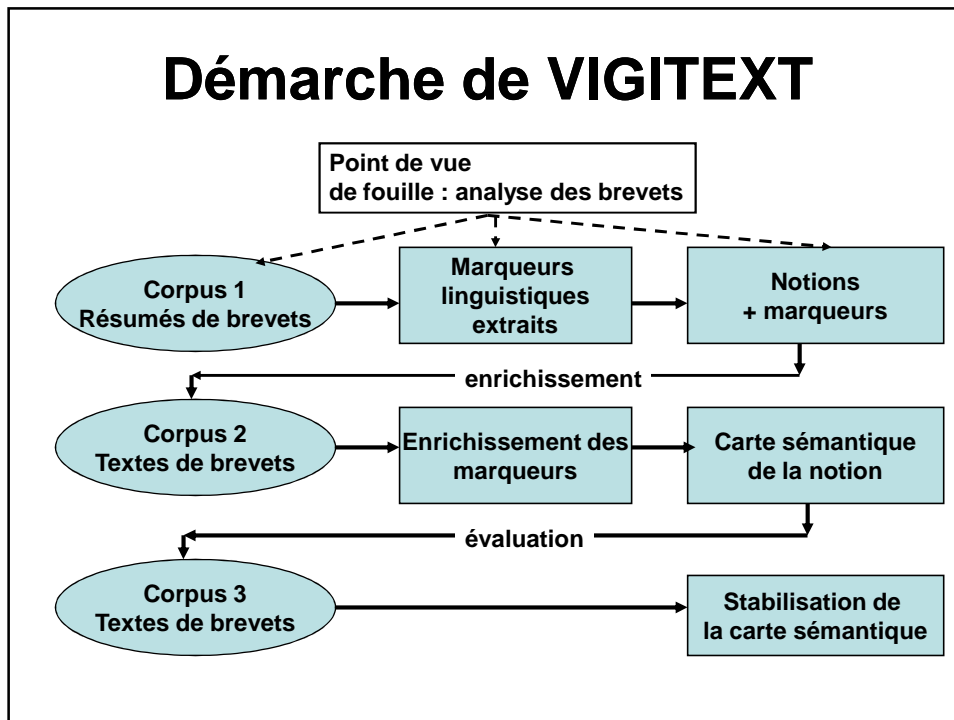
- Un **documentaliste** sera préoccupé par une extraction automatique des déclarations, plus ou moins récentes, de telle personnalité sur laquelle il est chargé de préparer un dossier de presse ;
- Un **patron de laboratoire**, pressé par le temps, cherchera à connaître rapidement les méthodes expérimentales employées dans l'identification d'une molécule en exploitant une base d'articles publiés dans des revues spécialisées ;
- Un **chercheur en sciences sociales** désirera relever comment différents auteurs ont appréhendé un même terme ;
- Un **économiste** voudra rechercher les commentaires textuels relatifs à un diagramme qui présente l'évolution comparée du chômage dans divers pays ;
- Un **étudiant** en sciences politiques cherchera à rassembler très rapidement certaines informations pertinentes pour développer et appuyer son exposé...

Premier cours : 24 octobre 2007

Deuxième cours

VIGITEXT

Démarche de VIGITEXT



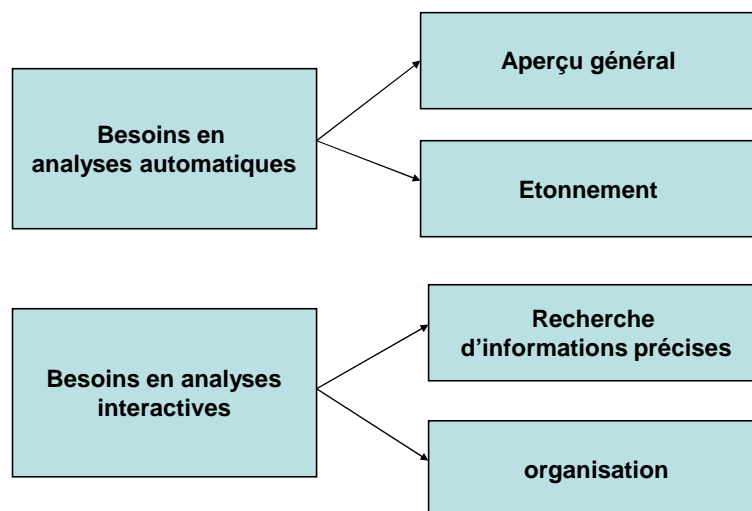
Analyse de la démarche des veilleurs

- Que cherche un veilleur ?
- Quelles informations ?
- Quelles réponses ?
- Les logiciels sont-ils satisfaisants ?

Etapes de la réalisation de l'étude (plantes transgéniques)

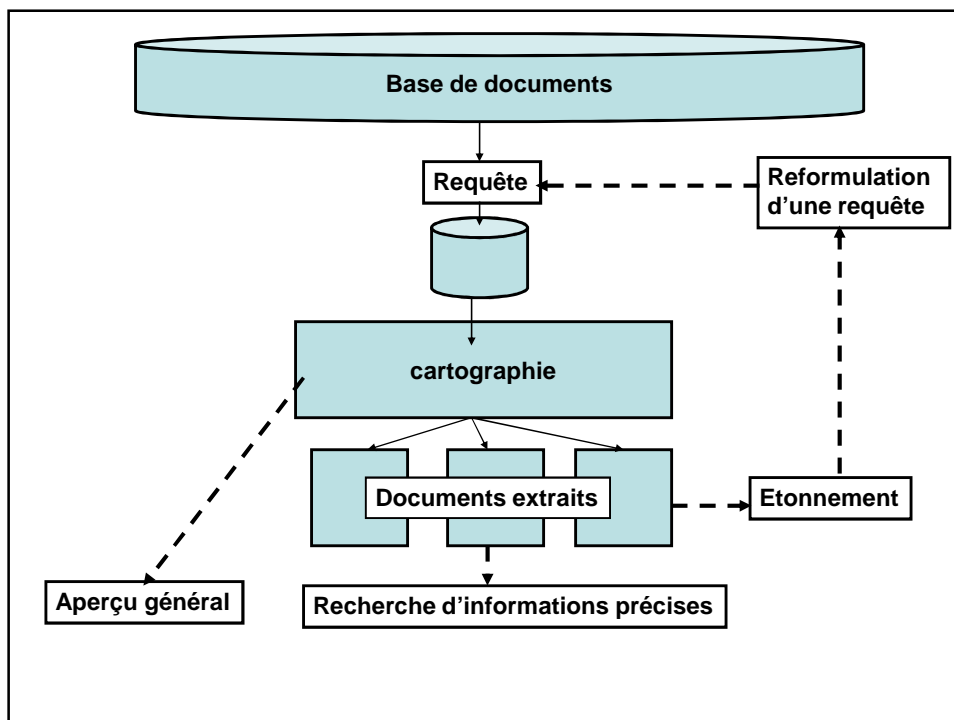
1. Choix du sujet
2. Choix des sources
3. Élaboration de requêtes pertinentes adaptées aux sources
4. Récupération des données
5. Formatage des données
6. Analyses statistiques sur les données structurées
7. Exploitation des résultats d'analyser
8. Mise en commun des résultats des analyse automatiques et humaines des documents
9. Identification des sous-thèmes, de sujets pertinents, plan général de l'étude
10. Réalisation de l'étude de veille

Analyse des besoins

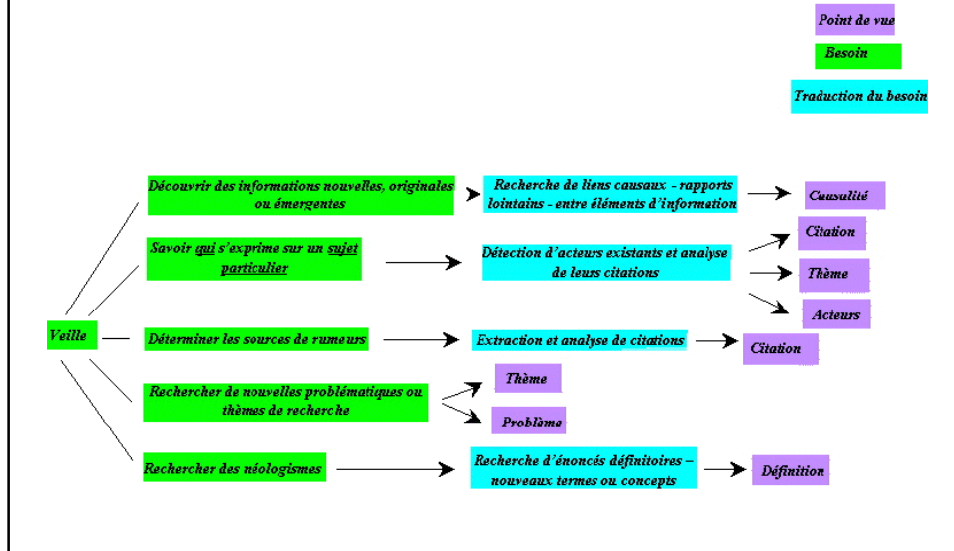


Besoins des veilleurs

- Analyses automatiques
 - Informations sur le contenu global des documents
 - Résultats bruts qui peuvent étonner ; l'extraction d'un mot n'a aucune valeur.
- Analyses interactives
 - reformulation de requêtes pour rechercher des informations plus précises
 - Regrouper des mots, des informations afin de classer des documents



Satisfaction de besoins complexes : exemple de plan de satisfaction pour les veilleurs



Notions de veille (1)

- / changement/ : *modify, alter*
- /amélioration/ : *enhanced, improve*
- /augmentation/ : *increased*
- /production/ : *produce, producing*
- /résistance/ : *defend ...a gainst, resistance to*
- /utilisation/ : *useful for, used for*
- /détermination/ : *degrade*
- /diminution/ : *reduce*

Notions de veille (2)

- **/maintien/ : *maintain***
- **/effet causé/ : *caused by***
- **/identification/ : *identify***
- **/destruction/ : *killing***
- **/contrôle/ : *controlling***
- **/analyse/ : *study***
- **/mesure/ : *quantify***
- **/nouveau/ : *new***

/amélioration/

- **Liste des indicateurs :**
to improve, to enhance, to ameliorate, to correct, to raise, improvement, amelioration, correction
- **Exemples typiques :**
 - « *improves* physical fitness* »
 - « *Similarity vegetable quality may be enhanced** »

/production/

- **Indicateurs :**

To produce, to create, to develop, to synthesize, to form, to design, to compose, to construct, to yield, to generate, to manufacture, to derive, synthesis, construct, production, creation, yield form, generation, manufacture.

- **Exemples typiques :**

« **produce* virus resistant plants using genetic engineering techniques »*

« **construct* a uniform distribution of individual particles over a very small target area »*

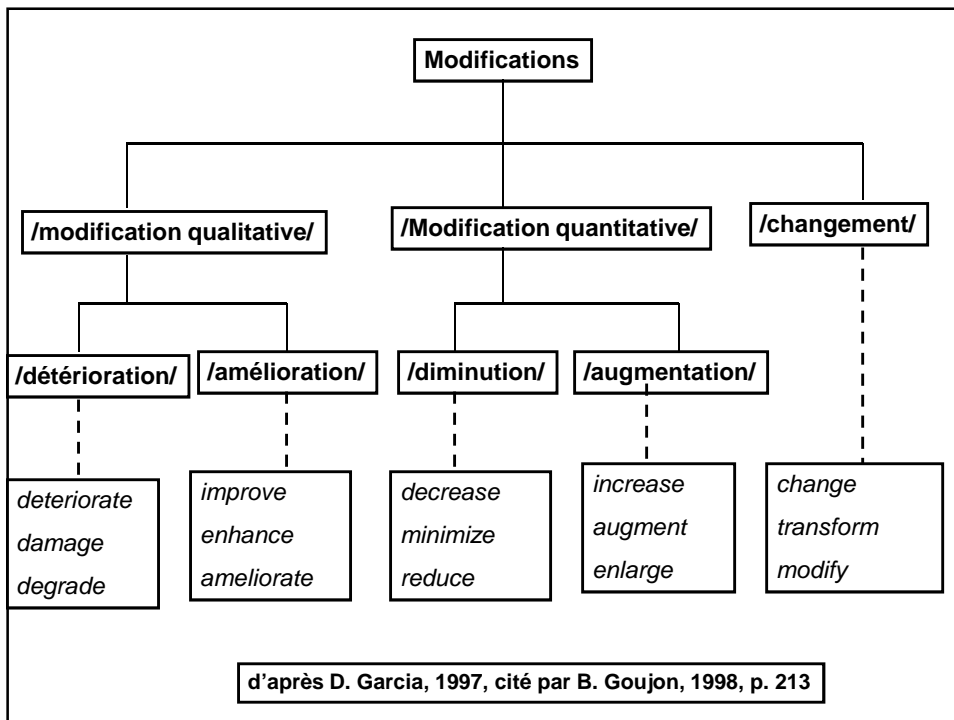
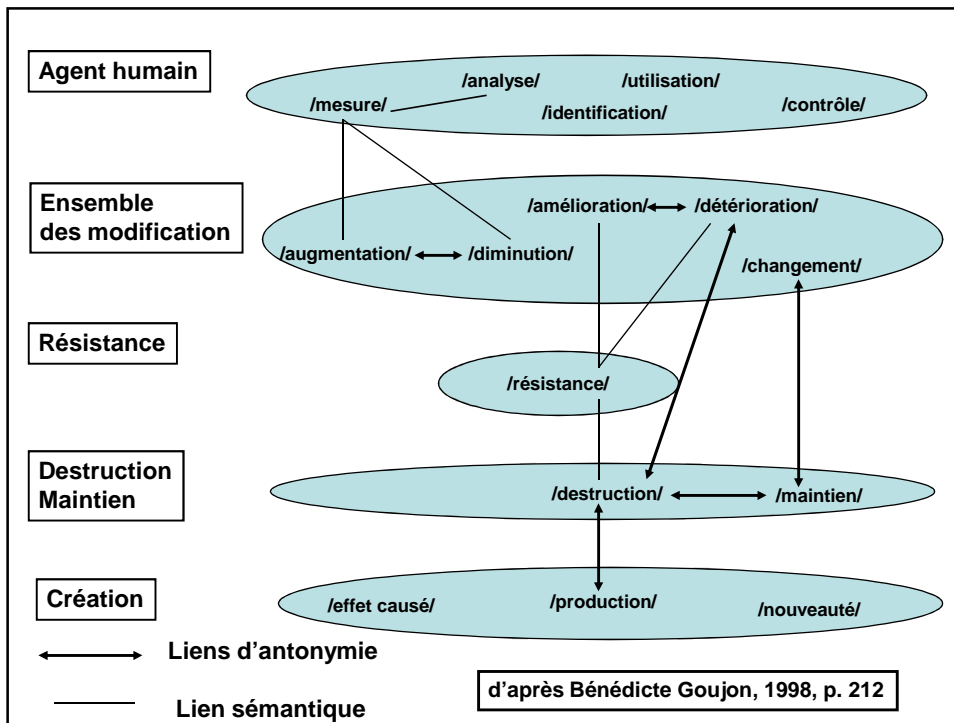
/ effet causé/

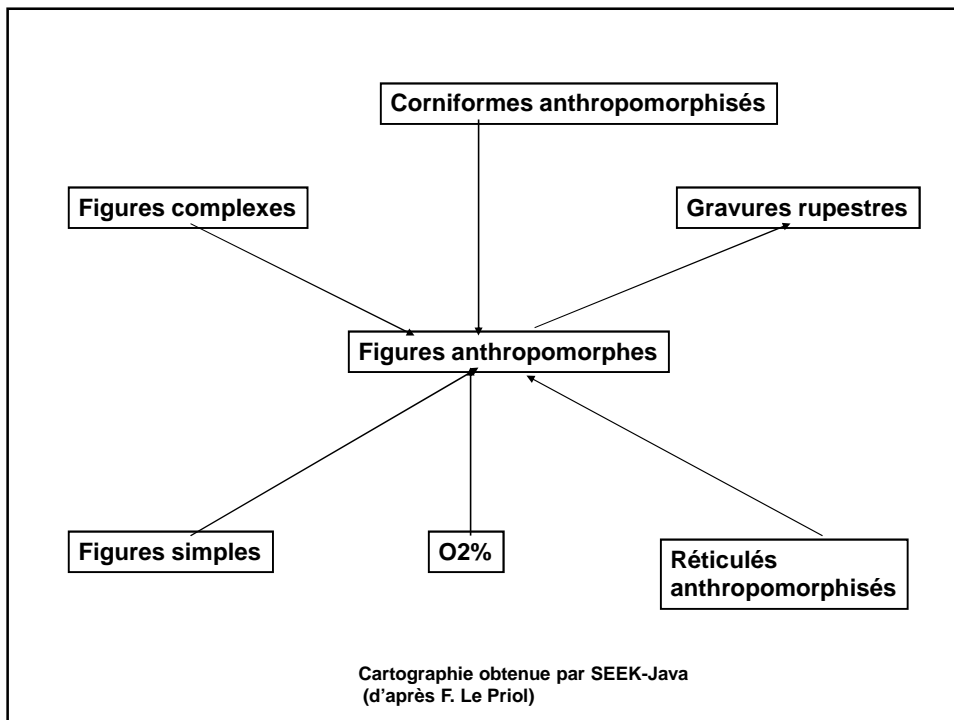
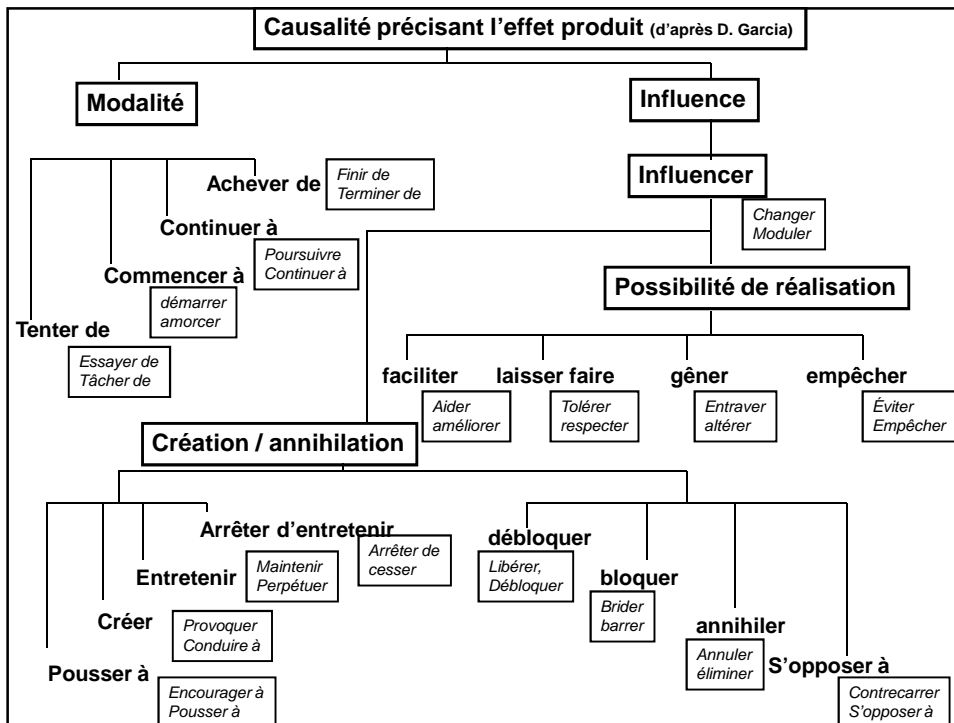
- **Indicateurs :**

To cause, to provoke, to induce, to result in, to give rise, to bring about

- **Exemple typique :**

« **causing* a substantial loss of AAA1 gene prod activities »*





Quelques éléments bibliographiques

- François Jakobiak, *L'information scientifique et technique*, PUF, 1995
- François Jakobiak, H. Dou, 1992, « De l'information documentaire à la veille technologique pour l'entreprise. Enjeux aspects généraux et définitions », *La veille technologique*, 1992, pp. 1-45
- Daniella Garcia, *Analyse automatique des textes pour l'organisation causale des actions. Réalisation, du système informatique COATIS*, Thèse de doctorat, Université de Paris-Sorbonne, 1998.
- Bénédicte Goujon, *Utilisation de l'exploration contextuelle pour l'aide à la veille technologique. Réalisation du système informatique VIGITEXT*, Thèse de doctorat, Université de Paris-Sorbonne, 1998.

**Comment introduire
« plus de sémantique » dans
les recherches d'informations ?**

Introduire plus de sémantique ?

Les **méthodes numériques** (réseaux de neurones formels, méthodes statistiques, méthodes probabilistes, N-grammes, automates markoviens ...) sont utilisées dans la fouille. Elles ne sont pas suffisantes.

« Introduire plus de sémantique » dans la recherche des informations pertinentes est devenue une nécessité.

Pour introduire « plus de sémantique » dans les fouille de textes, il devient urgent d'élaborer une sémantique plus discursive où le texte entier est segmenté, représenté et structuré en thèmes cohérents et homogènes .

Quelles sémantiques ?

Sémantique des mots (des lexies)

Sémantique du grammatical (morphèmes grammaticaux)

Sémantique de la phrase

 sémantique véri-conditionnelle (de la référence)

 sémantique interprétative

Sémantique cognitive

Sémantique discursive (du texte)

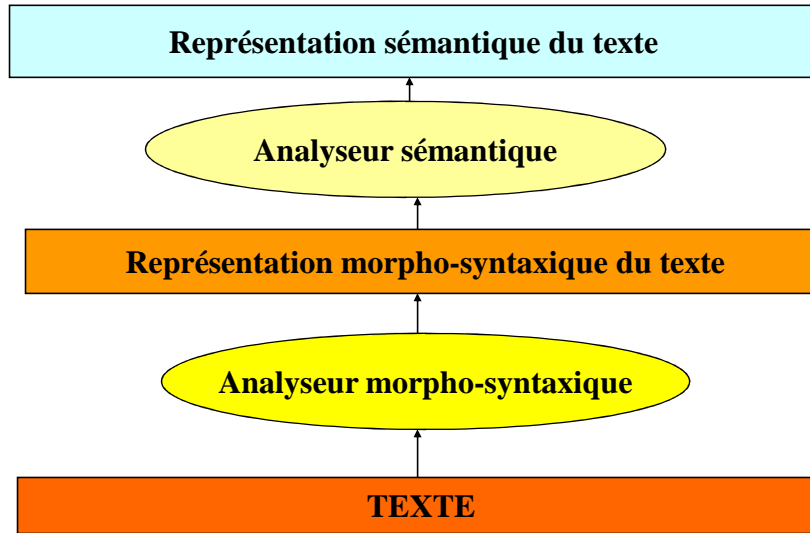
Quelles techniques sémantiques ?

- Sémantique des sèmes :
 - une combinaison booléennes de **sèmes** est la signification d'un mot
 - Sémantique logique :
 - une **formule logique** interprétées dans un ensemble exprime la référence d'une phrase (exemples phrases avec quantificateurs)
 - Sémantique cognitive :
 - un **schème sémantico-cognitif** représente la signification d'une phrase
 - Sémantique discursive :
 - un texte possède une structure qui contribue à sa signification
- => Faut-il avoir recours à la pragmatique et aux ontologies des domaines ?

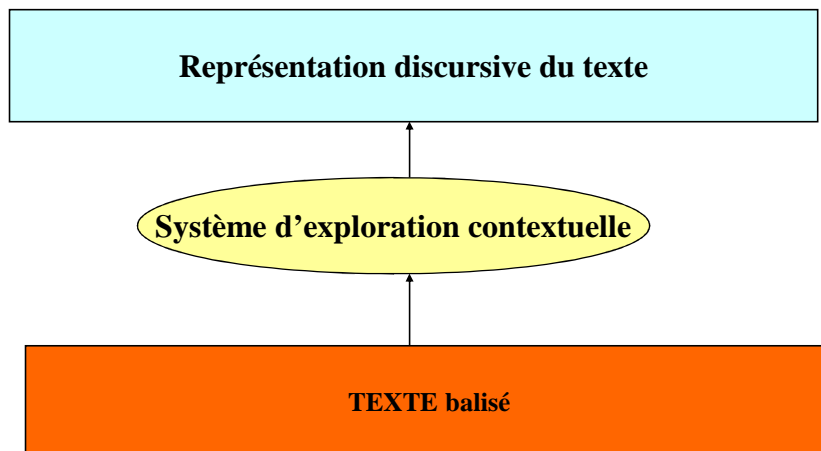
Faut-il nécessairement s'appuyer préalablement sur des découpages syntaxiques ?

- Beaucoup de textes (rapports techniques, notes administratives, textes juridiques...) ne respectent pas toujours une syntaxe normée.
- Les analyseurs syntaxiques automatiques échouent car ils n'arrivent pas à « couvrir » ce genre de documents => faire des analyseurs « robustes ».
- Certains analyseurs syntaxiques sont consommateurs de ressources qui doivent « couvrir » la globalité de la langue.

Paradigme du TAL : Syntaxe => Sémantique



Peut-on faire de la sémantique sans syntaxe ?



Une nouvelle technique : l'Exploration contextuelle

Règle d'exploration contextuelle

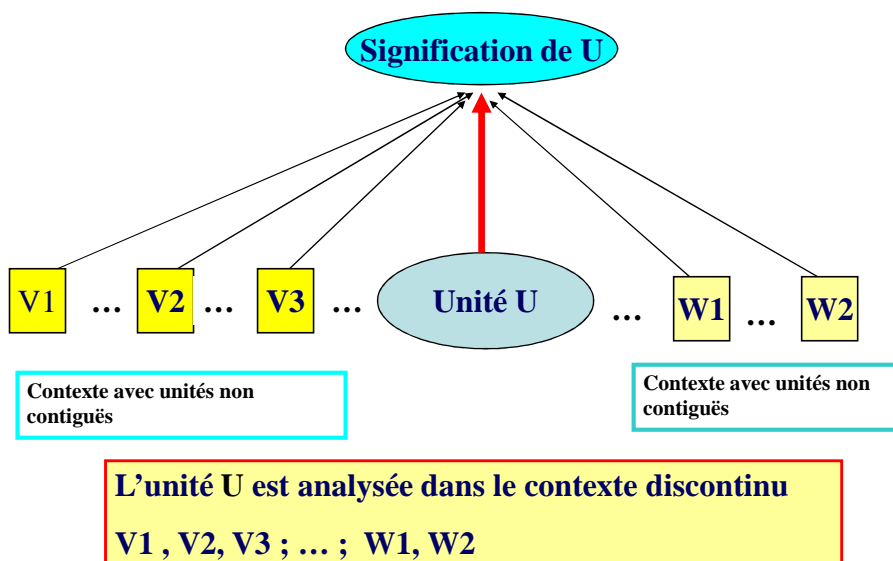
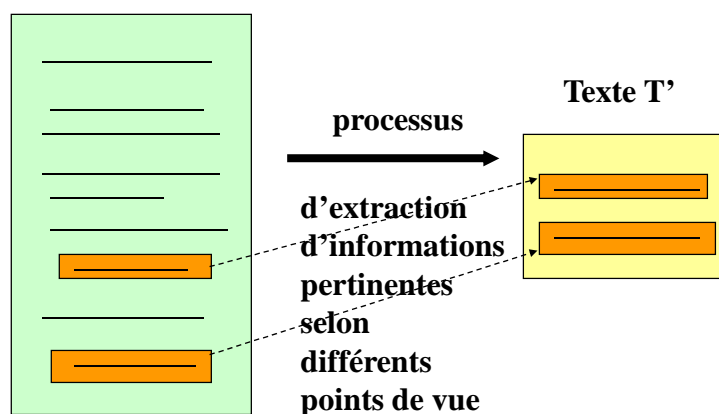


Plate-forme EXCOM

- 1/ Résumer les textes
- 2/ Identifier les changements de thématiques dans un texte
- 3/ Recherches d'informations sur la Toile

1/ Extraire des informations par des annotations sémantiques

Texte T



2/ Identifier des changements de thématiques dans un Texte

- Quelques Indices linguistiques :
- En ce qui concerne les impôts, je voudrais
- Mais, pour les abattements fiscaux, il faut

CT1

Au cours des temps géologiques,

le niveau absolu des mers fluctue en fonction du climat et de l'activité des dorsales océaniques

CT2

En période de haut niveau marin,

les mers s'étendent largement sur les continents et les eaux se réchauffent., car la surface qui capte le rayonnement solaire est grande.

Le plancton prolifère : les sédiments marins contiennent de la matière organique qui se transforme en hydrocarbures.

CT3

En période de bas niveau marin,

les mers régressent et le lit des fleuves se creuse à partir de leur embouchure.

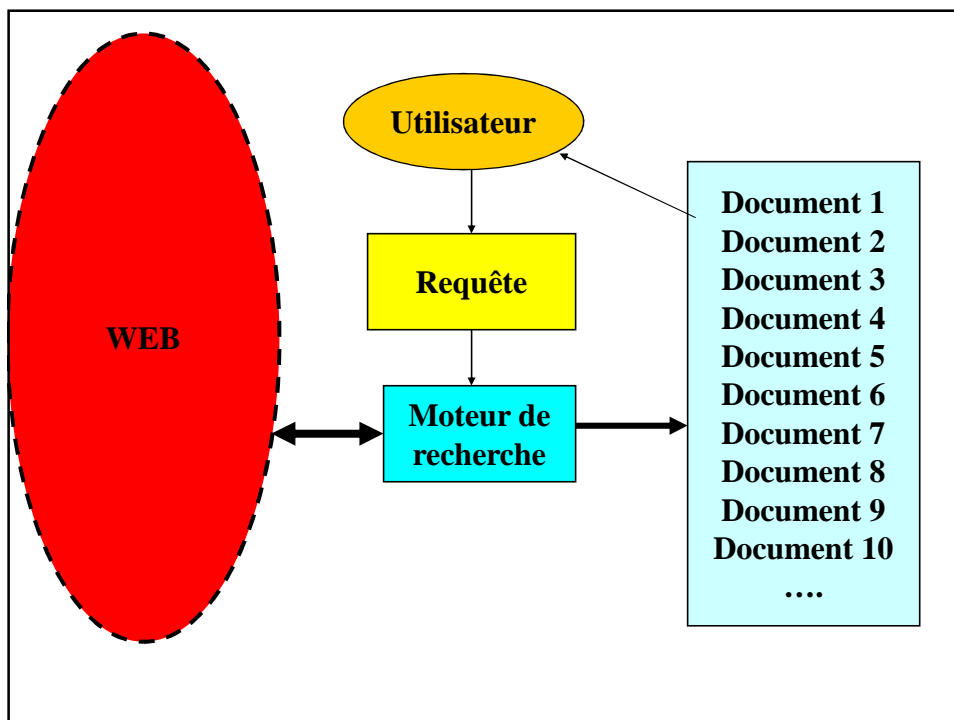
Cette érosion accumule des sables sur les fonds marins : ils y forment des roches poreuses qui pourront stocker les hydrocarbures.

Les prospections pétrolières sont facilitées lorsque l'on connaît précisément la succession de ces différentes périodes. Toutefois, si l'on sait bien déterminer l'âge des différentes couches de dépôts sédimentaires, on ignore trop souvent à quelle profondeur elles sont formées. »

3/ Rechercher des informations sur le Web

Problèmes :

- 1) Les moteurs actuels ramènent trop d'informations, donc ils sont difficilement utilisables => bruit.
- 2) On ne maîtrise pas assez la façon dont elles ont été sélectionnées ;
- 3) On a un aperçu trop sommaire des documents ramenés : sont-ils « pertinents » ?
- 4) Le travail de tri est laissé à l'utilisateur



Recherche assistée « intelligente » d'informations sur la Toile

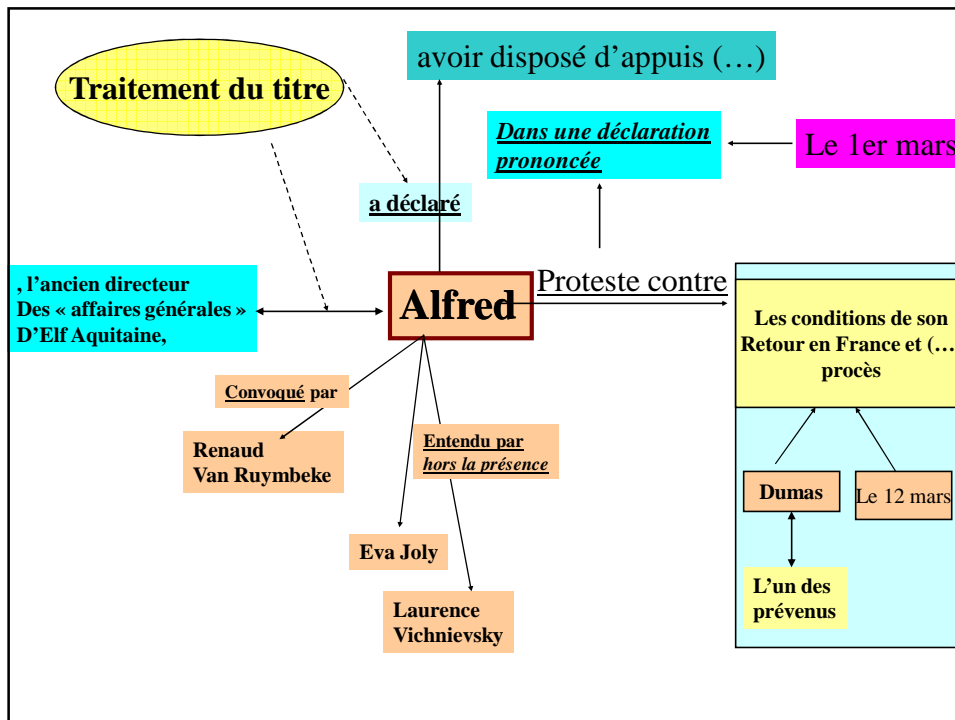
- **Identification sémantique des événements et des lieux saillants** dans des documents textuels (Web) ;
- Construire des **réseaux d'événements** localisés dans des lieux ;
- **Répondre à des questions** comme :
 - « Qui est qui ? »
 - « Qui est où ? »
 - « Où cela a-t-il eu lieu ? »

Qui a rencontré Qui ? Où ? Quand ?

Texte

TITRE : Alfred Sirven a déclaré avoir disposé d'appuis pour échapper à la justice (Le Mondes / 05.03.01/13h15)

Détenu à la maison d'arrêt de la Santé **depuis le 7 février**, l'ancien directeur des « affaires générales » d'Elf Aquitaine, **a protesté**, dans une déclaration prononcée le **1er mars**, **contre** les conditions de son retour en France et celles du procès de l'affaire **Dumas**, dont il est l'un des prévenus, et qui doit reprendre le **12 mars**. **Convoqué par Renaud Van Ruymbeke**, il **a été entendu hors la présence** des juges **Eva Joly** et **Laurence Vichnievsky**. (...)



Aider les utilisateurs à reformuler sa requête en fonction de ses intentions (points de vue)

L'utilisateur a un certain point de vue pour chercher des informations

CAUSE d'un phénomène

DEFINITION d'un terme

extraire les COMMENTAIRES d'une série de diagrammes

CITATION effectuée par une personnalité

QUI a rencontré QUI ? OU ? QUAND ?

Quelles sont les TRANSACTIONS entre compagnies

Qui DIRIGE telle entreprise ? QUI est QUOI ?

L'utilisateur n'a pas en tête les expressions générales relatives à l'objet de sa requête

Expressions de la CAUSE ?

Expressions des CITATIONS ?

Expressions de la CITATION ?

Stratégies de recherche

1. Augmenter automatiquement les requêtes par une précision accrue ;
2. Ramener des documents plus pertinents en moins grand nombre
3. Donner des extraits pertinents en rapport avec la requête
4. Revenir aux documents pertinents
5. Utiliser plusieurs moteurs de recherche
6. Sélectionner en synthétisant les réponses

Importance des expressions discursives

- 1) Beaucoup d'expressions discursives structurent un texte ;
- 2) Identifier automatiquement ces expressions, c'est
 - découvrir la structure discursive du texte ;
 - être capable d'attribuer des étiquettes discursives – annotations automatiques - à des phrases (propositions, paragraphes) ;
 - extraire les unités discursives en fonction des points de vue adoptés
 - créer des liens intra-texte et hyper-textes .

Annotation sémantique automatique

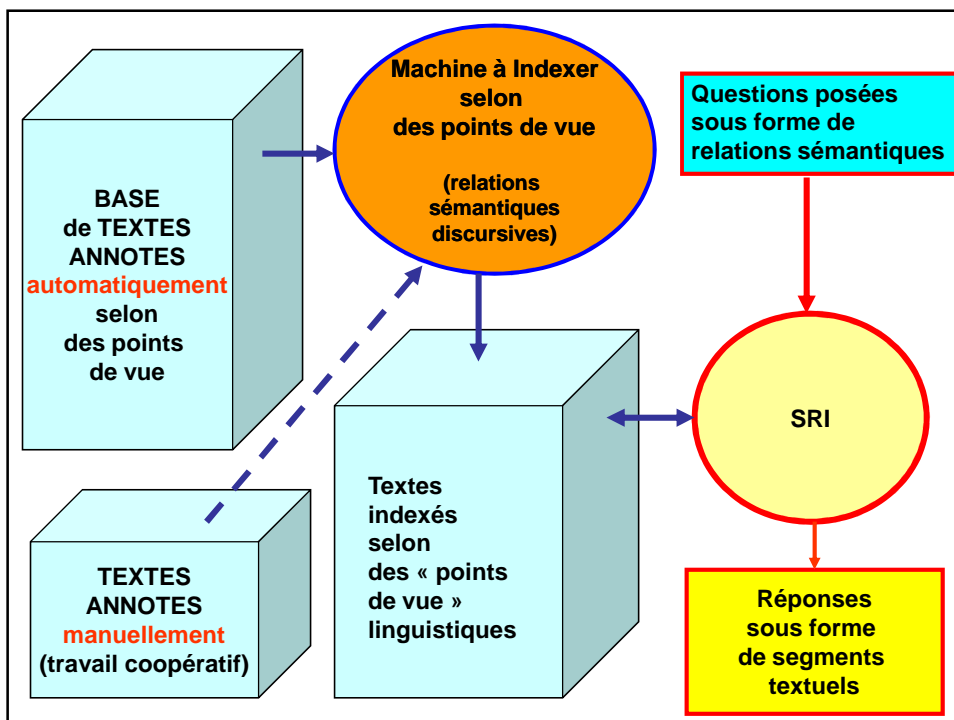
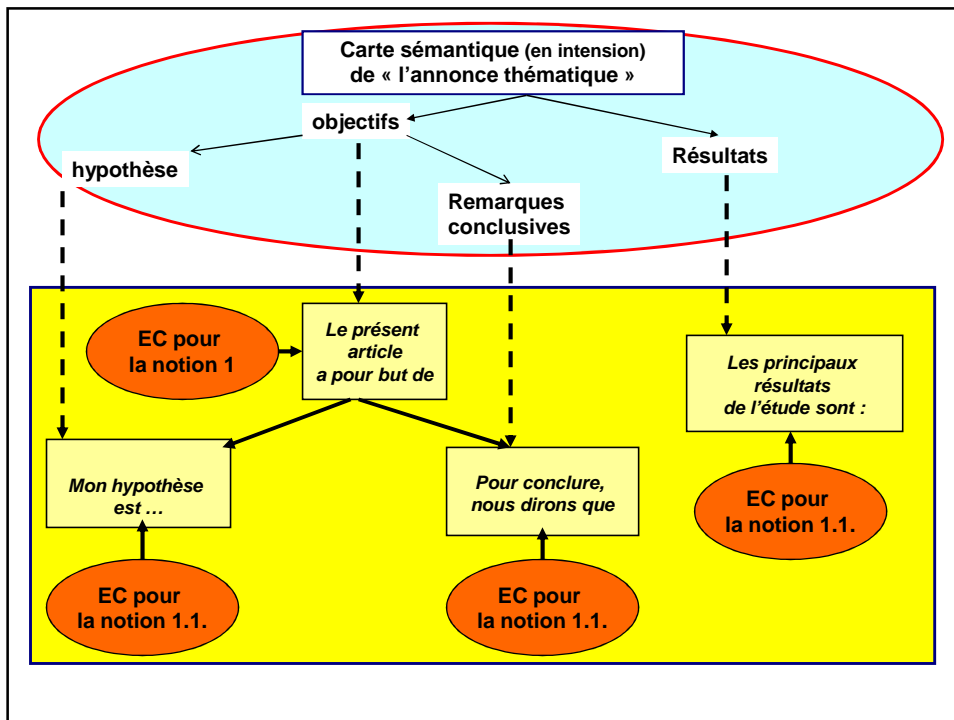
=> enrichissement du texte à partir du texte lui-même pour des traitements opératoires

Termes linguistiques vs Relations sémantiques

- Les SRI actuels exploitent essentiellement des réseaux de termes (nominaux) qui désignent des entités =>
 - les thésaurus, réseaux sémantiques statiques;
 - les ontologies des domaines sous forme de réseaux statiques;
 - les liens privilégiés avec l'approche documentaire;
 - groupes « IA- terminologie »;
 - poids des syntagmes nominaux (identification des lexies);
 - peu de sémantique verbale (domaines des actions dynamiques ...)
- Or, « **une langue n'est pas un système de nomenclature** » !
(Ferdinand de Saussure : *Cours de linguistique générale*)
≠ la conception des documentalistes, des terminologues (Ecole de Vienne) ... , ce qui bloque la conception des SRI actuels (et du Web-sémantique).
- En effet, une langue est constituée par des **systèmes de mises en relations sémantiques** exprimées par des verbes, des prépositions, des connecteurs, des marqueurs grammaticaux ...

Applications de l'annotation...

- **Fouille sémantique** de textes selon des points de vue;
- **Synthèses automatiques** de documents textuels avec filtrage dans des fiches;
- **Structuration des connaissances** et constructions d'Ontologies à partir de textes;
- **Articulations** entre textes et images: annoter les images par des annotations textuelles;
- **Bibliométrie avancée** par catégorisation (positive, négative ...) de citations...
- **Ordonnement temporel** des événements dans une narration;
- **Spécifications informatiques** à partir de textes ...



Accès multilingue aux informations de documents multilingues

1. Pouvoir interroger dans une langue une base de textes (annotés) dans une autre langue
2. Exploiter des informations et des connaissances exprimées dans différentes langues.

Une Application : l'interrogation multilingue

