

# Informational Validations for Knowledge Organization in Community Manager: **case study context in nano-sciences and technologies**

---

**Sahbi SIDHOM** (LORIA & University of Lorraine, France)  
& **Philippe LAMBERT** (C'NanoGrand Est & IJL/University of Lorraine, France)

eMails: **sahbi.sidhom@loria.fr, philippe.lambert@univ-lorraine.fr**

# Outline

Concepts et Validations:

1. Cognitive Observations
2. Adaptabilities
3. Informational Analysis & Validations

Approach:

4. Knowledge Organization (KO)

Applications:

5. *Case study in context of Nanometrology cluster*

# I.

Concepts et Validations:

- 1. Cognitive Observations**
- 2. Adaptabilities**
- 3. Informational Analysis & Validations**

## Cognitive Observations:

### Grille d'Analyse (1:4)

Indexation audiovisuelle à l'I.N.A.

Étude ORTF & SEMA (1970) :  
Méthode d'analyse des films d'actualités



*Préparation de l'indexation*



*Gain du temps d'analyse de l'audiovisuel*

Grille d'Analyse de l'audiovisuel (+1990) :



- Position des plans & Description plan par plan
- Mouvements de caméra
- Personnes identifiées & Lieux de l'action
- Distinction entre Image & Son

# Grille d'Analyse (2:4)

## Guide de l'indexation audiovisuelle à l'I.N.A.

### Exemple d'une GRILLE d'analyse :

NATURE DE PRODUCTION : PRODUCTION PROPRE (CODE 01)

descripteurs principaux :

DE Descripteurs thématiques et géographiques, personnes morales, personnes physiques évoquées.

descripteurs secondaires :

DES Descripteurs séquences et images.  
Nom des villes si images réutilisables.  
Nom des personnes si visibles à l'image.



**résumé court :**

RES / chapeau Chapeau précisant la forme du sujet et situant l'événement dans le temps et dans l'espace



**résumé développé :**

SEQ / résumé Description par séquence du sujet.

# Grille d'Analyse (4:4)

## Pratiques à l'I.N.A.

### Exemple dans la pratique :

Titre propre : Un lac venu de l'espace : le cratère du Nouveau Québec

Titre collection: France 2 Documentaire

**Descripteurs principaux:** météorite; lac (Nouveau Québec); Québec; expédition (scientifique); chercheur

Producteurs (aff) : FRANCE 2, 1995

**Nature de production (aff) :** PRODUCTION PROPRE



**Chapeau :** Ce documentaire retrace les travaux menés par une équipe de chercheurs dans le nouveau Québec, afin d'expliquer la présence d'un lac qui se serait formé suite à la chute d'une météorite.



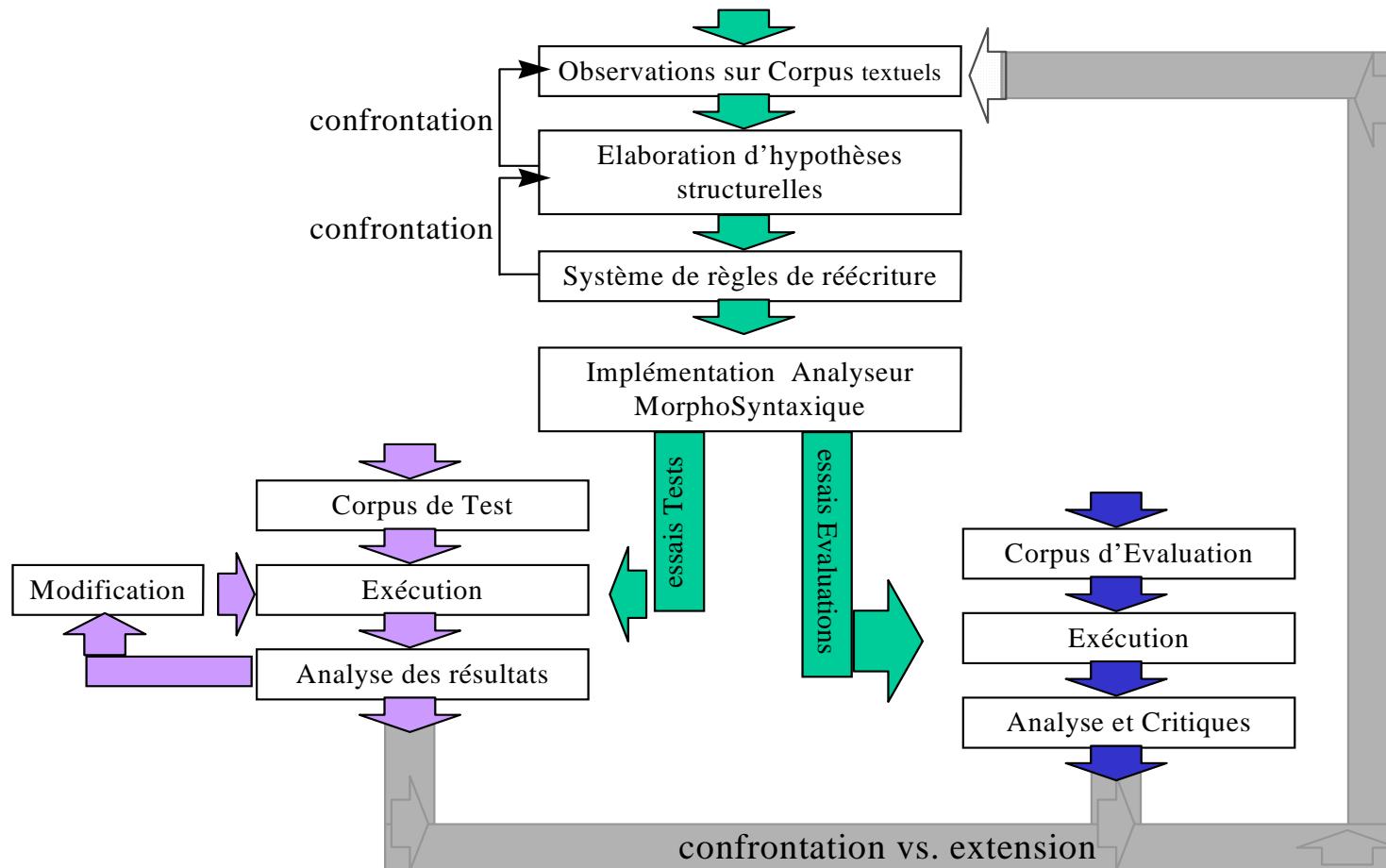
**Résumé :**

La chute d'une **météorite** venue de l'espace a créé un **lac** dans la Toundra du **Nouveau Québec**. Celui-ci mesure 2,7 km de diamètre, 267 mètres de profondeur et son cratère s'étend sur 3 km.

Une équipe multidisciplinaire de **chercheurs**, dirigée par le professeur Michel BOUCHARD, de l'université de Montréal, a monté une **expédition** sur le site afin de répondre aux questions essentielles que se pose le monde scientifique sur l'origine extra-terrestre de la météore et la datation de son impact. Le résultat de leurs études concernant l'importance de l'impact, démontre que sa puissance est équivalente à 8500 fois celle que détenait la bombe d'Hiroshima. Pour établir de façon certaine l'origine extra-terrestre de la météorite, les chercheurs tentent de retrouver des fragments d'impactible, ce qui est un succès.

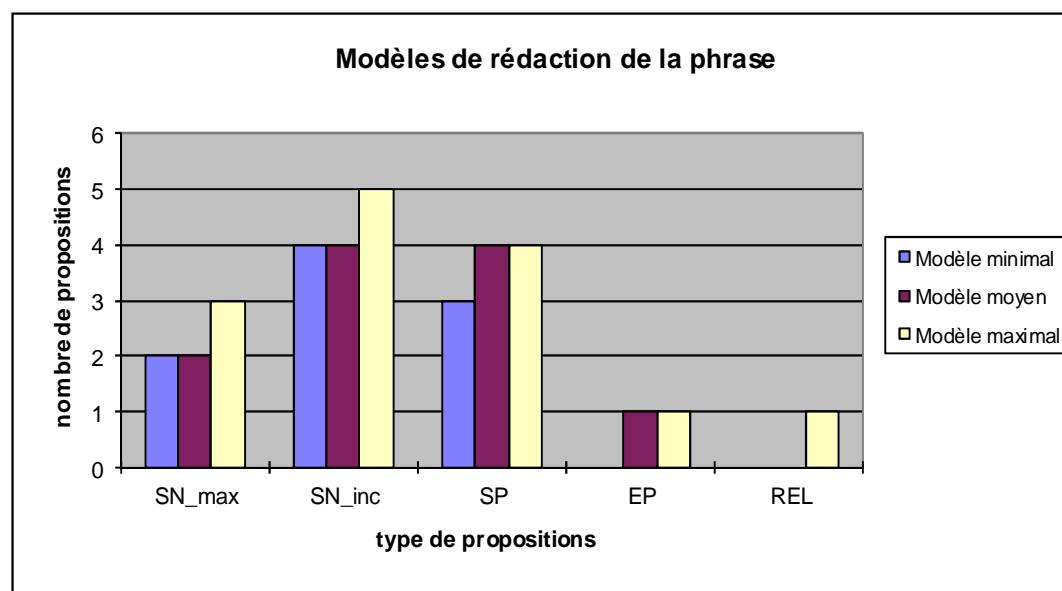
...

# Informational Analysis & Validations:



Objets étudiés	SN_max	SN_inc	SP	EP	REL
chapeau	2.56	4.30	4.01	0.38	0.37
résumé	2.05	4.37	3.35	0.66	0.46
moyenne	2.30	4.33	3.68	0.52	0.41

MODELE	SN_max	SN_inc	SP	EP	REL
Modèle minimal	2	4	3	0	0
Modèle maximal	3	5	4	1	1
Modèle intermédiaire	2	4	4	1	0

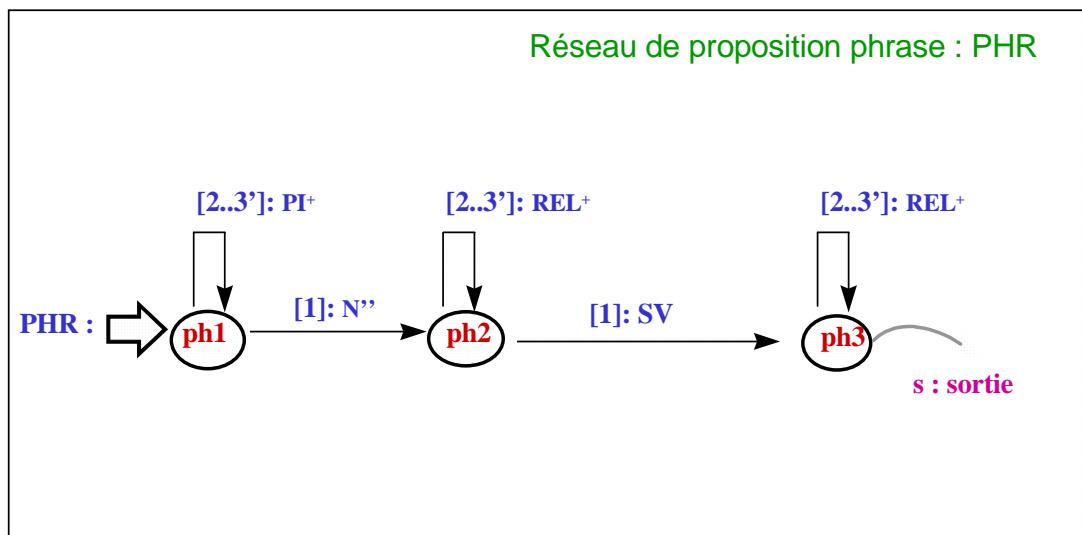


## Adaptabilities:

In INA context, the S syntactic structure is:

$$PHR \rightarrow [PI]^+ SN + [REL_{SN}]^+ SV + [REL_{SV}]$$

[x]: élément facultatif



In Nanometrology context, the S syntactic structure is:

$$PHR \rightarrow [V \inf]^+ SN + [REL_{SN}]$$

## II.

Approach:

### 4. Knowledge Organization (KO)

# Knowledge Management (KM)?

Knowledge Management is the name of a **concept**

in which an enterprise

**consciously and comprehensively** (= process)

**gathers,**

**organizes,**

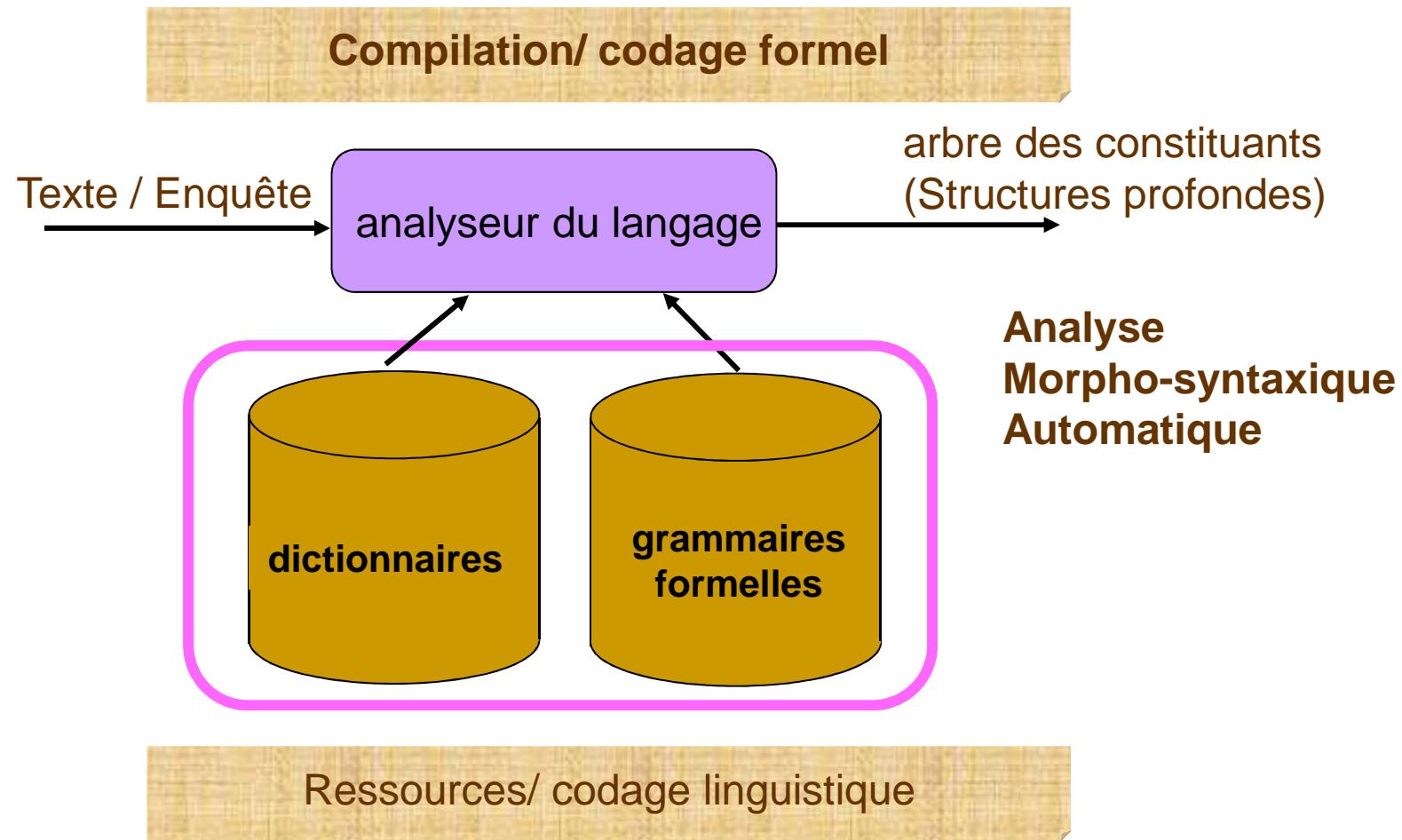
**shares,** and

**analyzes**

its **knowledge** in terms of **resources, documents, and people skills.**

(Jeff Angus and Jeetu Patel, 1998)

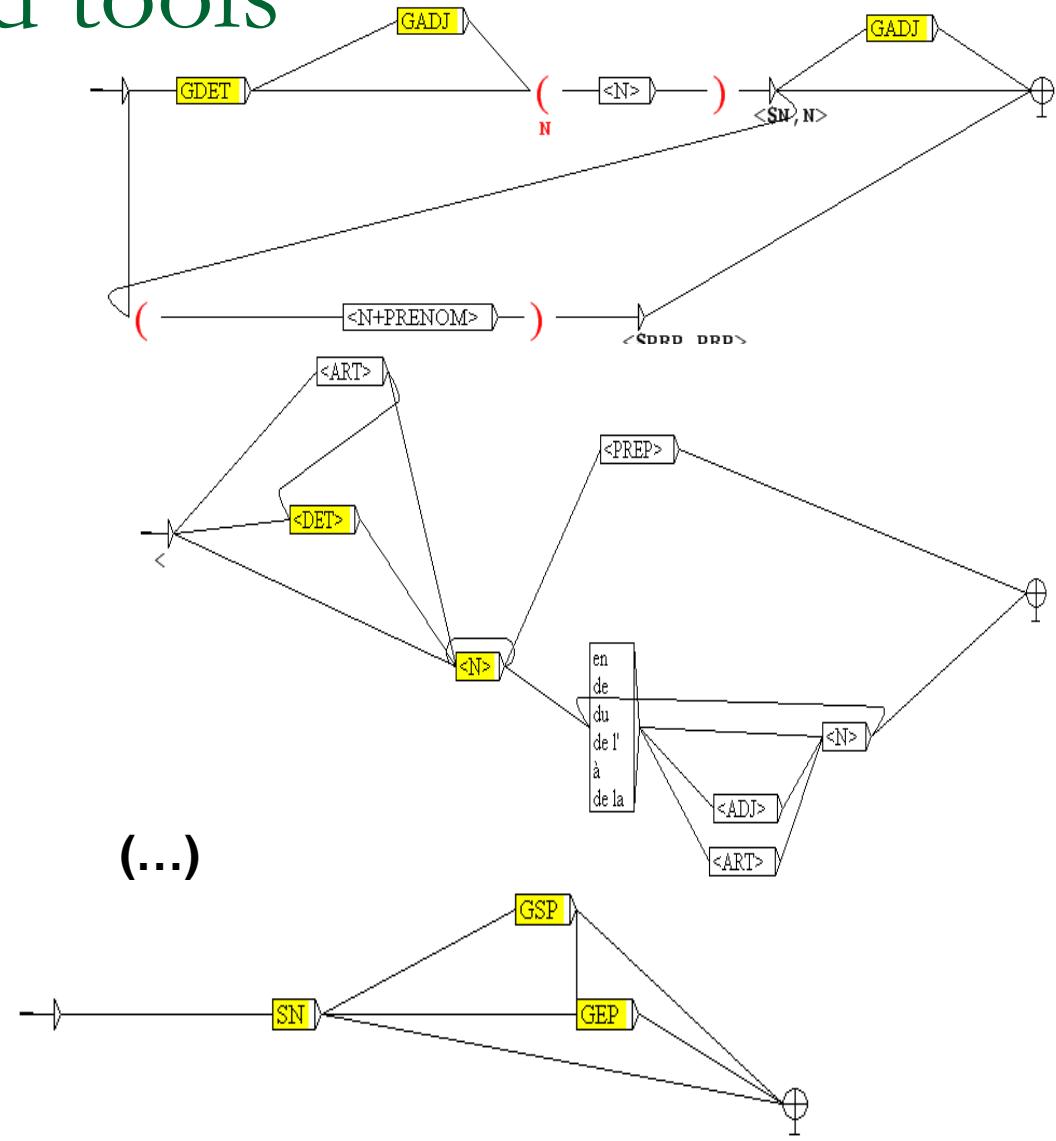
# *NLP Processing and NooJ*

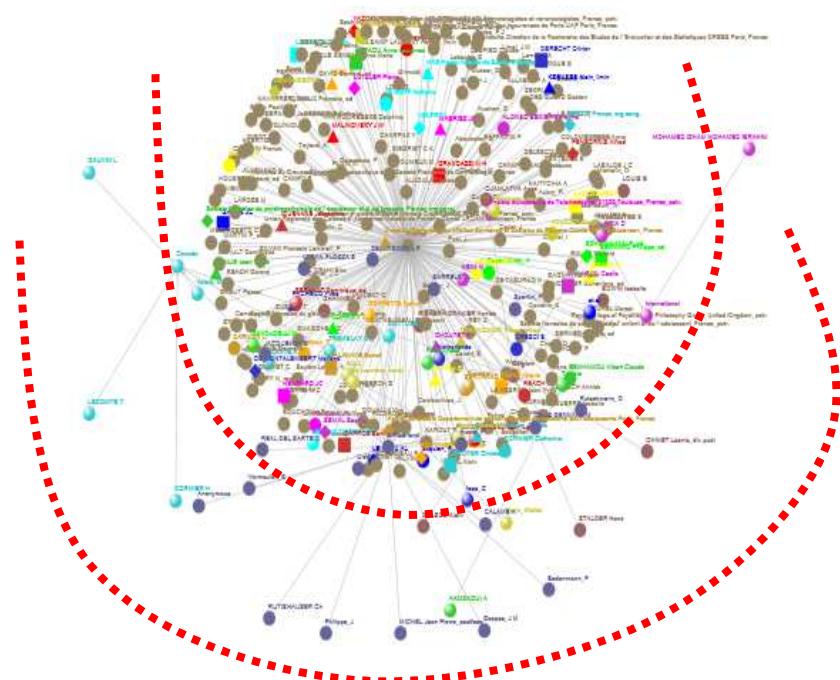
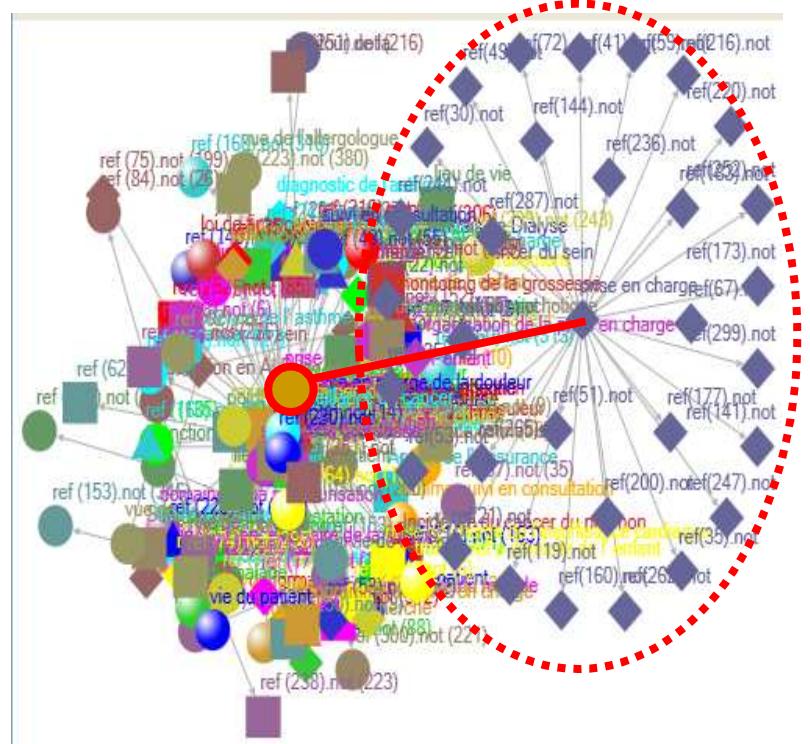


# Methodology and tools

## 1°/ NLP on NooJ platform

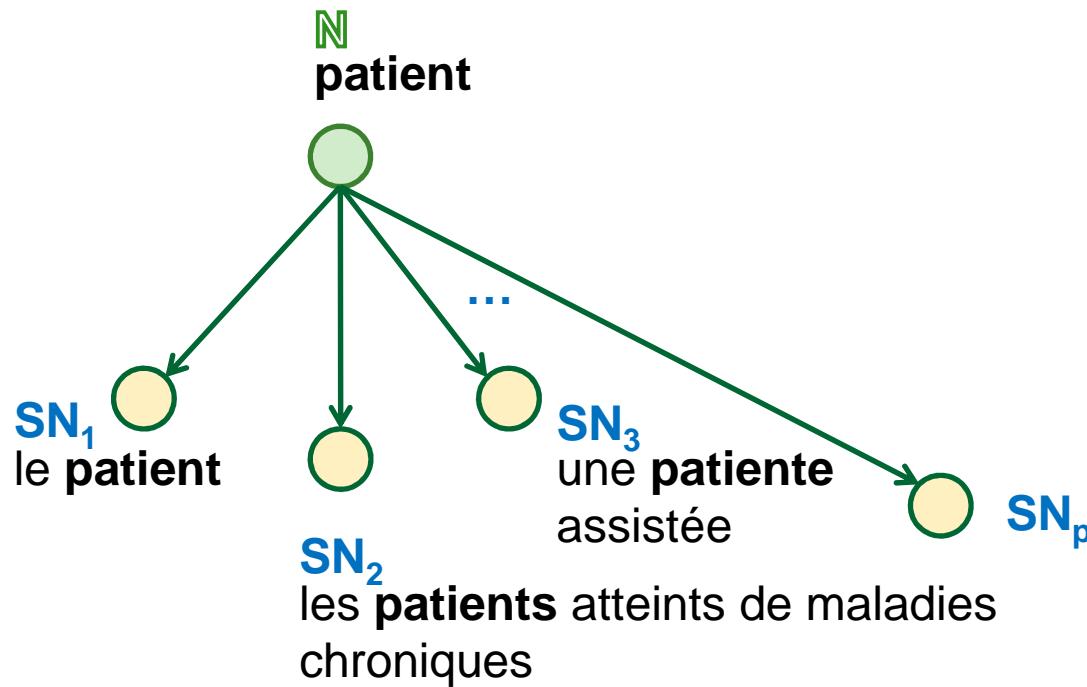
Description of graphs A. and B. in NooJ Main rule NP ::= N''	
Rules	Examples
$N'' \rightarrow D' + N'$	[FR] une+maladie [EN] a+disease
$N' \rightarrow N + A'$	[FR] maladie + chronique [EN] disease+chronic
$N \rightarrow A' + N$	[FR] pénible+maladie [EN] painful+disease
$N'' \rightarrow \text{NOM-PRP}$	Alzheimer
$N'' \rightarrow D' + N + \text{NOM-PRP}$	[FR] la+ maladie+ Alzheimer [EN] the+Alzheimer+ disease
$D' \rightarrow D   \text{ADV} + P   \dots$	[FR] un   beaucoup de   ... [EN] a   many of  ...
$N' \rightarrow N$	[FR] maladie   patient   ... [EN] disease   patient   ...
$\text{EP} \rightarrow P + N   (\text{EP})^n$	[FR] sans+guérison   (lésion dermatologique) en + forme + d' + anneau [EN] without + cure   (dermatological lesion) in + shape + of + ring
$\text{SP} \rightarrow P + N''   (\text{SP})^n$	[FR] par + des morsures (+de + (des tiques )) [EN] by + the bites (+of+the + ticks)
etc.	





# Knowledge?

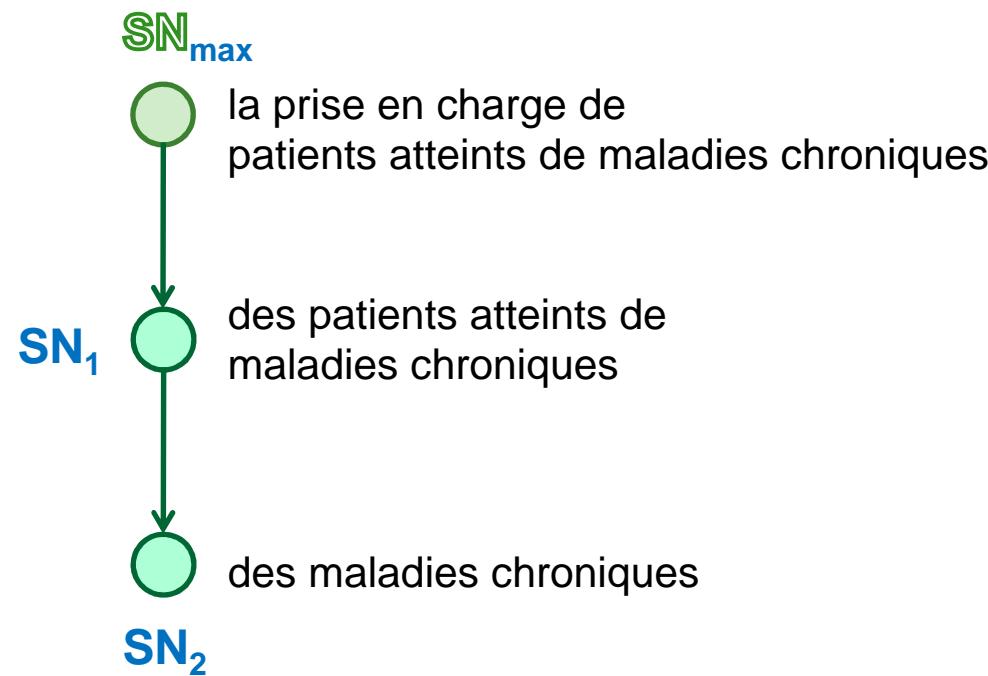
# Knowledge in Class relation



Class relation: *i.e.*  $\{SN\} \in N$

).

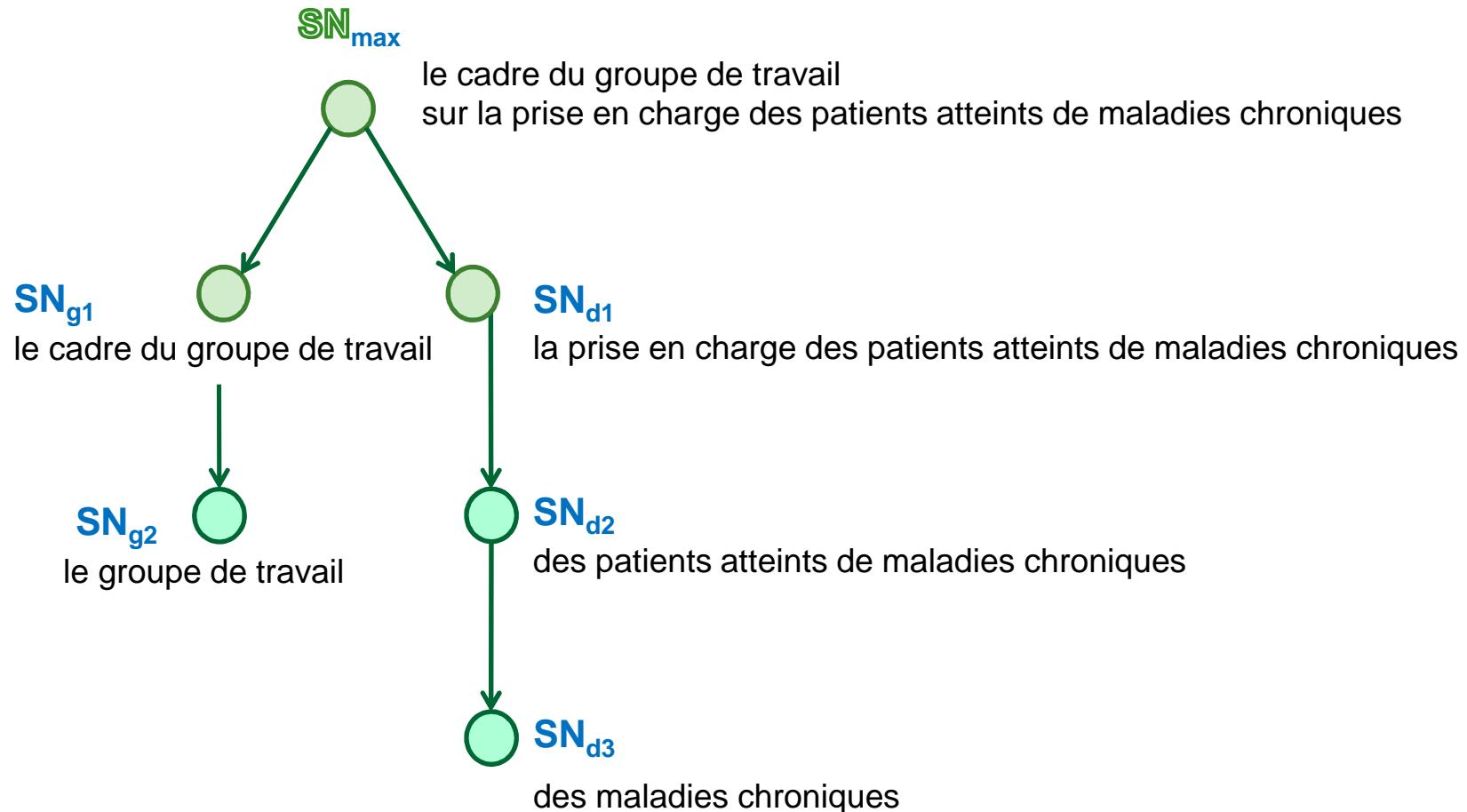
# Knowledge in Fitting relation



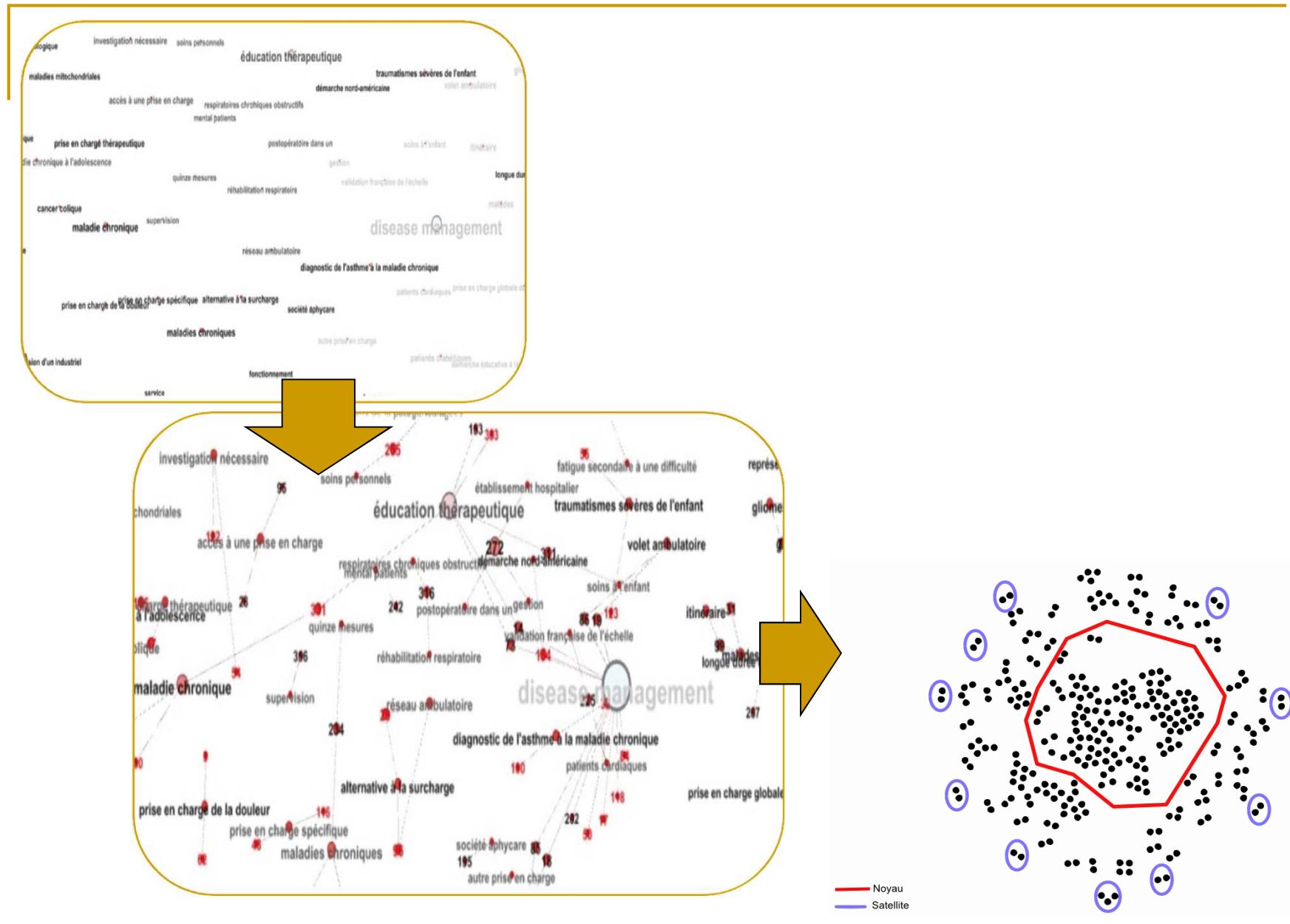
Fittig relation:

$$\text{SN}_{\max} \supseteq \text{SN}_1 \supseteq \text{SN}_2$$

# Knowledge in Tree relation



Tree relation:  $\text{SN}_{g1} \subseteq \text{SN}_{\max} \supseteq \text{SN}_{d1}$



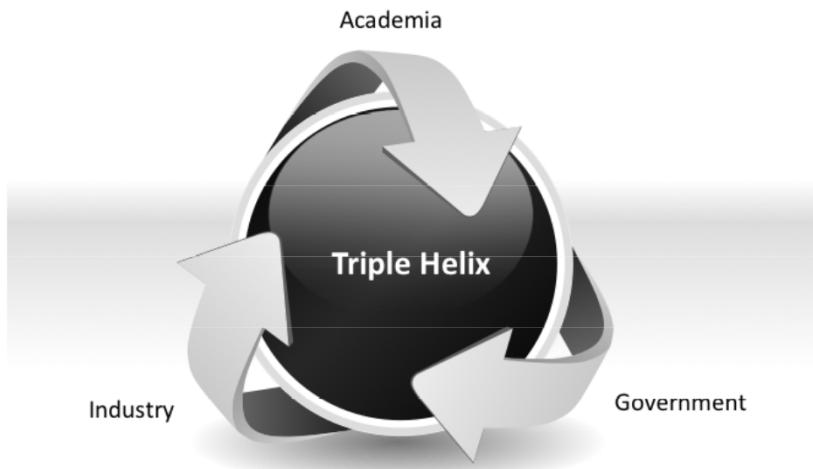
### III.

Applications:

***5. Case study in context of Nanometrology cluster***



## Project management by economic intelligence: **Nanometrology cluster**



### Two specialized networks in France:

1. C'Nano (nanoscience cluster)
2. LNE (metrology Lab.)

# Case Study : NooJ for Nanometrology Cluster

## Treatment for surveys

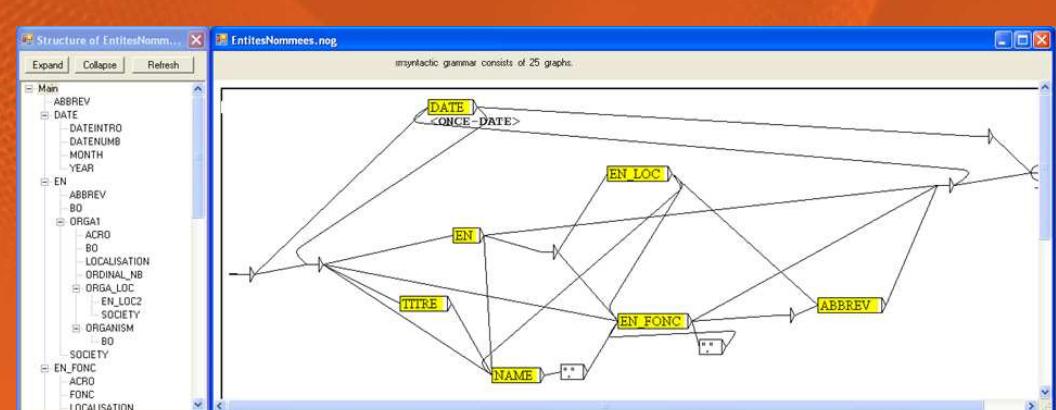


### Treatment of Responses

- Statistics (closed questions)
  - Statistics (open questions)
- Automatic Language Processing



### Projection (Social Network Analysis)



## Object processing:

Etablissement d'origine : \*

Taille de votre établissement \*

- 1-10 personnes
- 11 -100 personnes
- 100 - 500 personnes
- + de 500 personnes

Votre rôle au sein de l'établissement : \*

- Domaines d'activités de votre établissement (ex. chimie, environnement, automobile) : \*

Vos domaines d'activité en rapport avec les nanosciences et nanotechnologies :

Type d'activité : \*

- Recherche
- Développement
- Service

## Treatment of questions:

- Online questionnaire on nanoscience
- Determination skills of actors  
&
- main motivations



Strategic steering committee

## Actors survey respondents:

- hundreds researchers in R&D research laboratories (CNRS and universities...),
- Fields of Nano sciences (spintronics, photovoltaics, optoelectronics, hot plasmas, etc.).

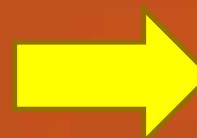
CSV Files

XSD mapping



**Data Formatting:**

- Selection of data
- Data Cleansing
- Construction of new data
- Data Integration
- Formatting data



## Data Formatting:

XLS source File:



XML File

	A	B	C	D	P	Q	R	S	T	U
1	Horodateur	Nom	Prénom	Etablissement d'origine :	Mots-clés (entre 3 et 10) correspondant à vos activités scientifiques/techniques :	d'équipements pour lesquels vous avez une	Nombre de personnes impliquées :	établissement possède-t-il une expertise en	établissement possède-t-il une expertise en	raison principale qui a motivé votre inscription au
46	1/5/2012 15:53:12			Recherche et innovation	Emissions industrielles : Valorisation énergétique des déchets ; chaudières Eau, effluents industriels, effluents gazeux Particules ultrafines, aérosols Impact environnemental Exposition professionnelle	Impacteurs Elpi SMPS	5	0	En tant qu'opérateur dans le traitement des déchets et le traitement de l'eau, Veolia est directement concerné par cette thématique et notamment les nanoparticules présents dans les produits en fin de vie.	
47	1/5/2012 17:05:51				Logiciels scientifiques Simulations Prédiction des propriétés physico-chimiques Science des matériaux	Simulation par ordinateur	1	0	La simulation à l'échelle atomique peut jouer un rôle important dans la nano-métrieologie participer à des travaux collectifs permettant de synthétiser ou d'accroître les connaissances dans le domaine de l'utilisation des	

indicators

```

743 </Entrees>
744 <Entrees>
745   <Nom>_____</Nom>
746   <Prénom>_____</Prénom>
747   <EtablissementDorigine>_____</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 calcul empirique: DISCOVER, LAMMPS</TypeEquipement>
752   <RaisonInscriptionClub>participer a des travaux collectives permettant de synthétiser ou d'accroitre les connaissances dans le domaine de l'utilisation des codes de calculs
753 </Entrees>
754 <Entrees>
755   <Nom>_____S</Nom>
756   <Prénom>_____S</Prénom>
757   <EtablissementDorigine>_____A&lt;/EtablissementDorigine>
758   <DomainesActivitesEtabl>MECANIQUE
759   <DomainesActivitesEtabl>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>_____ni</Nom>
768   <Prénom>_____W</Prénom>
769   <EtablissementDorigine>_____f</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 SYSTÈMES DANS LE DOMAIN 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>_____F</Nom>
780   <Prénom>_____W</Prénom>
781   <EtablissementDorigine>_____1</EtablissementDorigine>
782   <DomainesActivitesEtabl>Nanotechnologie</DomainesActivitesEtabl>

```

extraction of specific data:

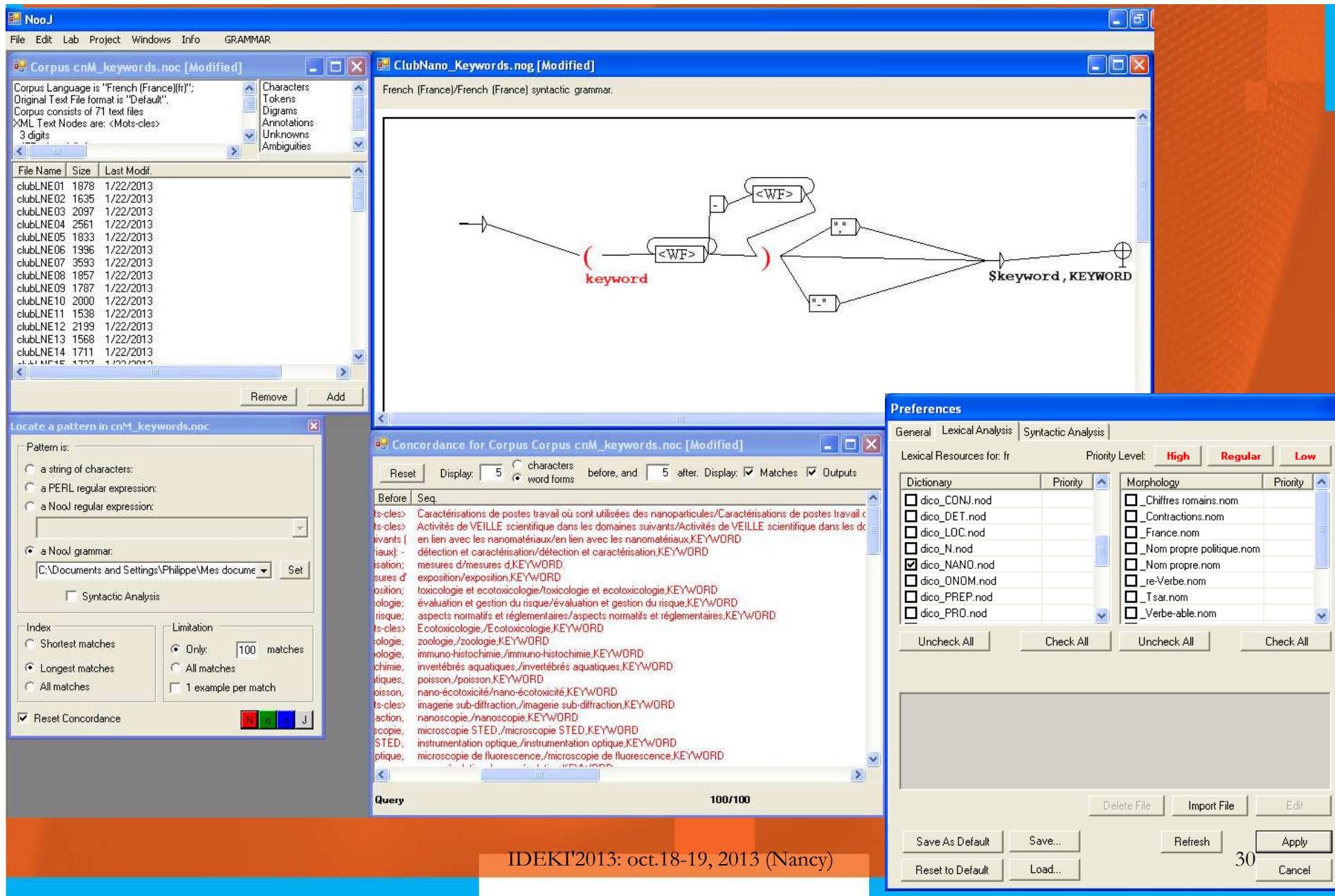
## SPLIT function : Retrieve data from documentary sources

Nom	Modifié le	Type	Taille	
clubLNE34.xml	15/01/2013 22:26	Document XML	2 Ko	<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
clubLNE35.xml	15/01/2013 22:26	Document XML	3 Ko	<Entrees>
clubLNE36.xml	15/01/2013 22:26	Document XML	2 Ko	<Nom>EN NACIRI</Nom>
clubLNE37.xml	15/01/2013 22:26	Document XML	1 Ko	<Prénom>AOTMANE</Prénom>
clubLNE38.xml	15/01/2013 22:26	Document XML	1 Ko	<EtablissementDorigine>Université de Lorraine
clubLNE39.xml	15/01/2013 22:26	Document XML	2 Ko	</EtablissementDorigine>
clubLNE40.xml	15/01/2013 22:26	Document XML	3 Ko	<DomainesActivitesEtabl>Physique : Optique
clubLNE41.xml	15/01/2013 22:26	Document XML	1 Ko	</DomainesActivitesEtabl>
clubLNE42.xml	15/01/2013 22:26	Document XML	1 Ko	<Mots-cles>Ellipsométrie-Nanomatériaux-couches
clubLNE43.xml	15/01/2013 22:26	Document XML	2 Ko	minces-propriétés optiques-Instrumentation en ellipsométrie-
clubLNE44.xml	15/01/2013 22:26	Document XML	2 Ko	Surface et interface. </Mots-cles>
clubLNE45.xml	15/01/2013 22:26	Document XML	2 Ko	<TypeEquipement>Ellipsométrie</TypeEquipement>
clubLNE46.xml	15/01/2013 22:26	Document XML	1 Ko	<RaisonInscriptionClub>Échanger les idées et le
clubLNE47.xml	15/01/2013 22:26	Document XML	1 Ko	savoir faire avec les collègues dans le domaine de la
clubLNE48.xml	15/01/2013 22:26	Document XML	1 Ko	maîtrise des mesures optiques sur des échantillons
clubLNE49.xml	15/01/2013 22:26	Document XML	1 Ko	nanométriques : vers une stratégie de nanométrieologie
clubLNE50.xml	15/01/2013 22:26	Document XML	1 Ko	</RaisonInscriptionClub>
clubLNE51.xml	15/01/2013 22:26	Document XML	1 Ko	<AccesBasesDonneesLesquelles>1
clubLNE52.xml	15/01/2013 22:26	Document XML	1 Ko	</AccesBasesDonneesLesquelles>
clubLNE53.xml	15/01/2013 22:26	Document XML	2 Ko	</Entrees>
clubLNE54.xml	15/01/2013 22:26	Document XML	1 Ko	
clubLNE55.xml	15/01/2013 22:26	Document XML	1 Ko	
clubLNE56.xml	15/01/2013 22:26	Document XML	1 Ko	
clubLNE57.xml	15/01/2013 22:26	Document XML	1 Ko	
clubLN	15/01/2013 22:26	Document XML	2 Ko	
clubLN	15/01/2013 22:26	Document XML	1 Ko	
clubLN	15/01/2013 22:26	Document XML	1 Ko	
clubLNcoo.xml	15/01/2013 22:26	Document XML	1 Ko	
clubLINE61.xml	15/01/2013 22:26	Document XML	2 Ko	
clubLINE62.xml	15/01/2013 22:26	Document XML	1 Ko	
clubLINE63.xml	15/01/2013 22:26	Document XML	3 Ko	
clubLINE64.xml	15/01/2013 22:26	Document XML	2 Ko	
clubLINE65.xml	15/01/2013 22:26	Document XML	1 Ko	
clubLINE66.xml	15/01/2013 22:26	Document XML	1 Ko	

## **Using of NooJ for Questionnaires by processing XML nodes**

- 1. Structured answers (closed questions) : reorganizing data**
- 2. Unstructured answers (open questions) : NLP and filtering**

Reorganizing Data (closed questions) : the keywords – Creation of thematic resources



## Reorganizing Data (closed questions) : the keywords – Creation of thematic resources

Specific  
dictionaries  
**&**  
Syntactic  
grammars  
for  
"Keywords"

The screenshot displays the NooJ software interface, which integrates corpus management, XML annotation, and syntactic grammar creation.

**Top Left Window:** Shows the corpus "Corpus cnM\_keywords.noc [Modified]" with details: Corpus Language is "French (France)(fr)", Original Text File format is "Default", Corpus consists of 71 text files, XML Text Nodes are: <EtablissementDorigine> <Mots-clés>, and Corpus contains 355 text units delimited by "".

**Top Right Window:** Shows the XML annotation for "clubLNE51.not" with the following code:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<Entrees>
    <Nom>          </Nom>
    <Prénom>Christophe</Prénom>
    <EtablissementDorigine>SCHAEFER TECHNIQUES</EtablissementDorigine>
    <DomainesActivitésEtabl>Instrumentation - Metrologie</DomainesActivitésEtabl>

```

A red box highlights the number 16 next to the word "SCHAEFER TECHNIQUES".

**Bottom Window:** Shows the syntactic grammar "ClubNano\_Keywords.nog [Modified]" with a diagram illustrating the structure of a keyword. The diagram shows a root node connected to a "keyword" node, which is then connected to a "WF" (Word Form) node. The "WF" node is enclosed in parentheses and has arrows pointing to opening and closing brackets "<" and ">". The "WF" node is also connected to a "KEYWORD" node, which is further connected to a plus sign node.

At the bottom of the interface, the text "IDEKI'2013: oct.18-19, 2013 (Nancy)" is visible, along with the page number "31".

## Reorganizing Data (closed questions) : the keywords – Creation of thematic resources

NooJ Data Extraction



Export Files

The screenshot shows the NooJ interface with three main windows:

- Corpus cnM\_keywords.noc**: Shows a list of 16 text files (clubLNE01.not to clubLNE16.not) with their sizes and last modification dates.
- Locate a pattern in cnM\_keywords.noc**: A search dialog where the pattern is set to "<EN>+<KEYWORD>".
- Concordance for Corpus Corpus cnM\_keywords.noc**: Displays a list of matches, with the first few entries being:
  - clubLNE01.not > DMNT (CEA/CNRS)
  - clubLNE01.not > Activités de VEILLE scientifique dans les domaines suivants
  - clubLNE01.not | en lien avec les nanomatériaux
  - clubLNE01.not détection et caractérisation;
  - clubLNE01.not mesures d'
  - clubLNE01.not exposition;
  - clubLNE01.not toxicologie et ecotoxicologie;
  - clubLNE01.not évaluation et gestion du risque;
  - clubLNE01.not aspects normatifs et réglementaires

At the bottom of the interface, it says "IDEKI'2013: oct.18-19, 2013 (Nancy)" and "Query 353/353".

## Reorganizing Data (closed questions) : the keywords – Creation of thematic resources

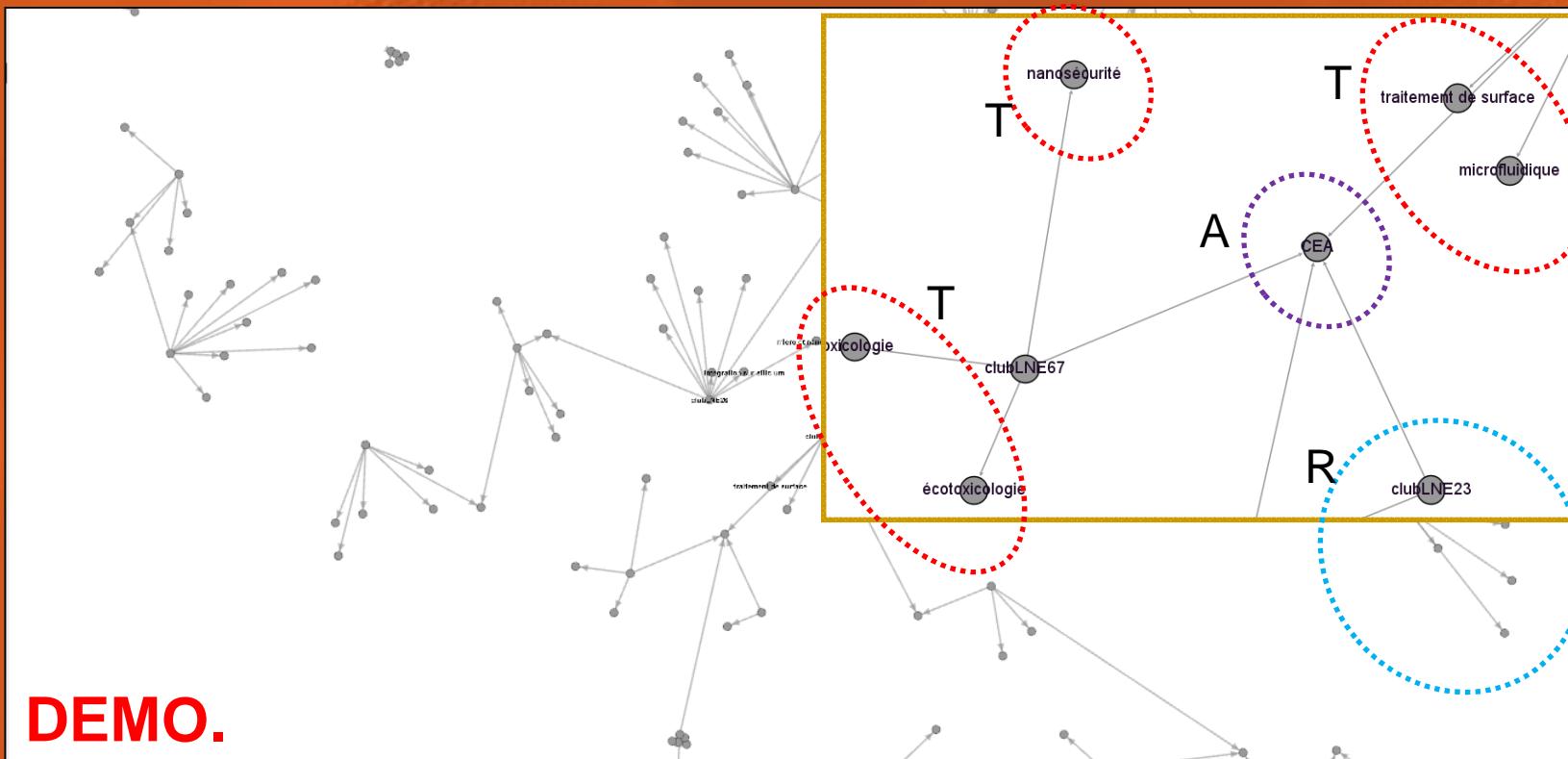
1	clubLNE71	CEA-Grenoble
2	clubLNE71	Caractérisations de postes travail où sont utilisées des nanoparticules
3	clubLNE01	OMNT (CEA/CNRS)
4	clubLNE01	Activités de VEILLE scientifique dans les domaines suivants
5	clubLNE01	en lien avec les nanomatériaux
6	clubLNE01	détection et caractérisation
7	clubLNE01	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	ellipsométrie
32	clubLNE04	couches minces
33	clubLNE04	photonique
34	clubLNE04	photo-voltaique

# Illustrations : needs in Nanometrology Custer

## Information Design



- Identification of weak signals: long-term “strategic” information
- Diagnostics in continuity: data + structures + NooJ resources



## NLP (open questions) : Identifying the needs of members

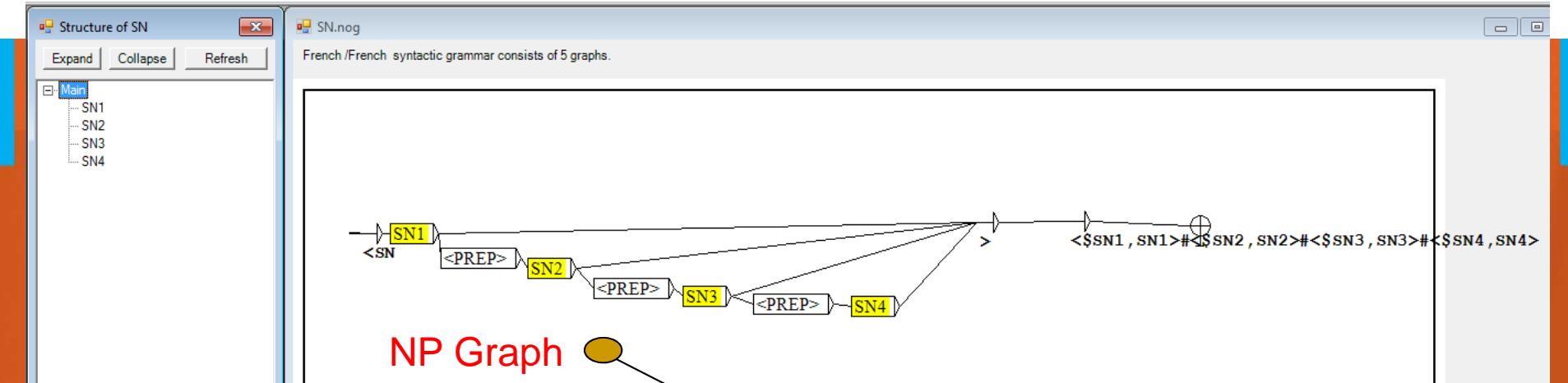
Open Response: "The reasons for your participation?"



NooJ Processing with specific XML node



Extracting the most relevant noun phrases (NP)



NPs Extraction  
&  
Locate specific structure  
Through Regex NooJ  
 $\langle V \rangle + \langle SN \rangle$   
Lemmatization + interlocking NP

établir des contacts avec la communauté	0,07811217	
réaliser des caractérisations de couches minces	0,0778794	
DÉVELOPPER DE NOUVELLES ACTIVITÉS	0,07638687	
Développement de compétences complémentaires	0,07566451	
une activité de recherche	0,07534191	
Structure française de métrologie pour définition	0,07405033	
optimiser la qualité des consortiums pour mes futurs projets	0,0733888	
autres projets collaboratifs à impact industriel	0,07256933	
Rencontrer des acteurs de ce secteur	0,07195017	
Avancées vers la métrologie	0,06616665	
pour définition de protocoles	0,06457878	
Faire partie du réseau	0,06431864	
connaître les acteurs français du monde	0,06356252	
REALISER DES ETUDES	0,06222717	
Avoir des contacts	0,0618805	
apporter ensemble les réponses les plus adéquates	0,06041966	
Recherche de nouveaux clients	0,05958986	
mise en place de collaboration	0,05934748	
analyse des nanoparticules	0,0581209	
un réseau industriels	0,05809941	
Etre informé des nouvelles méthodes	0,05720627	
Évaluer les besoins instrumentaux pour mesures	0,05679448	
les autres techniques	0,05672379	
analyse des risques	0,05592926	
développement de mesures	0,05495035	
réfèrent en métrologie dimensionnelle	0,05461787	
établir des collaborations	0,05402814	
Discussion avec divers partenaires	0,05341636	
opérateur dans le traitement des déchets	0,05322692	
enseignement des questions scientifiques	0,05296692	
des projets collaboratifs en cours de montage	0,05237262	
développe une technologie laser	0,0519196	
sur un sujet en plein développement	0,05174325	
utilisation des outils	0,05062894	
les questionnements dans ce domaine	0,05022253	
Acquérir des conseils	0,05015725	

TF-IDF

# Conclusion

- Key Performance Indicators (KPI):
  1. Strategic Orientations to nanometrology at t+12
  2. Network Diagnosis to nanometrology cluster
    - Common Thematics
    - Synergy Detection (or limits) to Community Manager
  3. Project Anticipations in time to



*Thanks...*



***Scientific Projects:***

ISKO-Maghreb int. symposium ([www.isko-maghreb.org](http://www.isko-maghreb.org))  
SIIIE int. conference ([www.siie.fr](http://www.siie.fr))