

Colloque Frontières Numériques'2022
Université Paris 8 : 9-10 nov. 2022.

Impacts de la traçabilité numérique dans le contexte
d'innovation et de veille technologique.

Par :

Sahbi SIDHOM

MCF. Université de Lorraine & Lab. LORIA (Eq. SMART)

<https://members.loria.fr/SSidhom/>

9-10 . 11. 2022

Mes travaux & projets de recherche



PROJETS

I am co-Founder & President of academic conferences:

- **2019+** : The **OCTA** international Multi-Conference on « Organization of Knowledge and Advanced Technologies » is a large-scale scientific event (<https://multiconference-octa.loria.fr/>).
- **2019+** : The **TBMS** international symposium on « Big-Data-Analytics Technologies for Strategic Management: innovation and competitiveness », (<https://tbms2019.loria.fr/>).
- **2018+** : The **CITED** international symposium on « *Advanced Technologies, Renewable Energies and Economic Development.* » (<https://cited2019.loria.fr/>)
- **2010+** : The **ISKO-Maghreb** international symposium on « *Knowledge Organization* » (<https://isko-maghreb2019.loria.fr>) and the international scholarly society of ISKO – Maghreb Chapter (www.isko-maghreb.org)
- **2008+** : The **SIIE** international conference on « *Information Systems & Economic Intelligence.* » (www.siie.fr & <https://siie2019.loria.fr>)

I was involved in International & European projects:

- **DEBORA**® project : Digital accEss to Books of the RenAissance (<http://cordis.europa.eu/libraries/en/projects/debora.html>).
- **ERUDIT**®, **MILANO**® & **BALAENA**© projects: Banques de Renseignements en Langage Naturel (France Télécom R&D Lannion – Technopole Anticipa).
- **TORJOMAN**® project: système de traduction assistée par ordinateur anglais-arabe, compilation et analyse du langage naturel, sémantique dans la conception des dictionnaires de traduction (<http://www.irsit.nrnt.tn>).

Mes projets au LORIA & Eq. SMaRT :

- L'objectif de l'équipe de recherche SMaRT dans le laboratoire LORIA est de modéliser le langage : écrit ou parlé (naturel ou altéré), en utilisant des techniques de Machine Learning.
- Mes propres contributions :
 - **TALN** dans les processus d'analyse et de génération
 - **Indexation** Automatique dans le processus **d'Organisation des Connaissances**
 - **Valeur Ajoutée** dans les systèmes de veille pour les processus métiers
 - Détection des **signaux faibles** pour le processus de décision
 - Evolution d'un **SRI intelligent** pour l'aide à la prise de décision

Plan de l'exposé :

1. **Genèse : Traces et Traçabilité Numériques ?** en engageant cette réflexion sur les « frontières numériques » ;
2. **Humanité** et sa place dans le **monde informationnel** ;
3. **Humanité** et sa place dans les **modalités de la connaissance** ;
4. Contexte d'une **recherche réflexive** : **Web dynamique, interactif et social ?**
5. **Monde de données** - de traces numériques, de documents composites et d'informations hétérogènes - : la notion de la **Valeur Ajoutée ?**
6. **Intelligence Humaine** et sa Traçabilité Numérique : la notion de **poly-représentation Système-Utilisateur ?**
7. Principaux **résultats et outils** : un zoom sur **des recherches applicatives.**

1. Genèse : contexte de recherche

Traces et Traçabilité Numériques ? en engageant cette réflexion sur les « frontières numériques »

Genèse :

Traitements de l' IST dans la production et l'organisation des connaissances.

Processus :

**Sources et Informations
hétérogènes :**

Corpus, Enquêtes, Contenus Web, Contenus liés au Multimédia, contenus liés à l'imagerie médicale, Traces d'usage, etc.

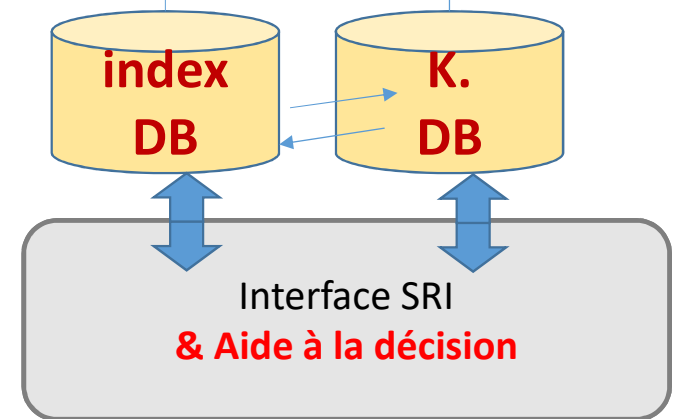
Moteur TALN :

Analyse,
Traitement &
Représentation.

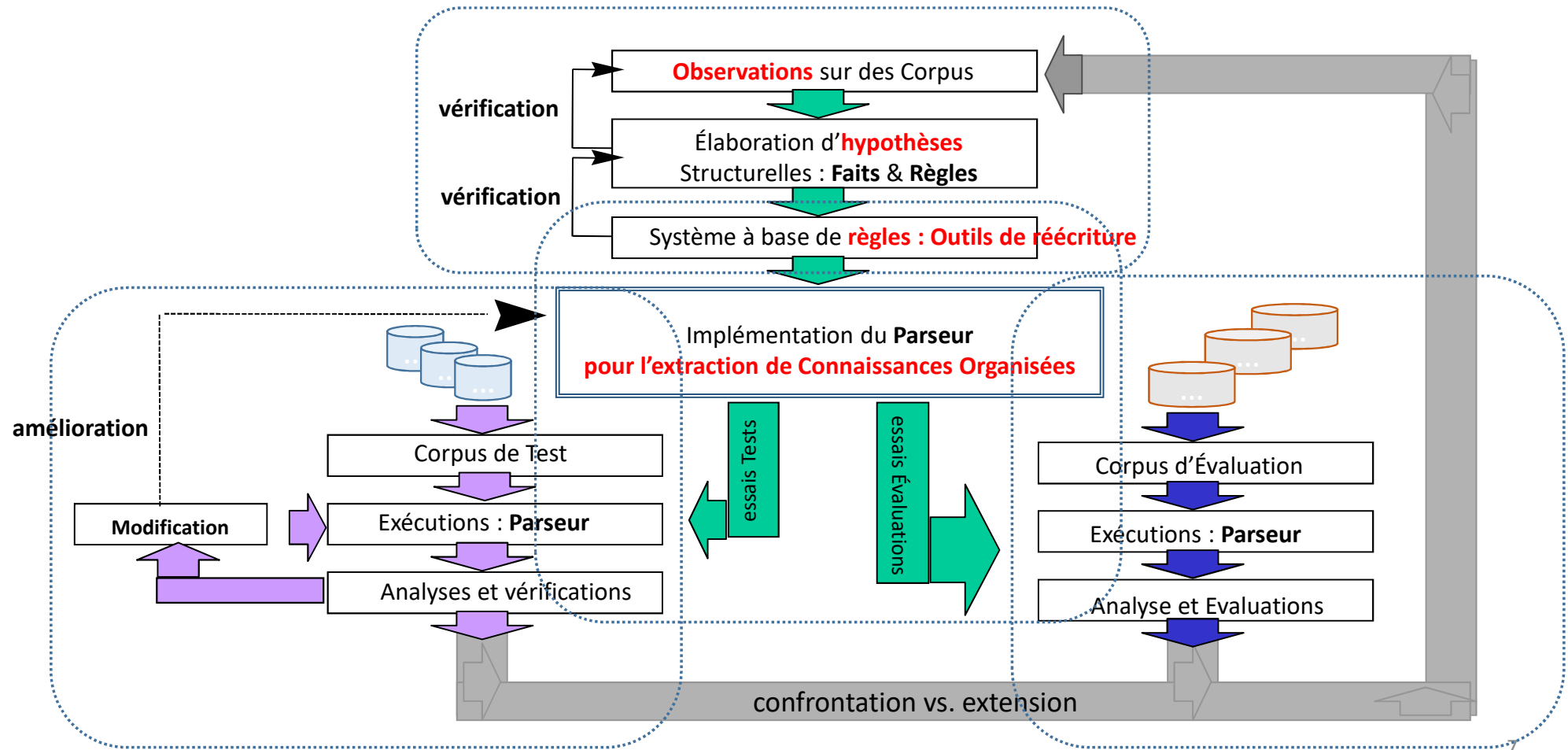
Formalisations :
**Informations à valeurs ajoutées
& Connaissances organisées.**



Expériences sur des Corpus & Données Bibliographiques



*Démarche structuraliste sur l'objet d'étude :
des contenus documentaires et textuels... aux traces numériques.*



TAL & Indexation : ex. groupe de travail SYDO

- Un groupe de travail et de chercheurs universitaires sur le thème des SYstèmes DOcumentaires (**SYDO, 1990**),
- installé au cœur d'un environnement scientifique, industriel et universitaire très riche : **Grenoble, Lyon, Saint-Etienne, Dijon et Fribourg**
- avec le financement de: Centre CEA-Grenoble, Université Claude Bernard, ENSSIB, Université Lumière, INSA-Lyon, etc.
- **Membres :**
 - [1978] R. Boucher, A. Berrendonner, Y. Chiaramella, M. Leguern,
 - [1990] J-P. Metzger, J. Kouloumdjian, M. Hassoun, S. Lainé-Cruzel, M. Debrito, R. Palermitti,
 - [1998] H. Curamoto, **S. Sidhom**, M. Hachani, ...

Le groupe SYDO a consacré l'essentiel de ses recherches au développement de solutions innovantes, dans l'analyse des ressources liées aux domaines de l'énergie, de la santé, de l'information et de la communication avec des approches TALN.

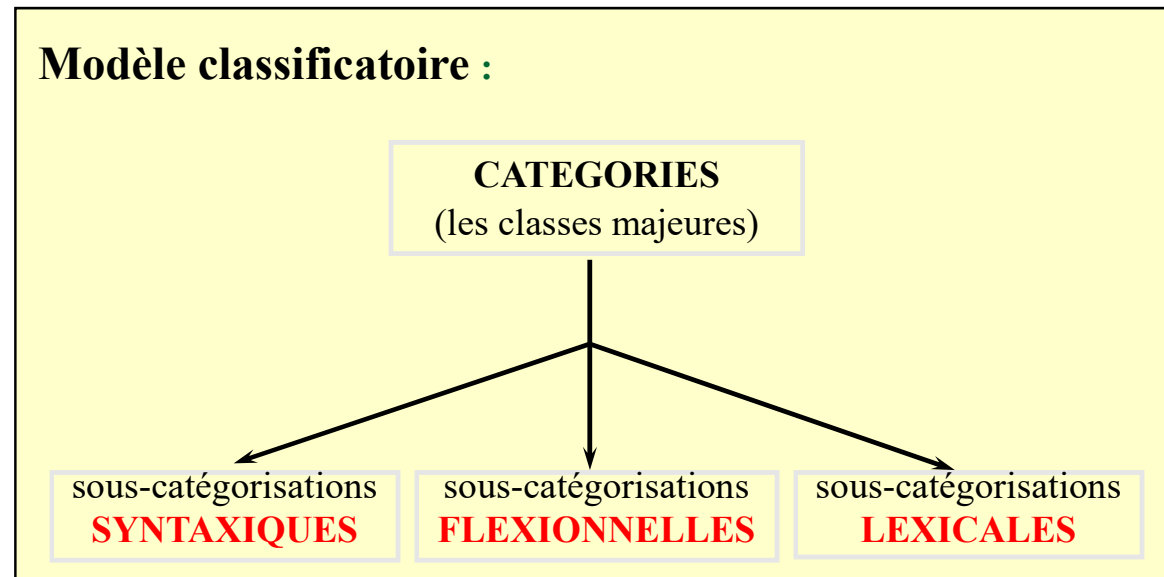
Intérêt :

- « L'une des hypothèses essentielles du modèle SYDO est l'importance des fragments de discours ayant une fonction référentielle, c'est-à-dire faisant référence à un objet de l'univers de discours. Cette fonction référentielle est majoritairement assumée par les groupes nominaux (plus précisément, par ce que les linguistes appellent les **syntagmes nominaux**). » (Lainé-Cruzel, 2001).
- « ... C'est autour de ces unités, de nature syntagmatique et non lexicale, que se forment ensuite les **énoncés du discours** sous la forme de propositions, de phrases, de « périodes »,... qui ont pour fonction d'apporter de l'« information » sur la réalité ainsi référencée. En aucun cas ces unités ne peuvent être assimilées à des unités lexicales, ni même à leurs occurrences, qui « symbolisent » des **concepts** ... » (Metzger, 1988).

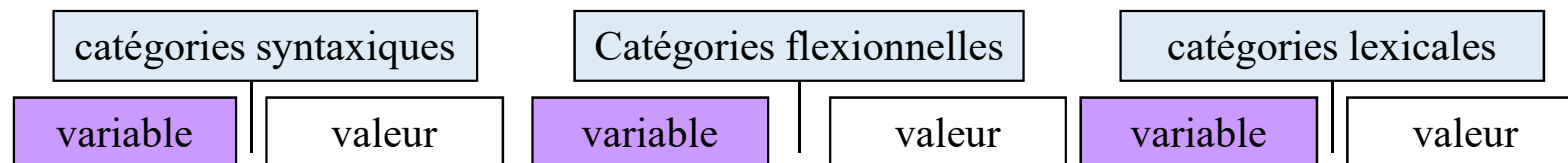
2. Modèle informationnel :

Humanité et sa place dans le **monde informationnel**

Modèle des FAITS : mots & morphologie classificatoire



Inventaire des catégories, variables et valeurs

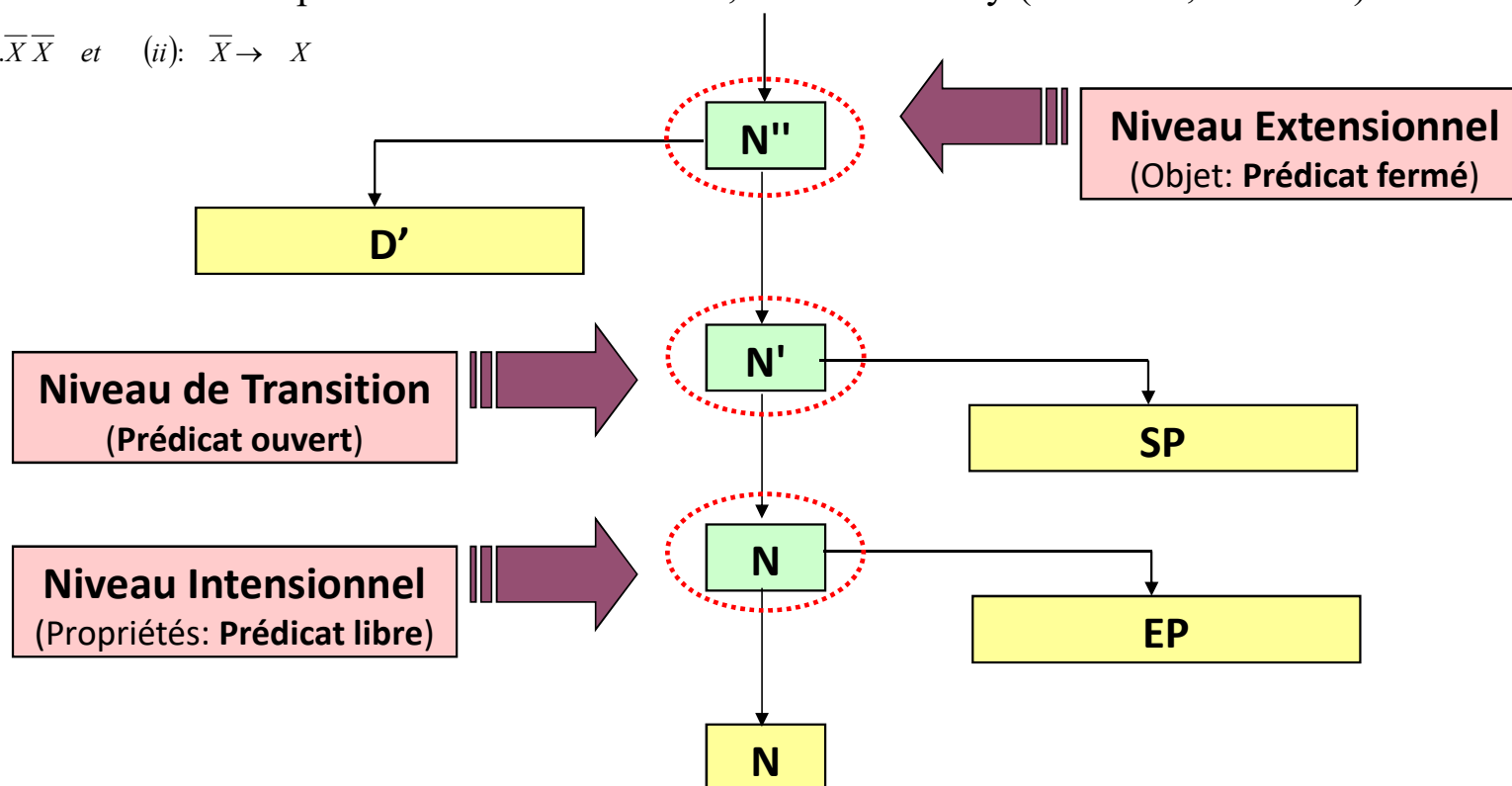


Modèle des REGLES : règles de réécriture (SN ou N'')

Modèle Logico-Sémantique du SN dans la dynamique de la phrase

La règle de réécriture est inspirée de la notion X-barre, de N. Chomsky (\bar{X} ou X' , $\bar{\bar{X}}$ ou X'') :

(i): $\bar{\bar{X}} \rightarrow \text{spéc.} \bar{X} \bar{X}$ et (ii): $\bar{X} \rightarrow X$

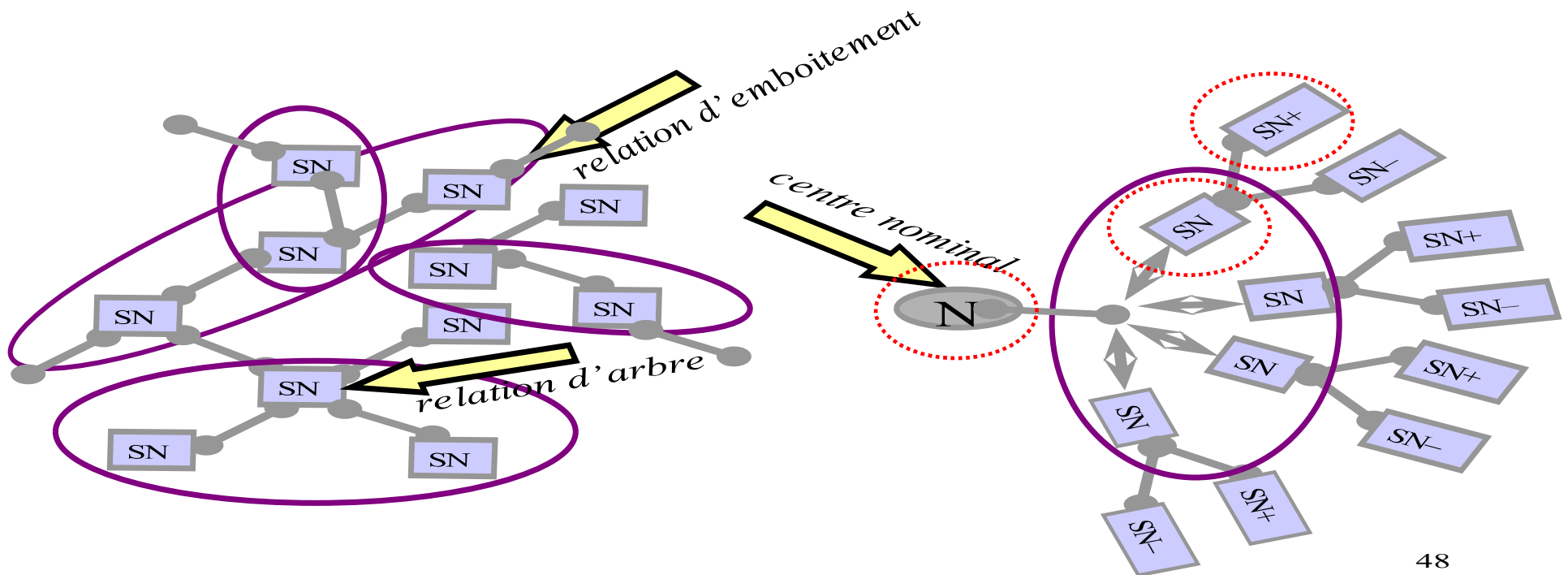


Modèle SN et ses propriétés :

Réseau de connaissances Organisées

Accès au réseau par le SN

Accès au réseau par le centre N



Modèle en application :

V_T : Vocabulaire Terminal du SN

Symbole	Catégorie
F-NOM	les noms
F-NOM-PRP	les noms propres
F-NOM-PRO	les noms pronoms
F-NAN	selon le contexte, nom ou adjectif
F-ADJ	les adjectifs
D	les prédéterminants
D-DEF	les prédéterminants définis
D-NUM	les prédéterminants numéraux , cardinaux et assimilées
D-IND	les autres prédéterminants
W-QUA	les adverbes de quantité
W-AAJ	les adverbes d'intensité (modificateurs d'adjectif)
P	les prépositions
P-DE	la préposition /de/
CI, LA	les mots /ci/ et /là/



V_N : Vocabulaire Non-terminal du SN {N'', N', N, A', A, D', Ep, Sp}

Symbole	Catégorie
N'', N', N, A', A, D', PHR N'' _c , A _c , P _c , PHR _c , W _c , SP _c , EP _c	l'axiome : N'' représente la catégorie des syntagmes nominaux (SN). N'' domine N' qui domine N.
EP	l'expansion prépositionnelle
SP	le syntagme prépositionnel



R_i : Les règles de réécriture SN (ou N'')

Description	N° règle	Règle
syntagmes nominaux N'' :	1	$N'' \rightarrow D' + N + F - PRP$
	2	$N'' \rightarrow D' + N'$
	3	$N'' \rightarrow NOM - PRO$
	4	$N'' \rightarrow NOM - PRP$
expressions nominales N' :	5	$N' \rightarrow N + SP^n$
	6	$N' \rightarrow N + A'$
	7	$N' \rightarrow N + CI$
	8	$N' \rightarrow N + LA$
	9	$N' \rightarrow N$



expressions prédéterminatives D' :	10	$D' \rightarrow D-DEF+D-NUM$
	11	$D' \rightarrow P-DE+D-DEF$
	12	$D' \rightarrow W-QUA+P-DE+D-DEF$
	13	$D' \rightarrow W-QUA+P-DE$
	14	$D' \rightarrow D$
centres adjectivaux A' :	15	$A' \rightarrow W-AAJ+A$
	16	$A' \rightarrow A+EP$
	17	$A' \rightarrow F-ADJ,REL$
	18	$A' \rightarrow A$
centres nominaux N :	19	$N \rightarrow N+EP$
	20	$N \rightarrow N+A(QUA)$
	21	$N \rightarrow A(QUA) + N$

nominaux N:	22	$N \rightarrow F\text{-NOM}$
	23	$N \rightarrow F\text{-NAN}$
	24	$A \rightarrow F\text{-NAN}, (\text{QUA})$
	25	$A \rightarrow F\text{-ADJ}, (\text{QUA})$
syntagme prépositionnel SP:	26	$SP \rightarrow P + N''$
expansion prépositionnelle EP:	27	$EP \rightarrow P + N'$

coordination des adjectifs A:		
- propriétés communes (QUA1)=(QUA2)	28	$A(QUA) \rightarrow A(QUA1)+C+A(QUA2)$
- propriétés différentes (QUA1) \neq (QUA2)	29	$A_C(QUA) \rightarrow A(QUA1)+C+A(QUA2)$
coordination de catégories :		
- syntagme nominal	a	$N''_C \rightarrow N'' + C + N''$
- syntagme verbal	b	$SV_C \rightarrow SV + C + SV$
- verbe	c	$V_C \rightarrow V + C + V$
- adjectif	d	$A_C \rightarrow A + C + A$
- adverbe	e	$W_C \rightarrow W + C + W$
- syntagme prépositional	f	$SP_C \rightarrow SP + C + SP$
- préposition	g	$P_C \rightarrow P + C + P$
- phrase	h	$PHR_C \rightarrow PHR + C + PHR$

Des règles et leur réalisation :

N°règle	Illustration	Règle
1	le + président + Macron	$N'' \rightarrow D'+N+F-PRP$
2	le + ministre	$N'' \rightarrow D'+N'$
3	lui	$N'' \rightarrow NOM-PRO$
4	Chirac	$N'' \rightarrow NOM-PRP$
5	(la) vente de produit + à la cantine + de l'université	$N' \rightarrow N+ SP^n, n : \text{entier}$
6	étudiant + <u>assidu aux cours</u>	$N' \rightarrow N+A'$
7	chien-ci	$N' \rightarrow N+CI$
8	chien-là	$N' \rightarrow N+LA$
9	ministre	$N' \rightarrow N$
10	les + trois (candidats)	$D' \rightarrow D-DEF+D-NUM$
11	de + ces (élections)	$D' \rightarrow P-DE+D-DEF$
12	beaucoup + de + leur (temps)	$D' \rightarrow W-QUA+P-DE+D-DEF$
13	peu + de (résultat)	$D' \rightarrow W-QUA+P-DE$
14	le	$D' \rightarrow D$

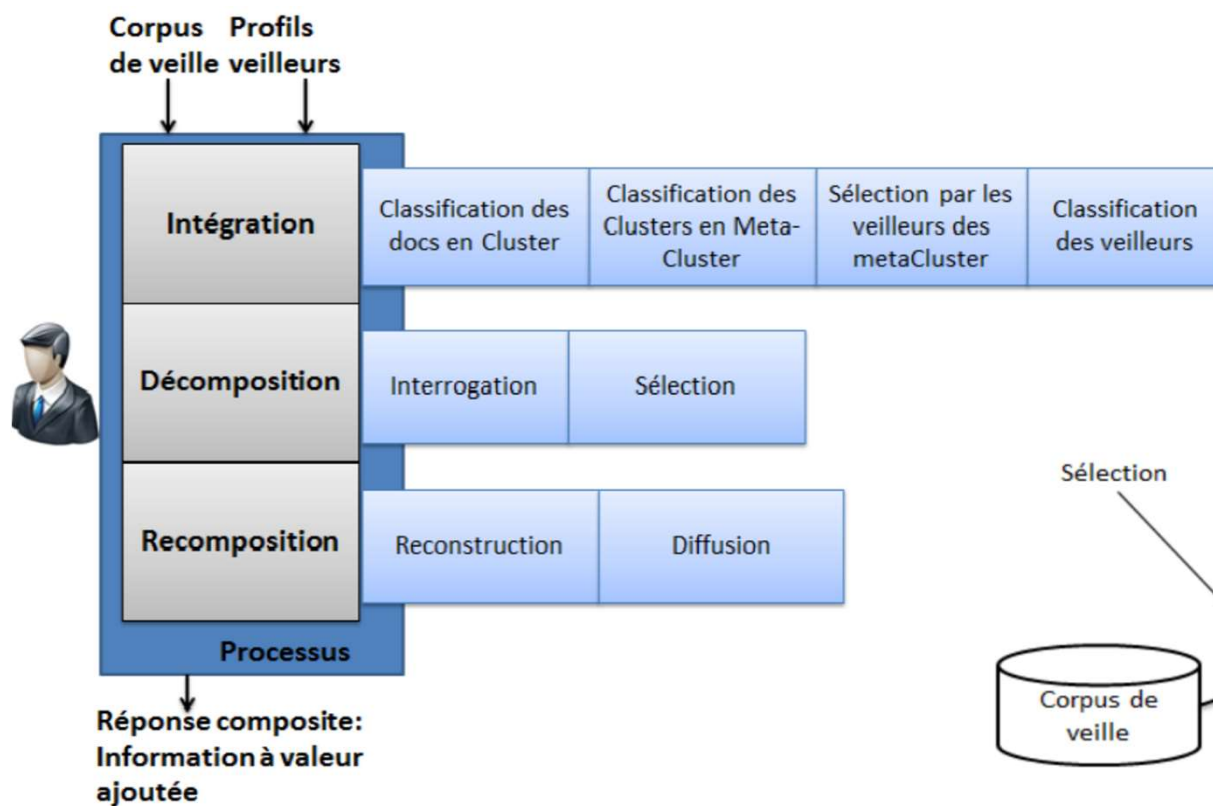


15	particulièrement + fidèle	$A' \rightarrow W-AAJ+A$
16	rouge + <u>de colère</u>	$A' \rightarrow A+EP$
17	rectoral	$A' \rightarrow F-ADJ,REL$
18	assidu + <u>aux cours</u>	$A' \rightarrow A + SP$
19	chef + <u>de gare</u>	$N \rightarrow N+EP$
20	drapeau + blanc	$N \rightarrow N+A(QUA)$
21	grand + sportif	$N \rightarrow A(QUA) + N$
22	ville	$N \rightarrow F-NOM$
23	fenêtre	$N \rightarrow F-NAN$
24	joli	$A \rightarrow F-NAN,(QUA)$
25	impartial	$A \rightarrow F-ADJ,(QUA)$

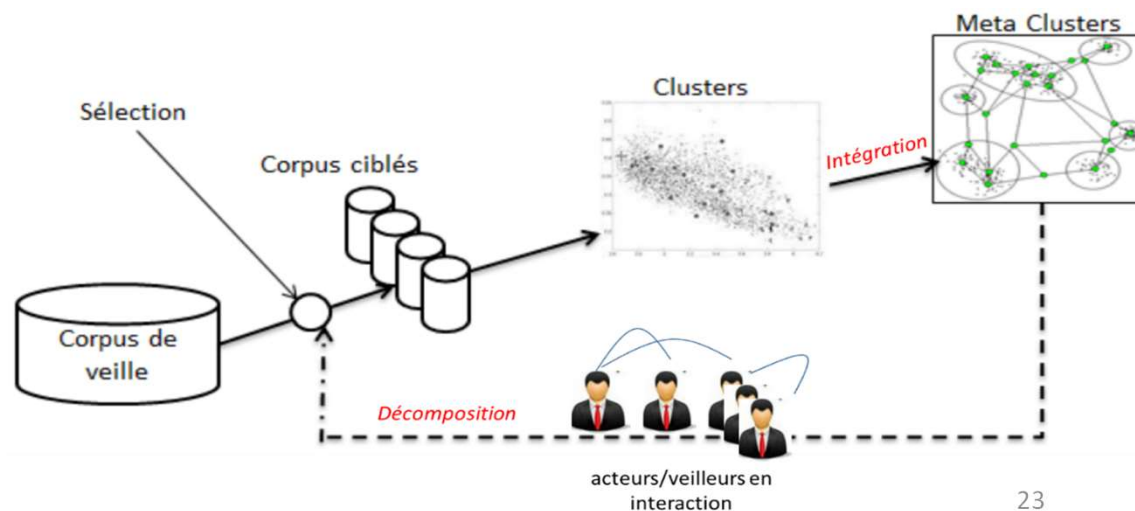
26	(le directeur) de + <u>la gare</u>	SP → P + N''
27	(le chef) de + <u>gare</u>	EP → P + N'
28	(le drapeau) blanc + et + rouge	A _c (QUA) → A(QUA1) + et + A(QUA2) c: propriétés communes
29	(un plat) chaud + et + parfumé	A _d (QUA) → A(QUA1) + et + A(QUA2) d: propriétés différentes

Autres modèles de représentation...

Processus de Veille Générique (Gwatch) et ses trois phases (Mseddi et Sidhom, 2018) :



Intégration et Décomposition des objets informationnels avec l'interaction des acteurs en veille



Pertinence :

- La pertinence d'un document pour une requête (P_d^q) :

$$P_d^q = \sum_{i=1}^t x_{(i,q)} * x_{(i,d)}$$

Avec $x_{(i,j)}$ représente les poids respectifs des index i allant de 1 à t .

- La pertinence d'un cluster

$$P_c^q = |D_c^q| * \sum_{d_i \in D} P_d^q(q, d_i)$$

avec $|D_c^q|$ est le nombre de document dans ce cluster ayant un score non nul

- La pertinence d'un méta-cluster

$$P_M^q = |C_M^q| * \sum_{c_i \in C_M^q} P_c^q(q, c_i)$$

avec $|C_M^q|$ est le nombre de clusters dans ce meta – cluster ayant un score non nul

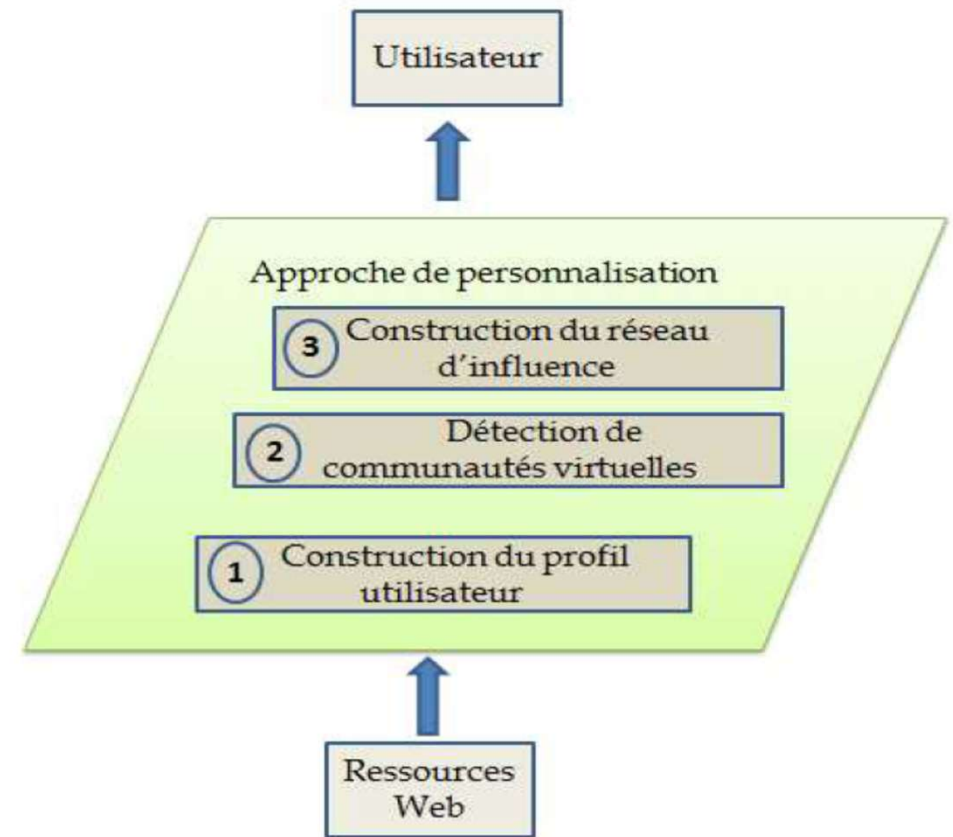
Méthodologique de la RI personnalisée :

se précise avec les objectifs suivants :

1. La définition de l'approche RI personnalisée qui intègre la construction des **profils utilisateurs** ;
2. La définition d'un processus collaboratif en RI qui permet de créer une communication entre deux processus : la RI et la **gestion des profils** utilisateurs ;
3. La définition d'une **ontologie** qui permet de construire l'appartenance d'un utilisateur au profil d'une **communauté virtuelle** détectée et son application dans la RI personnalisée ;
4. La définition de **réseaux d'influence** qui permet la détection des utilisateurs influents dans les communautés virtuelles ;
5. L'**évaluation** de l'approche proposée de la RI personnalisée.

Approche de la RI personnalisée évolue en conjuguant les aspects suivants :

- les **communautés virtuelles** ;
- la distance sémantique entre les **profils utilisateurs** avec la définition d'une méthode de déduction de réseaux d'influence par le calcul des taux de partage et des taux de confiance ;
- l'impact du profil utilisateur avec un modèle de calculs implicite et explicite en interrogeant l'**ontologie du domaine** (ODP) ;
- la **proximité sociale** entre les utilisateur induit l'application de mesures dédiées à la proximité sémantique entre les concepts dans les requêtes et les réponses aux besoins.



Approche RI personnalisée
(Harbaoui, Sidhom & Ben Ghézala, 2015).

Communautés virtuelles... réseaux d'influence

Algorithme 3 ALGORITHME DÉTECTION DES COMMUNAUTÉS VIRTUELLES

1: **Algorithm** LOUVAIN

Précondition : Soit G le graphe initial (ensembles de profils utilisateurs et les liaisons d'amitié entre eux)

Postcondition : Sortie : SC , c'est l'ensemble de sous communautés virtuelles

2: **répéter**

3: **jusqu'à fin**

4: Placer chaque sommet de G dans une unique communauté

5: Sauver la modularité de cette décomposition

6: **tant que** <il y a des sommets déplacés > **faire**

7: Chercher c la communauté voisine maximisant le gain de modularité

8: **pour tout** sommet n de G **faire**

9: Chercher c la communauté voisine maximisant le gain de modularité

10: **si** c induit un gain strictement positif **alors**

11: déplacer n de sa communauté dans c

12: **si** la modularité atteinte est supérieure à la modularité initiale **alors**

13: fin ← Faux

14: Afficher la décomposition trouvée

15: Transformer G en le graphe entre les communautés

16: **sinon**

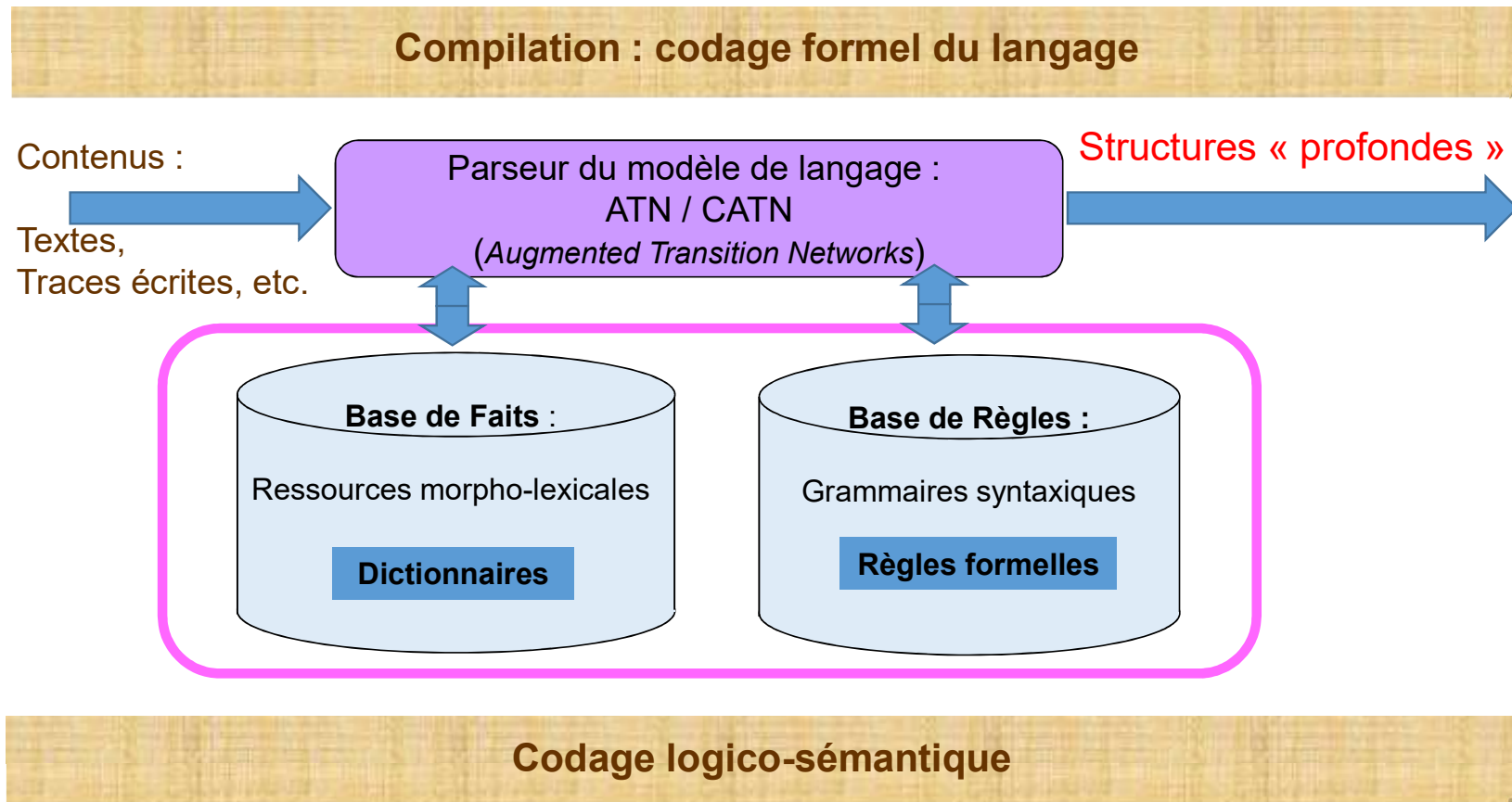
17: fin ← Vrai

18: **retour** < SC >

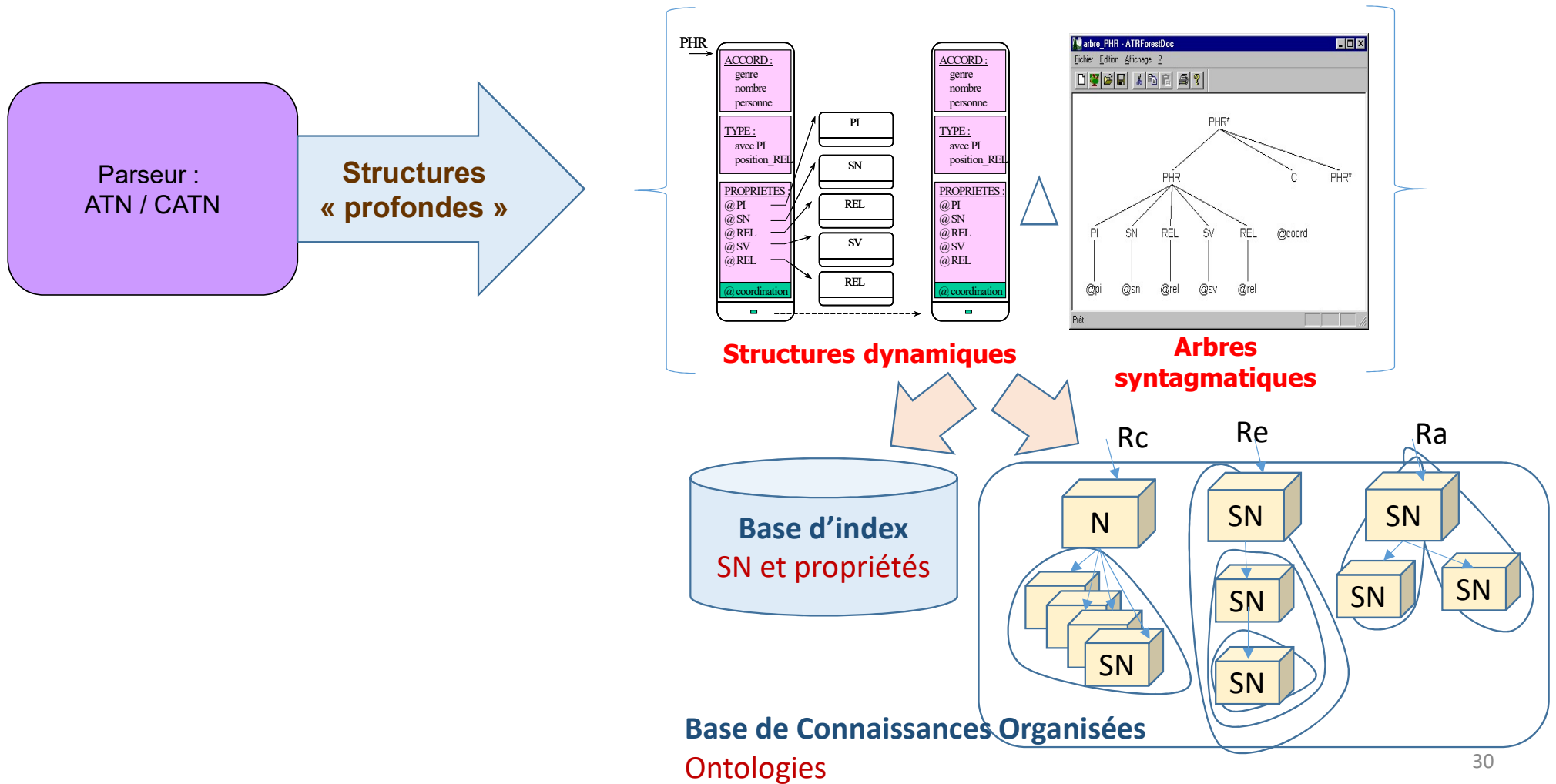
3. Modèle de connaissances organisées

Humanité et sa place dans les **modalités de la connaissance**

Parseur « Morpho-Syntaxique »

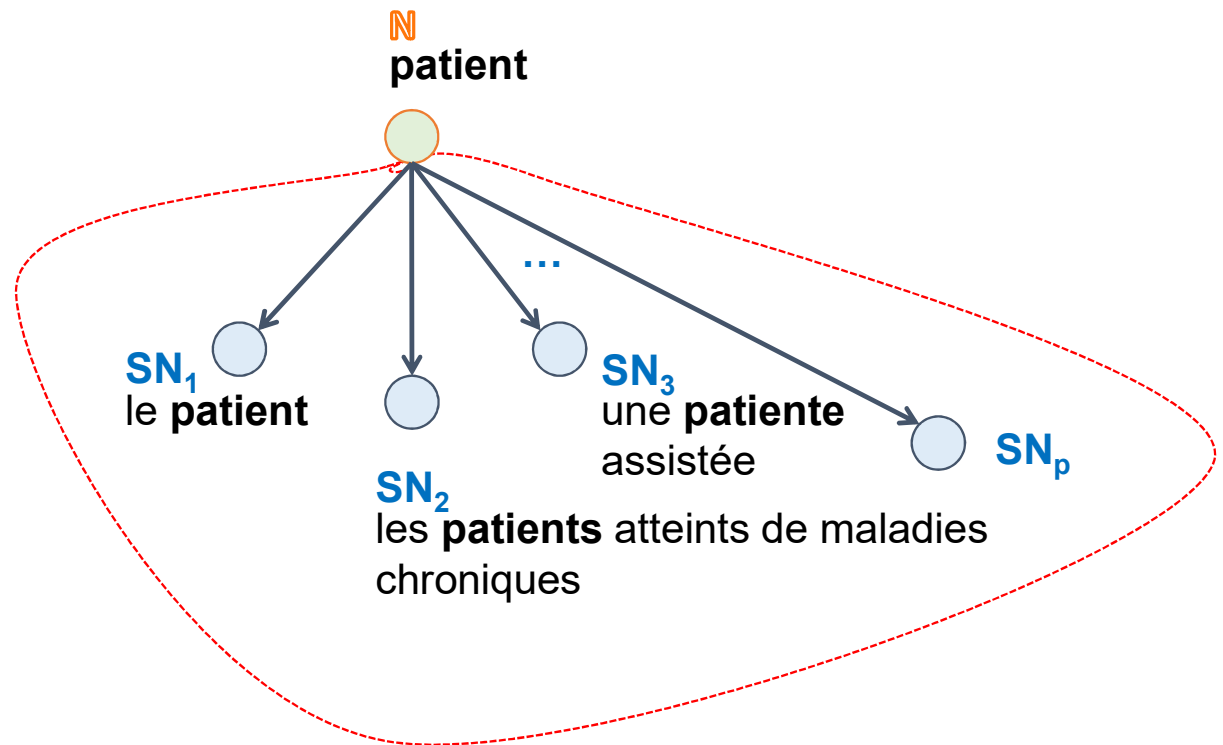


Structure profonde en exploitation...



Knowledge in Class relation:

Class_relation("patient") :
ie. {SN} ∈ N



Class relation: $N_1(NP_1, NP_2, NP_3 \dots)$

Example : (réf. INA.archives, publié le 06/06/05)

In : un convoi de troupes américain est attaqué par les escadrilles japonaises.

Out :

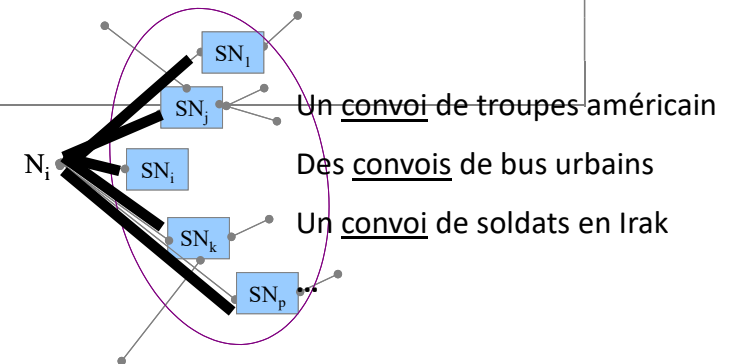
« [un **convoi** de [des **troupes**]^{SN} américain]^{SN} est attaqué par [les **escadrilles** japonaises]^{SN} ».

Indexing =

{ $N_i \in SN_i$ },

With :

- convoi \in un convoi de troupes américain
- troupe \in des troupes
- escadrille \in les escadrilles japonaises

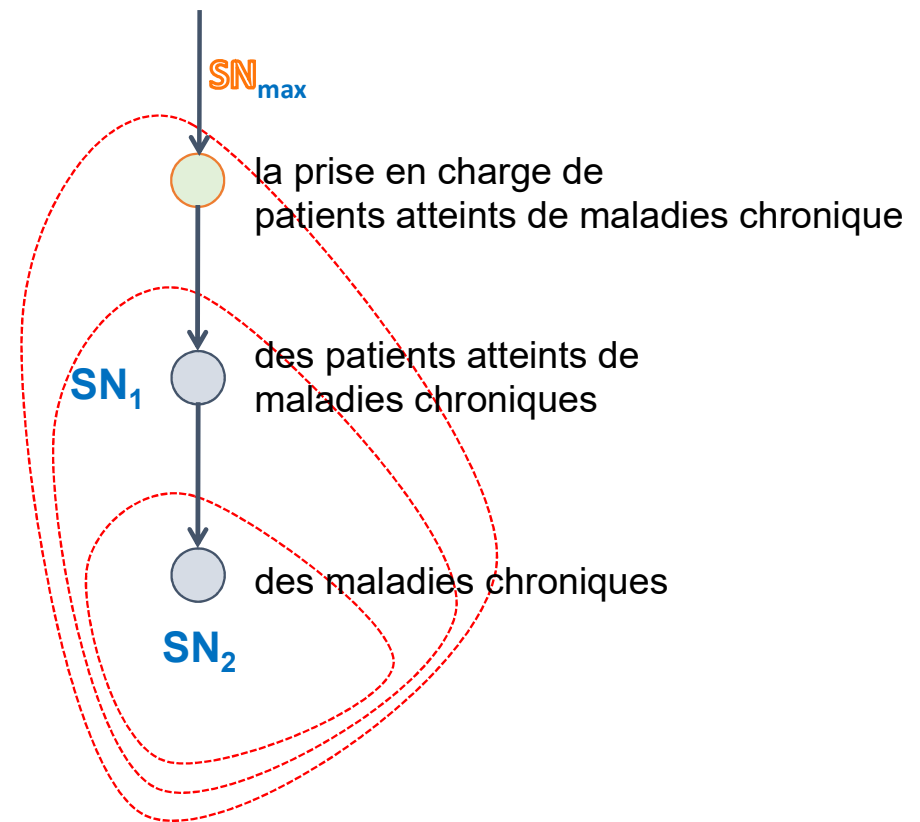


).

Knowledge in Fitting relation:

Fittig_relation(SNmax):

ie. $SN_{MAX} \supseteq SN_1 \supseteq SN_2$



Fitting relation: $NP_1 (NP_2 (NP_3 (...)))$

Example : (réf. France 3, JT. publié le 06/06/05 à 21:27)

In : B. Clinton a aussi évoqué le résultat du référendum en France .

Out : [B. Clinton]^{SN} a aussi évoqué [le résultat de [le référendum en [France] ^{SN₁}] ^{SN₂}] ^{SN₃} ».

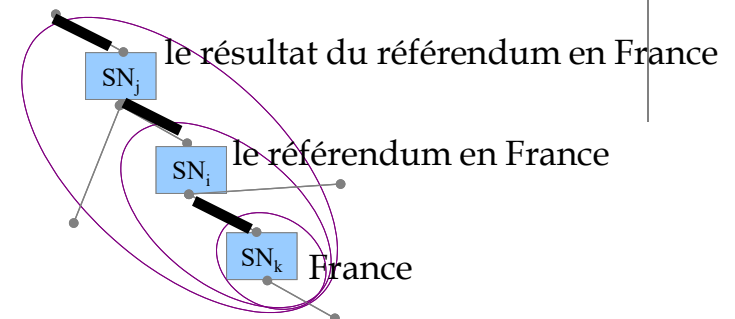
Indexing =

$SN_3 \supset SN_2 \supset SN_1$; **With :**

SN_3 = « le résultat du référendum en France »,

SN_2 = « le référendum en France »,

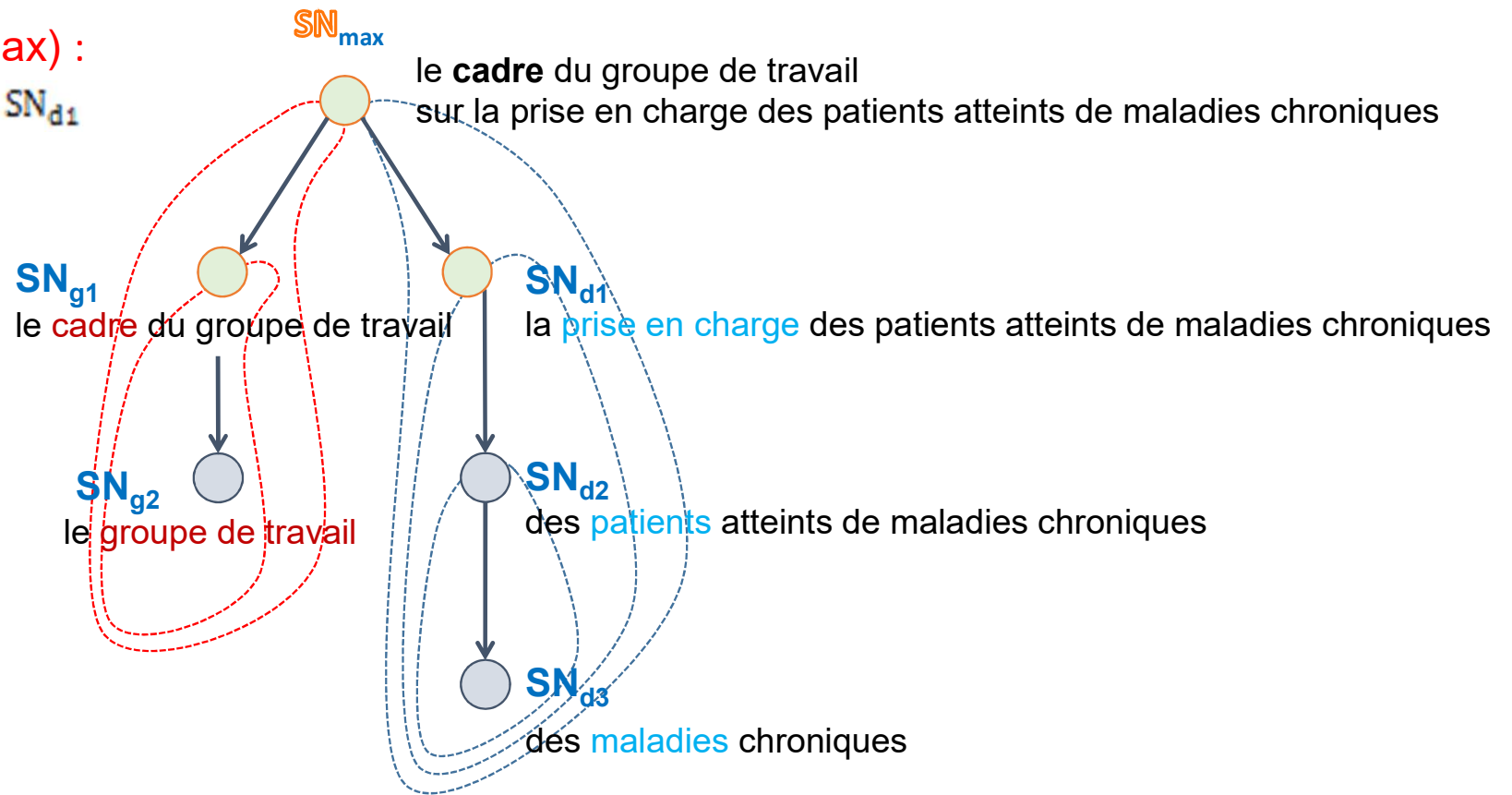
SN_1 = « France »,



Knowledge in Tree relation:

Tree_relation(SNmax) :

ie. $SN_{g1} \subseteq SN_{max} \supseteq SN_{d1}$



Tree relation: $NP_1 (NP_a (...), NP_b (...))$

Example : (réf. ouest-france.fr, publié le 06/06/05)

In : l'annonce contre la DCA allemande du débarquement allié du 6 juin 1944.

Out :

« [l'annonce contre [la DCA allemande]^{SN_{1g}} de [le débarquement allié de [le 6 juin 1944]^{SN_{1d}}]^{SN_{2d}}]^{SN_{max}} ».

Indexing =

$(SN_{max} \supset SN_{2d} \supset SN_{1d})$

With:

^

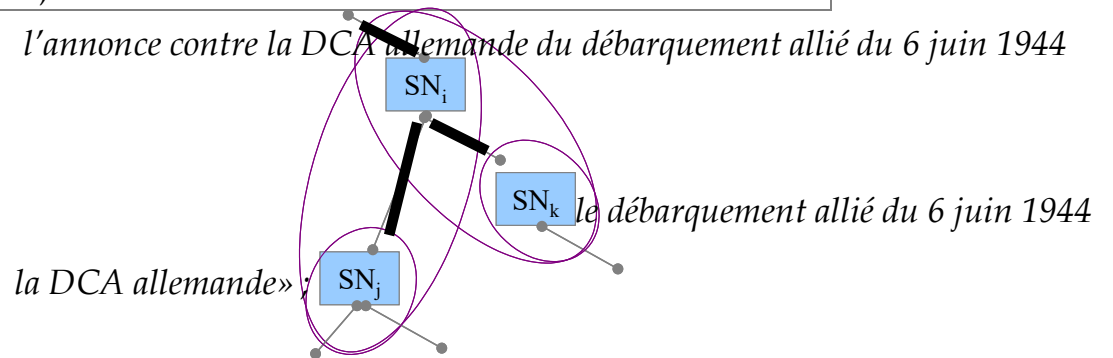
$(SN_{max} \supset SN_{1g})$;

SN_{max} = « l'annonce contre la DCA allemande du débarquement allié du 6 juin 1944 » ;

SN_{1g} = « la DCA allemande » ;

SN_{1d} = « le débarquement allié du 6 juin 1944 » ;

SN_{2d} = « le 6 juin 1944 » ;



Formalisme : C'ATN (Woods, 1970-1980)

Au sommet de la hiérarchie de la classe des grammaires formelles (N. Chomsky) et équivalent à la **Machine de Turing** :

- $A.T.N \supset R.T.N$ (Recursive Transition Network)
- $ATN = RTN +$ **adjonction de Registres dynamiques**

Pouvoir des grammaires transformationnelles :

- Arbre de dérivation et Structure profonde du texte

Robustesse :

- Conditions et Actions au niveau des arcs de l'automate

Évolution : des ATN vers les **C'ATN** (= des ATN en cascade)

Formalisme ATN : machine SIMBAD

<ATN> ::= (<machinename> (accepts <phrasetype>*) <statespec>*)

; un ATN est une liste de nom de machines (automates), une spécification de phrases types qui seront acceptées par l'ATN, et une liste de spécification des états.

<statespec> ::= (<statename> { optional <initialspec> } <arc>*)

; la spécification d'un état dans l'automate consiste à l'attribution d'un nom à cet état et des arcs de transition, et en option s'il s'agit d'un état initial.

<initialspec> ::= (initial <phrasetype>*)

; indique un état initial pour une phrase type indiquée.

<arc> ::= (<phrasetype> <nextstate> <act>*)

; une transition qui consomme une phrase type indiquée.

::= (<pattern> <nextstate> <act>*)

; une transition qui consomme un élément en entrée et qui vérifie un modèle



<pattern>.

::= (J <nextstate> <act>*)

; *une transition vers un nouvel état sans consommer un élément d'entrée.*

::= (POP <phrasetype> <form>)

; indique un état final pour une phrase type indiquée et spécifie une forme à retourner à sa structure.

<nextstate> ::= <statename>

; *spécifie l'état suivant pour une transition.*

<pattern> ::= (<pattern>*)

; vérifie une liste dont les éléments correspondent à la succession de pattern spécifié.

::= <wordlist>

; *correspond à tout élément dans la liste.*

::= ϵ

; *correspond à n'importe quel élément.*

::= --

; *correspond à n'importe quelle sous-séquence.*

::= <form>

; *correspond à une valeur de <form>.*

::= <<classname>>

; *correspond à n'importe quel élément qui a ou hérite les traits de la classe.*

<wordlist> ::= { ' <word> | ' <word>, <wordlist> }

<act> := (transmit <form>)
 ; *transmettre la valeur de <form> comme une sortie.*
 := (setr <registername> <form>)
 ; *fixe le registre à la valeur de <form>.*
 := (addr <registername> <form>)
; *ajouter la valeur de <form> à la fin de liste dans le registre indiqué*
; *(initialement a pour valeur NIL quand le registre n'est pas fixé).*
 := (require <proposition>)
 ; *abandonner le chemin si la proposition est fausse.*
 := (dec <flaglist>)
 ; *fixe la liste des booléens indiqués.*
 := (req <flagproposition>)
 ; *abandonner le chemin si la proposition est fausse.*
 := (once <flag>)
 ; *équivalent à (req (not <flag>)) (dec <flag>).*

<flagproposition> ::= <boolean combination of flag registers>

<proposition> ::= <form>

; la proposition est fausse si la valeur de <form> est NIL.

<form> ::= !<registername>

; retourne les contenus du registre.

::= ' <liststructure>

; retourne une copie de la structure en liste à l'exception de toutes les expressions précédées par ! sont remplacées par leur valeur et celles précédées par @ ont leur valeur insérée comme une sous-liste.

::= !c

; contenus des constituants courants du registre.

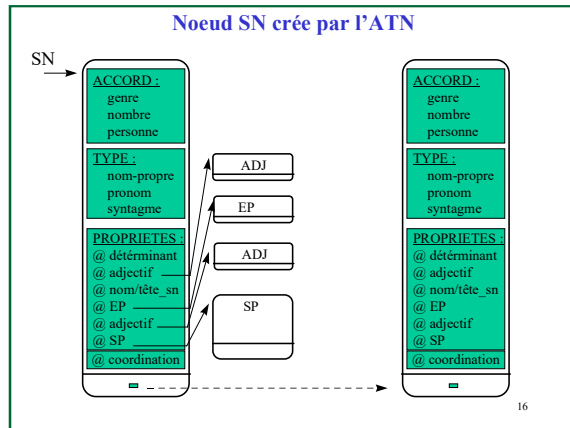
::= !<liststructure>

; retourne la valeur de la structure en liste interprétée comme une expression fonctionnelle.

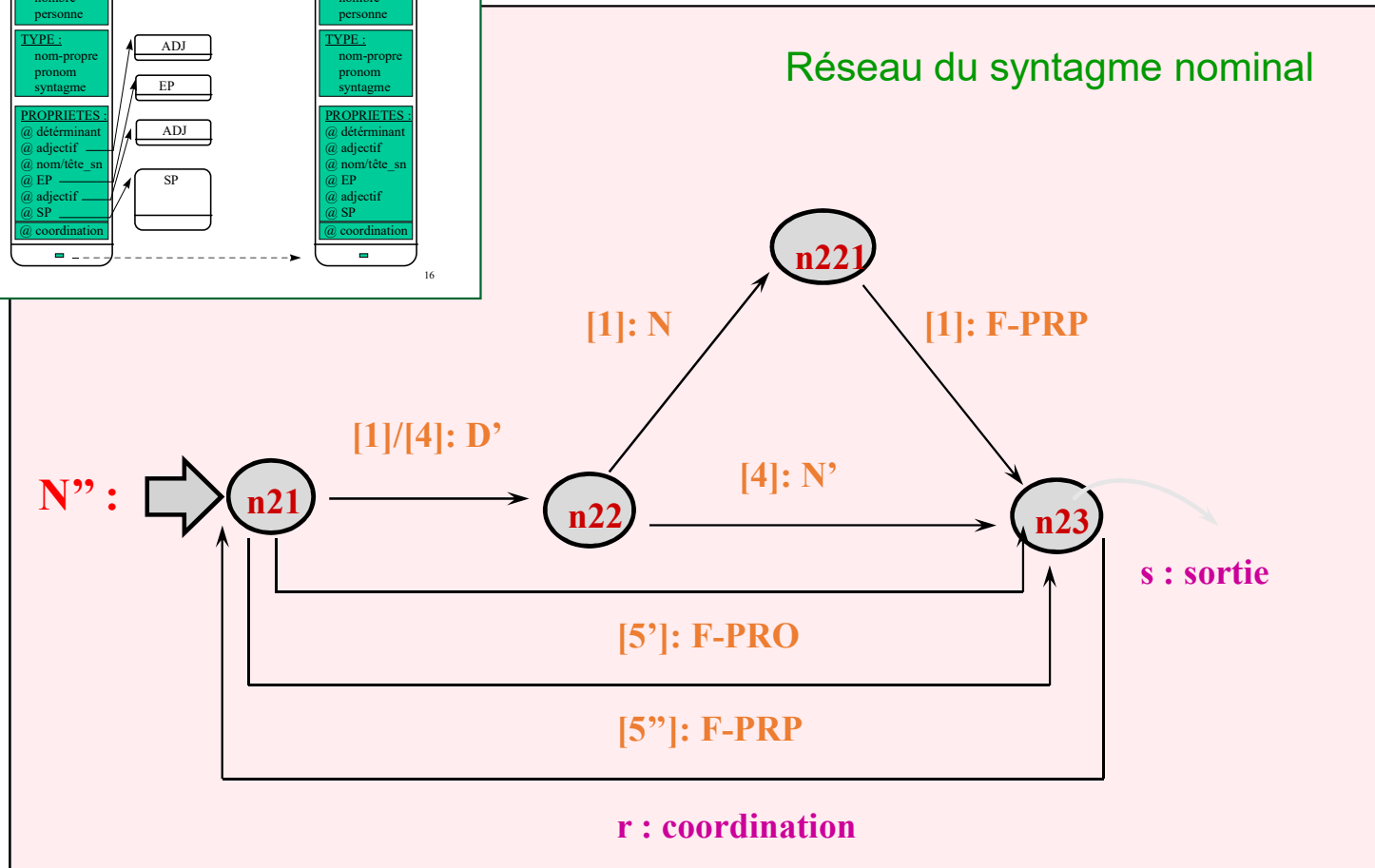
Généralisation :

<C'ATN> ::= (<machinename> (accepts **<ATN>***) <statespec>*)

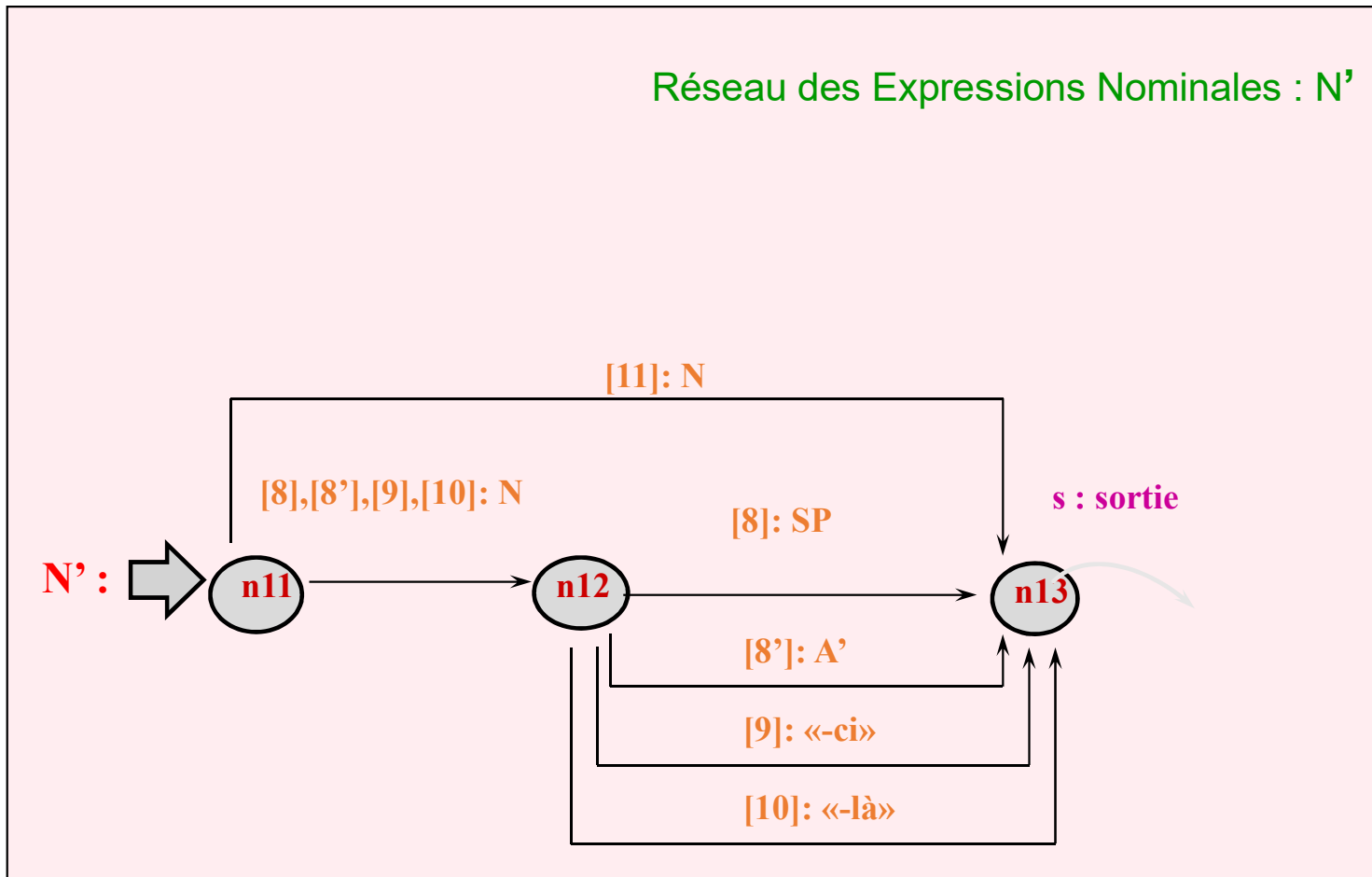
Quelques exemples d'automates :



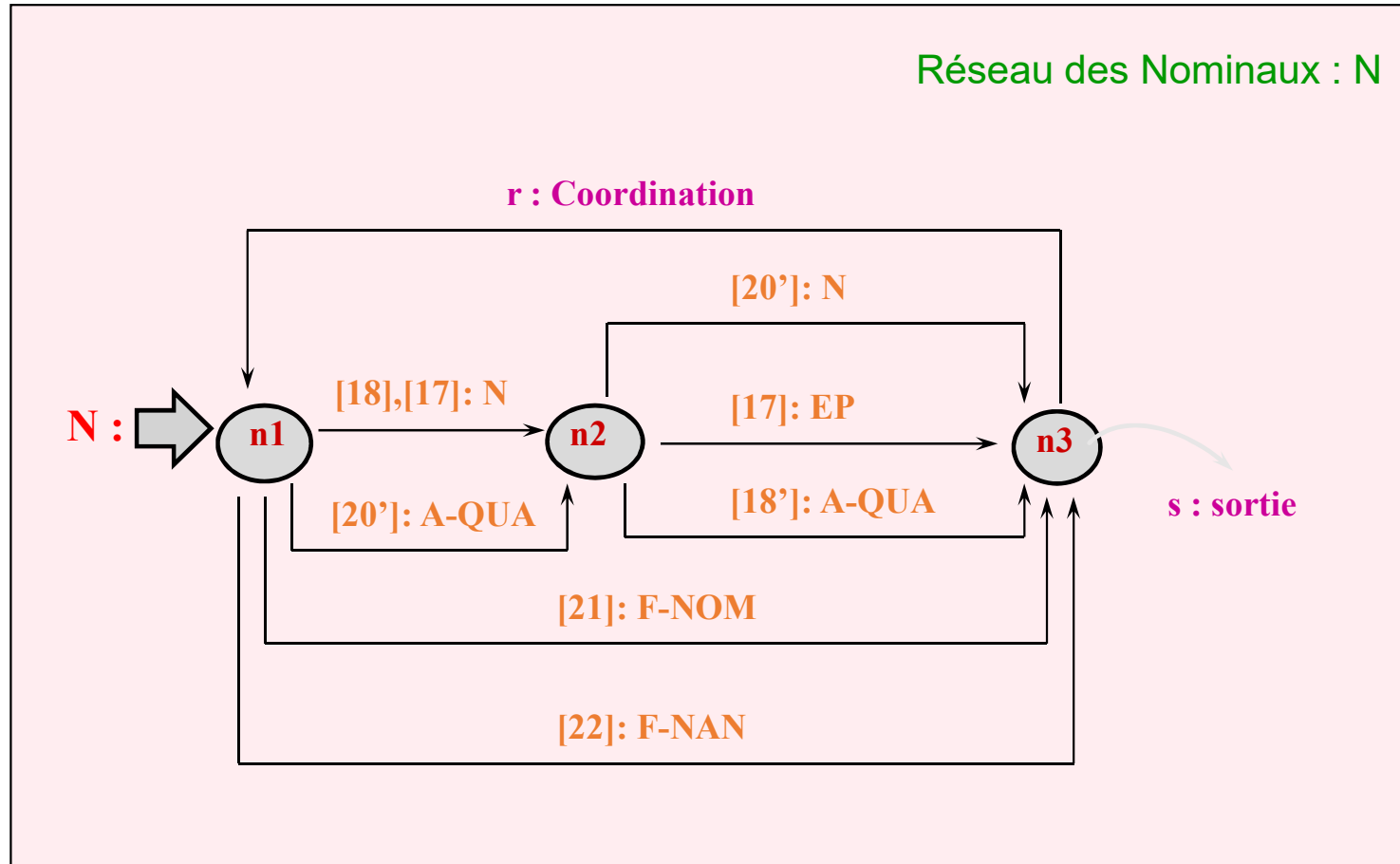
L'AUTOMATE ATN de : SN ou N''

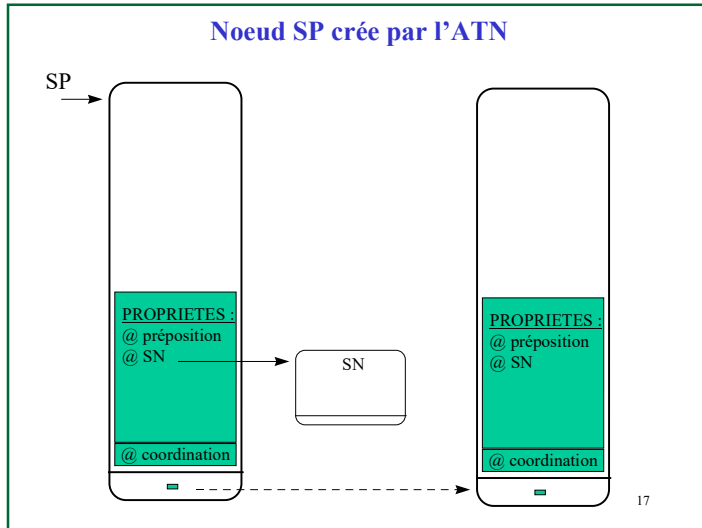


L' AUTOMATE ATN : N'



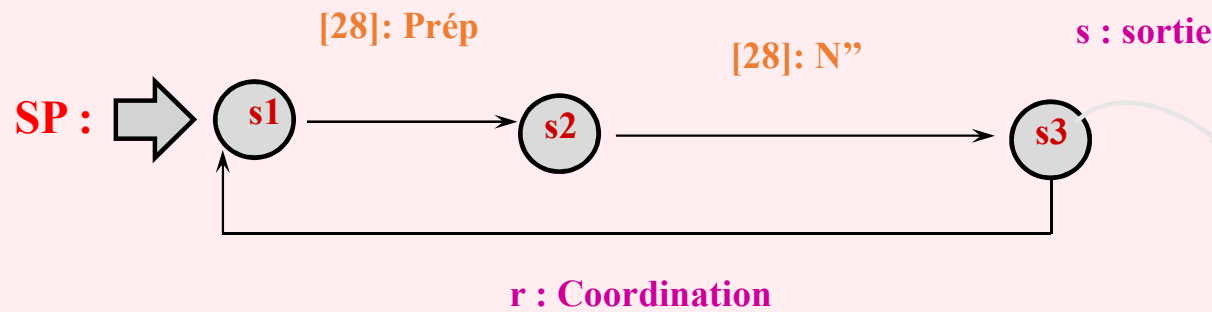
L' AUTOMATE ATN : N





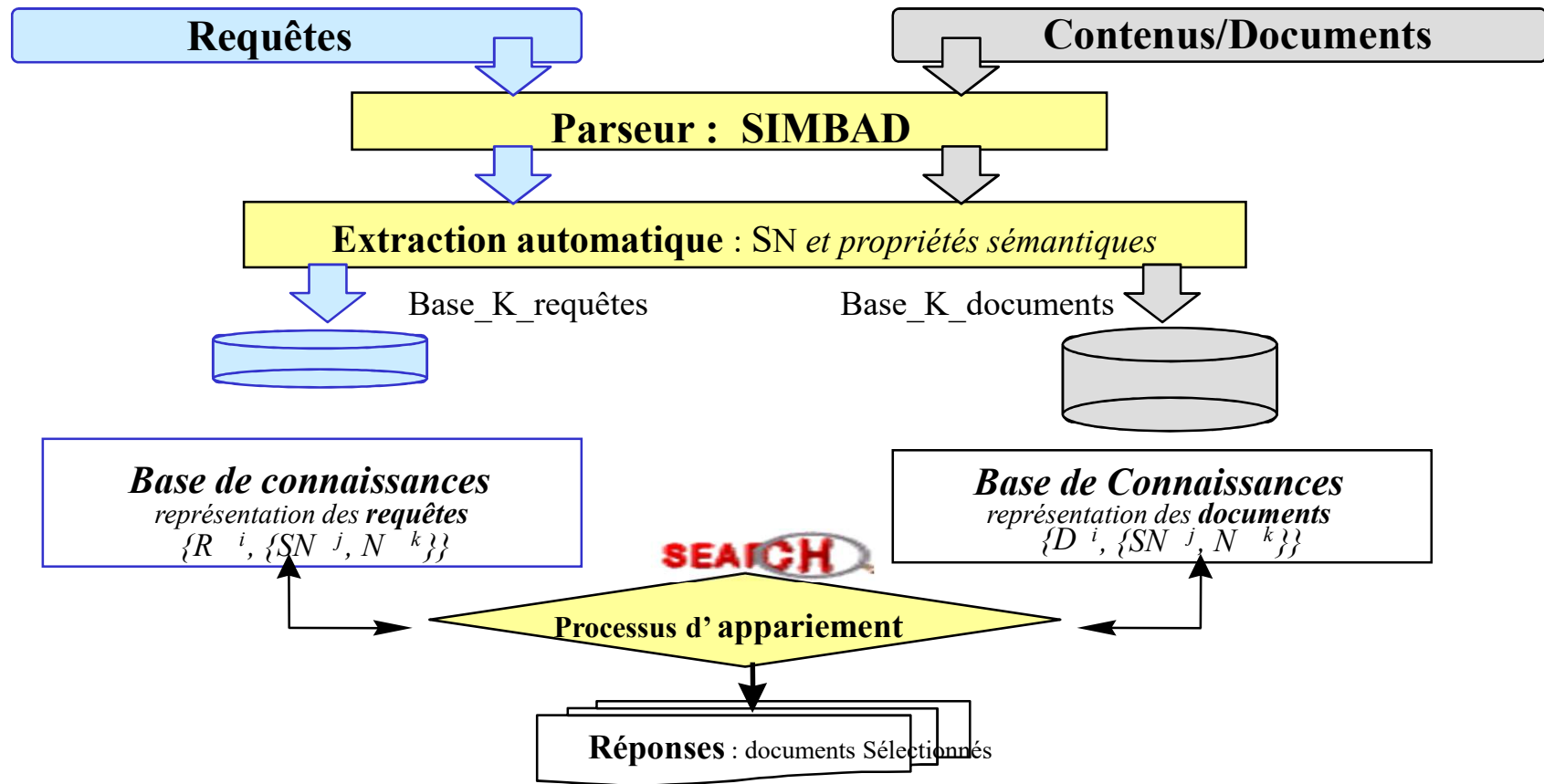
L' AUTOMATE ATN : SP

Réseau du syntagme prépositional



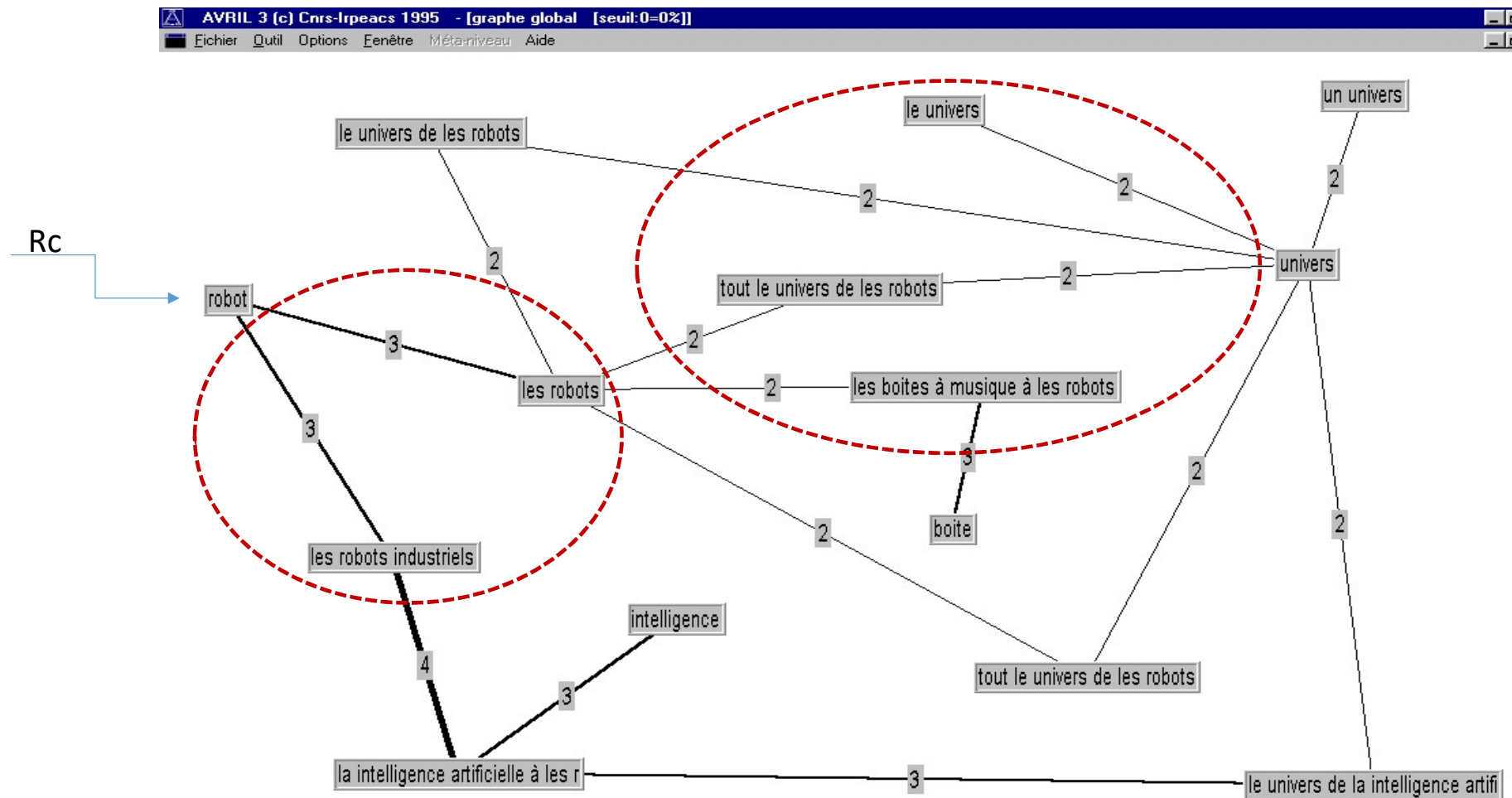
SIMBAD en exécution :

la manipulation des connaissances de type SN et ses propriétés dans un SRI.

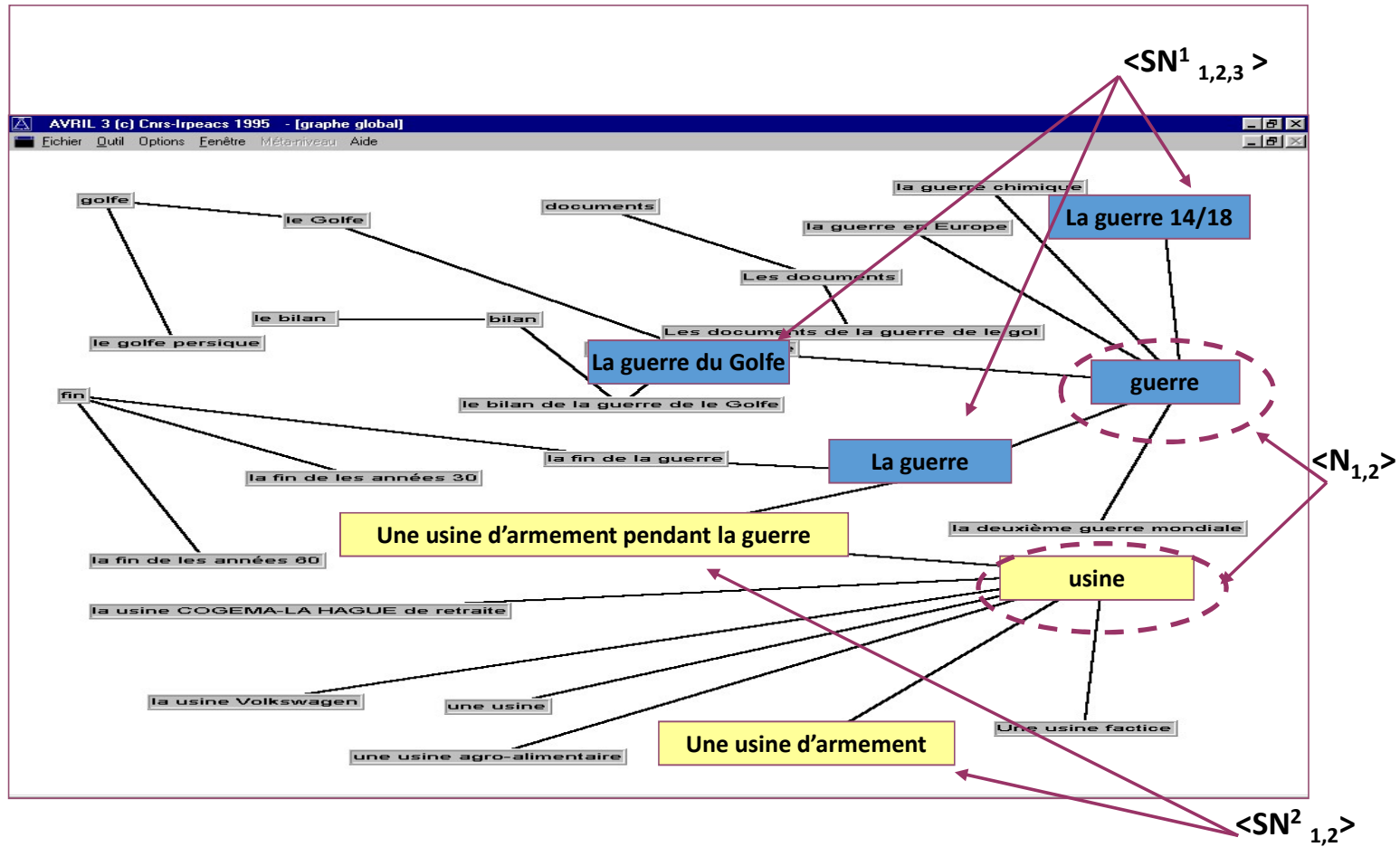


Réseau de connaissances : SN et propriétés

(1/2)



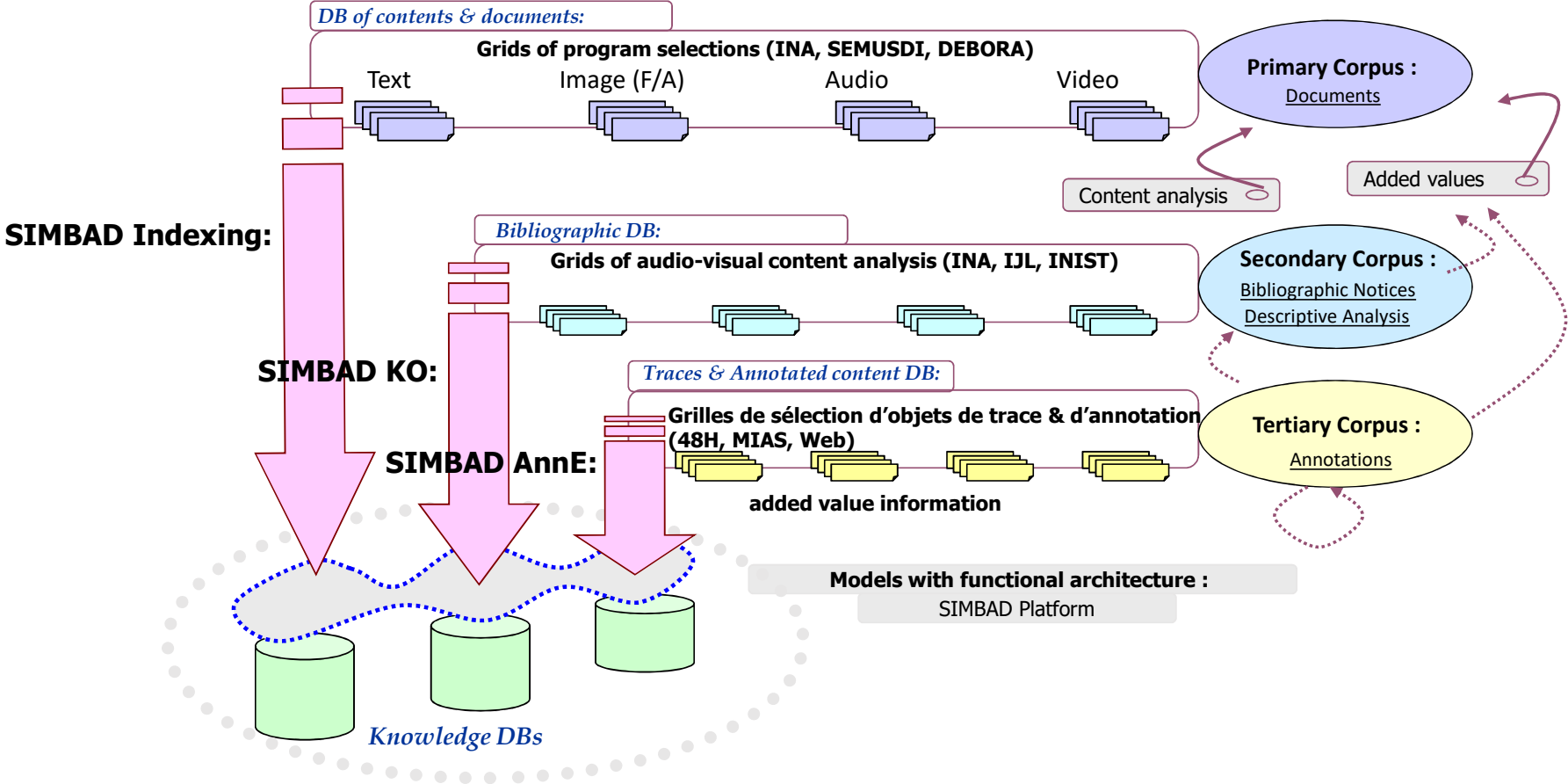
(2/2)



4. Technologie-s évolutive-s :

Contexte d'une recherche réflexive : **Web dynamique, interactif et social**

Evolution du contexte applicatif : veille IST et exploitation des traces dans les Réseaux sociaux.



Nouvelle technique : NooJ©

www.nooj-association.org/downloads.html

Corpus : C'nano (IJL), ChroniSanté (INIST), LRRPT (e-Patrimoine), Web et Réseaux Sociaux

Figure 1: Simple SN graph in NooJ.

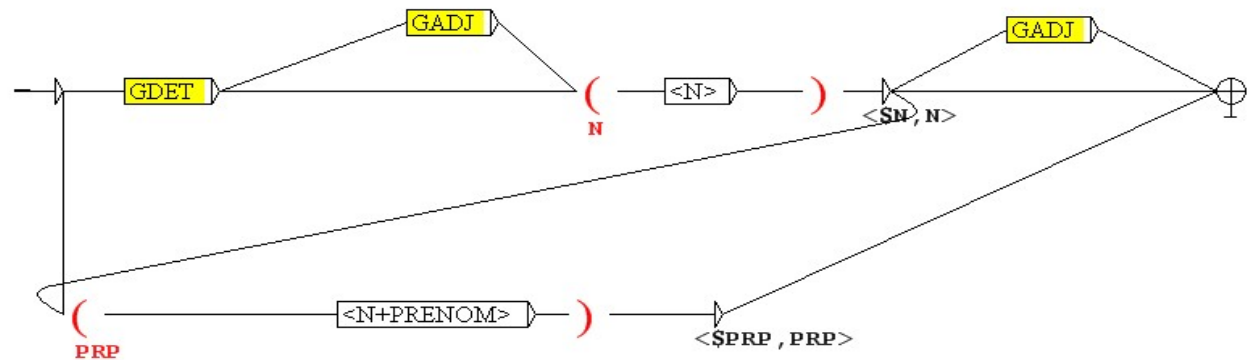
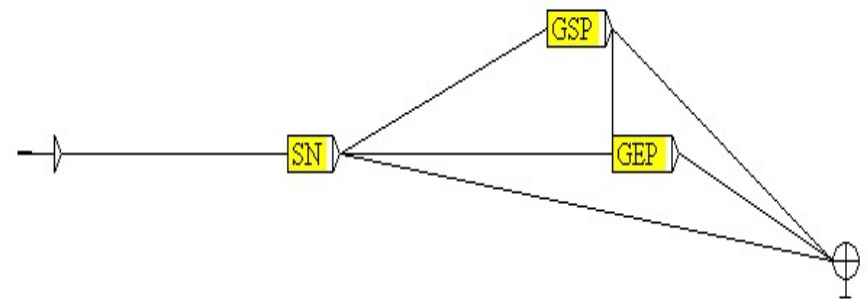


Figure 2: Complex SN graph in NooJ.



Machine NooJ©

The screenshot displays the NooJ software interface with several windows open:

- Corpus cnM_keywords.noc [Modified]:** Shows corpus metadata including file names (e.g., klubNE01 to klubNE15), sizes, and last modification dates. It also lists characters, tokens, digrams, annotations, unknowns, and ambiguities.
- ClubNano_Keywords.nog [Modified]:** Displays a syntactic grammar diagram for French. The diagram shows a root node branching into a left part (starting with a parenthesis '(' and a node labeled '<WF>' with 'keyword' below it) and a right part (starting with a node labeled '<WF>' and a node labeled 'keyword, KEYWORD').
- Locate a pattern in cnM_keywords.noc:** A dialog box for searching patterns. It offers options for a string of characters, a PERL regular expression, a NooJ regular expression, or a NooJ grammar. The 'a NooJ grammar' option is selected, with the path 'C:\Documents and Settings\Philippe\Mes docum...' shown.
- Concordance for Corpus Corpus cnM_keywords.noc [Modified]:** Shows search results for a query. The 'Query' field contains 'keyword'. The results list various text segments with their corresponding keywords, such as 'Caractérisations de postes travail où sont utilisées des nanoparticules/Caractérisations de postes travail' and 'Activités de VEILLE scientifique dans les domaines suivants/Activités de VEILLE scientifique dans les d...'. The 'Query' field shows '100/100' matches.
- Preferences:** A dialog box for configuring the software. It has tabs for 'General', 'Lexical Analysis', and 'Syntactic Analysis'. The 'Lexical Resources for: fr' section shows a list of dictionaries with checkboxes and priority levels (High, Regular, Low). The 'Morphology' section also shows a list of morphological resources with checkboxes and priority levels. Buttons for 'Uncheck All', 'Check All', 'Delete File', 'Import File', 'Edit', 'Save As Default', 'Save...', 'Refresh', 'Apply', 'Reset to Default', 'Load...', and 'Cancel' are visible.

Machine NooJ©

Specific dictionaries

&

Syntactic grammar for SN concepts

The screenshot displays the NooJ software interface with three main windows:

- Corpus cNM_keywords.noc [Modified]:** Shows corpus metadata and a list of text units. The list includes columns for File Name, Size, and Last Modified. The units are labeled from clubLINE45 to clubLINE60.
- clubLINE51.not:** Displays XML text annotation for a specific text unit. The XML structure includes tags for <Entrees>, <Nom>, <Prénom>, <EtablissementDorigine>, and <DomainesActivitesEtabl>. A red box highlights the number '16' in the text: "SCHAEFER TECHNIQUES,EN+ORGANISM+NANO".
- ClubNano_Keywords.nog [Modified]:** Shows a syntactic grammar diagram for French. The diagram features a root node <WF> (in a cloud shape) which branches into a <WF> node and a <keyword, KEYWORD> node. The <WF> node further branches into a <WF> node and a <keyword, KEYWORD> node. The <keyword, KEYWORD> node branches into three <keyword, KEYWORD> nodes. The <keyword, KEYWORD> node branches into three <keyword, KEYWORD> nodes. The <keyword, KEYWORD> node branches into three <keyword, KEYWORD> nodes. The <keyword, KEYWORD> node branches into three <keyword, KEYWORD> nodes.

5. Modèle d'analyse des activités : Exemple du Community Management en nanosciences

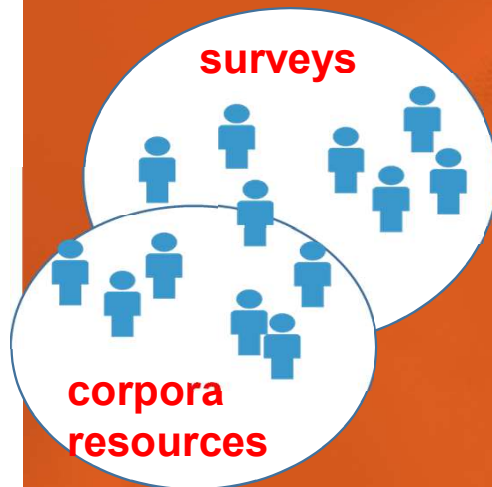
Nouveau monde foisonnant de données de traces numériques, de documents composites et d'informations hétérogènes : la **notion de la Valeur Ajoutée ?**

Projet n°1 : C'nano (pôle de compétitivité)

(Lambert & Sidhom, 2018)

Analysis on surveys & corpora

Club **nano**Métrie



Analysis on :

- Scientific resources (authors by domain)

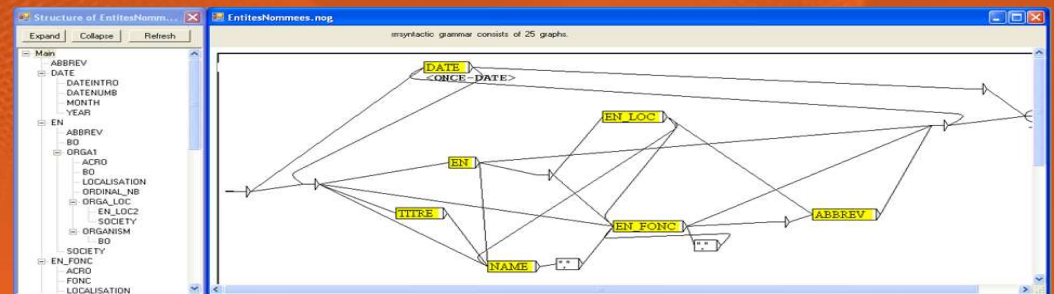
Analysis on :

- Statistics (closed questions)
- Statistics (open questions)



Automatic Language Processing

Projection
(Social Network Analysis)



Extraction de connaissances :

SN et propriétés



Thèmes et
Connaissances (SN)
dans les Ressources :
(Organisations, Projets,
Activités-acteurs)

1	clubLNE71	CEA-Grenoble
2	clubLNE71	Caractérisations de postes travail où sont utilisées des nanoparticules
3	clubLNE01	OMNT (CEA/CNRS)
4	clubLNE01	Activités de VEILLE scientifique dans les domaines suivants
5	clubLNE01	en lien avec les nanomatériaux
6	clubLNE01	détection et caractérisation
7	clubLNE01	mesures d
8	clubLNE01	exposition
9	clubLNE01	toxicologie et ecotoxicologie
10	clubLNE01	évaluation et gestion du risque
11	clubLNE01	aspects normatifs et réglementaires
12	clubLNE02	Université de Provence
13	clubLNE02	Ecotoxicologie
14	clubLNE02	zoologie
15	clubLNE02	immuno-histochimie
16	clubLNE02	invertébrés aquatiques
17	clubLNE02	poisson
18	clubLNE02	nano-écotoxicité
19	clubLNE03	Institut Pasteur
20	clubLNE03	imagerie sub-diffraction
21	clubLNE03	nanoscopie
22	clubLNE03	microscopie STED
23	clubLNE03	instrumentation optique
24	clubLNE03	microscopie de fluorescence
25	clubLNE03	super-résolution
26	clubLNE03	imagerie dynamique du neurone
27	clubLNE03	imagerie du vivant
28	clubLNE04	JLS sarl
29	clubLNE04	optique UV Visible
30	clubLNE04	spectroscopie
31	clubLNE04	ellipsometrie
32	clubLNE04	couches minces
33	clubLNE04	photonique
34	clubLNE04	photo-voltaïque


```

743 </Entrees>
744 <Entrees>
745   <Nom>[redacted] </Nom>
746   <Prénom>[redacted] </Prénom>
747   <EtablissementDorigine>[redacted] </EtablissementDorigine>
748   <DomainesActivitesEtabl>école d'ingénieur dans le domaine de la mécanique, matériaux et management</DomainesActivitesEtabl>
749   <Mots-cles>modélisation atomistique </Mots-cles>
750   <TypeEquipement>codes de calculs DFT: VASP, CASTEP, GAUSSIAN, DMOL
751   code de caculs empirique: DISCOVER, LAMMPS</TypeEquipement>
752   <RaisonInscriptionClub>participer a des travaux collectives permettant de synthétiser ou d'accroître les connaissances dans le domaine de l'utilisation c
753 </Entrees>
754 <Entrees>
755   <Nom>[redacted] S</Nom>
756   <Prénom>[redacted] S</Prénom>
757   <EtablissementDorigine>[redacted] A</EtablissementDorigine>
758   <DomainesActivitesEtabl>MECANIQUE
759   ELECTRICITE</DomainesActivitesEtabl>
760   <Mots-cles>conception mécanique,
761   études électriques,
762   calculs,
763   simulations numériques de phénomènes physiques</Mots-cles>
764   <RaisonInscriptionClub>NOUS SOMMES AMENES A REALISER DES ETUDES ET DES CONCEPTIONS EN MICRO ET NANO MECANIQUE</RaisonInscriptionClub>
765 </Entrees>
766 <Entrees>
767   <Nom>[redacted] r</Nom>
768   <Prénom>[redacted] </Prénom>
769   <EtablissementDorigine>[redacted] fi</EtablissementDorigine>
770   <DomainesActivitesEtabl>AERONAUTIQUE, AUTOMOBILE, NUCLÉAIRE, BIOTECHNOLOGIE, </DomainesActivitesEtabl>
771   <Mots-cles>Mesures manométriques sans contact capacitives et optiques de déplacement e tde positions</Mots-cles>
772   <TypeEquipement>CAPTEURS ET SYSTEMES DANS LE DOMAINE DES MESURES DE DÉPLACEMENT ET DE DISTANCES MANOMÉTRIQUES</TypeEquipement>
773   <QuelleExpertiseMetrologie>MESURE DE NANODEPLACEMENTS</QuelleExpertiseMetrologie>
774   <QuelleExpertiseNanometro>MESURE DE NANODEPLACEMENTS</QuelleExpertiseNanometro>
775   <RaisonInscriptionClub>DÉVELOPPER DE NOUVELLES ACTIVITÉS</RaisonInscriptionClub>
776   <GroupequestionsApproGT3>DÉVELOPPER DE NOUVEAU CONCEPT DE MESURE PAR FUSION DE CAPTEURS</GroupequestionsApproGT3>
777 </Entrees>
778 <Entrees>
779   <Nom>[redacted] </Nom>
780   <Prénom>[redacted] </Prénom>
781   <EtablissementDorigine>[redacted] </EtablissementDorigine>
782   <DomainesActivitesEtabl>Nanotechnologie</DomainesActivitesEtabl>

```

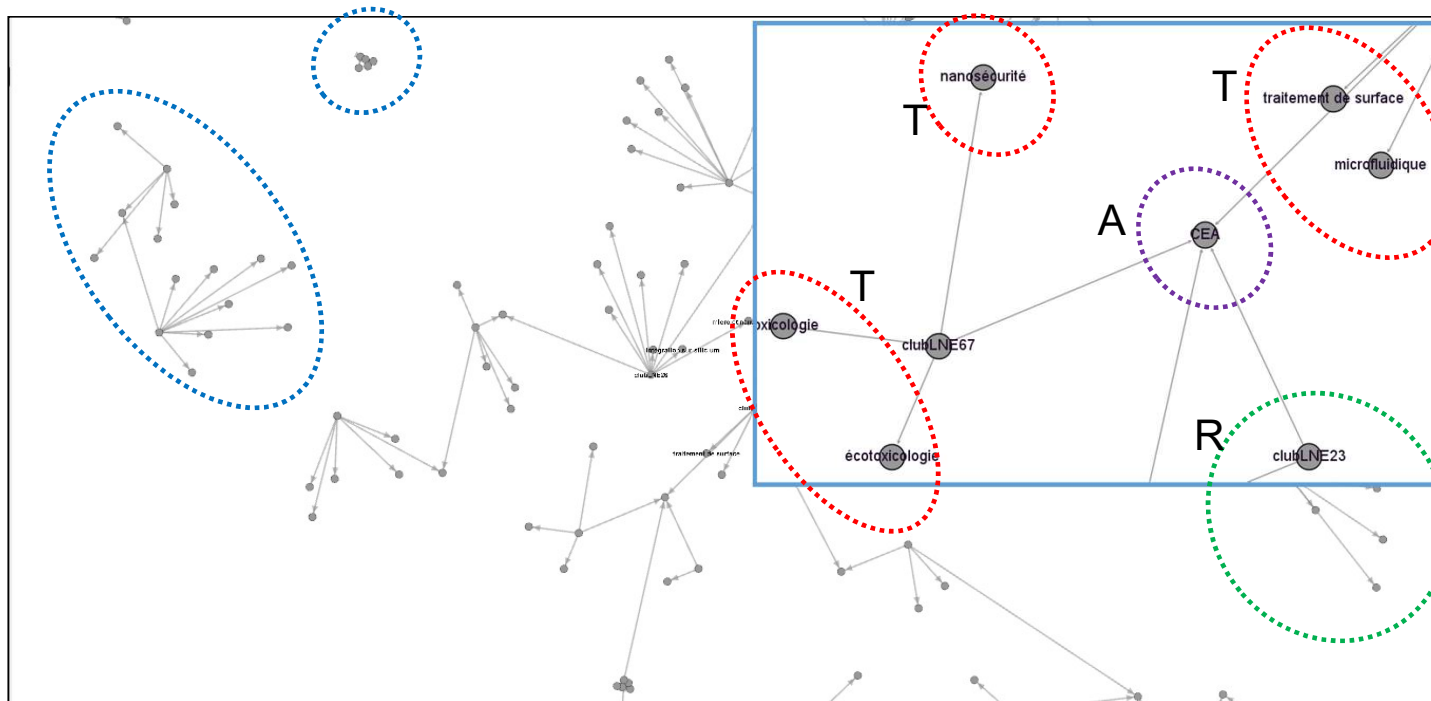
Pôle de compétitivité C'nano :

Nanometrology Cluster



Identification of **weak signals**: long-term “strategic” information in C'nano

Diagnostics in continuity: K in Data and Structures + Resources



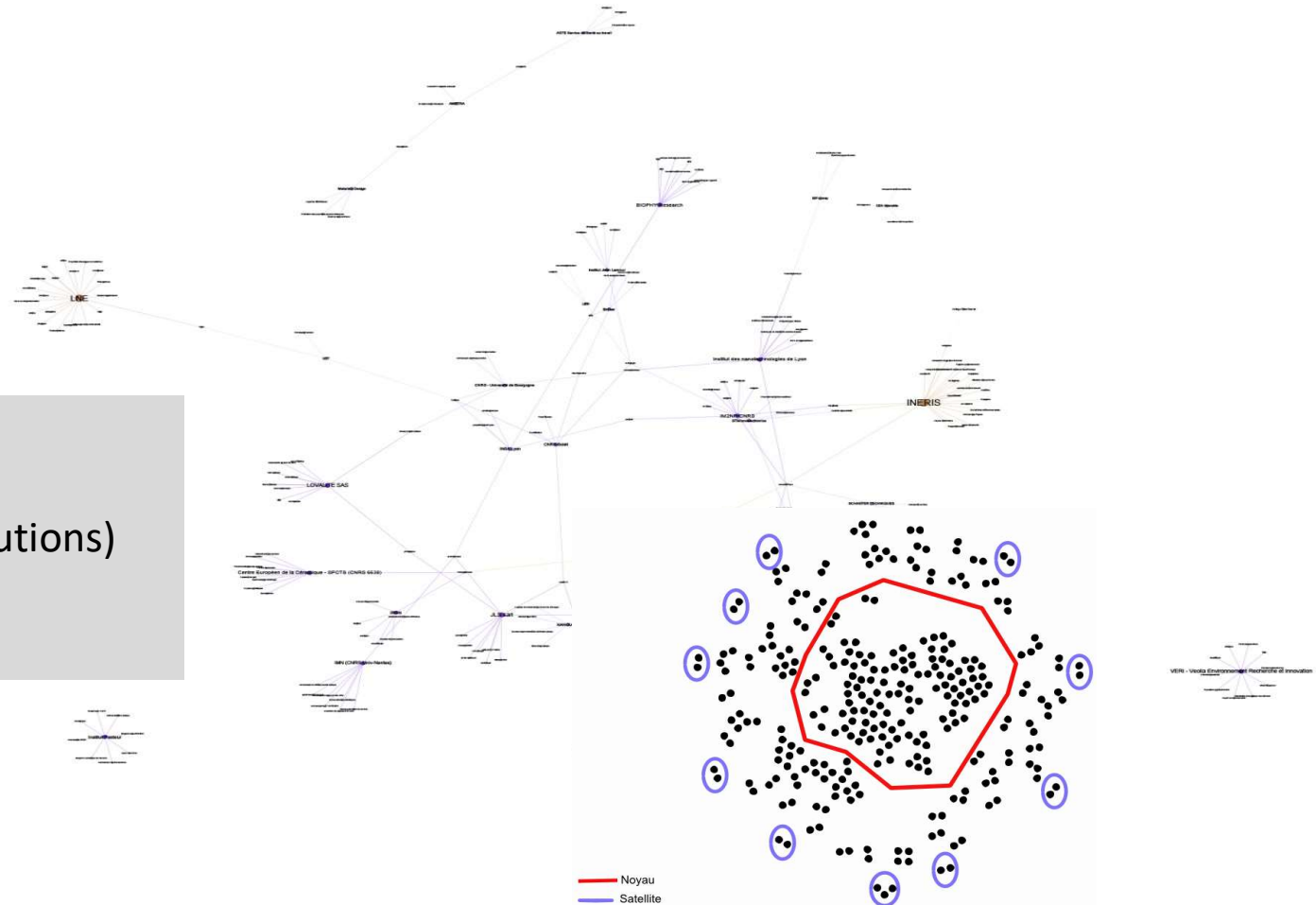
Détection de nouveaux concepts en nanosciences

Processus :

1. Cartographie conceptuelle {N,SN} par des analyse TALN
2. Veille IST
3. Aide à la décision sur les projets

Principaux résultats :

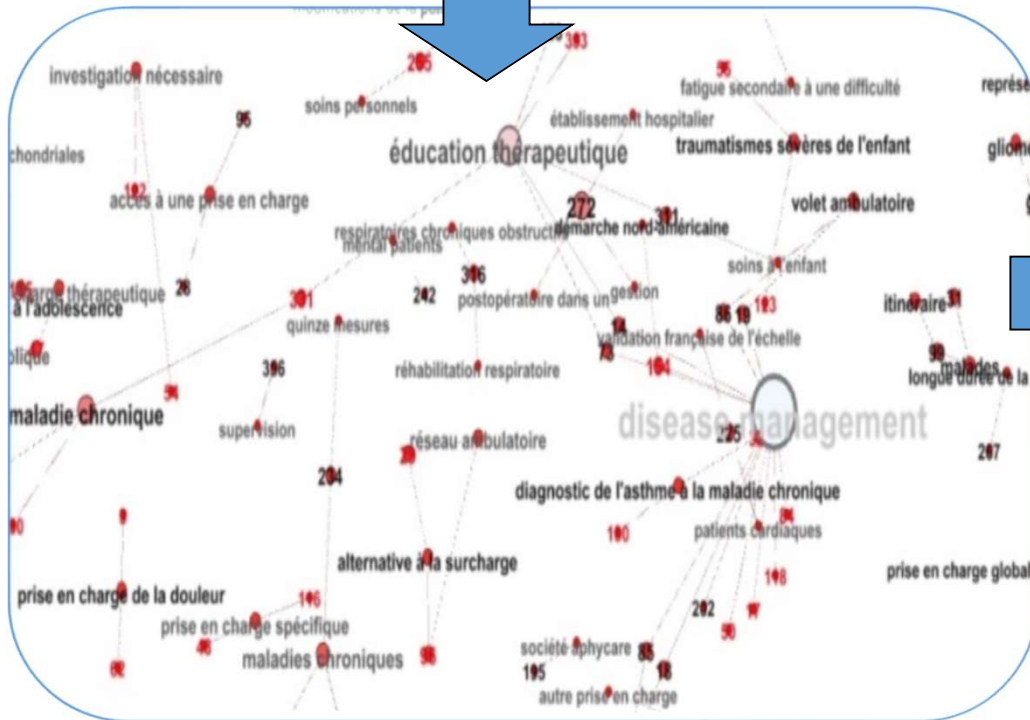
1. Détection des **signaux faibles**
2. Diagnostics (acteurs, projets, institutions)
3. Identification des communautés (compétences)



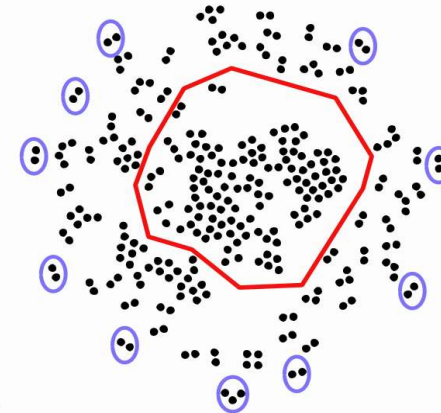
Processus :



(N,SN)
Identifiés et propriétés



(SN) et (n)
et relations sémantiques : Rc, Ra, Re



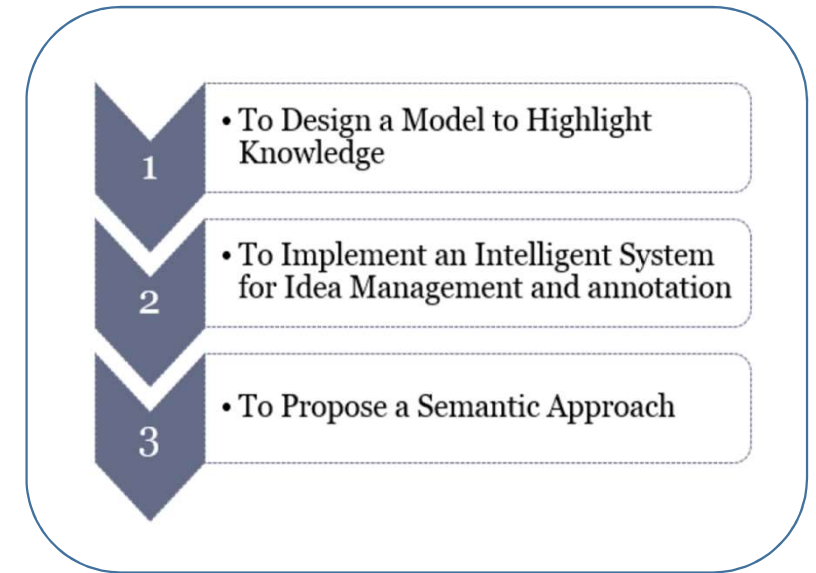
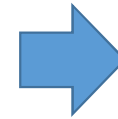
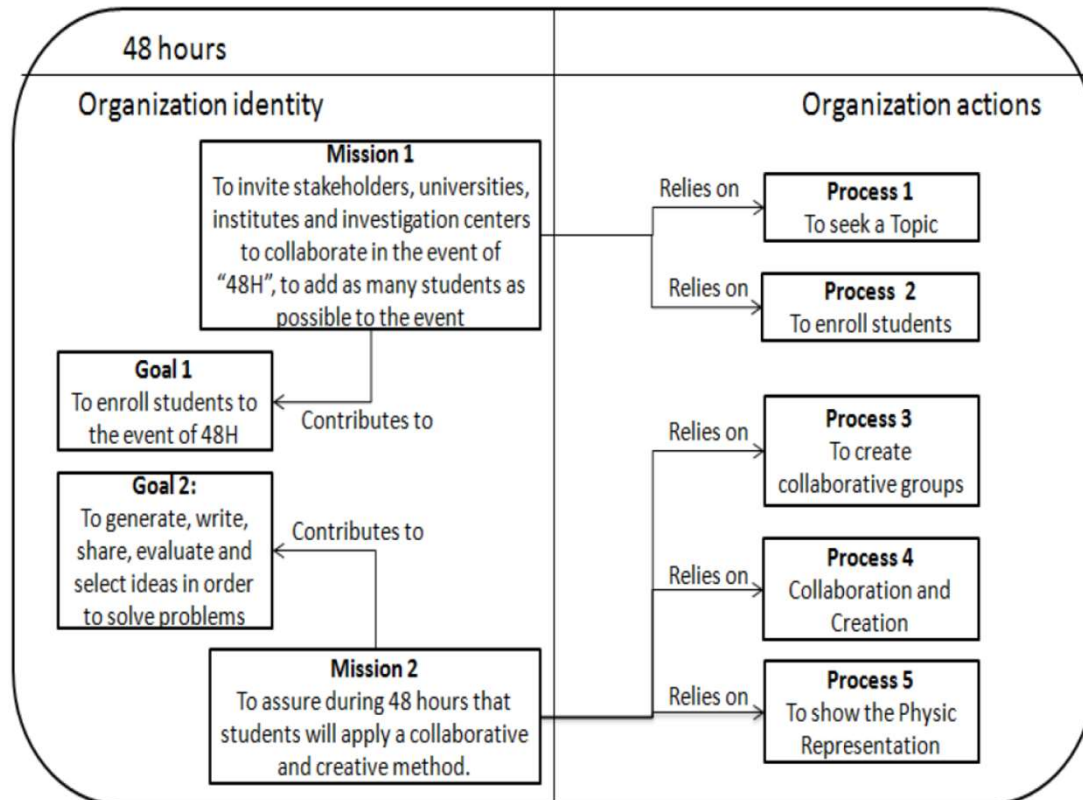
détection des signaux faibles

6. Poly-représentation Système-Utilisateur ?

Intelligence Humaine : Traces et Traçabilité Numérique en **poly-morphisme**

Projet n°2 : Ateliers 48H de créativité

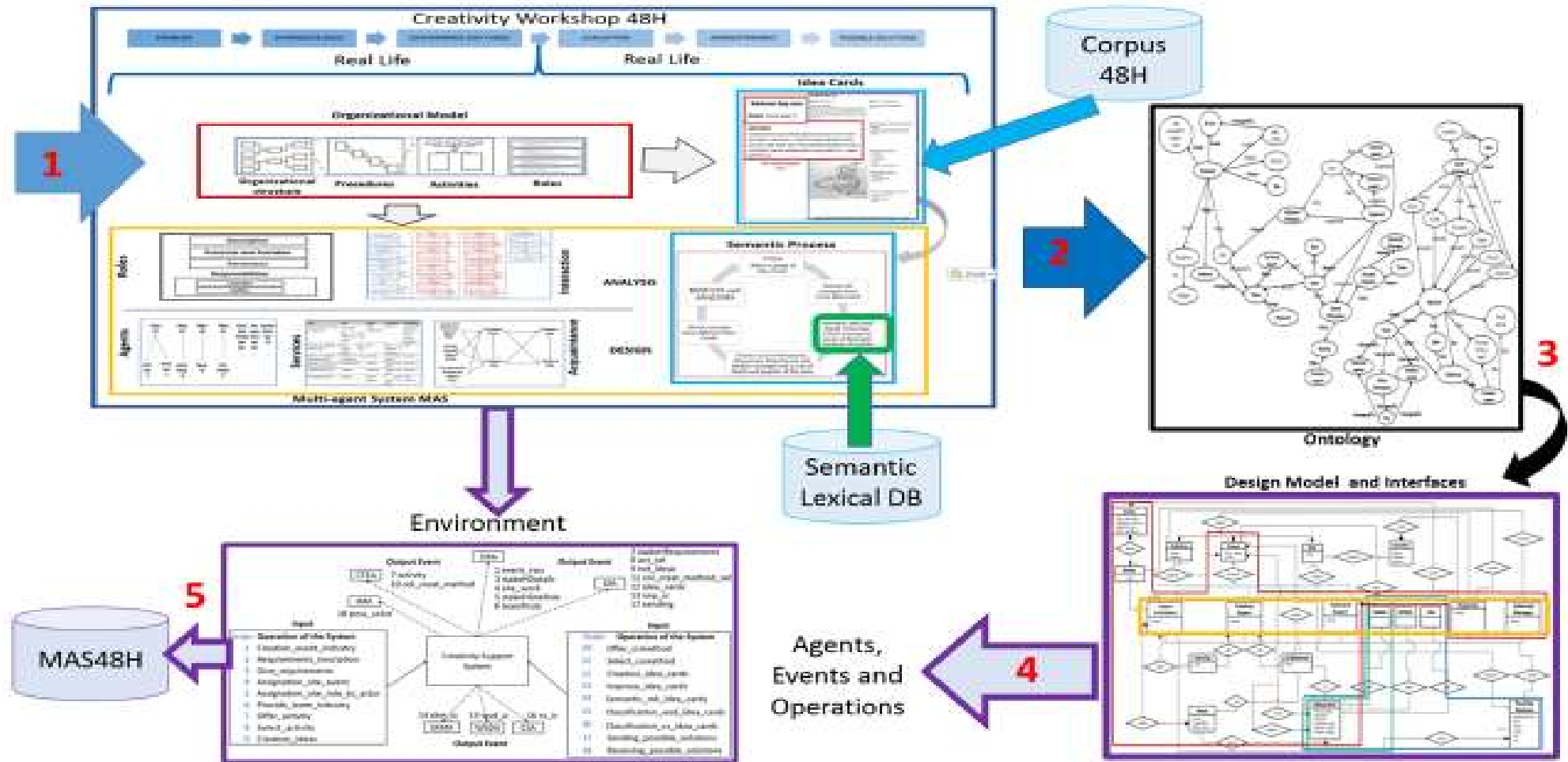
(Barrios, Monticolo & Sidhom, 2019)



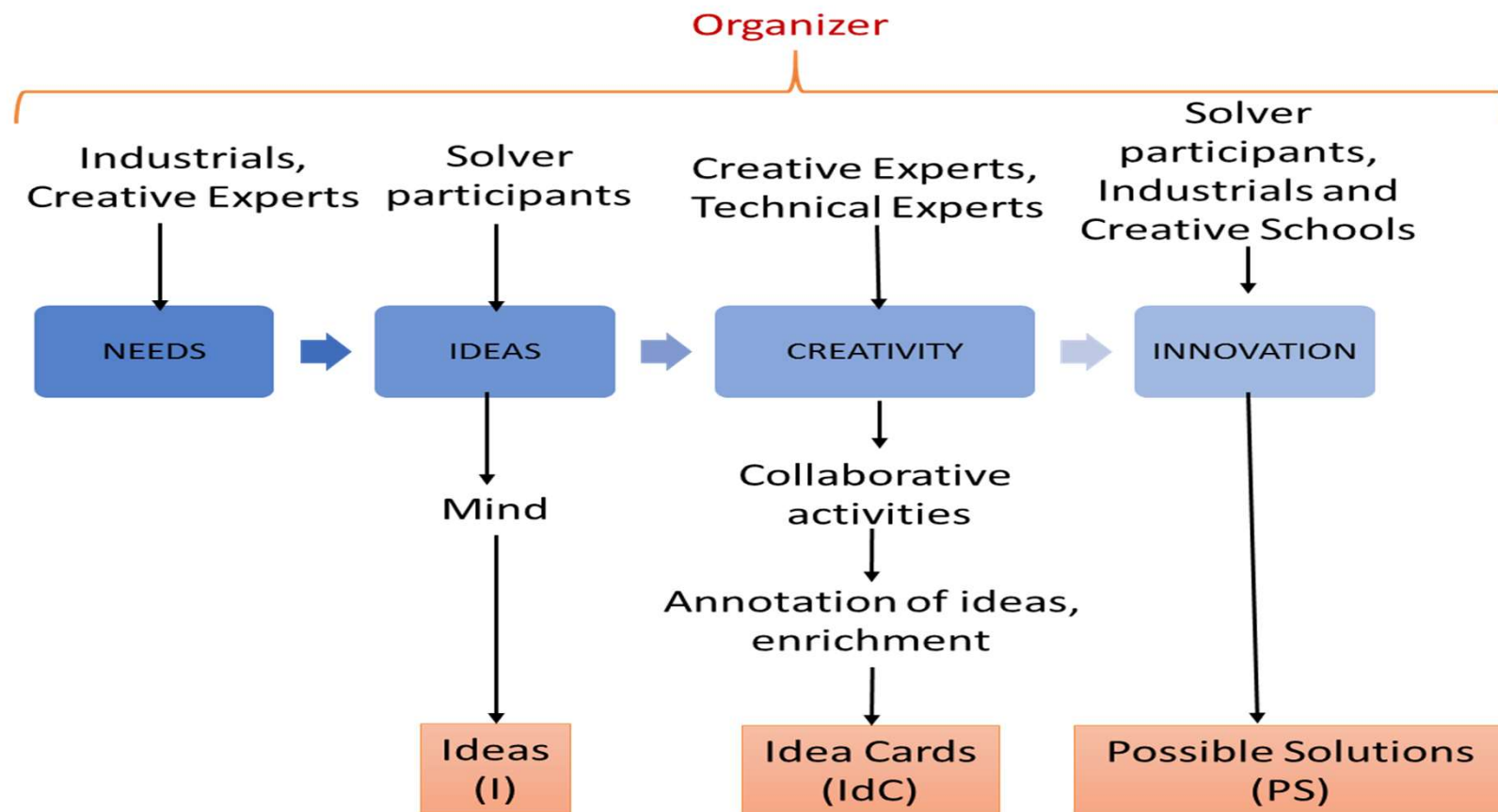
Processus de créativité
selon (Barrios, Moticolo & Sidhom, 2019)

Structure organisationnelle des ateliers 48H de créativité.

Management de l'innovation : 48H de créativité

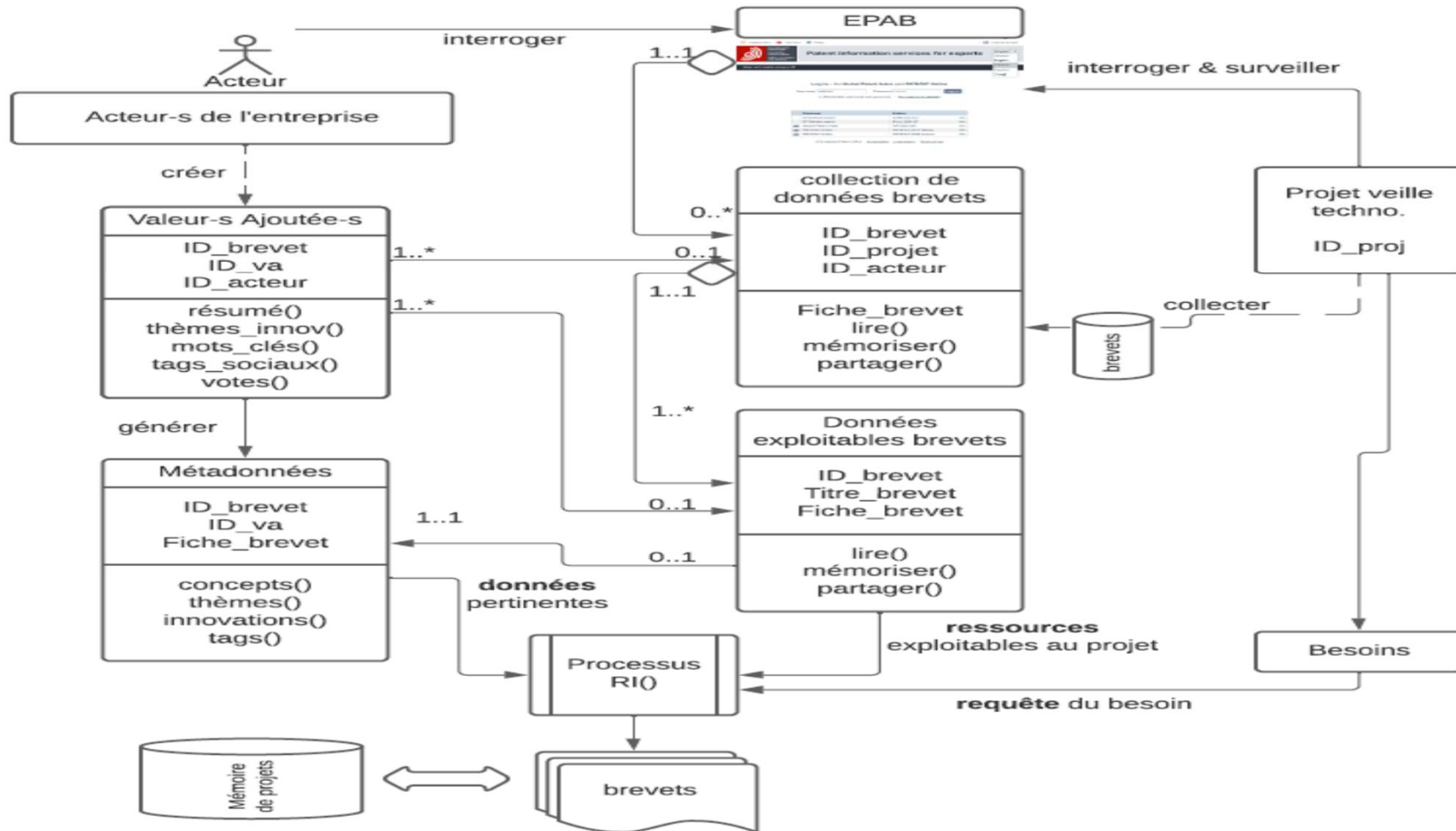


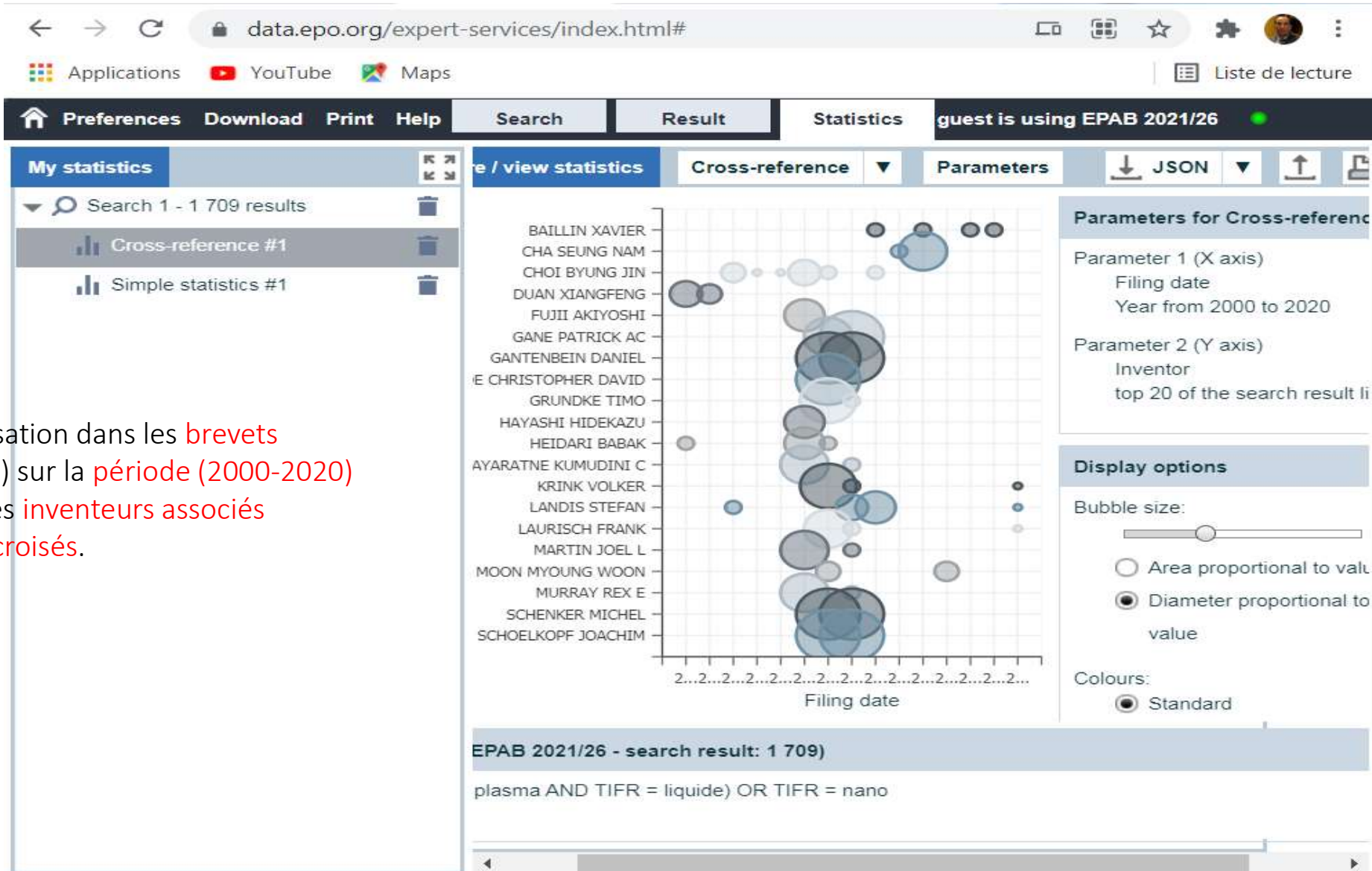
Processus de créativité et son organisation dans les ateliers 48H :



Projet n°3 : brevets en veille technologique.

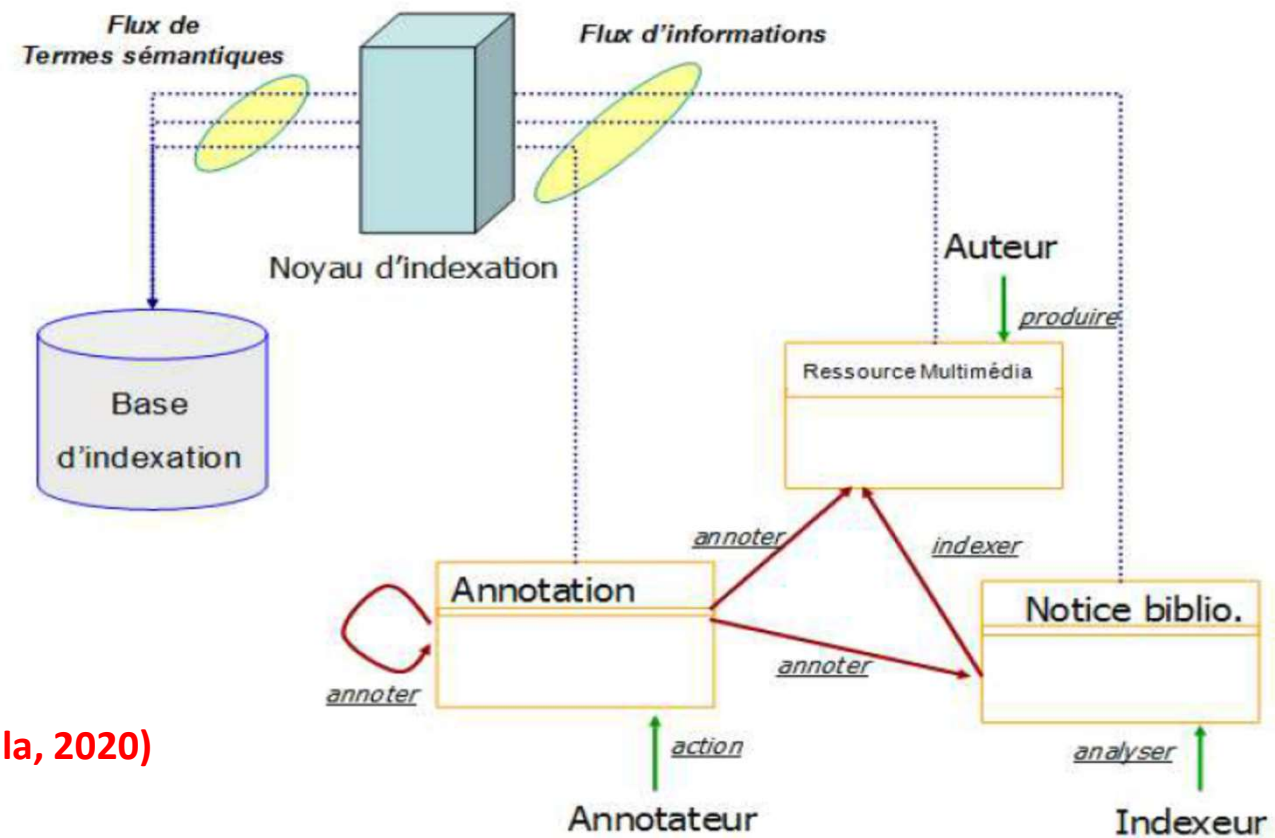
(Henri, Sidhom & Saleh, 2019)





Filtres de visualisation dans les brevets EPAB (EPO, 2017) sur la période (2000-2020) pour identifier les inventeurs associés sur des thèmes croisés.

Projet n°4 : valorisation informationnelle dans les ressources hétérogènes



(Khémiri, Sidhom & Ben Ghézala, 2020)

Projet n°6 : convergence des BigData et IA - vers l'automatisation des prises de décisions

Definition Random variables X, Y, Z are said to form a Markov chain in that order (denoted by $X \rightarrow Y \rightarrow Z$) if the conditional distribution of Z depends only on Y and is conditionally independent of X . Specifically, X, Y , and Z form a Markov chain $X \rightarrow Y \rightarrow Z$ if the joint probability mass function can be written as

$$p(x, y, z) = p(x)p(y|x)p(z|y).$$

Theorem 2.8.1 (Data-processing inequality) If $X \rightarrow Y \rightarrow Z$, then $I(X; Y) \geq I(X; Z)$.

Proof: By the chain rule, we can expand mutual information in two different ways:

$$I(X; Y, Z) = I(X; Z) + I(X; Y|Z) \quad (2.119)$$

$$= I(X; Y) + I(X; Z|Y). \quad (2.120)$$

Since X and Z are conditionally independent given Y , we have $I(X; Z|Y) = 0$. Since $I(X; Y|Z) \geq 0$, we have

$$I(X; Y) \geq I(X; Z). \quad (2.121)$$

Constat : « *intelligence de l'information (i^2)* »

- Dans les workflows de la Data Intelligence, l'espace de données contient beaucoup plus de variations en termes d'entropie que l'espace de décision, en conséquence : la perte d'information n'est pas seulement inévitable mais peut s'avérer significative (Chen & Golan, 2016).

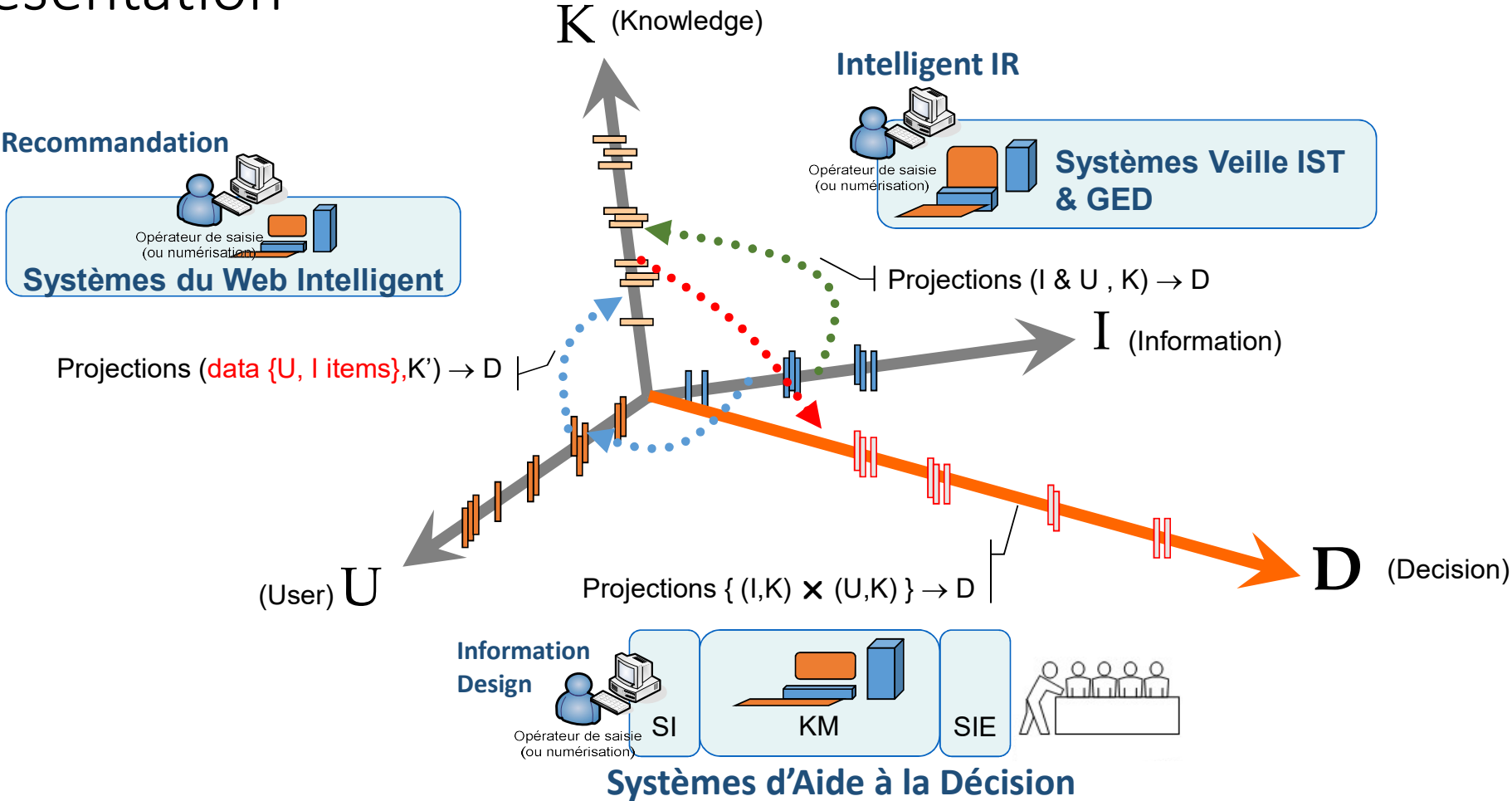
Chen, M. & Golan, A. (2016). What May Visualization Processes Optimize?. In IEEE Transactions on Visualization and Computer Graphics, 12 (2016), 2619–2632.

Conclusion :

En démonstration :

1. **Processus d'Analyse** : les propriétés du langage dans l'analyse automatique des ressources ;
2. **Usage des Connaissances Organisées** : la sémantique des propriétés syntagmatiques et des représentations dans les données pour fonder l'organisation des connaissances ;
3. **Expériences sur les Corpus et Usages (Traces et Traçabilités numériques)** : pour observer les (contenus, activités, profils), pour construire des recommandations par la détection des signaux faibles et servir la décision.

Dispositifs info-communicationnels : poly- morphisme et représentation



Merci pour votre attention.