

## Exemple 1

WordNet est construit sous la forme  
d'une hiérarchie de

`<terme>synsets</terme>`

`<definition>(ensemble de lexies  
synonymes entre elles)</definition>` ;

## Exemple 2

Nous montrons comment améliorer globalement les listes de voisins en prenant en compte <terme>la réciprocité de la relation de voisinage</terme>, <definition>c'est-à-dire le fait qu'un mot soit un voisin proche d'un autre et vice-versa.</definition>

Cette étape est réalisée grâce à <terme>un classifieur d'entropie</terme>, <definition>c.a.d, un outil statistique qui permet de calculer une adéquation (une entropie) entre les données qu'il reçoit à l'évaluation et les données sur lesquelles il a été entraîné (Daumé III, 2004).</definition>

Les marqueurs et leur fréquence  
dans le corpus annoté

une notion d'	1
sont	9
,	8
consiste à	5
est	15
.....	
, c'est-à-dire	10
:	14
(	48
)	57
.....	
.....	

## Marqueur 1, métalinguistique

- Les marqueurs métalinguistiques à utiliser indépendamment (au nombre de 9) :
- appeler,
- baptiser,
- définir comme,
- dénommer,
- dénoter,
- désigner,
- nommer,
- signifier,
- vouloir dire ;

## Marqueur 2, métalinguistiques nominaux

- Les marqueurs métalinguistiques nominaux :
  - appellation, acception, concept, dénomination, désignation, expression, mot, nom, notion, terme, vocable, à associer à un verbe support parmi :
    - appliquer,
    - donner,
    - employer,
    - prendre,
    - porter,
    - recevoir,
    - référer, renvoyer, réserver, utiliser ;

## Marqueur 3, de reformulation

**3, – Les marqueurs lexicaux n'étant pas explicitement métalinguistiques, ou ceux de reformulation (21) :**  
**c'est-à-dire,**  
**en d'autres termes,**  
**soit,**  
**à savoir,**  
**en quelques sortes,**  
**une sorte de,**  
**il s'agit de, entendre par,**  
**vouloir dire,**  
**indiquer, comme, dit,**  
**par exemple, autrement dit, même chose que, ;**

## Marqueur 4, ponctuation

- 4, – Les ponctuations :
- parenthèses,
- guillemets et
- tirets d'incise.



# marqueurs—>contextes lexicaux et syntaxiques

La finalité d'un patron lexico-syntaxique est de préciser les contextes lexicaux et syntaxiques dans lesquels un marqueur introduit bien l'une des relations sémantiques recherchées.

Par exemple, autour du marqueur «comme», nous en avons défini trois :

- SN définir MOTS\* comme SN ;
- SN comme DET{1,3} SN ;
- SN comme (celui|ceux|celle|celles) VERBE SN.



Ressource

# INFOMATIQUE et LINGUISTIQUE

## INFORMATIQUE:

Phrase Tokenizer (NLTK)

Chunker en français (Treetagger)

## LINGUISTIQUE:

Lexique des marqueurs

# Prétraitement

1, Lire un fichier——> `fichier.split('Introduction')`

2, Tokenize en phrase, filtrer par des conditions(longueur, présence de marqueur).

3, Obtenir des phrases candidates (559 phrases).

# Pseudocode

- 1, Passer chaque phrase au chunker
- 2, Supprimer ‘\t’, ‘\n’ dans le résultat
- 3, Split le résultat avec le marqueur
- 4, Condition: où sont mes chunks ‘<NP>.....</NP>’ ?  
Proche ( splitlist [0] [-70:] et splitlist [1] [70:]) ou loin ?
- 5, Conditions spéciales pour parenthèse:
  - a, pas de <VN>...</VN>;
  - b, marqueur ‘,’ ‘est’, plus exigeant au niveau de la distance marqueur-NP
  - c, est-ce qu’il est a des marqueurs de reformulation? —> 4

# Exemple qui marche I

La vectorisation est une technique de plongement(embedding) permettant de projeter un calcul de similarité quelconque entre deux documents (ou un document et une requête pour la RI) dans un espace vectoriel.

<s><NP>La DET:ART levectorisation NOM <unknown></NP><VN>est VER:pres être</VN><NP>une DET:ART untechnique NOM technique</NP><PP>de PRP de<NP>plongement(embedding) NOM <unknown></NP></PP><VN>permettant VER:ppre permettre</VN><VPinf>de PRP de</VPinf><VN>projeter VER:infi projeter</VN><NP>un DET:ART uncalcul NOM calcul</NP><PP>de PRP de<NP>similarité NOM similarité</NP></PP><AP>quelconque ADJ quelconque</AP><PP>entre PRP entre<NP><NP>deux NUM deuxdocuments NOM document( PUN (</NP><COORD>ou KON ou</COORD><NP>un DET:ART undocument NOM document</NP></NP></PP><COORD>et KON et</COORD><NP>uneDET:ART unrequête NOM requête</NP><PP>pour PRP pour<NP>la DET:ART leRI NOM <unknown>) PUN )</NP></PP><PP>dans PRP dans<NP>un DET:ART unespace NOM espace</NP></PP><AP>vectoriel ADJ vectoriel</AP>. SENT .</s>

# Exemple qui marche II

<s><NP>La DET:ART levectorisation NOM  
<unknown></NP><VN>est VER:pres être</  
VN><NP>une DET:ART untechnique NOM technique</  
NP><PP>de PRP de<NP>plongement(embedding)  
NOM <unknown></NP></PP><VN>permettant  
VER:ppre permettre</VN><VPinf>de PRP de</  
VPinf>..... SENT .</s>

# Exemple qui ne marche pas

Nous montrons comment améliorer globalement les listes de voisins en prenant en compte la réciprocité de la relation de voisinage, c'est-à-dire le fait qu'un mot soit un voisin proche d'un autre et vice-versa.

<s><VN>Nous PRO:PER nousmontrons VER:pres montrer</VN><VPinf>comment ADV comment</VPinf><VN>améliorer VER:infi améliorererglobalement ADV globalement</VN><NP>les DET:ART lelistes NOM listede PRP devoisins NOM voisin</NP><VPpart>en PRP en</VPpart><VN>prenant VER:ppre prendre</VN><PP>en PRO:PER en<NP>compte VER:pres compter</NP></PP><NP>la DET:ART leréciprocité NOM réciprocité</NP><PP>de PRP de<NP>la DET:ART lerelation NOM relation</NP></PP><PP>de PRP de<NP>voisinage NOM voisinage, PUN ,</NP></PP><VN>c' NOM <unknown>est-à-dire ADV est-à-dire</VN><NP>le DET:ART lefait NOM fait</NP><AP>qu' ADJ <unknown></AP><NP><NP>un DET:ART unmotNOM mot</NP><COORD>soit KON soit</COORD><NP>unDET:ART unvoisin NOM voisinproche ADJ proche</NP></NP><AP>d' ADJ <unknown></AP><NP>un DET:ART unautre ADJ autre</NP><COORD>et KON etvice-versa ADV vice-versa</COORD>. SENT .</s>

# Evaluation I

15 articles tirés au sort;  
109 définitions gold-standard.

Paramètre : distance 70 pour les marqueurs; 50 au sein de la parenthèse.

85 phrases trouvées, 15 qui contiennent la définitions.

Précision = 17.64%

Rappelle = 13.76%

F-mesure = 15.46%



# Evaluation II

**15 articles tirés au sort;  
109 définitions gold-standard.**

**Paramètre : distance 120 pour les marqueurs; 70 au sein de la parenthèse.**

**273 phrases trouvées, dont 50 qui contiennent la définitions.**

**Précision = 18.31%**

**Rappelle = 45.87%**

**F-mesure = 25.92%**