



Vers une extraction automatique des événements dans les textes

Béatrice ARNULPHY Xavier TANNIER Anne VILNAT

{Beatrice.Arnulphy, Xavier.Tannier, Anne.Vilnat}@limsi.fr

*Colloque international
Langage, discours, événements*

31 mars, 1-2 avril 2011

1. Événements

2. Entités Nommées Événement

3. Utilisation de lexiques

4. D'autres indices à suivre :

1. Indicateurs temporels

2. Verbes

Les événements verbaux

De Vendler [1967] à TimeML

Le verbe est porteur d'informations qui permettent de caractériser des événements : aspect, temps, etc.

TimeML [Pustejovsky, 2003]

Formalisme d'annotation des événements

But : extraire les événements (principalement verbaux), les expressions temporelles et la relation entre les deux

Annotation manuelle / Extraction automatique sans résultats très satisfaisants

Corpus annotés :

TimeBank (anglais)

FR-TimeBank (français)

IT-TimeBank (italien)

Événements verbaux vs nominaux

Verbes

Rôle principal de représentation d'événements (hors verbes d'état)

Noms

- Non porteurs de toutes les informations véhiculées par le verbe
- Indiquer un événement : pas une qualité unanime du nom
- Bipolarité de l'événementialité des noms événements

Les événements nominaux

Intérêt à détecter les noms événements

La nominalisation des événements est à priori lié à leur importance.

Exemple : la dénomination « 11 septembre »

Des bombes ont éclatées / Une explosion a retenti

Des avions se sont écrasés dans le WTC

Un attentat terroriste

Les attaques du 11 septembre / Le 11 septembre

Intérêt pour le TAL

Extraction d'information

Analyse d'un texte en surface dans le but d'une application précise

Exemple d'une application pratique : projet Chronolines (ANR)

Partenaires : MoDyCo, LIMSI, AFP, XRCE, Exalead

Buts :

- Ordonner les événements sur un axe temporel
- Proposer, dans la chronologie d'un thème, les événements les plus importants

1. Événements

2. **Entités Nommées Événement**

3. Utilisation de lexiques

4. D'autres indices à suivre :

1. Indicateurs temporels

2. Verbes

Quaero

Projet Quaero : projet de recherche européen sur 5 ans

ILES : Questions Réponses (3.5) et Entités Nommées (3.2)

La tâche Entités Nommées

Définition : Entités nommées

Éléments « notables » des textes, comme les noms de personnes et de lieux

- Détection d'entités nommées ou constitution de ressources (patrons ou règles d'extraction, modèles, dictionnaires) en rapport avec les entités nommées
- Reconnaissance des différentes mentions d'une même entité dans un document ("le président de la France" = "Nicolas Sarkozy")

Reconnaissance des EN-É

Définition : Entités nommées étendues [Grouin et al., 2011]

Extension des entités nommées à de nouveaux types
(e.g. civilisations, les fonctions, etc.)

Extension de la définition des entités nommées à des expressions
construites autour de noms communs : autorisation d'inclusion
d'expressions ne contenant aucun nom propre

- Les EN à proprement parler
 - monoréférentialité (Enjalbert, Vicente) ;
 - entité du monde concret (Ester) ;
 - définies comme des noms propres (MUC-7)

Le festival de Cannes

La seconde guerre mondiale

- Les désignations nominales des événements

Le festival s'est déroulé en juillet.

L'explosion d'un réacteur nucléaire

Composition des EN-É

Trois possibilités :

1. Des noms déverbaux / dérivés de verbes

Le 21 juin, c'est la **fête de la musique**.

L'**adoption** par le Parlement d'une loi [...]

2. Des éléments qui évoquent des événements de façon non ambiguë

le **Festival du film de Berlin**

Composition des EN-É

3. Des mots qui prennent un caractère événementiel en contexte

(polysémie)

La **cinquième édition du salon de l'éducation**
L'**organisation du procès** dans les 60 jours

(métonymie)

Les frégates de Taïwan s'invitent à Lorient.

(- toponyme)

Personne ne veut d'un nouveau **Tchernobyl**.
Copenhague se solde par un échec.

(- héméronyme)

Les indemnisations pour le **11 Septembre**
On pourrait assister à un **21 avril** à l'envers.

Quelques exemples EN-É

Les entités uniques

La Grande Guerre

Celles plus récurrentes

Le festival de Cannes

L'instanciation de ces phénomènes

Les JO de 96

Les événements plus anodins / flous

la descente de police de demain

le branle-bas de combat mondial

le débat

la décision

Annotation manuelle de corpus

Corpus journalistique

Le Monde (2001-2002) + L'Est Républicain

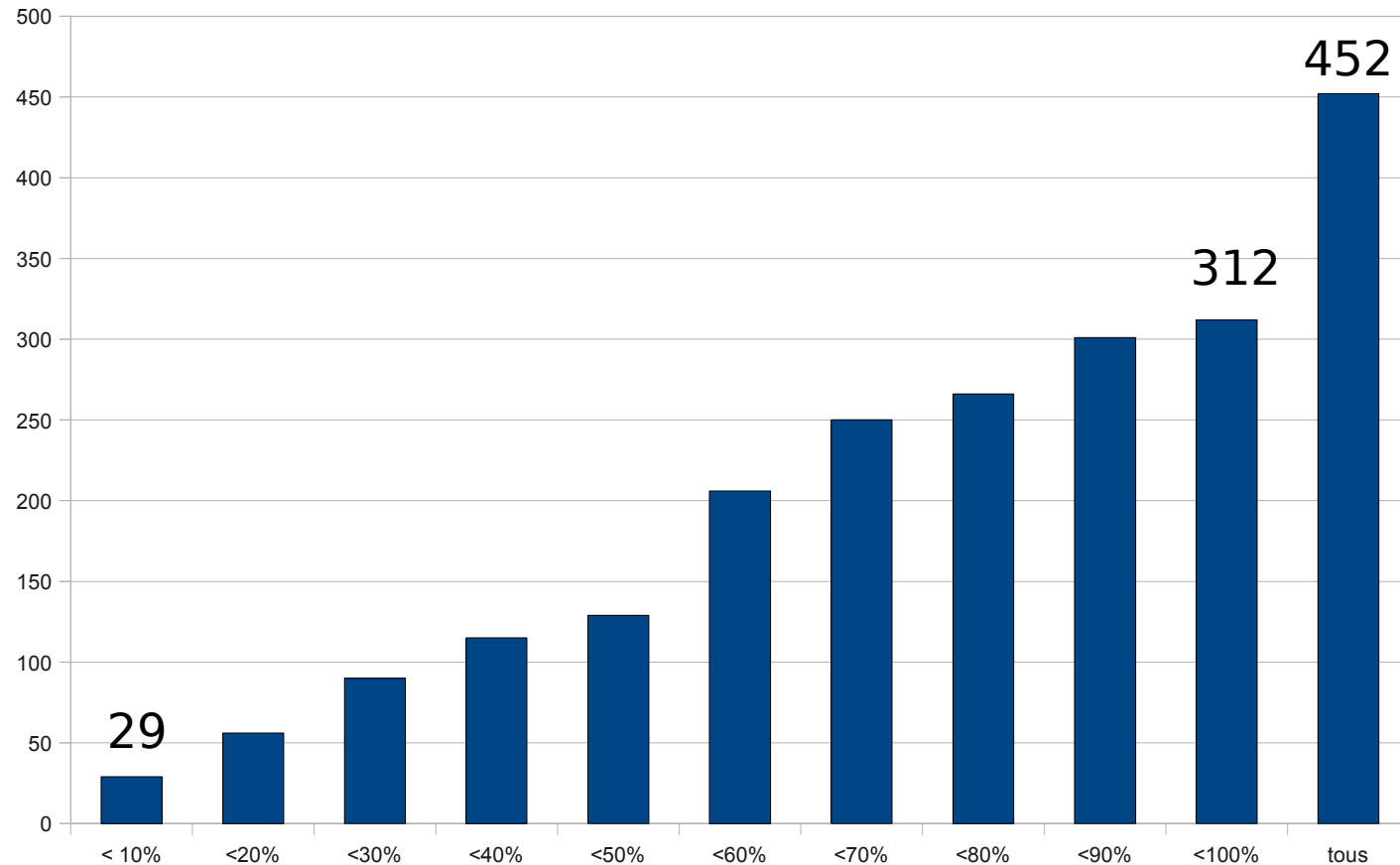
725 noms différents et 269 noms d'une seule occurrence

	Le Monde	<u>L'Est</u> Républicain	total
articles	83	109	192
mots	31449	16197	47646
événements	1105	736	1841
événement/mots	3,51%	4,54%	3,86%

Répartition du corpus et des annotations

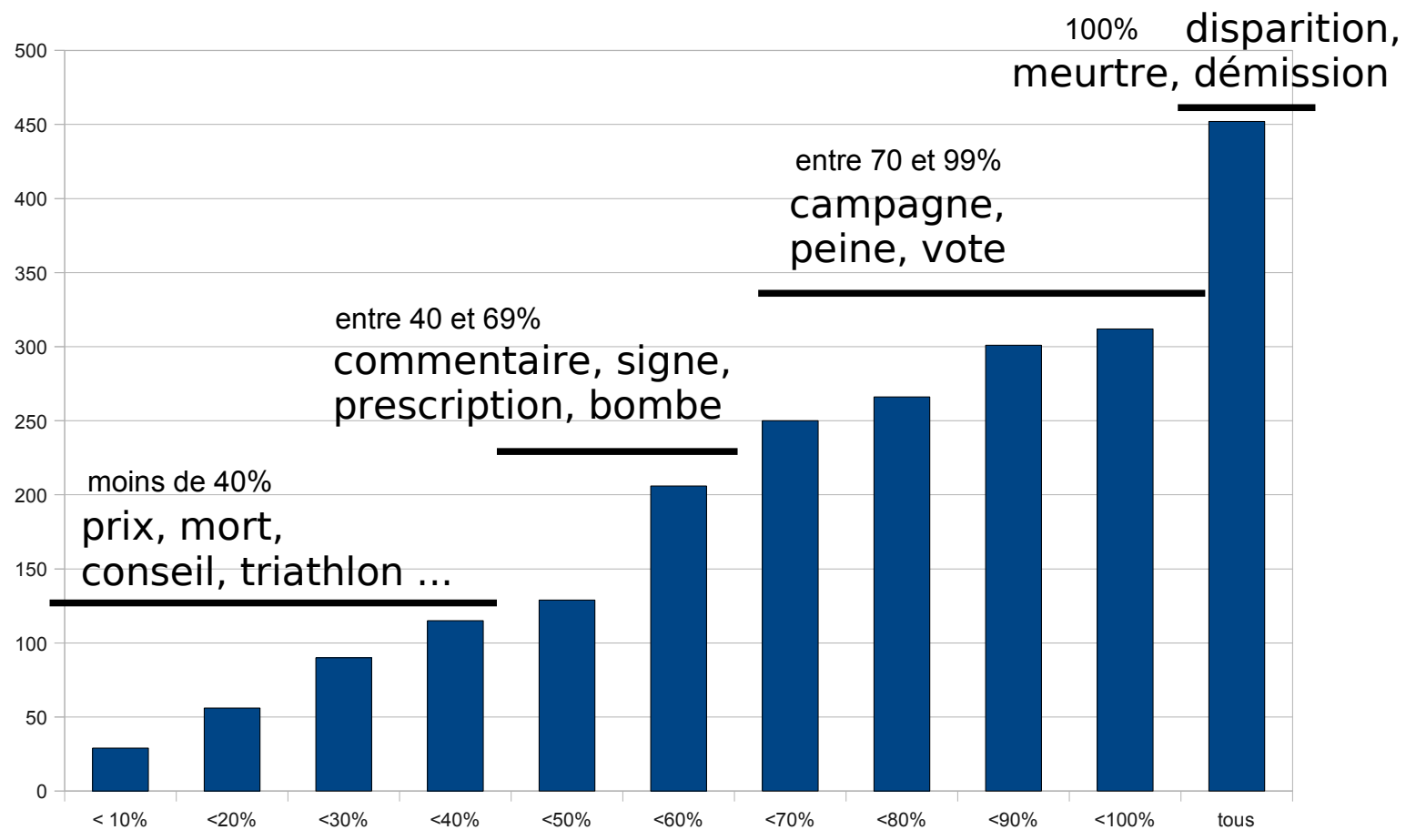
Guide d'annotation

Statistiques sur le corpus annoté



Progression du nombre de noms événements
par rapport au nombre d'occurrences total de noms événements

Statistiques sur le corpus annoté



Progression du nombre de noms événements par rapport au nombre d'occurrences total de noms événements



1. Événements

2. Entités Nommées Événement

3. Utilisation de lexiques

4. D'autres indices à suivre :

1. Indicateurs temporels

2. Verbes

Les lexiques existants

Lexique de noms déverbaux (nominalisations de verbes d'action ou de procès)

VerbAction [Tanguy et Hathout, 2002]

9393 couples verbes-lemme – 9200 lemmes nominaux uniques

(action/événement)

sensibilisation, victimisation

(ambigus)

aération, étalage, shampoing, voeu

Lexique complémentaire au VerbAction : noms ayant au moins une fois une interprétation événementielle

Lexique des noms événementiels [Bittar, 2010]

746 noms d'événement

(non-déverbaux)

anniversaire, grève

(lexique spécifique)

anticoagulothérapie

(ambigus)

apéritif

Les lexiques existants

Deux difficultés :

– Ambiguïté :

- Les déverbaux peuvent désigner l'événement ou le résultat.

La **construction** du port a duré 50 ans. (événement)

Cette **construction** fait 150 mètres de haut. (résultat)

- Difficulté même pour l'annotateur humain

L'**étalage de marchandise** a été interdit sur le port.

Soumettre/ présenter une **proposition de loi**

– Les mots autres que polysémiques qui prennent leur caractère événementiel en contexte (toponymes, héméronymes, etc.)

Constat : les lexiques seuls ne suffisent pas.

1. Événements

2. Entités Nommées Événement

3. Utilisation de lexiques

4. **D'autres indices à suivre :**

1. Indicateurs temporels

2. Verbes

D'autres indices à suivre

- Les indicateurs temporels
- Les verbes
- Utilisation de XIP (XRCE) : analyseur syntaxique robuste
 - Dépendances syntaxiques
 - Reconnaissance d'EN (personne, lieu, organisation)

Possibilité pour l'utilisateur d'implémenter ses propres règles de grammaire en plus de celles existantes

Les indicateurs temporels (IT)

- Ancrés dans le temps, les événements peuvent être utilisés avec des prépositions temporelles, dans des compléments de temps.
 - le fait que l'événement se produise
à l'occasion de, lors de
A Jérusalem , lors de la **réunion du gouvernement israélien**
 - usage référentiel de l'événement
pendant, après, la veille de, le lendemain de
La population a été évacuée avant l'**arrivée de la lave**.
 - un moment de l'événement
à l'issue de, au commencement de
les activistes qu'ils ont libérés au début de l'**Intifada**.

Les verbes (VB)

- Les verbes d'événement et de cause/conséquence
[Arnulphy et al., 2010]

Des événements comme conséquence ou cause d'autres

Les **crises** ont pour origine des problèmes de défaillance technique.

Cette élection entraînera-t-elle la **mise en sourdine des intérêts communaux** ?

[...] a provoqué un **tollé** chez les organisations amérindiennes.

Les verbes qui introduisent des événements

Le **Salon de l'Agriculture** est organisé Porte de Versailles

Les **matches de huitièmes et de quart de finale** ont eu lieu devant plus de 5000 personnes.

Le général, qui assistait à une **cérémonie** au côté du chef de l'État, s'est déclaré très content de sa décision.

Les règles XIP : VB et IT

Les règles sur les IT et les VB permettent de récupérer de nouveaux noms d'événements potentiels.

mise en sourdine, tollé

depuis le **couac de mardi**, la presse japonaise s'est largement étendue sur les négociations en cours.

La période ouverte par le **11 septembre** signe la fin de cette hégémonie sur le monde.

Mai 68 a précipité *sa disparition*.

	Proportion d'événements corrects par rapport aux annotés	Proportion d'événements corrects par rapport aux attendus
IT	81%	6%
VB90	84%	1%
VB90 + IT	82%	7%

Résultats obtenus au moyen des règles XIP

Récapitulatif des tests

	Proportion d'événements corrects p/r aux annotés	Proportion d'événements corrects p/r aux annotés
VerbAction	38%	67%
VerbAction + Bittar	39%	84%
IT	81%	6%
VB90	84%	1%
VB90 + IT	82%	7%

- Règles : précision +, couverture -
- Lexiques : précision -, couverture +
- Combinaison des deux

Perspectives

Désambiguïsation

Avant d'intégrer dans un lexique ces nouveaux déclencheurs et désignations d'événements, il faut pouvoir les désambiguïser.

Accorder un indice de confiance pour chaque mot/expression

Outil fonctionnel...

Merci !

- Questions ?
- Collaborations ?

Tableaux en plus ...

[Grimshaw, 1990] Détection de l'événementialité

Les pluriels des déverbaux sont des événements :

pas significatif pour les noms d'événements (déverbaux et autres)

Forme	Noms événements	Tous les noms
singulier	80,10%	83,40%
pluriel	19,90%	16,60%

Taux de singuliers et de pluriels des noms

Tableaux en plus ...

[Grimshaw, 1990] Détection de l'événementialité

L'article défini indique un événement, les démonstratifs des résultats :
pas significatif pour les noms d'événements (déverbaux et autres)

Forme	Noms événements	Tous les noms
article défini	27,90%	19,90%
article indéfini	14,30%	6,20%
démonstratif	4,00%	1,70%
posséssif	6,10%	3,30%

Taux d'événement selon le type de déterminant

Tableaux en plus ...

Détail des occurrences

	REF total	HYP total	Commun	Précision	Rappel	F-mesure
IT	1841	138	112	81%	6%	0,11
VB90	1841	25	21	84%	1%	0,02
VB90 + IT	1841	163	133	82%	7%	0,13
VB	1841	85	58	68%	3%	0,06
VB + IT	1841	223	170	76%	9%	0,16

Résultats obtenus au moyen des règles XIP

Étude des événements

Histoire, philosophie ou journalisme :
de nombreuses études sur la notion d'événement

Linguistique :
événements nominaux : typologies et définitions
événements verbaux

Informatique linguistique :
description et annotation des événements de type verbal

Guide

Conseil suivi pour l'annotation :

- Permutation des mots ambigus avec des mots événements plus sûrs.
Preuve : formalité (non événement) ou surprise (événement)
Les trente-sept journalistes français attendus ici en sont , s'il en était besoin , une **preuve supplémentaire**.
- Les items d'énumérations sont souvent (mais pas toujours) du même type.
- Lors d'une ambiguïté trop difficile, préférer ne pas annoter le mot.