

Annotation temporelle de textes en français

*André Bittar – ALPAGE, Université Paris Diderot
Séminaire de linguistique de l'Université de Marne-la-Vallée
19 janvier 2009*

Plan

- Introduction
 - Informations temporelles
 - Cas intéressants et pas intéressants
 - En TAL
- ISO-TimeML
 - la norme et les balises
- Projet
 - ressources ISO-TimeML pour le français
- Guide d'annotation pour le français
- Corpus de référence
- Modules d'annotation automatique
 - Module <TIMEX3>
 - Module <EVENT>
 - Evaluation
- Perspectives

Plan

- Introduction
 - Informations temporelles
 - Cas intéressants et pas intéressants
 - En TAL
- ISO-TimeML
 - la norme et les balises
- Projet
 - ressources ISO-TimeML pour le français
- Guide d'annotation pour le français
- Corpus de référence
- Modules d'annotation automatique
 - Module <TIMEX3>
 - Module <EVENT>
 - Evaluation
- Perspectives

Introduction

- Informations temporelles

Expressions de date	<i>le 19 janvier, 19.01.2009, lundi</i>
Expressions d'heure	<i>10h30, midi, treize heures quinze</i>
Expressions de durée	<i>5 minutes, quarante ans, un siècle</i>
Les événements	
- verbes	<i>atterrir, manger, venir, parler...</i>
- noms déverbaux	<i>atterrissage, débordement, naissance...</i>
- autres noms	<i>guerre, repas, crise...</i>
Les états	
- adjectifs	<i>malade, fâché, enneigé...</i>
- verbes	<i>penser, croire, habiter...</i>

- Importantes pour la bonne compréhension de textes car...

- les entités (personnes, organisations, pays...) peuvent changer dans le temps, elles sont impliquées dans des événements
- environ 30% des phrases du corpus des *Document Understanding Conferences* ont un adverbial temporel (Muller et al 2008)

Introduction

- Cas intéressants

*Au moins 22 Palestiniens ont **péri ces dernières heures** dans et autour de Gaza-ville, ce qui **porte** à au moins 940 le bilan des morts Palestiniens dans l'**offensive** israélienne **lancée le 27 décembre**, selon des sources médicales.*

- Pas intéressants

*Carter, **21 ans**, est le plus jeune joueur de l'équipe. (épithète)*

*Max est un chauffeur du **dimanche**. (expression figée)*

*Les invités ont mangé un bon **repas**. (objet concret)*

*Il est déconseillé de **manger** en conduisant. (événement générique)*

- En TAL

Questions – réponses

- Traiter des questions d'ordre temporel
- *Quand? Pendant combien de temps?*

Extraction d'information

- Extraire les entités et les relations dans le temps

Résumé automatique

- Prendre les événements les plus pertinents/importants

Plan

- Introduction
 - Informations temporelles
 - Cas intéressants et pas intéressants
 - En TAL
- **ISO-TimeML**
 - la norme et les balises
- **Projet**
 - ressources ISO-TimeML pour le français
- Guide d'annotation pour le français
- Corpus de référence
- Modules d'annotation automatique
 - Module <TIMEX3>
 - Module <EVENT>
 - Evaluation
- Perspectives

- TimeML : James Pustejovsky (Brandeis) et al
- Langage d'annotation pour la normalisation des informations temporelles
- Vise 4 points :
 - 1) « Time-stamping » des événements
 - 2) Ordonnancement chronologique des événements entre eux
 - 3) Raisonnement avec des expressions temporelles sous-spécifiées
 - 4) Raisonnement sur la durée des événements
- Annotation de surface
 - on annote uniquement ce qui est lexicalisé
 - pas Jean est tombé. Luc l'a poussé.

- **<EVENT>** : *une conception large des événements*

Verbes	<i>manger, vouloir, affirmer, voir</i>
Noms	<i>chute, atterrissage, manifestation</i>
Adjectifs	<i>Jean était <u>triste</u>, il est <u>possible</u> que</i>
GP	<i>Jean était <u>en retard</u></i>

- **Attributs**

<u>Attribut</u>	<u>Valeurs</u>	<u>Exemples</u>
class	ASPECTUAL	<i>commencer, terminer, interrompre...</i>
	PERCEPTION	<i>voir, entendre, apercevoir...</i>
	REPORTING	<i>dire, affirmer, annoncer...</i>
	I_ACTION	<i>accepter, essayer, empêcher...</i>
	I_STATE	<i>vouloir, savoir, être possible</i>
	STATE	<i>être/sembler/paraître/avoir l'air + ADJ</i>
	OCCURRENCE	Le reste

- **<EVENT>**

- **Attributs**

<u>Attribut</u>	<u>Valeurs</u>
aspect	PROGRESSIVE, PERFECTIVE, IMPERFECTIVE...
modality	SHOULD, MUST...
polarity	POS, NEG
pos	ADJECTIVE, NOUN, PREPOSITION, VERB
tense	PAST, PRESENT, FUTURE, IMPERFECT
mood	SUBJUNCTIVE, CONDITIONAL
vform	INFINTIVE, PASTPART, PRESPART

- **<TIMEX3>** :

<u>Type</u>	<u>Forme</u>	<u>Valeur</u>
Dates	<i>le 3 janvier</i> <i>19/12/2006</i>	XXXX-01-03 2006-12-19
Heures	<i>17h30</i> <i>5h du matin</i>	T1730 T0500
Durées	<i>3 jours</i> <i>100 ans</i>	P3D P100Y
Expressions déictiques	<i>l'année dernière</i> <i>lundi prochain</i>	2008 2009-01-26

ISO-TimeML

- **<SIGNAL>** : marqueurs de relation *pendant, avant, après, lors de*
- **<ALINK>** : relations aspectuelles *BEGINS, ENDS, CONTINUES...*
- **<TLINK>** : temporelles *DURING, AFTER, BEFORE...*
- **<SLINK>** : de subordination *INTENSIONAL, EVIDENTIAL, FACTIVE*

ex. *Jean est arrivé après 18h.*

Jean est

```
<EVENT id="e1" class="OCCURRENCE">arrivé</EVENT>
```

```
<SIGNAL id="s1">après</SIGNAL>
```

```
<TIMEX3 id="t1" val="T1800">18h</TIMEX3>.
```

```
<TLINK relType="AFTER" eid="e1" tid="t1" sid="s1"/>
```

- Établissement d'une norme
- Travaux existants sur l'anglais
 - Pustejovsky et al : www.timeml.org
 - Guide d'annotation
 - Outils d' annotation manuelle et automatique
 - Divers corpora de référence
- Chinois, coréen, italien, allemand
 - Divers outils d'annotation automatique
 - Consignes d'annotation dans guide ISO
- Sur le français
 - Complexité du calcul des relations temporelles
Muller & Tannier (2004)
Muller & Reymonet (2005)
 - Repérage d'événements et d'expressions temporelles
Parent, Gagnon & Muller (TALN 2008)
Bittar (RECITAL 2008)

Plan

- Introduction
 - Informations temporelles
 - Cas intéressants et pas intéressants
 - En TAL
- ISO-TimeML
 - la norme et les balises
- **Projet**
 - ressources ISO-TimeML pour le français
- Guide d'annotation pour le français
- Corpus de référence
- Modules d'annotation automatique
 - Module <TIMEX3>
 - Module <EVENT>
 - Evaluation
- Perspectives

Projet

- Fédération de ressources pour le français
- Guide d'annotation pour le français
 - prendre en compte les spécificités de cette langue
- Corpus de référence
 - « Gold Standard » pour évaluation
 - « Silver Standard » de grande taille pour faire de l'apprentissage
- Modules d'annotation automatique
 - identification des événements (EVENT, SIGNAL)
 - repérage des expressions temporelles (TIMEX3, SIGNAL)
 - (à terme) calcul des relations (ALINK, TLINK, SLINK)

Plan

- Introduction
 - Informations temporelles
 - Cas intéressants et pas intéressants
 - En TAL
- ISO-TimeML
 - la norme et les balises
- Projet
 - ressources ISO-TimeML pour le français
- Guide d'annotation pour le français
- Corpus de référence
- Modules d'annotation automatique
 - Module <TIMEX3>
 - Module <EVENT>
 - Evaluation
- Perspectives

Guide d'annotation pour le français

• Modalité

Verbes

devoir, falloir, se pouvoir, risquer de...

`class="MODAL"`

- en FR peuvent être conjugués à tous les temps

`tense="..."`

- conséquences temporelles

- modifient les conditions de vérité des événements enchâssés

ex. *Jean devait venir.*

Il se peut que Jean vienne.

Adjectifs

Il est nécessaire, possible, probable que...

- modaux peuvent avoir une portée sur des événements

ex. *Il est possible que Jean vienne.*

Guide d'annotation pour le français

• Modalité

Une classification

POSSIBILITY	<i>il est possible/impossible que, il se peut que...</i>
PROBABILITY	<i>il est probable/improbable que, Jean risque de...</i>
CERTAINTY	<i>il est vrai/faux/certain/sûr que...</i>
OBLIGATION	<i>il est obligatoire/facultatif de, Jean doit...</i>
NECESSITY	<i>il faut, il est nécessaire de...</i>
PERMISSION	<i>il est permis/interdit de, Jean peut...</i>
CONJECTURAL	<i>Selon les sources, Arcelor serait impliqué dans un trafic de déchets toxiques.</i>

*Il a <EVENT class="MODAL">fallu</EVENT>
<EVENT class="OCCURRENCE" modality="NECESSITY">agir</EVENT>
vite pour empêcher des dégâts.*

Guide d'annotation pour le français

- Aspect

Groupes prépositionnels

en plein + N, en cours de + N, en train de + VInf
en passe de, sur le point de + VInf

aspect="PROGRESSIVE"

aspect="PROSPECTIVE"

en voie de + N (achèvement)
ex. Les vidéocassettes sont en voie d'extinction.

aspect="PROSPECTIVE"

en voie de + N (activité/processus)
ex. Ce pays est en voie de développement.

aspect="PROGRESSIVE"

- Auxiliaires aspectuels

venir de + VInf
aller + VInf

aspect="PERFECTIVE"

aspect="PROSPECTIVE"

Guide d'annotation pour le français

- Annotation des **EVENT** nominaux

- manque de consignes dans le guide général ISO

- N génériques : ne pas annoter

- ex. *La guerre est un conflit armé.*

- N sans déterminant : ne pas annoter

- ex. *Les prisonniers de guerre ont été libérés.*

- N en prédication nominale : annoter avec "**class=STATE**" quand temporel (temps verbaux, modifieurs)

- ex. *Dupont est le président de la compagnie depuis 2002.*

Guide d'annotation pour le français

- Annotation des **EVENT** adjectivaux

- peu de consignes dans le guide général ISO
- adjectifs modaux : annoter quand ils introduisent un **EVENT**
- adjectifs qualitatifs :
 - Distinction INDIVIDUAL-LEVEL versus STAGE-LEVEL (G. Carlson 1980)
peut être pertinente pour l'annotation temporelle

- Prédicats INDIVIDUAL-LEVEL

- propriétés permanentes / statiques
- ne pas annoter
- *Jean est intelligent.*

- Prédicats STAGE-LEVEL

- propriétés temporaires / dynamiques
- annoter comme **EVENT**
- *Jean est malade.*

Guide d'annotation pour le français

- Ajout de 4 nouvelles classes de **EVENT**

- **MODAL** : *devoir, falloir, se pouvoir, il être + ADJ MODAL*

- déterminent les conditions de vérité d'une proposition événementielle

ex. Il a fallu décaler la conférence.

```
Il a <EVENT class="MODAL">fallu</EVENT>  
<EVENT class="OCCURRENCE" modality="NECESSITY">  
décaler</EVENT> la conférence.
```

- **EVENTIVE** : *avoir lieu, se passer, se produire, se tenir*

- servent à situer un événement (pro)nominal dans le temps ou dans l'espace

ex. La conférence aura lieu le 23 août.

```
La <EVENT class="OCCURRENCE">conférence</EVENT>  
<EVENT class="EVENTIVE">aura lieu</EVENT> le 23 août.
```

Guide d'annotation pour le français

- Ajout de 4 nouvelles classes de **EVENT**
- **TEMPORAL** : *précéder, suivre, s'ensuivre, coïncider, accompagner*
 - verbes de discours (Danlos 2006)
 - réalisent des relations temporelles entre deux événements

ex. *Un pot d'accueil a précédé la conférence.*

Un `<EVENT class="OCCURRENCE">pot</EVENT>` d'accueil a
`<EVENT class="TEMPORAL">précédé</EVENT>`
la `<EVENT class="OCCURRENCE">conférence</EVENT>` le 23 août.

- **CAUSE** : *causer, engendrer, faire + Vinf, laisser + Vinf*
 - établissent une relation causale (et donc temporelle) entre événements

ex. *L'accident a causé la mort du conducteur.*

L' `<EVENT class="OCCURRENCE">accident</EVENT>` a
`<EVENT class="CAUSE">causé</EVENT>`
la `<EVENT class="OCCURRENCE">mort</EVENT>` du conducteur.

Plan

- Introduction
 - Informations temporelles
 - Cas intéressants et pas intéressants
 - En TAL
- ISO-TimeML
 - la norme et les balises
- Projet
 - ressources ISO-TimeML pour le français
- Guide d'annotation pour le français
- Corpus de référence
- Modules d'annotation automatique
 - Module <TIMEX3>
 - Module <EVENT>
 - Evaluation
- Perspectives

Corpus de référence

- 30 articles journalistiques *Le Monde*
 - annotation manuelle (très longue)
 - ~18k mots
 - 1155 verbes, 476 noms, 61 adjectifs, 7 GP <EVENT>
 - 393 <TIMEX3>
- 21 textes biographiques de *Prix Nobel*
 - annotation manuelle
 - ~4.5k mots
 - 286 verbes, 90 noms, 1 adjectif, 1 GP <EVENT>
 - 147 <TIMEX3>
- A venir : *L'Est Républicain*
 - 125M mots
 - annotation automatique avec correction manuelle

Plan

- Introduction
 - Informations temporelles
 - Cas intéressants et pas intéressants
 - En TAL
- ISO-TimeML
 - la norme et les balises
- Projet
 - ressources ISO-TimeML pour le français
- Guide d'annotation pour le français
- Corpus de référence
- Modules d'annotation automatique
 - Module <TIMEX3>
 - Module <EVENT>
 - Evaluation
- Perspectives

Modules d'annotation automatique

- Module `<TIMEX3>`

Ensemble très exhaustif de graphes Unitex (Gross 2002)

- ajout de balises `<TIMEX3>` → transducteurs
- transducteurs → conversion en grammaires SxPipe (Sagot & Bouiller 2007)
- SxPipe applique les transducteurs sur texte brut

Script Perl pour la normalisation des valeurs des expressions annotées et calcul des valeurs des expressions déictiques

<u>Forme</u>	<u>Valeur</u>
<i>le 9 août 1978</i>	1978-08-09
<i>11 heures du soir</i>	T2300
<i>deux siècles</i>	P200Y
<i>il y a 4 jours</i>	2009-01-15

Modules d'annotation automatique

- Module <EVENT>

Entrée : texte taggué, analyse morpho, chunking « augmenté » avec Macaon (Alexis Nasr)

Sortie : Fichier XML annoté en ISO-TimeML

Ressources :

- Lexique des noms événementiels (**9 969** entrées)
 - VerbAction (Hatout et al 2002) : couples verbe ↔ nom déverbal (**9 200**)
 - Requêtes automatisées Google par patrons (**769**)
ex. "lors de la *", "le * a eu lieu"
→ *anniversaire, blocus, tsunami, anticoagulothérapie*
- Ambiguïté : *repas, construction* (événement ou objet?)
- Lexique de verbes
 - organisé selon la classification ISO-TimeML
 - 174 verbes pour 6 des 7 classes
 - classe par défaut OCCURRENCE
 - Non-EVENT : *durer, sembler, suffire...*

Modules d'annotation automatique

- Module <EVENT>

Approche : méthode symbolique

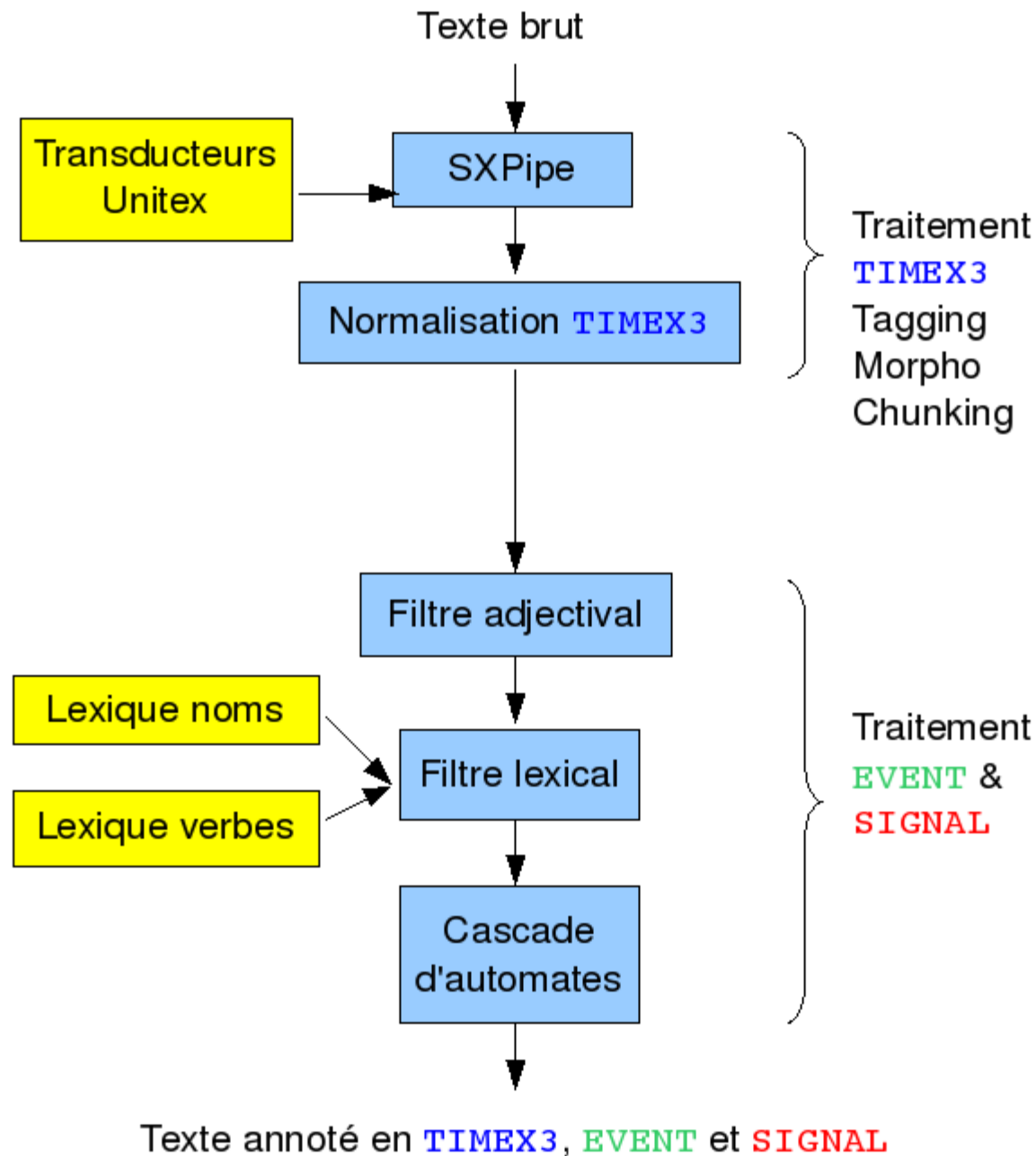
1) Niveau lexical

- recherche lexicale pour noms et verbes
- recherche lexicale pour les <SIGNAL>
- annotation des têtes de chunks (pas auxiliaires ex. *être* et *avoir*)

2) Niveau contextuel

- Application de 37 heuristiques sur le contexte local
- élimination de noms sans déterminant : *prisonnier de guerre*, *voie d'accélération*
- détection des adjectifs attributs, des participes passé en fonction d'adjectif (états à annoter)
- détection d'expressions aspectuelles *venir de/aller + Vinf*, *en voie de + N...*
- détection de constructions verbales complexes *avoir lieu*, *faire/laisser + VInf...*
- traitement de la polarité et la modalité

Modules d'annotation automatique



Modules d'annotation automatique

- Evaluation <TIMEX3>

<u>Rappel</u>	<u>Précision</u>	<u>F-Score</u>
81.00	96.00	88.5

- Erreurs

un homme de 76 ans
une quinzaine de jours
les prochains jours

âge pas durée
pas dans grammaire!
idem!

- Evaluation <EVENT>

	<u>Rappel</u>	<u>Précision</u>	<u>F-Score</u>
<u>Verbes</u>	92.03	87.35	89.69
<u>Noms</u>	92.62	49.02	70.82

- Erreurs

ils figurent parmi les plus grands
l'entrée se fait entre deux haies
NOMINAUX : en général

statif, non EVENT
générique
DESAMB INSUFFISANTE!

Plan

- Introduction
 - Informations temporelles
 - Cas intéressants et pas intéressants
 - En TAL
- ISO-TimeML
 - la norme et les balises
- Projet
 - ressources ISO-TimeML pour le français
- Guide d'annotation pour le français
- Corpus de référence
- Modules d'annotation automatique
 - Module <TIMEX3>
 - Module <EVENT>
 - Evaluation
- Perspectives

Perspectives

- Intégration des consignes d'annotation pour le français dans le guide ISO-TimeML
- Intégration d'autres ressources (ex. *WOLF*) dans les modules
- Création du corpus de référence de grande taille
 - tester des méthodes d'apprentissage
 - validation à grande échelle
- Annotation des relations *<ALINK>*, *<TLINK>*, *<SLINK>*
 - tâche très difficile
 - explorer le lien avec la réalisation des relations de discours
- Désambiguïsation lexicale pour les noms
 - explorer les méthodes probabilistes de WSD

The End

- Merci!