

# Reconnaissance automatique de désignations nominales d'événements

Séminaire du Cental

Béatrice Arnulphy

Univ. Paris-Sud 11 LIMSI-CNRS

18 novembre 2011  
Louvain-La-Neuve, Belgique



Béatrice Arnulphy

<http://www.limsi.fr/Individu/arnulphy/>

- Doctorante au LIMSI-CNRS
- École Doctorale d'Informatique de l'Université Paris-Sud 11
  - Directrice de thèse : Anne Vilnat
  - Encadrant de Recherche : Xavier Tannier
- Sujet de thèse :  
Reconnaissance automatique de désignations nominales  
d'événements

## Extraction d'Information

- But général : Analyse de surface d'un texte dans le but d'une application précise.
  - L'**explosion de la centrale** = Le **festival de Cannes**
- La détection des descriptions d'événements est importante.
- Les entités nommées plus générales comme les noms de lieux ou de personnes sont plus fréquemment étudiées en extraction d'information que les noms d'événements

# Intérêt pour le TAL

## Exemple d'application pratique : le projet ANR Chronolines

- 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

D'autres domaines du TAL sont aussi intéressés par la reconnaissance des événements :

- Résumé automatique, veille, agrégation de résultats (de news par exemple), ...

# Outline

- 1 État de l'art
  - Définition de l'événement
    - En linguistique
    - En TAL
  - Extraction automatique
  - Ressources existantes
    - Lexiques existants
    - Les corpus disponibles
- 2 Les événements nominaux
- 3 Notre Corpus annoté
- 4 Un lexique pondéré

# Définitions

## L'événement médiatique

- Neveu & Quéré (1996)
  - Une occurrence singulière, imprévue, non répétable et survenu dans un passé plus ou moins proche
  - Aucun intérêt n'est porté à la dénomination.
- nous* Les événements prévus ou non, tout autant que ceux se tenant à d'autres moments.

## L'objet linguistique

- Velde (2000) : Introduction à la notion générale de "triade" je-ici-maintenant (en lien avec les entités nommées)
- Les événements métonymiques :
 

Lecolle (2004) :	les toponymes événementiels	Tchernobyl
Calabrese (2008) :	les héméronymes	11 septembre

# Définitions

## TimeML (Pustejovsky, 2003)

- Un langage de spécification pour les événements et les expressions temporelles créé au départ pour améliorer les performances de systèmes questions-réponses
  - L'événement est : "a cover term for situations that happen or occur".
    - Qualité de ponctualité ou de durée
    - Peut décrire des états
- Un intérêt principalement porté sur les verbes  
*nous* Que les noms, pas les états
- De nombreux corpus ont été développés sur plusieurs langues.

# Définitions

## ACE (Automatic Content Extraction)

- Une classification précise et détaillée des événements
- Un nombre limité de catégories thématiques  
life ← be-born, die, be-injured, ...
- Des arguments spécifiques sont liés à des événements particuliers, comme "origin" et "destination" pour le thème "transportation".
- Objectifs de ACE : détection d'événements issus de thèmes précis et leur classification

*nous* Toutes les mentions d'événements sans classe thématique donnée

## Les campagnes ESTER (entités nommées)

- Une typologie simple  
historiques et uniques vs. répétitifs

*nous* Une vision des événements proche de la nôtre

# Extraction automatique d'événements nominaux

- Les applications pour l'extraction automatique sont principalement pour l'anglais
  - Selon les spécifications TimeML ou les définitions ACE des événements
- Des études pour la désambiguïsation des noms événementiels dans plusieurs langues
- Les méthodes suivies :
  - Modules statistiques / Approches à base de règles linguistiques
- Références:

EN :	(Sauri et al., 2005)	DE :	(Eberle, 2009)
	(Creswell et al., 2006)	IT :	(Russo, 2011)
		ES :	(Peris, 2010)

# Ressources existantes

- Les lexiques existants
  - VerbAction
  - Le lexique alternatif des noms événementiels
  - WordNet
  
- Corpus
  - TimeBank1.2
  - FR-TimeBank

# Les lexiques existants en français

## VerbAction

- Une liste de verbes d'action accompagnée des **noms déverbaux** morphologiquement apparentés à ceux-ci [Tanguy & Hathout, 2002]
- 9 393 couples verbes – lemmes  
9 200 lemmes nominaux uniques  
fêter – fête
- Composé de façon semi-automatique et validé manuellement
  - (événement/action) ablation, sensibilisation, victimisation
  - (ambigus) aération, étalage, shampooing, vœu

# Les lexiques existants en français

## Le lexique alternatif des noms événementiels [Bittar, 2010]

- Complémentaire au *VerbAction* pour l'extraction d'événements nominaux
- Noms ayant **au moins une** interprétation événementielle
- 804 noms événementiels

(non-déverbaux)

anniversaire, grève

(lexique spécifique)

anticoagulothérapie

(ambigus)

apéritif

# Les lexiques existants en français

## Deux difficultés

- Ambiguïté :
  - Les déverbaux peuvent désigner l'événement ou le résultat.  
`La construction(evt) du port a duré 50 ans.`  
`Cette construction(resultat) fait 150 mètres de haut.`
  - Difficulté même pour l'annotateur humain  
`L'étalage de marchandise(?) a été interdit sur le port.`  
`Soumettre/ présenter une proposition(?) de loi`
- Absence des mots, autres que polysémiques, qui prennent leur caractère événementiel **en contexte** (toponymes, héméronymes, ...)

# Le lexique disponible en anglais

## Les noms d'action et d'événements de WordNet

- 5 903 noms étiquetés dans WordNet comme act (action) ou event (événement)
- Une liste comparable aux lexiques français (VerbAction et Bittar)

des noms fortement événementiels  
 des expressions ambiguës  
 des expressions multi-mots (MWE)  
 des noms dévénements

des expressions autres

war, election, show  
 arts and crafts, bet  
 a cappella singing  
 Arab-Israeli War,  
 Battle of Britain,  
 laser trabecular surgery  
 ? Attorney General, judo,  
 ? industry

# Les corpus disponibles

## Des corpus annotés selon TimeML

- Des corpus d'articles journalistiques
- TimeBank 1.2
  - En anglais
  - Sources : *Wall Street Journal*, *ABC*, *CNN*, *Voice Of America*
  - 1 722 événements nominaux qui ne sont pas étiquetés dans le corpus comme des états
- FR-TimeBank
  - En français
  - Source : *L'Est-Républicain*
  - 663 événements nominaux

# Outline

- 1 État de l'art
- 2 Les événements nominaux
  - Notre définition
  - Composition du nom événementiel
- 3 Notre Corpus annoté
- 4 Un lexique pondéré

# Notre objet d'étude : l'événement

## Définition de la notion d'événement

- L'événement est :
  - Ce qui survient
  - Le changement d'état opéré
- Il peut être :
  - Récurrent ou unique
  - Prévu ou non
- Il peut durer ou être instantané.
- Il a peut aussi se produire dans le passé, le présent ou le futur.

# Événements verbaux vs. nominaux

La plupart des événements → dans des verbes et des noms.

- Les événements verbaux (Vendler, 1967), (Pustejovsky, 2005)
  - Les verbes désignant intrinsèquement une éventualité
  - Plus nombreux, plus simples à identifier
  - Simples à lier aux expressions temporelles
  - Désignent souvent des événements plus “communs” et moins pertinents. Objet d'étude principal en TAL.
- Les événements nominaux
  - Nominalisation des événements liée à leur importance
  - Plus difficiles à repérer que les événements verbaux
  - Les études en Sciences Humaines (sociologie, linguistiques) s'intéressent principalement aux événements nominaux.

## Notre constat général

- Les lexiques existants sont des listes de mots.
- Presque tous les noms peuvent acquérir une lecture événementielle en contexte.

## Notre objectif

- Évaluer les lexiques existants
- Construire un lexique de désignations nominales d'événements qui fournirait **une information quantitative de l'événeementialité** des noms du lexique (aide à la désambiguisation)

## Composition du nom événementiel

# Composition du nom événementiel

# Composition du nom événementiel

## ① Nominalisation apparentée à un verbe d'action

fêter – fête

La fête de la musique

adopter – adoption

l'adoption d'une loi au Sénat

→ La désambiguïsation est nécessaire dans toutes les langues entre action et résultat/instrument/lieu de l'action

construction, exil

→ action ou résultat

shampooing

→ action ou instrument (produit)

arrivée

→ action ou lieu

# Composition du nom événementiel

## ② Nominalisation autre que dérivée de verbes

Des noms qui évoquent intrinsèquement des événements

le Festival du film de Berlin

le match PSG-OM

→ La désambiguïsation est parfois nécessaire en français

le salon<sub>(evt)</sub> du livre

le salon<sub>(obj)</sub> de ma grand-mère

Il a rapporté le triathlon<sub>(evt)</sub> des lacs.

Il est le nouveau champion du monde de triathlon<sub>(sport)</sub>.

# Composition des noms d'événements

## ③ Nominalisation métonymique

Des noms ou groupes nominaux qui prennent un caractère événementiel en contexte

→ lieu **Tchernobyl** → toponyme [Lecolle2009]  
(l'accident nucléaire de 1986)

→ date **11 septembre** → héméronyme [Calabrese2008]  
(l'attaque terroriste de 2001 aux États-Unis)

→ objet Les **frégates de Taïwan** s'invitent à Lorient.

# Outline

- 1 État de l'art
- 2 Les événements nominaux
- 3 **Notre Corpus annoté**
  - Guide d'annotation
    - Les Entités Nommées Quaero
    - Notre typologie des événements
  - Présentation du corpus
    - Observations sur le corpus annoté
- 4 Un lexique pondéré

# Guide d'annotation

## Les entités nommées

Éléments “notables” des textes, comme les noms de personnes et de lieux.

## Guide d'annotation entités nommées dans Quaero

- Les entités nommées définies de façon étendues :
  - Classiques :  
lieux, personnes, organisations
  - Raffinées par des composants :  
org.adm, org.ent
  - De nouvelles entités : noms de fonctions et de métiers, événements

```
les <func.coll> <kind>maires</kind> de  
<loc.adm.nat>France</loc.adm.nat> </func.coll>
```

# Guide d'annotation

## Complément au guide d'annotation en entités nommées Quaero

- Annotations Imbriquées
  - Avec les autres entités
  - Entre événements



- D'autres entités nommées devenues événement

- Lieux



- Dates



# Guide d'annotation

## Aide à l'annotation

- Tenter de remplacer par un nom non-ambigu.
- S'inspirer des définitions événementielles et non-événementielles du mot et des exemples de réalisation.
- Se rappeler que souvent les items d'une énumération sont de même type.
- Par défaut ne pas annoter, si en contexte, la désambiguïsation est trop ardue.

# Typologie

# Typologie

# Typologie

## Modalité

- Réel, réalisé → type="factual"  
L'orientation d'une nouvelle expérience après `<event type="factual" frequence="instance" temp="future"> les prochaines élections </event> [...]`
- Potentiel, hypothétique, probable → type="hypothetical"  
pour ceux qui souhaitent une `<event type="hypothetical" frequence="unique" temp="after"> victoire de la gauche </event>`
- Événement qui n'a pas eu lieu → type="nonfactual"  
`<event type="nonfactual" frequence="unique" temp="before"> Cette prétendue agression </event> avait [...]`
- Événement abstrait → type="abstract"  
La `<event type="abstract" frequence="unique" temp="now"> crise </event>` suit une période de confiance excessive.

# Typologie

## Fréquence de l'événement

- Unique → `frequence="unique"`

Et pour accroître le désarroi, on apprend qu' un `<event type="hypothetical" frequence="unique" temp="before"> autre attentat </event>` se serait produit à moins de 100 km d'Alger.

- Récurrent → `frequence="recurring"`

Les `<event type="abstract" frequence="recurring" temp="now"> Jeux paralympiques </event>` se tiennent toujours en marge de [...]

- Instanciation d'un phénomène récurrent  
→ `frequence="instance"`

Les `<event type="factual" freq="instance" temp="before"> <event type="abstract" type="recurring" temp="before"> Jeux Olympiques </event> de 1996 </event>` ont été un succès.

# Typologie

## Moment de la réalisation

- Passé → temp="before"

Ils aboutissent à un `<event type="factual" frequency="unique" temp="before"> constat d'échec </event>`

- Présent → temp="now"

`<event type="factual" frequency="unique" temp="now"> Cette rentrée-ci </event>` se place sous le signe de la contestation sociale.

- Futur → temp="after"

Le `<event type="factual" frequency="unique" temp="after"> sommet Chine - Union Européenne </event>` se tiendra à Prague.

# Notre corpus manuellement annoté

- 192 articles de presse française :
- un corpus de 48K mots
- *Le Monde* et *L'Est Républicain*
- 1 844 noms événement
- 2 annotateurs / un bon accord ( $\kappa = 0,808$ )

# Notre corpus manuellement annoté

- Un corpus de taille comparable aux autres corpus existants

	Corpus	Nbre de noms d'événements	
<b>FR</b>	<b>Notre corpus</b>	<b>1 844</b>	
EN	TimeBank 1.2	1 722	[Pustejovsky, 2005]
FR	FR-TimeBank	663	[Bittar, 2010]
IT	IT-TimeBank	3 695	[Russo et al., 2011]
EN	Creswell 2006	1 579	[Creswell et al., 2006]

## Notre corpus comparé à FR-TimeBank

- 109 documents sont communs avec FR-TimeBank
  - Annotation avec un but différent, mais annotations similaires ( $\kappa = 0,704$ )

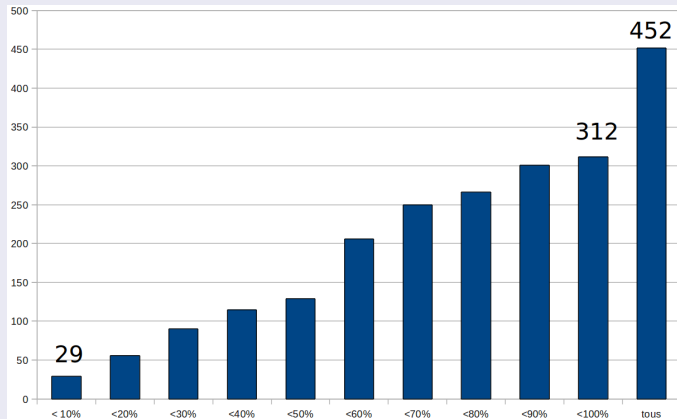
# Observations sur le corpus annoté

Têtes de syntagmes annotés	1 844
Lemmes différents	725
apparaît seulement une fois	269
apparaît plus d'une fois	452
apparaît plus d'une fois et toujours événement	140
<i>disparition, démission</i>	

- Type de nominalisation des noms annotés :

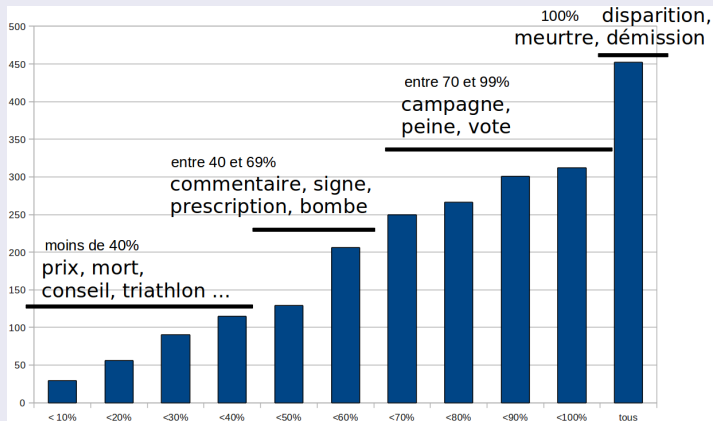
Nominalisation apparentée à un verbe d'action	67 %
Nominalisation autre que dérivée de verbes	32 %
Nominalisation métonymique	0,4 % (7)

# Observations sur le corpus annoté



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

# Observations sur le corpus annoté



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

# Observations sur le corpus annoté

## Comportement des lexiques disponibles

	Précision	Rappel	F-mesure
<i>VerbAction</i>	<b>48,7 %</b>	<b>66,8 %</b>	0,56
<i>VA + Bittar</i>	<b>48,3 %</b>	<b>84,1 %</b>	0,61

- Uniquement avec *VerbAction* :
  - Les déverbaux ont plus souvent une lecture non-événementielle.
  - 1/3 des noms d'événements ne proviennent pas d'un déverbal.
- En ajoutant le lexique de *Bittar* :
  - Le rappel augmente sans affecter la précision.
  - 15 % des noms d'événements manquent.

→ **Constat** : les lexiques seuls ne suffisent pas.

# Outline

- 1 État de l'art
- 2 Les événements nominaux
- 3 Notre Corpus annoté
- 4 Un lexique pondéré
  - Règles d'extraction
  - Construction du lexique
    - Lexiques développés
  - Évaluation

# XIP

## XIP (XRCE) (Aït-Mokhtar et al., 2002)

- Analyseur syntaxique robuste XIP (XRCE) pour l'anglais et le français
  - Dépendances syntaxiques
  - Reconnaissance d'EN (personne, lieu, organisation)
  - Possibilité pour l'utilisateur d'implémenter ses propres règles de grammaire en plus de celles existantes
- Outil utilisé pour effectuer les expérimentations et évaluations

# Nos règles d'extraction

## Les verbes d'événement et de cause/conséquence (Arnulphy et al., 2010)

- Les verbes qui introduisent des événements ou des noms d'événements

*avoir lieu, se tenir, assister à*

Le **sommet du G8** est *organisé* à Deauville.

- Des événements comme conséquence ou cause d'autres

*entraîner, provoquer, avoir pour origine*

La **crise économique** *entraînera* la **famine** dans de nombreux pays sous-développés.

- *VB90*: Règles utilisant les verbes ayant un argument événementiel, dans au moins 90 % de ses utilisations et prenant en considération la position des arguments.

# Nos règles d'extraction

## Indicateurs temporels

- Les événements sont ancrés dans le temps
- Utilisation des prépositions temporelles et dans des contextes temporels
  - le fait qu'un événement se produise
    - *à l'occasion de, lors de*
  - l'usage référentiel de l'événement
    - *avant/après, la veille/le lendemain de*
  - un moment de l'événement
    - *à l'issue de, au début de*
- Certains de ces indices sont ambigus (lieu ou temps).
- *IT*: Règles fondées sur les marqueurs temporels

# Les indices verbaux et temporels

- Annotation automatique du corpus au moyen de XIP
  - Tous les noms sujet ou argument des verbes de la liste *VB90*.
  - Tous les noms apparaissant dans un contexte décrit par les règles *IT rules*.

	Précision	Rappel	F-mesure
<i>IT</i>	81,2 %	6,1 %	0,11
<i>VB90</i>	<b>84,0 %</b>	<b>1,1 %</b>	0,02
<i>VB90 + IT</i>	<b>81,6 %</b>	<b>7,2 %</b>	0,13

- Une bonne précision
- Un rappel très bas
- Les règles implémentées ont pour but de repérer certaines désignations d'événement et de manière précise.

# Combinaison des lexiques et des règles

	Précision	Rappel	F-mesure
<i>VerbAction</i>	48,7 %	66,8 %	0,56
<i>VA + Bittar</i>	<b>48,3 %</b>	<b>84,1 %</b>	0,61
<i>IT</i>	81,2 %	6,1 %	0,11
<i>VB90</i>	84,0 %	1,1 %	0,02
<i>VB90 + IT</i>	81,6 %	7,2 %	0,13
Combinaison			
Lexiques + Règles	<b>48,0 %</b>	<b>85,9 %</b>	

- Combinaison → Pas de changement
  - Augmentation du rappel de 1,8 points
  - Baisse de la précision de 0,3 points

# Construction du lexique pondéré

## Nos règles d'extraction

- Les règles d'extraction sont précises (précision  $> 80\%$ ).
    - **Utilisons ces règles pour construire un lexique de manière automatique**
  - Les règles sont peu couvrantes (rappel  $< 10\%$ )
    - **Appliquons les règles sur un corpus de grande taille**
- Extraire une liste de désignants nominaux d'événement
- Apporter une information sur le degré d'ambiguïté de chaque mot du corpus (événement ou non)
- Prédire quantitativement l'événementialité d'un nom

# Construction du lexique pondéré

## Un ratio est calculé pour chaque nom extrait

- Pour  $m$ , chaque mot étiqueté plus d'une fois comme événement par nos règles

$$ratio_{(m)} = \frac{\text{nombre d'occurrences de } m}{\text{nombre total d'occurrences de } m} \quad (1)$$

- Remarques :
  - Le  $ratio_{(m)}$  n'est pas une probabilité de la lecture événementielle du mot (le rappel des règles est bas).
  - Le  $ratio$  est une valeur relative.
  - Le  $ratio$  est spécifique à un corpus.
  - Le  $ratio$  peut aider à la classification des noms.

# Les lexiques développés

## Sur des corpus journalistiques

Nbre de tokens			Nbre de lemmes dans le lexique
taille totale	extraits	différents	
(FR) <b>LM</b> (2001-2002) (Arnulphy, 2011–RANLP-Stud)			
61 920 573	19 767	4 843	1 559
(FR) <b>AFP</b> (2005-2011)			
	166 077	8 053	3 538
(EN) <b>AFP</b> (2004-2011)			
120 091 099	543 394	14 619	3 452

# Résultats en français sur *Le Monde* (2001-2002)

Noms d'événement <i>Français</i>	Nbre détecté / total occurrences	Ratio
<b>chute</b>	434 / 2620	<b>0,166</b>
clôture	63 / 470	<b>0,134</b>
<b>élection</b>	1243 / 9713	<b>0,128</b>
bousculade	12/115	<b>0,104</b>
crise	286 / 6185	<b>0,046</b>
<b>tension</b>	16 / 1595	<b>0,001</b>
<b>subvention</b>	2 / 867	<b>0,002</b>
Anschluss	3 / 4	<b>0,750</b>
méchoui	3 / 5	<b>0,600</b>
<b>krach</b>	20 / 169	<b>0,118</b>
RTT	14 / 166	<b>0,084</b>
demi-finale	35 / 553	<b>0,063</b>
cessez-le-feu	15 / 440	<b>0,034</b>
<b>accès</b>	9 / 2828	<b>0,003</b>
<b>11 septembre</b>	12 / 4354	<b>0,003</b>

# Résultats en anglais sur *AFP* (2004-2011)

Noms d'événement <i>Anglais</i>	Nbre détecté / total occurrences	Ratio
overthrow	383 / 448	<b>0,855</b>
<b>intifada</b>	7 / 11	<b>0,636</b>
bombardement	6 / 12	<b>0,500</b>
<b>testimony</b>	426 / 13109	<b>0,032</b>
sleepover	3 / 27	<b>0,111</b>
publication	154 / 9337	<b>0,016</b>
marathon	52 / 8070	<b>0,006</b>
<b>play-off</b>	73 / 75	<b>0,973</b>
breastfeeding	3 / 4	<b>0,750</b>
overheat	3 / 7	<b>0,428</b>
stopover	372 / 1345	<b>0,276</b>
cross-examination	53 / 416	<b>0,127</b>
distillery	4 / 126	<b>0,032</b>
welcome	66 / 3884	<b>0,017</b>
<b>influenza</b>	37 / 6019	<b>0,006</b>

## Observations : les lexiques pondérés

- Certains de ces mots ne sont pas dans les lexiques standards.
- Noms d'événement non-ambigus → valeur de ratio élevée  
`chute` (0,166), `élection` (0,128), `krach` (0,118)  
`play-off` (0,973), `intifada` (0,638)
- Noms très ambigus → valeur de ratio faible  
`tension` (0,001), `subvention` (0,002), `access` (0,003)  
`testimony` (0,032), `influenza` (0,006)  
  
`11 septembre` (0,003) → les dates sont rares, et celle-ci a de loin le meilleur ratio.
- Une désambiguïsation claire de certains mots en corpus  
`clôture` (0,134)

# Évaluation

## Évaluation par apprentissage

- Classifieur à base de règles J48 dans Weka
  - $M_l$  se base sur les lexiques standards :
    - (FR) VerbAction et Bittar (EN) WordNet
  - $M_r$  ne tient compte que des valeurs de ratio :
    - $M_r^{LM}$  ← le lexique pondéré extrait du corpus *Le Monde*.
    - $M_r^{AFP}$  ← les lexiques pondérés extraits des corpus *AFP*.
  - $M_{rl}$  fondé sur les informations issues des lexiques standards et pondérés.

	Corpus de développement	Corpus de test
FR	1 263 YES / 5 226	566 YES / 2 700
EN	1 092 YES / 2 182	453 YES / 3 246

# Évaluation

	$M_r^{LM}$	$M_r^{AFP}$	$M_l$	$M_{lr}^{LM}$	$M_{lr}^{AFP}$
Précision	49 %	55 %	53 %	54 %	60 %
Rappel	89 %	77 %	88 %	89 %	84 %
F-mesure	0,63	0,64	0,66	0,67	0,70

En français

	$M_r^{AFP}$	$M_l$	$M_{lr}^{AFP}$
Précision	36 %	30 %	36 %
Rappel	71 %	64 %	77 %
F-mesure	0,476	0,414	0,493

En anglais

# Évaluation

	$M_r^{LM}$	$M_r^{AFP}$	$M_l$	$M_{lr}^{LM}$	$M_{lr}^{AFP}$
Précision	49 %	55 %	53 %	54 %	60 %
Rappel	89 %	77 %	88 %	89 %	84 %
F-mesure	<b>0,63</b>	<b>0,64</b>	<b>0,66</b>	0,67	0,70

En français

	$M_r^{AFP}$	$M_l$	$M_{lr}^{AFP}$
Précision	36 %	30 %	36 %
Rappel	71 %	64 %	77 %
F-mesure	<b>0,476</b>	<b>0,414</b>	0,493

En anglais

- Des résultats similaires en utilisant les lexiques pondérés ( $M_r$ ) et les lexiques classiques disponibles ( $M_l$ ).

# Évaluation

	$M_r^{LM}$	$M_r^{AFP}$	$M_l$	$M_{lr}^{LM}$	$M_{lr}^{AFP}$
Précision	49 %	55 %	53 %	<b>54 %</b>	<b>60 %</b>
Rappel	89 %	77 %	88 %	<b>89 %</b>	<b>84 %</b>
F-mesure	0,63	0,64	0,66	<b>0,67</b>	<b>0,70</b>

En français

	$M_r^{AFP}$	$M_l$	$M_{lr}^{AFP}$
Précision	36 %	30 %	<b>36 %</b>
Rappel	71 %	64 %	<b>77 %</b>
F-mesure	0,476	0,414	<b>0,493</b>

En anglais

- La combinaison → une amélioration faible mais appréciable de la précision et du rappel.

# Évaluation

	$M_r^{LM}$	$M_r^{AFP}$	$M_l$	$M_{lr}^{LM}$	$M_{lr}^{AFP}$
Précision	49 %	55 %	53 %	54 %	60 %
Rappel	89 %	77 %	88 %	89 %	84 %
F-mesure	<b>0,63</b>	<b>0,64</b>	<b>0,66</b>	<b>0,67</b>	<b>0,70</b>

En français

	$M_r^{AFP}$	$M_l$	$M_{lr}^{AFP}$
Précision	36 %	30 %	36 %
Rappel	71 %	64 %	77 %
F-mesure	<b>0,476</b>	<b>0,414</b>	<b>0,493</b>

En anglais

- Nos lexiques pondérés extraits automatiquement sont aussi précis que les lexiques validés manuellement.

# Évaluation - Observations

	$M_r^{LM}$	$M_r^{AFP}$	$M_l$	$M_{lr}^{LM}$	$M_{lr}^{AFP}$
Précision	49 %	55 %	53 %	54 %	60 %
Rappel	89 %	77 %	88 %	89 %	84 %
F-mesure	0,63	0,64	0,66	0,67	0,70

En français

	$M_r^{AFP}$	$M_l$	$M_{lr}^{AFP}$
Précision	<b>36 %</b>	<b>30 %</b>	<b>36 %</b>
Rappel	<b>71 %</b>	<b>64 %</b>	<b>77 %</b>
F-mesure	<b>0,476</b>	<b>0,414</b>	<b>0,493</b>

En anglais

- Les performances absolues médiocres P/R au français.
  - La progression relative entre les lexiques sur les deux langues est équivalente: La qualité du lexique ne semble pas en cause.
  - Ceci révèle une tâche plus difficile.
- La traduction des règles du français vers l'anglais est possible.

# Évaluation - Observations

	$M_r^{LM}$	$M_r^{AFP}$	$M_l$	$M_{lr}^{LM}$	$M_{lr}^{AFP}$
Précision	<b>49 %</b>	<b>55 %</b>	<b>53 %</b>	<b>54 %</b>	<b>60 %</b>
Rappel	<b>89 %</b>	<b>77 %</b>	<b>88 %</b>	<b>89 %</b>	<b>84 %</b>
F-mesure	<b>0,63</b>	<b>0,64</b>	<b>0,66</b>	<b>0,67</b>	<b>0,70</b>

En français

	$M_r^{AFP}$	$M_l$	$M_{lr}^{AFP}$
Précision	36 %	30 %	36 %
Rappel	71 %	64 %	77 %
F-mesure	0,476	0,414	0,493

En anglais

- Le lexique pondéré extrait de l'*AFP* ( $P = 55\%$ ) est plus précis que les lexiques standards ( $P = 53\%$ ) et celui sur *Le Monde* ( $P = 49\%$ ).

# Évaluation - Observations

## En français

- À titre comparaison, appliqué séparément sur le FR-TimeBank et sur notre corpus annoté, le modèle  $M_r^{AFP}$  a des performances comparables.

	FR-TimeBank	Notre corpus annoté
Précision	56 %	55 %
Rappel	77 %	77 %
F-mesure	0,648	0,642

# Conclusion

- La notion d'événement a été traitée précédemment
  - Pas dans une approche directement applicable en TAL
  - Ou peu adaptée aux désignations nominales
- Nous avons proposé :
  - Une définition de l'événement qui nous semble appropriée pour les applications TAL,
  - Un guide d'annotation décrivant une typologie des événements,
  - Un corpus annoté en événements nominaux.

# Conclusion

- Nous avons proposé :
  - Un méthode d'extraction automatique de lexique fondée sur des indications linguistiques contextuelles,
    - Offrant une information quantitative relative sur le taux d'évenementialité d'un nom : le *ratio*
    - Dont les performances sont équivalentes aux lexiques créés manuellement
    - Applicable sur l'anglais et le français
    - Proposant des termes absents des lexiques standards et pour certains proches des entités nommées

# Perspectives

- Étude du problème de performance des lexiques sur l'anglais.
- Développement d'un modèle par apprentissage qui permettrait une extraction automatique des noms d'événements.
- Un lexique pondéré qui prendrait en compte les dates d'apparition des noms d'événements. Problématique en lien avec l'évolution des noms d'événements.