



# *Préliminaire à la construction d'un réseau de signalisation en biologie systémique*

Flavie Landomiel, Anubhav Gupta, Denis Maurel, Anne Poupon

Le 24/01/2017 à Grenoble  
Atelier - Fouille de textes - Text mine

# *Projet*

## **ISTEX**

- bibliothèque numérique : rétrospective de la Science

# *Projet*

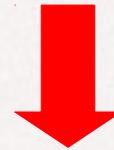
## **ISTEX**

- bibliothèque numérique : rétrospective de la Science
- tend à offrir tous les moyens accessibles de consultation et d'analyse aujourd'hui disponibles dans toutes les communautés de la Science

# Projet

## ISTEX

- bibliothèque numérique : rétrospective de la Science
- tend à offrir tous les moyens accessibles de consultation et d'analyse aujourd'hui disponibles dans toutes les communautés de la Science

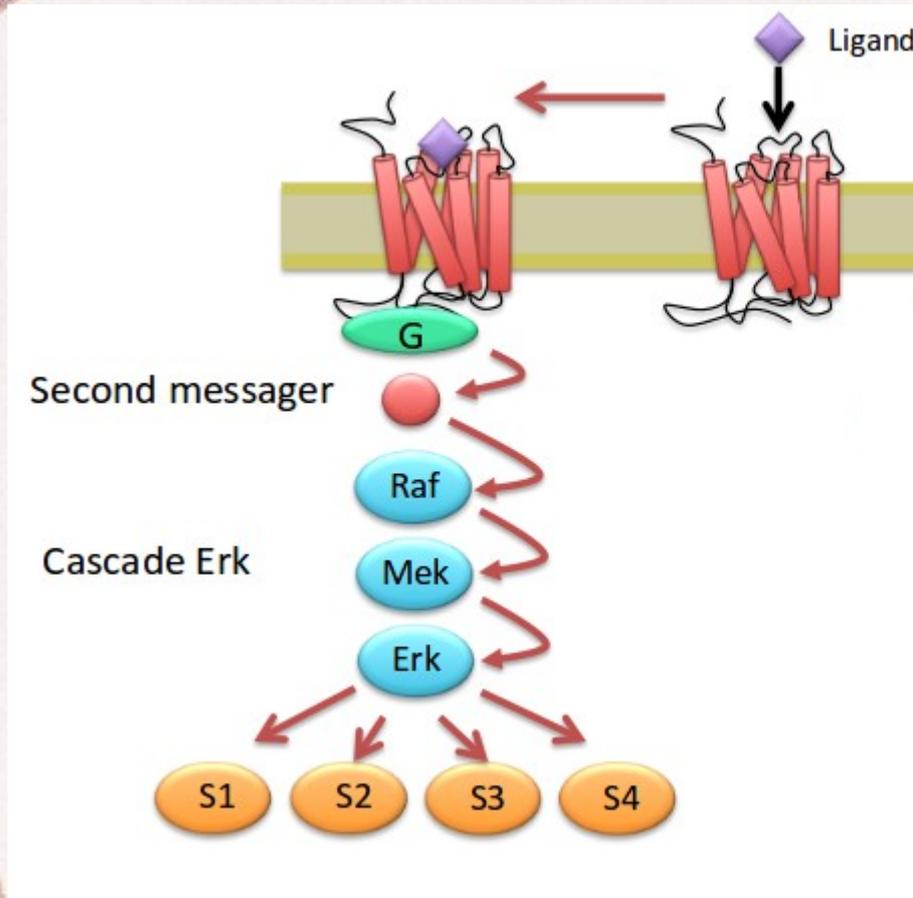


## Projet Biosystémique

*recherche de résultats expérimentaux dans les publications scientifiques concernant la biologie systémique afin de détecter les voies de signalisation des récepteurs couplés aux protéines G*

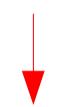
# ***1) Réseau de signalisation***

# Réseau de signalisation



Simple

A



1



2



3

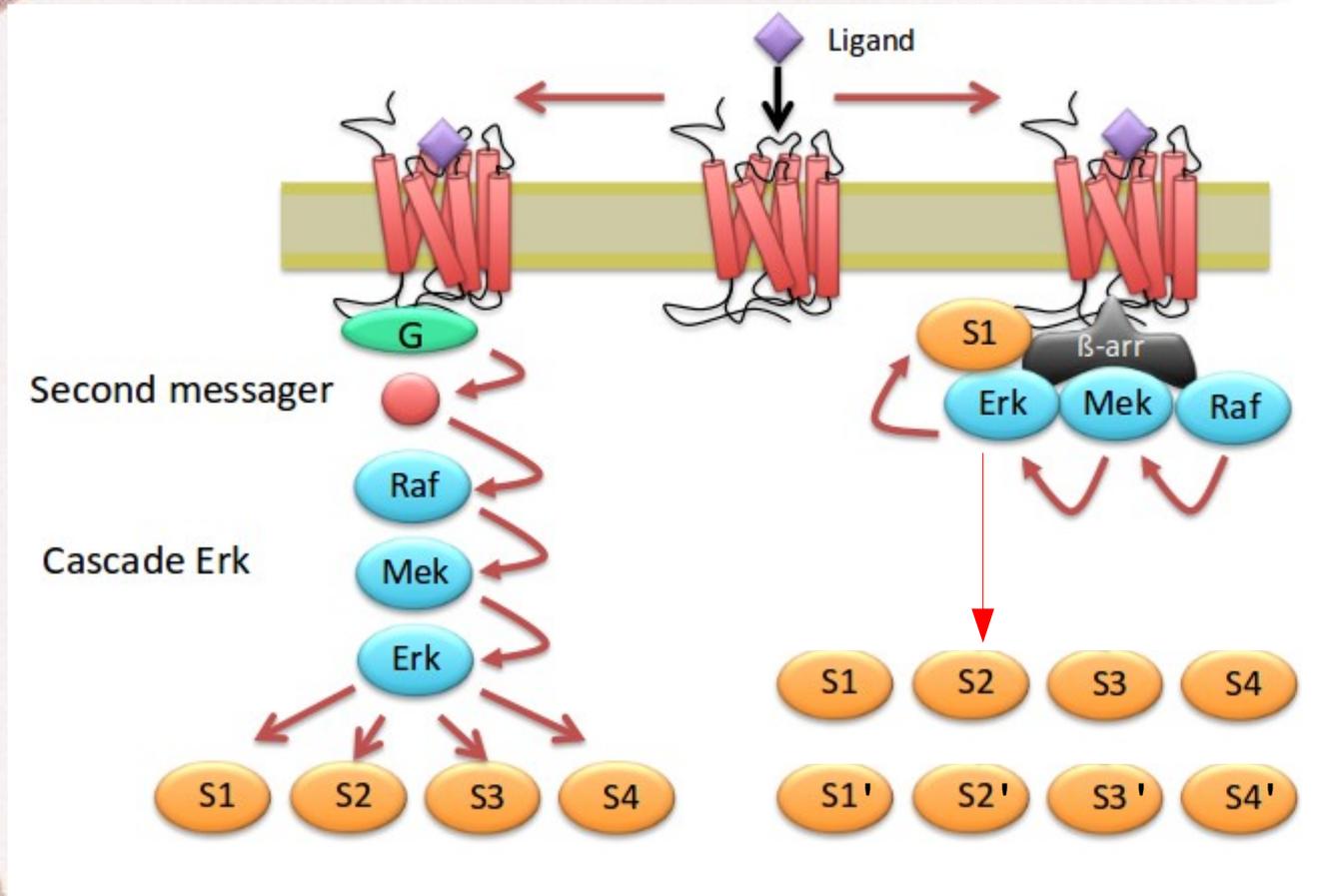


n

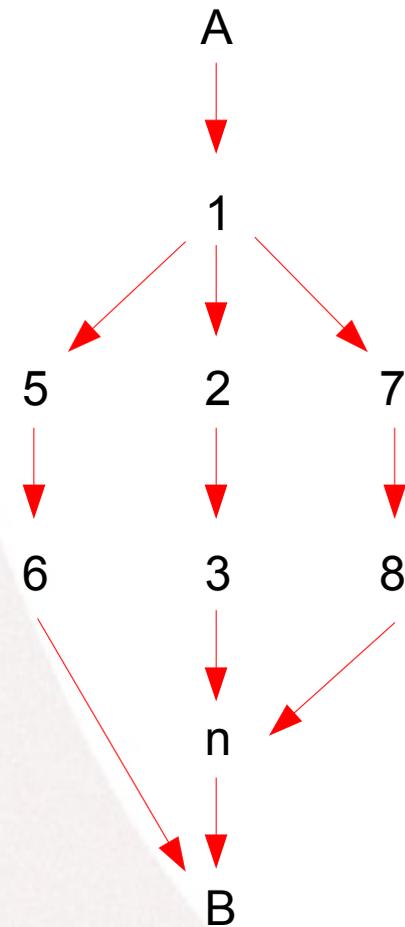


B

# Réseau de signalisation



## Complexe



# ***Article scientifique***

## **Organisation d'un article scientifique en biologie:**

- 1) Résumé => principaux résultats
- 2) Introduction => contexte de l'étude + état de l'art
- 3) Matériel et méthodes => mise en pratique des expériences
- 4) Résultats => liste et détail des expériences réalisées
- 5) Discussion => interprétation des expériences et perspectives

# *Article scientifique*

## Organisation d'un article scientifique en biologie:

- 1) Résumé => principaux résultats
- 2) Introduction => contexte de l'étude + état de l'art
- 3) Matériel et méthodes => mise en pratique des expériences
- 4) Résultats => liste et détail des expériences réalisées**
- 5) Discussion => interprétation des expériences et perspectives



**Meilleure fouille** car hors résumé et mots-clés  
(*habituel*)

# Identification de phrases d'intérêt

- Identification de **verbes**
- Identification des **molécules**
- Identification des **actions**
- Identifications des **méthodes expérimentales**

## © ERK phosphorylates Nup50 at Ser221 and Ser315



We then focused on Nup50 (also called Npap60), a component of the NPC, for the following reasons. First, ERK activation induced a marked acidic shift in the isoelectric point of endogenous Nup50, suggesting highly stoichiometric phosphorylation (Fig. 2b). Second, Nup50 is efficiently phosphorylated by ERK *in vitro* (Figs. 3a and 4a). Third, activated ERK is reported to inhibit nuclear protein import<sup>31</sup>. Fourth, ERK enters the nucleus via direct interaction with NPC proteins<sup>32,33</sup>. Nup50 seems to function at the terminal stages of nuclear protein import to coordinate import complex disassembly and importin recycling<sup>34</sup>. It has also been reported that Nup50 acts as a cofactor for importin- $\alpha$ -importin- $\beta$ -mediated nuclear import to stimulate classical nuclear localization signal (NLS)-directed import<sup>35</sup>. Nup50 contains an N-terminal domain that binds to importin- $\alpha$  (N), a central FG repeat domain that binds to importin- $\beta$  (FG) and a C-terminal Ran-binding domain (R)<sup>35</sup>. ERK phosphorylated Nup50 exclusively in the FG repeat domain *in vitro* (Fig. 4a). LC-MS and tandem MS (MS/MS) analyses of ERK-phosphorylated and unphosphorylated GST-tagged full-length Nup50 demonstrated that ERK phosphorylates Nup50 stoichiometrically at Ser221 and Ser315 *in vitro* (Supplementary Fig. 3). *In vitro* kinase assays with wild-type and Ser-to-Ala mutants of Nup50 confirmed this finding (Fig. 4b).

# ***Réseau de signalisation***

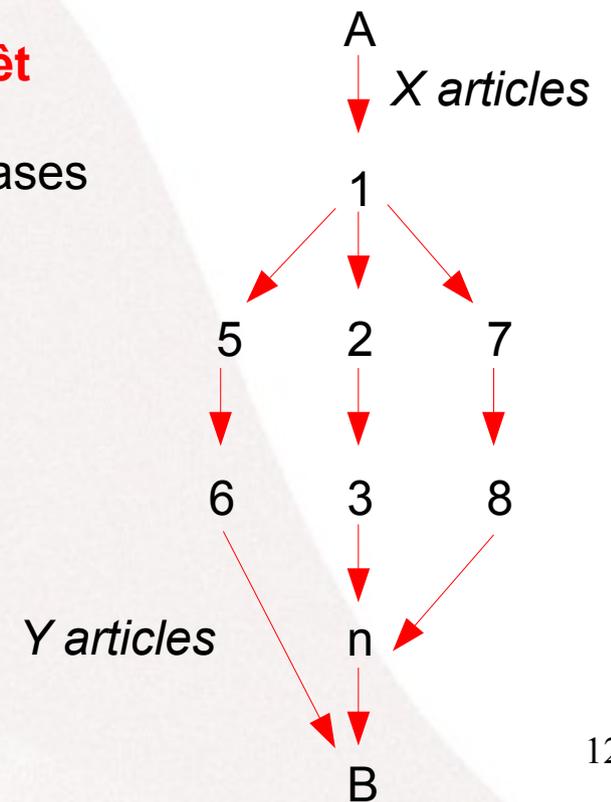
Réseau de signalisation réalisé au sein de l'équipe BIOS par **Gloaguen et al (2011)** :

Manuel	1) Recherche de phrases d'intérêt
	2) Création des faits issus des phrases
Automatique	3) Moteur d'inférence

# Réseau de signalisation

Réseau de signalisation réalisé au sein de l'équipe BIOS par **Gloaguen et al (2011)** :

- |             |   |
|-------------|---|
| Manuel      | 1) Recherche de phrases d'intérêt       |
|             | 2) Création des faits issus des phrases |
| Automatique | 3) Moteur d'inférence                   |



## ***II) Recherche automatique des phrases d'intérêt***

# *Procédure*

① Documents « test »

# Format des textes

**FIGURE 1. Mobilization of calcium, DNA synthesis, and phosphorylation of ERK stimulated by ANG and SII in VSMCs.** *A*, VSMCs were loaded with the calcium-binding dye Fura-2, and stimulated either with ANG (100 nM) or SII (10  $\mu$ M) in the presence or absence of pretreatment with the AT1R antagonist (*AT1RB*) valsartan (50  $\mu$ M) or AT2R antagonist (*AT2RB*) PD123319 (30  $\mu$ M). Calcium fluorimetric traces are shown with the 340/380 nm excitation ratio (y axis) plotted as a function of time (x axis). Results displayed are mean  $\pm$  S.E. of three independent experiments. *B*, VSMCs were serum starved for 24 h to arrest cycling and pretreated with dimethyl sulfoxide (*DMSO*) or MEK inhibitor PD98056 (20  $\mu$ M). ANG (100 nM), SII (10  $\mu$ M), or EGF (10 ng/ml) along with  $^3$ H-labeled thymidine were then added to the media, and 24 h later cells were harvested as described under "Experimental Procedures." NS indicates no stimulation. Results depicted represent the mean  $\pm$  S.E. of count per minute (cpm) values from four independent experiments. Statistical analysis was done using a one-way ANOVA (PRISM software) to correct for multiple comparisons (Bonferroni's multiple comparison test) with post test. The PD98056-pretreated condition shows significant reduction compare with the dimethyl sulfoxide-pretreated condition for each stimulant (\*,  $p < 0.05$ ). *C*, VSMCs with endogenous AT1R were treated with 100 nM ANG or 10  $\mu$ M SII for the indicated times. Equal amounts of cell lysate were separated by SDS-PAGE and analyzed for phosphorylated ERK (*p-ERK*) and total ERK (*ERK*) by Western blotting. *IB*, immunoblot. *D*, signals were quantified by densitometry and p-ERK was normalized to a loading control (ERK). p-ERK activation was expressed as percentage of the maximal phosphorylated ERK obtained by ANG stimulation at 5 min. Each data point represents the mean  $\pm$  S.E. from eight independent experiments.

contribution of  $\beta$ -arrestin-mediated ERK activation to VSMC proliferation, we studied ANG- and SII-mediated DNA synthesis in VSMCs as an indicator of cell proliferation. VSMCs were pretreated or not with a pharmacological inhibitor (PD98059) of mitogen-activated protein kinase kinase (MEK) followed by stimulation with 100 nM ANG, 10  $\mu$ M SII, or 10 ng/ml EGF. Although SII cannot activate classical G-protein-dependent  $Ca^{2+}$  fluxes (Fig. 1A), it significantly increased DNA synthesis (thymi-

# Format des textes

**FIGURE 1. Mobilization of calcium, DNA synthesis, and phosphorylation of ERK stimulated by ANG and SII in VSMCs.** *A*, VSMCs were loaded with the calcium-binding dye Fura-2, and stimulated either with ANG (100 nM) or SII (10  $\mu$ M) in the presence or absence of pretreatment with the AT1R antagonist (*AT1RB*) valsartan (50  $\mu$ M) or AT2R antagonist (*AT2RB*) PD123319 (30  $\mu$ M). Calcium fluorimetric traces are shown with the 340/380 nm excitation ratio (y axis) plotted as a function of time (x axis). Results displayed are mean  $\pm$  S.E. of three independent experiments. *B*, VSMCs were serum starved for 24 h to arrest cycling and pretreated with dimethyl sulfoxide (*DMSO*) or MEK inhibitor PD98056 (20  $\mu$ M). ANG (100 nM), SII (10  $\mu$ M), or EGF (10 ng/ml) along with  $^3$ H-labeled thymidine were then added to the media, and 24 h later cells were harvested as described under "Experimental Procedures." NS indicates no stimulation. Results depicted represent the mean  $\pm$  S.E. of count per minute (cpm) values from four independent experiments. Statistical analysis was done using a one-way ANOVA (PRISM software) to correct for multiple comparisons (Bonferroni's multiple comparison test) with post test. The PD98056-pretreated condition shows significant reduction compare with the dimethyl sulfoxide-pretreated condition for each stimulant (\*,  $p < 0.05$ ). *C*, VSMCs with endogenous AT1R were treated with 100 nM ANG or 10  $\mu$ M SII for the indicated times. Equal amounts of cell lysate were separated by SDS-PAGE and analyzed for phosphorylated ERK (*p-ERK*) and total ERK (*ERK*) by Western blotting. *IB*, immunoblot. *D*, signals were quantified by densitometry and p-ERK was normalized to a loading control (ERK). p-ERK activation was expressed as percentage of the maximal phosphorylated ERK obtained by ANG stimulation at 5 min. Each data point represents the mean  $\pm$  S.E. from eight independent experiments.

contribution of  $\beta$ -arrestin-mediated ERK activation to VSMC proliferation, we studied ANG- and SII-mediated DNA synthesis in VSMCs as an indicator of cell proliferation. VSMCs were pretreated or not with a pharmacological inhibitor (PD98059) of mitogen-activated protein kinase kinase (MEK) followed by stimulation with 100 nM ANG, 10  $\mu$ M SII, or 10 ng/ml EGF. Although SII cannot activate classical G-protein-dependent  $Ca^{2+}$  fluxes (Fig. 1A), it significantly increased DNA synthesis (thymi-

Reconnait comme 1 phrase



**FIGURE 1. Mobilization of calcium, DNA synthesis, and phosphorylation of ERK stimulated by ANG and**

contribution of  $\beta$ -arrestin-mediated

# Procédure

1 Documents « test »

2 bases de données :

NCBI

NCBI Resources How To Sign in to NCBI

PMC US National Library of Medicine National Institutes of Health

ERK AND arrestin AND phosphorylation Search

Save search Journal List Advanced Help

**Search results**

Items: 1 to 20 of 2739

<< First < Prev Page 1 of 137 Next > Last >>

ISTEX

Bienvenue sur le démonstrateur ISTEDEX

En savoir plus

restin AND phosphorylation Options

Recherche avancée

Requête `https://api.istex.fr/document/?q=ERK AND arrestin AND phosphorylation &facet=corpusName[*],host.genre[*]>genre[*],pdfVersion[*],refE` Réponse brute complète

Résultats : 578 ( 822 ms) 1/ 58



3 255 documents

# Procédure

① Documents « test »



② Pré-traitement

*Présence/absence mots*  
*Revue/résumé*

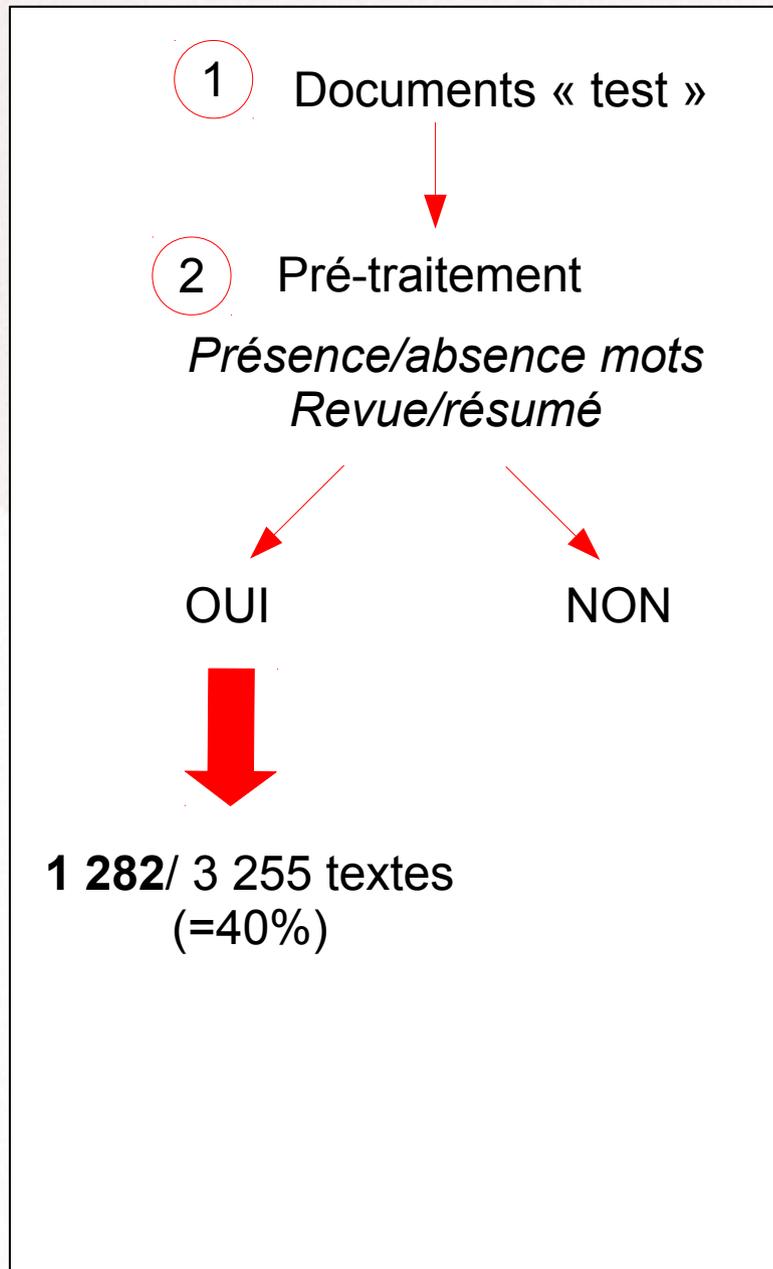


OUI

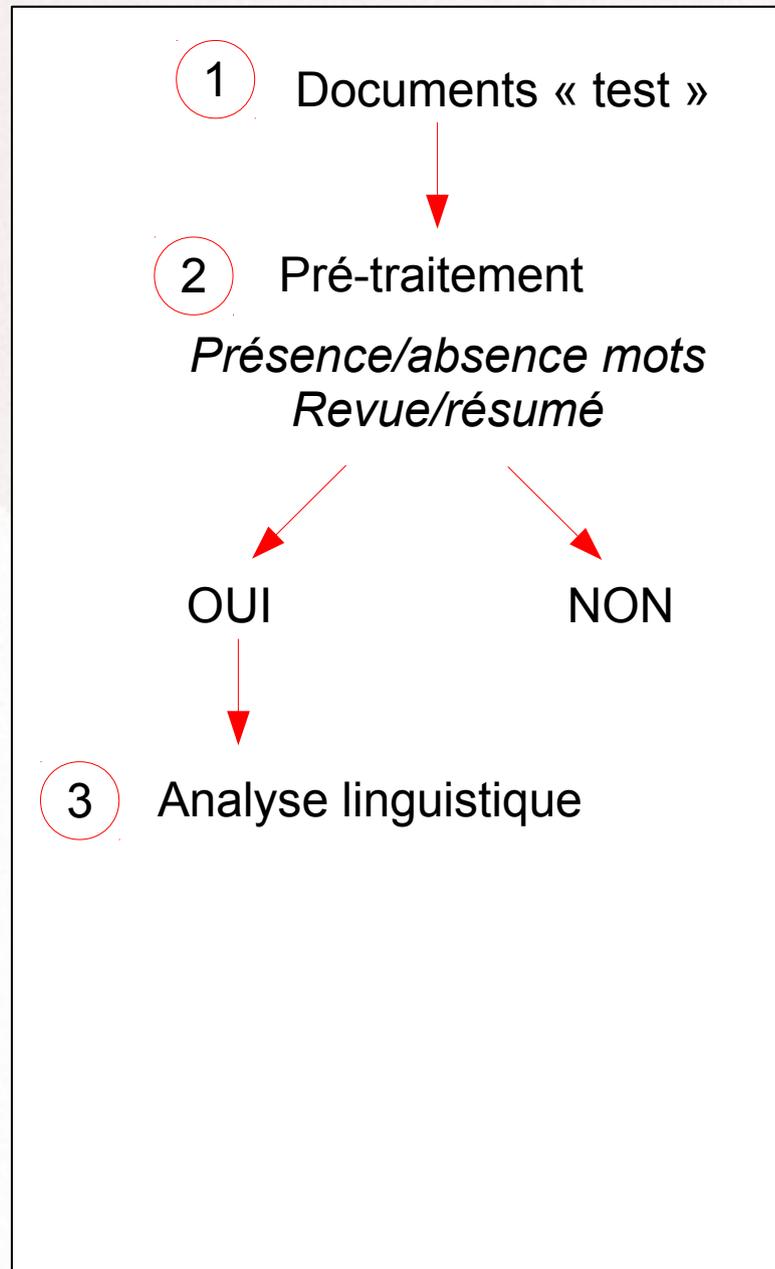


NON

# Procédure



# Procédure





# Analyse linguistique



## 1)Création des dictionnaires

Dictionnaire	Entrée
Biosystemic	co-elutions, co-elution.N+Biosystemic+experimentation:p confirm,confirm.V+Biosystemic+demonstration:P
Cell	CAKI-1,.Cell+kidney_carcinoma CCRF-CEM,.Cell+T_cell_leukemia
Compound	BAPTA/AM,.Compound:s:p carvedilol,.Compound:s:p
Protein	Spy1,Speedy protein A.Protein+Biosystemic:s:p Src family-associated phosphoprotein 1, Src kinase-associated phosphoprotein 1.Protein+Biosystemic:s:p



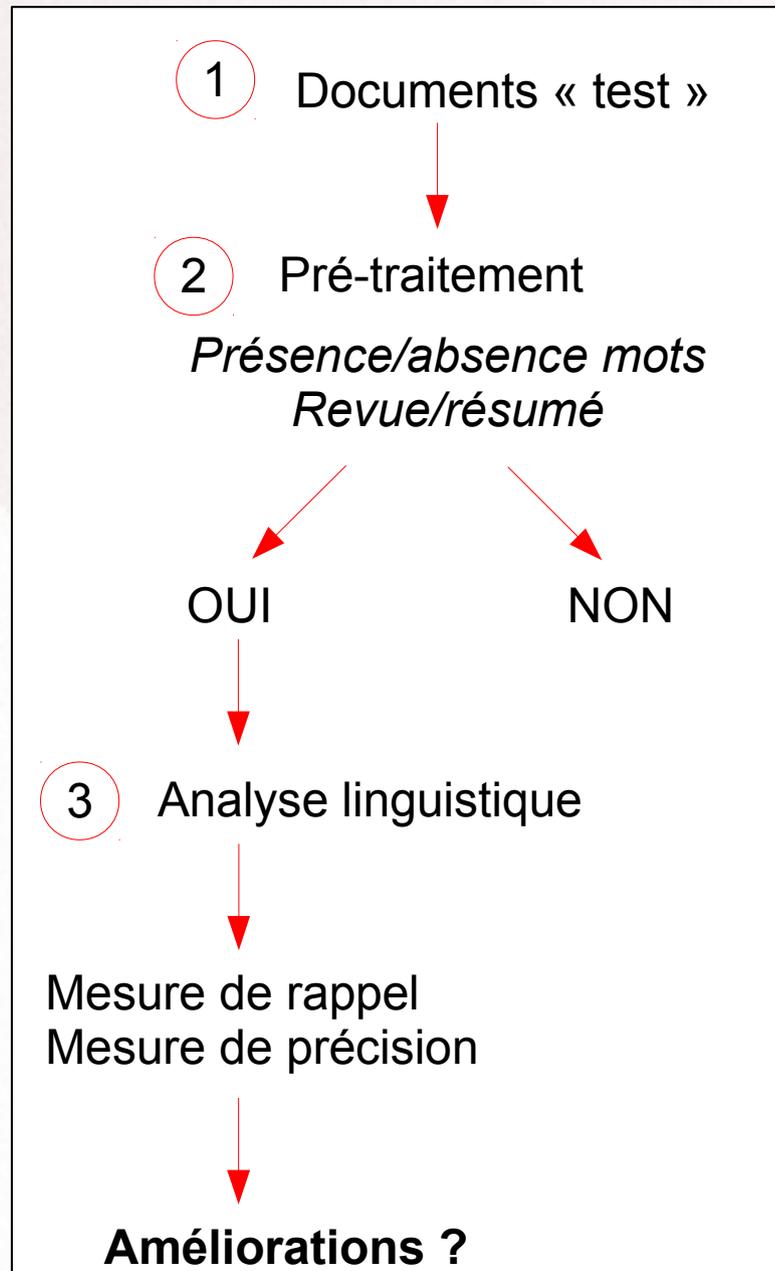
Seul Biosystemic a été créé **manuellement**







# Procédure



# *Rappel et précision*

## **Sélection des articles :**

- 27 textes choisis au hasard (divers en terme d'année de parution et journaux de publication) représentatifs du corpus;
- Analysés et vérifiés manuellement

n=27	Textes
Rappel	0,90
Précision	0,81

# Améliorations

## Analyse des phrases manquantes :

Mot erreur	Nombre de fois
Tested	5
et al	9
Analyzed	4
Assessed	4
When	7
Cell/cells	4
Previously	11
Because	4
Examined	3
Was/were used	3
Reported	4
To + infinitif	3
Already	3
We also	4

# Améliorations

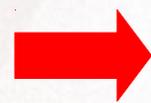
## Analyse des phrases manquantes :

**1/3 erreurs totales**

Mot erreur	Nombre de fois
Tested	5
<b>et al</b>	<b>9</b>
Analyzed	4
Assessed	4
<b>When</b>	<b>7</b>
Cell/cells	4
<b>Previously</b>	<b>11</b>
Because	4
Examined	3
Was/were used	3
Reported	4
To + infinitif	3
Already	3
We also	4

# Résultats

n=27	avant	après
Rappel	0,90	0,92
Précision	0,81	0,82



**Augmentation** du rappel de 2%  
**Sans perte** de la précision

# *Conclusion*

## Travaux réalisés :

- Textes au format XML/ HTML
- 1 282 documents => 62 655 phrases d'intérêt extraites
- Mesures de rappel et de précision peuvent être améliorées avec précaution

# *Conclusion*

## Travaux réalisés :

- Textes au format XML/ HTML
- 1 282 documents => 62 655 phrases d'intérêt extraites
- Mesures de rappel et de précision peuvent être améliorées avec précaution

## Travaux en cours :

- Transformer les phrases d'intérêt sous forme de faits reconnus par les règles
- Lancer le moteur d'inférence

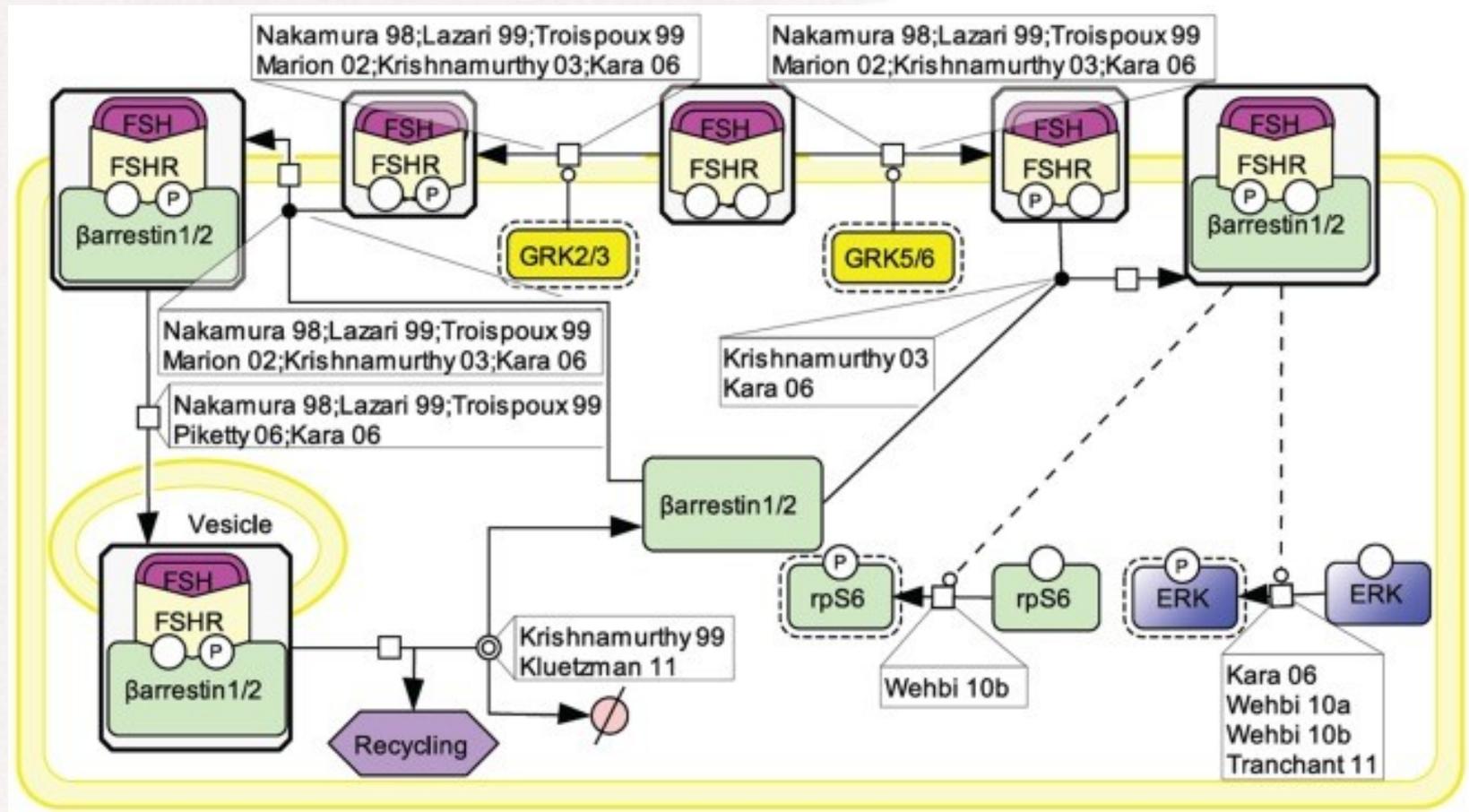
# *Perspective*

Analyse **transposable** à d'autres domaines car seuls **3/12** graphes sont spécifiques



***Merci pour votre attention***

# Contexte et Réseau de signalisation



# ***Format des textes***

n=15	PDF		
	brut	avec légende	sans légendes
Rappel	0,66	0,9	0,9
Précision	0,48	0,6	0,8

# ***Article scientifique***

## **Organisation d'un article scientifique en biologie:**

1) Abstract => principaux résultats

# ***Article scientifique***

## **Organisation d'un article scientifique en biologie:**

1) Abstract => principaux résultats

2) Introduction => contexte de l'étude + état de l'art

# ***Article scientifique***

## **Organisation d'un article scientifique en biologie:**

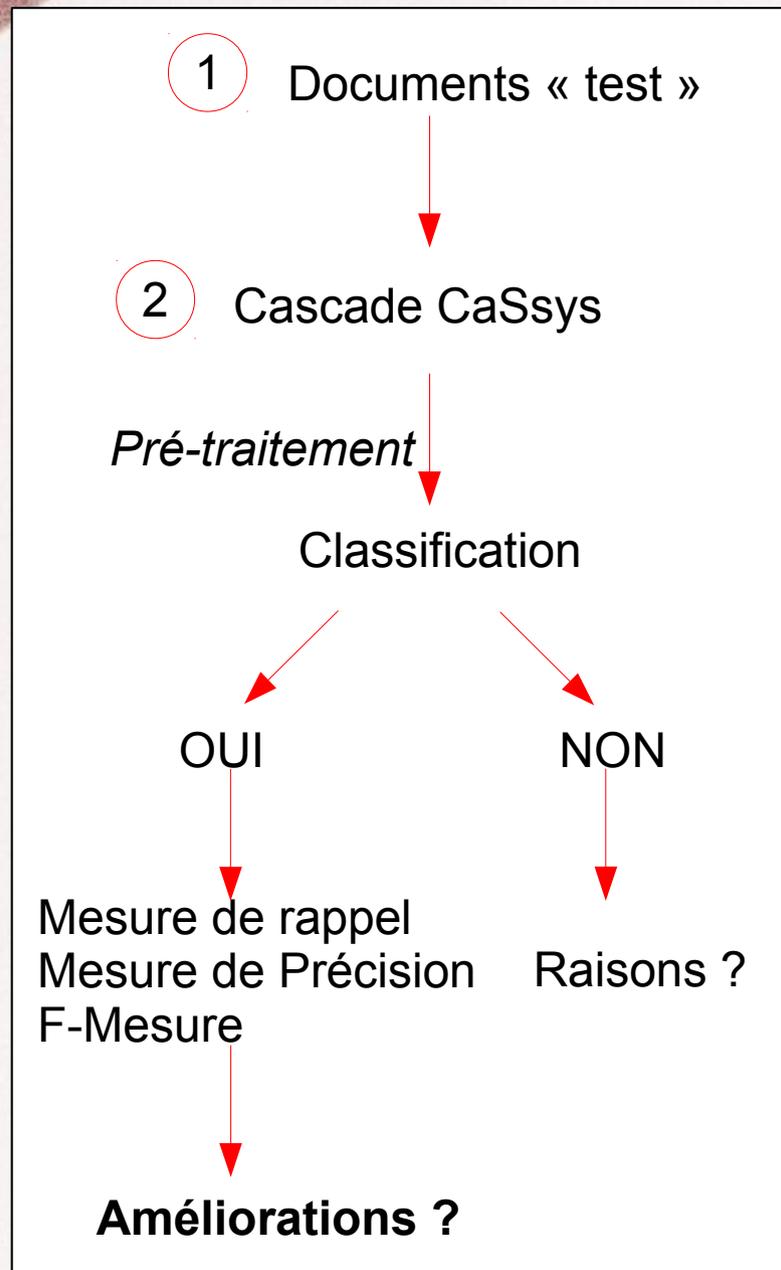
- 1) Abstract => principaux résultats
- 2) Introduction => contexte de l'étude + état de l'art
- 3) Résultats => liste et détail des expériences réalisées

# ***Article scientifique***

## **Organisation d'un article scientifique en biologie:**

- 1) Abstract => principaux résultats
- 2) Introduction => contexte de l'étude + état de l'art
- 3) Résultats => liste et détail des expériences réalisées
- 4) Discussion => interprétation des expériences

# Procédure



**Automatisation :**  
**Moteur d'inférence**





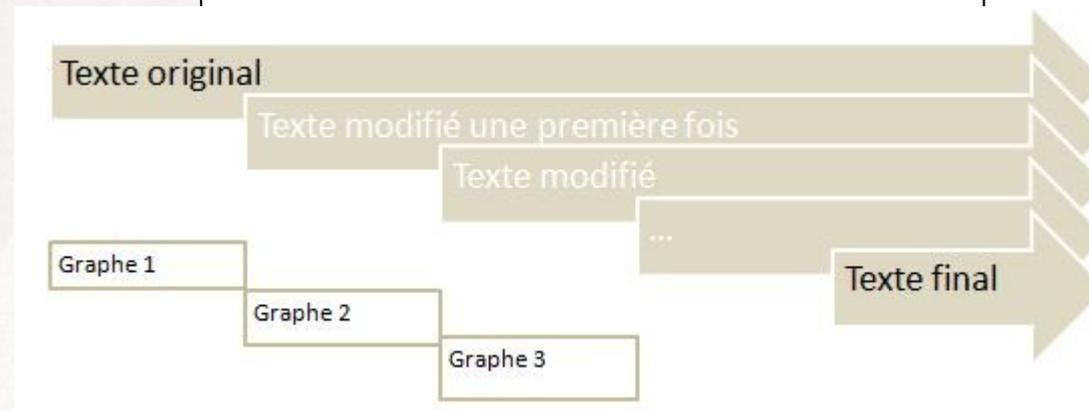
# Procédure



1 Documents « test »



2 Création des dictionnaires  
Création des graphes  
Création de la cascade



# Procédure



1 Documents « test »



2 Création des dictionnaires  
Création des graphes  
Création de la cascade



Analyse	Synthèse	Extraction
toolPhraseTei	suppression	extract
verbs	balisage	final
polylexical	exclusion	
test_bio		
no_verb		
no_protein		
sentenceResult		

# Procédure



1 Documents « test »



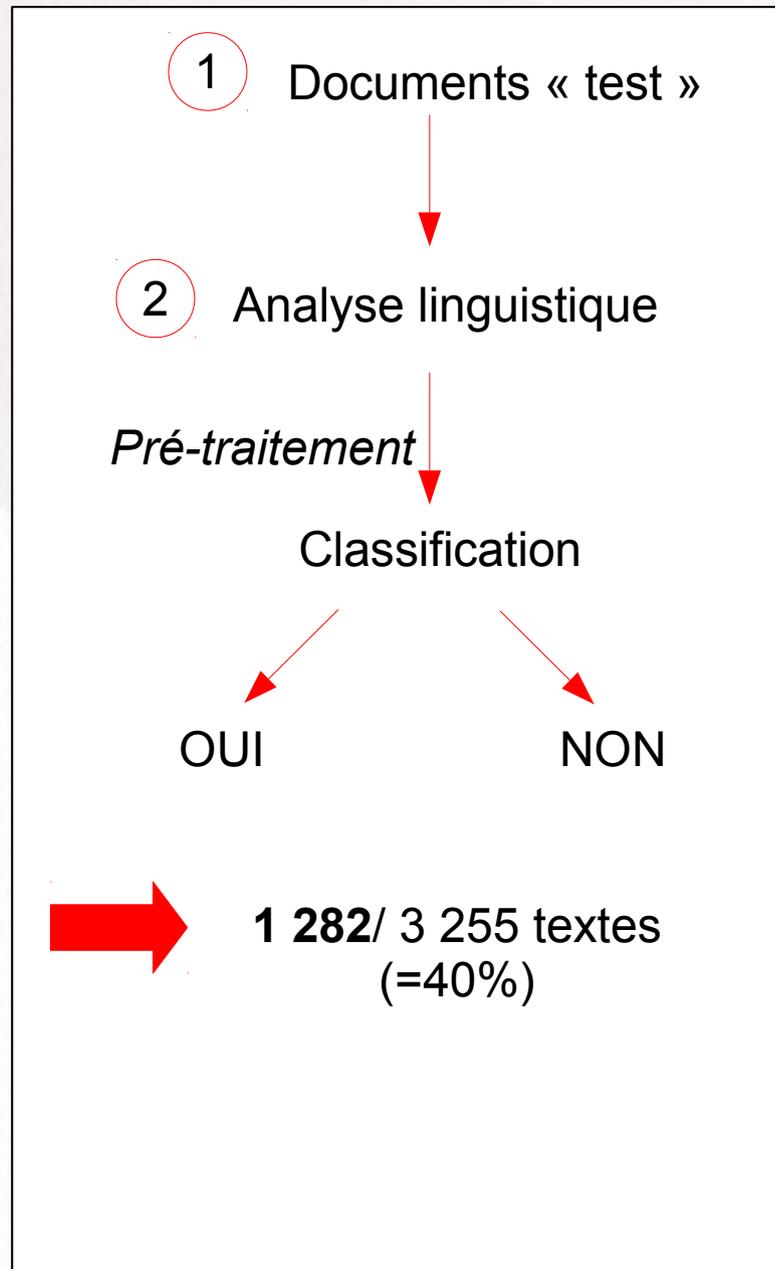
2 Création des dictionnaires  
Création des graphes  
Création de la cascade



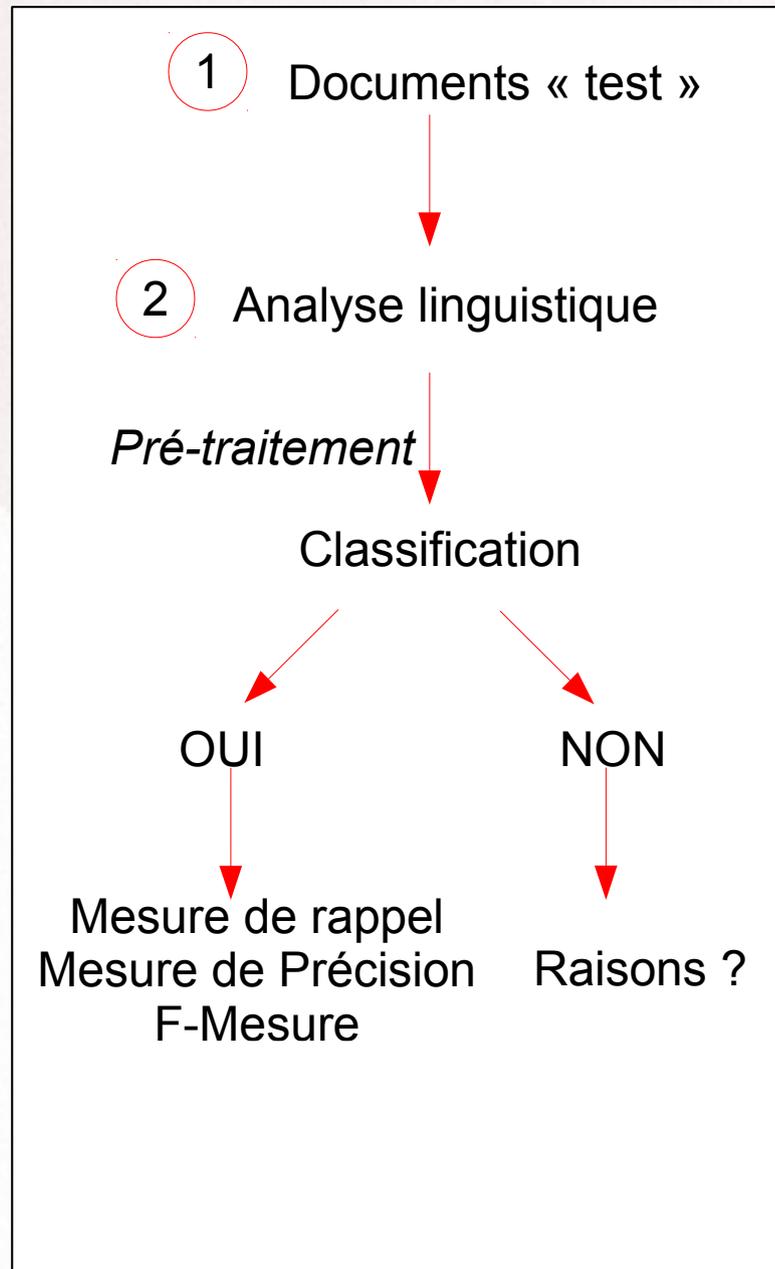
Analyse	Synthèse	Extraction
toolPhraseTei	suppression	extract
verbs	balisage	final
polylexical	exclusion	
test_bio		
no_verb		
no_protein		
sentenceResult		

**3/12**  
**spécifiques**

# Procédure



# Procédure



# Procédure

