
Extraction de dates saillantes pour la construction de chronologies thématiques

Rémy Kessler* — **Xavier Tannier*,**** — **Caroline Hagège***** —
Véronique Moriceau*,** — **André Bittar*****

* *LIMSI-CNRS, Orsay*

prenom.nom@limsi.fr

** *Université Paris-Sud, Orsay*

*** *Xerox Research Centre Europe, Meylan*

caroline.hagege@xrce.xerox.com, andre.bittar@gmail.com

RÉSUMÉ. Nous présentons ici une approche pour la détection de dates saillantes (importantes) dans les textes dans le but de construire automatiquement des chronologies événementielles à partir de requêtes thématiques (ex. le nom d'un événement, d'une personne, etc.). Ce travail a été mené sur un corpus de dépêches en français et en anglais fourni par l'Agence France Presse (AFP). Pour extraire les dates saillantes qui méritent de figurer dans une chronologie événementielle, les expressions temporelles dans les textes doivent, dans un premier temps, être reconnues et normalisées. Nous utilisons ensuite une approche par apprentissage pour extraire les dates saillantes pour un thème donné. Nous ne nous intéressons pour le moment qu'à l'extraction des dates et non aux événements associés.

ABSTRACT. We present an approach for detecting salient (important) dates in texts in order to automatically build event timelines from a search query (e.g. the name of an event or person, etc.). This work was carried out on a corpus of newswire texts in English provided by the Agence France Presse (AFP). In order to extract salient dates that warrant inclusion in an event timeline, we first recognize and normalize temporal expressions in texts and then use a machine-learning approach to extract salient dates that relate to a particular topic. For the time being, we have focused only on extracting the dates and not the events to which they are related.

MOTS-CLÉS : dates saillantes, événements, chronologies.

KEYWORDS: salient dates, events, timelines.

1. Introduction

Une chronologie événementielle thématique est une liste de descriptions d'événements précis, associés à leur date d'occurrence, et considérés comme importants du point de vue d'un thème particulier (par exemple, « le Printemps arabe », « Michael Jackson », « les attentats à Bagdad »).

Notre but final est de construire des chronologies événementielles thématiques à partir d'une requête utilisateur en utilisant un corpus de dépêches en français et en anglais fourni par l'Agence France Presse (AFP). Cette tâche, qui implique l'extraction des événements importants, est proche des tâches de *Retrospective Event Detection* (Yang *et al.*, 1998) ou *New Event Detection*, telles que définies dans les campagnes *Topic Detection and Tracking* (TDT) (Allen, 2002).

L'objectif du travail présenté dans cet article est intermédiaire, à savoir l'extraction thématique de *dates saillantes*. Nous allons décrire dans cette introduction les tenants et les aboutissants de cette tâche.

La majorité des systèmes conçus pour la construction de chronologies événementielles utilisent des approches par sac de mots, et très peu d'informations temporelles : généralement, seules les métadonnées du document, comme sa date de création, sont utilisées. Les quelques systèmes qui utilisent des informations temporelles n'extraient que des dates absolues (celles dont la lecture hors contexte suffit à les placer sur l'axe des temps, comme « 14 juillet 1789 », contrairement à « 14 juillet »). Dans le corpus de dépêches journalistiques que nous utilisons et que nous présentons en section 3.1, nous n'avons trouvé que 7 % d'expressions temporelles qui sont des dates absolues pour l'anglais et 12 % pour le français, d'où l'intérêt de traiter aussi les autres types d'expressions temporelles.

Notre travail se distingue des recherches précédentes dans la mesure où l'extraction d'événements importants est guidée par le traitement temporel. En effet, nous considérons que les événements importants, ceux que nous souhaitons retrouver dans les chronologies, ont lieu à des dates que nous pouvons également juger comme importantes – du point de vue du thème imposé par la requête de l'utilisateur. D'autre part, l'expression de ces événements dans les textes sera souvent accompagnée d'une expression temporelle, et répétée plusieurs fois si l'événement a une certaine importance – d'un point de vue journalistique. Nous voulons donc montrer qu'il est pertinent de rechercher ces dates en premier lieu, pour pouvoir ensuite associer un ou plusieurs événements à chacune d'elles – cette toute dernière partie n'étant cependant pas le sujet de cet article. Pour cela, nous nous fondons essentiellement sur un traitement linguistique temporel des textes. Les autres types de contenus (textuels par exemple) sont utilisés pour la recherche préalable de documents.

Pour extraire ces dates importantes, la première étape est la collecte d'un maximum d'informations temporelles dans les textes. Ensuite, nous créons la notion de *date saillante*, définie à l'aide de critères essentiellement statistiques, avec pour but de nous rapprocher des dates présentées dans les chronologies de référence. Nous

proposons ainsi plusieurs définitions de la saillance d'une date, en partant de la simple redondance de cette date – pouvant avoir différentes formes lexicales – dans des textes concernant la requête, jusqu'à une modélisation apprise automatiquement à partir d'un certain nombre de caractéristiques plus fines. De plus, nous mettons à profit des éléments linguistiques simples tels que le discours rapporté, la modalité ou le temps du verbe pour raffiner ce processus en filtrant les phrases pouvant introduire du bruit dans la chaîne.

Ainsi que nous l'avons dit plus haut, dans la mesure où les dates apparaissant dans les textes viennent ancrer temporellement l'expression linguistique des événements, si une date est considérée comme saillante pour une thématique donnée, alors nous supposons que les événements auxquels ces dates sont rattachées sont certainement importants d'un point de vue informatif pour cette thématique. Elles seront donc de bonnes candidates pour la constitution des chronologies événementielles.

En sortie, notre système fournit une liste de dates classées de la plus à la moins importante par rapport au thème de la requête. Chaque date est accompagnée d'un ensemble de phrases pertinentes pour le thème donné (voir plus loin la figure 5). L'évaluation d'un tel système consiste à vérifier dans quelle mesure les dates retournées sont effectivement celles des événements que l'on peut trouver dans des chronologies événementielles de référence – construites de façon indépendante par des journalistes de l'AFP.

Dans la section suivante, nous commençons par présenter des travaux liés à notre étude. La section 3 présente les ressources utilisées et le système. L'analyse temporelle des textes est détaillée en section 4. Enfin, les approches choisies pour l'indexation et l'extraction des dates saillantes, ainsi que les résultats obtenus, sont présentés dans la section 5.

2. État de l'art

2.1. Importance de la composante temporelle dans des tâches relevant du TAL

Le langage ISO-TimeML (Pustejovsky *et al.*, 2010) est un langage de spécification pour l'annotation des informations temporelles dans les textes mais, à notre connaissance, il n'a pas encore été utilisé dans des systèmes de recherche d'information. Néanmoins, Alonso *et al.* (2007), Alonso (2008), Kanhabua (2009) et Mestl *et al.* (2009), parmi d'autres, ont souligné le fait que l'analyse des informations temporelles est souvent un composant essentiel dans la compréhension de textes et utile dans un grand nombre d'applications de recherche d'information.

Ainsi, Harabagiu et Bejan (2005) et Saquete *et al.* (2009) soulignent l'importance du traitement des expressions temporelles pour les systèmes de question-réponse. Par exemple, dans la campagne d'évaluation TREC-10 pour les systèmes de question-réponse, plus de 10 % des questions nécessitaient un traitement temporel pour pouvoir être traitées correctement (Li *et al.*, 2005a).

Dans les applications de résumés multidocuments, le traitement temporel permet à un

système de détecter les extraits redondants à partir de plusieurs textes traitant du même thème et de présenter les résultats dans un ordre chronologique pertinent (Barzilay et Elhadad, 2002).

De plus, les grandes bases de connaissances constituées automatiquement ou semi-automatiquement grâce à l'extraction de l'information contenue dans les textes et qui peuvent être utilisées comme source d'informations encyclopédiques pour des tâches relevant du TAL s'enrichissent aujourd'hui d'une composante temporelle (Wang *et al.*, 2010).

Enfin, l'analyse temporelle est également utile pour des applications telles que l'aide à la décision notamment médicale. Par exemple, Kim et Choi (2011) proposent une approche pour l'extraction d'information temporelle dans des récits cliniques. De la même manière, Jung *et al.* (2011) présentent un système qui analyse les dossiers cliniques, détecte les événements et construit une chronologie de l'histoire médicale des patients.

2.2. L'annotation temporelle automatique des textes

L'importance de cette composante temporelle a conduit au développement de systèmes d'annotation de l'information temporelle, en grande majorité pour l'anglais. Plusieurs de ces systèmes sont présentés et évalués dans le cadre des campagnes TempEval (Verhagen *et al.*, 2007 ; Verhagen *et al.*, 2010). Ils sont dans la plupart des cas basés sur l'apprentissage à partir de corpus annotés selon la norme ISO-TimeML. Alors que dans (Verhagen *et al.*, 2007) les tâches d'annotation principales étaient centrées sur la détermination de relations temporelles entre événements et/ou expressions temporelles pré-annotées, dans (Verhagen *et al.*, 2010) deux tâches consistent en la reconnaissance et le typage automatique des expressions temporelles et des événements sans annotation préalable. Les efforts de la communauté tendent ainsi vers une tâche d'annotation temporelle partant du texte brut et allant jusqu'à une analyse fine de la structure temporelle d'un document. Le travail de Uzzaman et Allen (2011) est une illustration de cette préoccupation. Il présente un système combinant analyse syntaxique profonde et classifieurs pour procéder à l'annotation et la normalisation des événements et des expressions temporelles à partir de textes tout venant. La plupart des systèmes existants et ayant participé aux compétitions mentionnées ci-dessus traitent des textes en anglais. Pour le français, nous pouvons citer les travaux de Parent *et al.* (2008) et de Battistelli *et al.* (2008). Le premier annotateur repère les événements et utilise des automates à états finis pour reconnaître et annoter automatiquement les expressions de localisation temporelle. L'annotation consiste ici au typage de ces expressions en fonction de leur référence. Le deuxième annotateur repère et normalise un sous-ensemble d'expressions de localisation temporelle (les expressions calendaires). La normalisation se fait en terme d'opérations s'effectuant sur une base calendaire. Nous avons développé pour l'extraction des dates saillantes un outil d'annotation et de normalisation pour l'anglais et le français. Comme dans (Uzzaman et Allen, 2011), il bénéficie d'une analyse syntaxico-sémantique fine. Nous nous intéresserons cepen-

dant uniquement à un sous-ensemble des expressions temporelles envisagées dans le travail cité.

2.3. Détection d'événements

L'analyse du temps est étroitement liée à celle des événements, et cela de manière générale mais également, plus particulièrement, dans le domaine de la presse. Les différentes éditions de la tâche TDT ont permis le développement de différents systèmes qui détectent la nouveauté dans les fils d'actualité (Allen, 2002 ; Kumaran et Allen, 2004 ; Fung *et al.*, 2005). La plupart de ces systèmes s'appuient sur des modèles statistiques à base de sac de mots utilisant des mesures de similarité pour déterminer la proximité entre les documents (Li *et al.*, 2005b ; Brants *et al.*, 2003). Smith (2002) utilise les informations spatio-temporelles des textes pour détecter les événements dans une bibliothèque numérique. Sa méthode s'appuie sur les collocations temps/lieu et classe les événements selon des mesures statistiques.

Des efforts ont également été faits pour construire automatiquement des chronologies textuelles et graphiques. Par exemple, Allen *et al.* (2001) présentent un système qui utilise des mesures de pertinence et de nouveauté pour construire des chronologies constituées d'une phrase par date. Chieu et Lee (2004) proposent un système similaire qui extrait d'une collection de documents les événements pertinents par rapport à une requête. Les événements considérés comme importants sont ceux qui sont mentionnés dans un grand nombre d'articles et chaque événement est construit à partir d'une seule requête et représenté par un ensemble de phrases (voir à la section 5.1).

Swan et Allen (2000) proposent une approche pour la génération de chronologies graphiques qui implique l'extraction de clusters de syntagmes nominaux et d'entités nommées. Plus récemment, Yan *et al.* (2011b) et Yan *et al.* (2011a) ont mis en œuvre une approche par résumé pour générer automatiquement des chronologies en prenant en compte les caractéristiques évolutives des actualités. Par ailleurs, les développeurs d'interfaces graphiques se sont récemment intéressés à la dimension temporelle, comme le montre Time Explorer¹, Dipity² ou Time-Surfer (Llorens *et al.*, 2011). Cependant, la plupart de ces applications ne traitent que les expressions temporelles absolues dans les textes.

1. fbmya01.barcelonamedia.org:8080/future/

2. www.dipity.com

3. Description des ressources et du système

3.1. Corpus de l'AFP

Pour ce travail, nous utilisons un corpus de dépêches pour la période 2004-2011 fourni par l'AFP. Chaque dépêche est un document XML contenant un titre, une date de création (DCT), un ensemble de mot-clés et un texte découpé en paragraphes.

Le corpus anglais est composé de 1,3 million de documents (en moyenne 511 documents par jour et 426 millions de mots). Le corpus français est lui composé de 1 million de documents (en moyenne 499 documents par jour et 390 millions de mots).

3.2. Chronologies de l'AFP

Parmi les dépêches de l'AFP de notre corpus, on trouve un type spécifique d'articles : ce sont des chronologies événementielles textuelles écrites par des journalistes de l'AFP dans le but de contextualiser des événements médiatiques. Elles sont sous forme d'une liste de dates (généralement entre 10 et 20) associées à un texte décrivant l'événement ayant eu lieu à cette date. La figure 1 présente un exemple de chronologie produite par l'AFP. D'autres exemples sont donnés dans la figure 2.

Puisque notre objectif est de produire automatiquement des chronologies de ce type, nous avons choisi d'utiliser les chronologies de l'AFP pour l'apprentissage et l'évaluation de notre système. Pour cela, nous avons sélectionné un certain nombre de chronologies satisfaisant les contraintes suivantes :

- toutes les dates des chronologies doivent être comprises entre 2004 et 2011 afin de s'assurer que les événements associés sont bien présents dans le corpus. Par exemple, une chronologie sur la guerre du Vietnam a été éliminée car les dates la composant n'apparaissent pas dans le contenu des articles de notre corpus ;
- toutes les dates d'une chronologie doivent être antérieures à la date de création de la chronologie. Par exemple, la chronologie « *Space in 2005 : A calendar* », publiée en janvier 2005 et listant les événements prévus pendant l'année, a été éliminée car la plupart des lancements de fusées n'ont finalement pas eu lieu à la date prévue ;
- la granularité temporelle des chronologies est le jour. Nous ne nous intéressons pas aux chronologies relatant, par exemple, des événements heure par heure. Dans le cadre de l'élaboration de notre système, nous nous plaçons sur ce niveau de granularité.

Ces chronologies sont au nombre de 91 pour l'anglais et 94 pour le français. Dans la suite, nous les appellerons « chronologies manuelles ».

Ces chronologies sont dans un format très hétérogène, avec en particulier des dates non balisées. Pour assurer une exploitation aisée pour l'apprentissage et l'évaluation de notre système, chaque chronologie est donc convertie dans un format XML plus précis et homogène. D'autre part, nous lui associons manuellement des mots-clés,

```

<NewsML Version="1.2">
  <NewsItem xml:lang="en">
    <HeadLine>Key dates in Thailand's political crisis</HeadLine>
    <DateId>20100513T100519Z</DateId>
    <NameLabel>Thailand-politics</NameLabel>
    <DataContent>
      <p>The following is a timeline of events since the protests began, soon after Thailand's Supreme Court confiscated 1.4 billion dollars of Thaksin's wealth for abuse of power.</p>
      <p>March 14 : Tens of thousands of Red Shirts demonstrate in the capital calling for Abhisit's government to step down, saying it is elitist and undemocratic. The premier and key ministers hole up in an army barracks.</p>
      <p>March 28 : The government and the Reds enter into talks but hit a stalemate after two days, as Abhisit refuses to meet a 15-day deadline for polls.</p>
      <p>April 3 : Tens of thousands of protesters move from Bangkok's historic district into the city's commercial heart, raising the stakes in the standoff.</p>
      <p>April 7 : Abhisit declares state of emergency in capital after Red Shirts storm parliament.</p>
      <p>April 8 : Authorities announce arrest warrants for protest leaders.</p>
      ...
    </DataContent>
  </NewsItem>
</NewsML>

```

Figure 1. Exemple de chronologie manuelle de l'AFP

```

- Chronologies de 18 mois de troubles en Côte d'Ivoire
- Histoire des prises d'otages par des rebelles tchetchènes
- Les désordres politiques en Irak depuis l'élection du 7 mars
- Athlétisme : les records mondiaux du 800m
- Les accidents majeurs dans les mines chinoises
- L'espace en 2005 : un calendrier
- L'impasse sur le nucléaire iranien
- Chronologie de la guerre du Vietnam
- Chronologie de l'affaire Dominique Strauss-Kahn
- Le déroulement des attentats dans les transports londoniens

```

Figure 2. Exemples de thèmes de chronologies manuelles de l'AFP

représentant la requête de l'utilisateur, par laquelle celui-ci souhaite obtenir la chronologie thématique. L'utilité de cette requête sera expliquée par la suite.

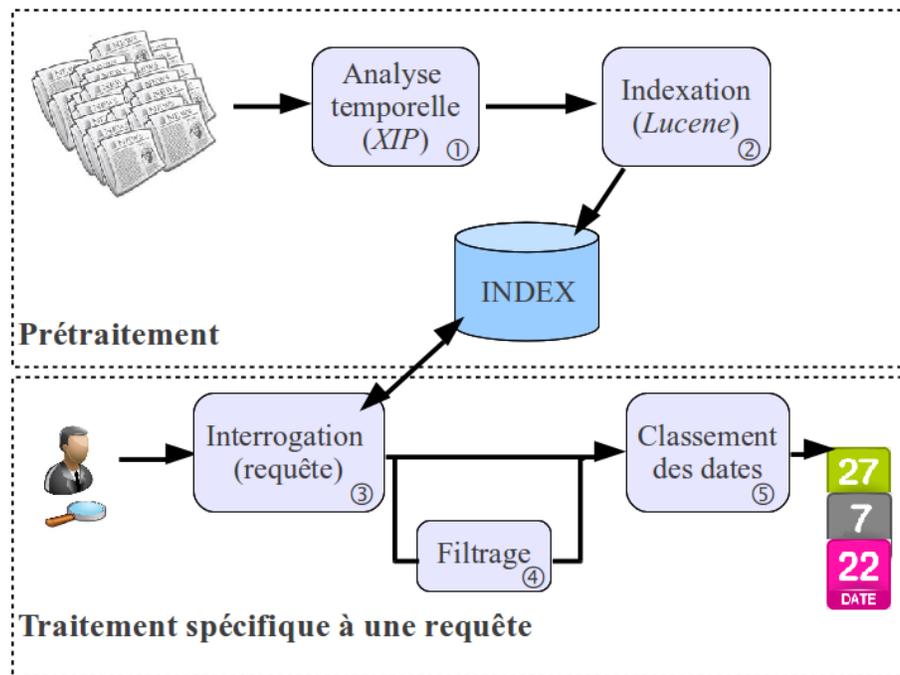


Figure 3. *Présentation du système*

3.3. *Présentation du système*

La figure 3 présente l'architecture générale de notre système. Dans un premier temps, un prétraitement du corpus AFP permet d'annoter et de normaliser les expressions temporelles de chaque article (étape ① sur la figure). Le corpus ainsi enrichi est indexé par le moteur de recherche Lucene³ (étape ②). L'ensemble des informations temporelles extraites est de cette façon associé au contenu textuel lors du processus d'indexation.

Étant donné une requête, un certain nombre de documents est renvoyé par Lucene (③). Ces documents peuvent être filtrés (④), et les dates sont extraites des documents restants. Chaque date est ainsi accompagnée d'un ensemble de phrases pertinentes pour la requête donnée. Ces dates sont ensuite classées afin de montrer les plus importantes à l'utilisateur (⑤) accompagnées des phrases qui les contiennent. Toutes ces étapes sont détaillées dans les sections suivantes.

3. <http://lucene.apache.org>

4. Analyse temporelle des textes

Dans cette section, nous décrivons les informations linguistiques et temporelles extraites pendant la phase de prétraitement du corpus et comment cette extraction est menée. Nous nous appuyons sur l'analyseur XIP (Aït-Mokhtar *et al.*, 2002) que nous avons adapté pour nos besoins.

4.1. Présentation de XIP

L'analyseur XIP réalise une analyse syntaxique profonde des textes. Il prend en entrée des fichiers XML et analyse le contenu textuel à l'intérieur de différentes balises XML selon des règles précisées dans un guide XML (un fichier donnant les instructions à l'analyseur, pour plus de détails voir (Roux, 2004)). XIP réalise un traitement linguistique complet allant de la segmentation jusqu'à l'analyse profonde en dépendances grammaticales. Il réalise aussi une reconnaissance des entités nommées les plus courantes ainsi que des expressions temporelles. Les unités linguistiques manipulées par l'analyseur sont soit des catégories terminales soit des syntagmes noyaux⁴. Chaque unité est associée à une matrice attribut-valeur qui contient les informations morphologiques, syntaxiques et sémantiques de l'unité en question. Les unités linguistiques sont reliées par des relations n-aires orientées dénotant les propriétés syntaxiques ou sémantiques du texte en entrée. Une API Java est fournie avec l'analyseur pour que toutes les structures et relations linguistiques puissent être facilement manipulées.

Dans les sections suivantes, nous détaillons les informations linguistiques que nous utilisons pour la détection des dates saillantes.

4.2. Reconnaissance des entités nommées

La reconnaissance des entités nommées est une des sorties fournies par XIP. Les entités nommées y sont représentées par une relation unaire. Nous utilisons les modules de reconnaissance existants pour le français et l'anglais qui annotent les types suivants : **noms de lieu**, **noms de personne** et **noms d'organisation**. L'ambiguïté pouvant exister entre certains types (entre le type **lieu** ou **organisation** pour les noms de pays par exemple) est conservée dans l'annotation.

4.3. Reconnaissance des expressions temporelles

Un module pour l'analyse temporelle avait déjà été développé et intégré à la grammaire anglaise (Hagège et Tannier, 2008) et évalué au cours de la campagne TempE-

4. Nous désignons par syntagme noyau des syntagmes non récursifs dont le dernier constituant est la tête syntaxique du syntagme. Ils correspondent aux *chunks* de Abney (1991).

val (Verhagen *et al.*, 2007). Cet analyseur d'expressions temporelles était uniquement développé pour l'anglais et suivait la norme TimeML (Pustejovsky *et al.*, 2005). Le but de cet analyseur était d'établir des relations temporelles entre dates et événements. Pour le travail présenté dans cet article, nous avons voulu privilégier la précision pour la normalisation et la reconnaissance d'un sous-ensemble des expressions temporelles considérées dans TimeML. Nous avons également développé un module de normalisation qui est commun aux deux langues (français et anglais). Enfin, le guide d'annotation que nous avons adopté dans ce travail se distingue de TimeML sur plusieurs points (Bittar *et al.*, 2012).

Le but de l'analyse temporelle est d'être capable de reconnaître, d'annoter et de normaliser⁵ un sous-ensemble d'expressions temporelles que nous considérons pertinentes pour notre tâche. Ce sous-ensemble d'expressions est décrit dans les sections suivantes.

En première approximation, dans la mesure où nous disposons d'un large corpus dans lequel la redondance existe, nous avons estimé que nous devons à tout prix privilégier la précision et non le rappel. En effet, nous pensons que l'introduction de bruit dans la détection des dates saillantes serait plus préjudiciable que l'omission de certaines dates. Nous avons ainsi volontairement ignoré les cas *difficiles* comme ceux qui nécessiteraient la résolution d'anaphore entre éventualités, et nous avons donc considéré les dates non récurrentes, complètes (absolues) ou relatives par rapport au moment de l'énonciation qui est donné comme métadonnée et correspond à la date de création de la dépêche.

4.3.1. Dates absolues

Les dates absolues sont des dates qui peuvent être normalisées sans connaissance externe ou contextuelle. C'est le cas, par exemple, de l'expression « *le 5 janvier 2003* ». Dans ce type d'expressions, toutes les informations nécessaires à la normalisation sont contenues dans l'expression linguistique. Cependant, les dates absolues sont relativement peu fréquentes dans notre corpus (7 % de toutes les dates extraites en anglais, 12 % en français). Il est donc indispensable de considérer les dates relatives, beaucoup plus fréquentes, afin d'élargir la couverture pour la détection des dates saillantes.

4.3.2. Dates relatives à la DCT

Les expressions temporelles relatives à la DCT (date de création du document) représentent 40 % des dates extraites du corpus AFP en anglais et en français. Contrairement aux dates absolues, leur seule considération ne permet pas de procéder à leur normalisation. D'autres informations, en particulier la date correspondant au moment

5. Nous appelons *normalisation* l'opération qui consiste à transformer une expression temporelle en une représentation formatée et entièrement spécifiée (par exemple, trouver la valeur absolue qui correspond à une date relative).

d'énonciation, sont nécessaires. Dans les dépêches, cette date correspond à la DCT. On peut distinguer deux sous-classes d'expressions temporelles relatives :

- la première nécessite seulement de connaître la valeur de la DCT pour effectuer la normalisation. C'est le cas d'expressions comme *vendredi prochain* qui correspond à la date calendaire du premier vendredi suivant la DCT ;

- la seconde nécessite des connaissances contextuelles pour effectuer la normalisation. Par exemple, l'expression *vendredi* peut correspondre soit à *vendredi dernier* soit à *vendredi prochain* selon le contexte où cette expression apparaît (par exemple, *Il viendra vendredi* correspond à *vendredi prochain* alors que *Il est arrivé vendredi* correspond à *vendredi dernier*). Dans de tels cas, le temps du verbe qui gouverne l'expression temporelle est essentiel pour la normalisation. Cette information nous est fournie par XIP.

4.4. Modalité, futur et discours rapporté

Dans le cadre de la création automatique de chronologies événementielles, on s'intéresse aux événements avérés, laissant de côté ceux qui sont d'une nature hypothétique ou incertaine, ou encore ceux qui n'ont pas lieu. Pour notre tâche, nous avons alors voulu ignorer les dates associées à de tels événements dans les textes. Un certain nombre de propriétés linguistiques peuvent indiquer qu'un événement est non avéré, notamment la négation, la modalité et certains temps et modes verbaux. De même, on peut estimer qu'un événement décrit dans un discours rapporté, n'est pas forcément avéré. Nous avons intégré dans notre traitement certaines de ces informations linguistiques. Plutôt que de nous appuyer sur une base théorique particulière, nous avons implémenté un ensemble d'heuristiques afin de pouvoir filtrer les dates associées à des événements non avérés. Nous présentons dans cette section notre traitement de la modalité, du temps futur, et du discours rapporté.

La modalité associée aux événements datés dans les textes est un point important qui peut affecter la détection des dates saillantes. Intuitivement, une date associée à un événement non avéré ne doit pas être considérée de la même manière qu'une date associée à un événement décrit comme certain par l'auteur de la dépêche, dans la mesure où les dates saillantes correspondront à des dates de faits avérés et non à des dates de faits hypothétiques. Par exemple, dans les phrases « *La réunion a eu lieu vendredi* », « *La réunion devrait avoir lieu vendredi* » et « *La réunion aura lieu vendredi, a déclaré M. Hong* », les dates n'ont pas le même statut de saillance. L'événement daté *réunion a lieu* est factuel dans le premier exemple et peut être considéré comme certain. Au contraire, dans les deux autres exemples, les événements ne se produisent pas nécessairement. Ceci est exprimé par la modalité (*devoir* dans le deuxième exemple) ou par l'utilisation du futur ou de discours rapporté (troisième exemple).

Nous avons voulu prendre en compte ces éléments pour l'extraction des dates saillantes. Il ne nous a cependant pas semblé nécessaire dans ce contexte d'attribuer des valeurs précises de modalité, mais simplement de marquer le fait qu'une expres-

sion temporelle que nous considérons pour l'extraction des dates saillantes modifie un prédicat ayant une valeur modale. Nous avons fait de même pour les expressions temporelles se rapportant à des prédicats au futur. Dans les deux cas, l'intuition est la suivante : dans le cas d'une éventualité non factuelle ou au futur, une incertitude existe quant à la réalisation effective de cette éventualité. La date associée à une telle éventualité devra donc avoir un statut particulier qui permettra postérieurement un filtrage pour la détection des dates saillantes. De la même façon, nous avons également voulu marquer les éventualités correspondant à du discours rapporté ou introduites par un verbe de discours rapporté.

XIP prend ainsi en compte les traits suivants pour l'annotation des expressions temporelles.

– **Verbes mentionnés au futur** : si une expression temporelle modifie un verbe au futur ou un nom gouverné par un verbe au futur, alors nous ajoutons l'attribut FUTURE avec la valeur + à l'annotation de cette expression temporelle correspondante.

– **Verbes décrits avec un verbe modal** : si un événement daté est introduit par un verbe modal (par exemple, *should* ou *would* pour l'anglais, *pouvoir*, *devoir* pour le français), alors on donne la valeur MODAL à l'attribut MODALITY de l'annotation pour l'expression temporelle correspondante.

– **Verbes de discours rapporté** : les verbes de discours rapporté introduisent un discours indirect ou rapporté. Dans le cas d'événements datés gouvernés par un verbe de discours rapporté ou apparaissant dans un discours rapporté, XIP marque les verbes de discours rapporté et segmente les phrases complexes en propositions.

- Si une expression temporelle modifie un tel verbe, on ajoute l'attribut DECLARATION="YES" à l'annotation de l'expression temporelle.

Par exemple, dans la phrase ci-dessous du corpus de l'AFP, l'expression temporelle *mercredi* aura dans son annotation l'attribut DECLARATION="YES" dans la mesure où elle modifie le verbe *annoncer*.

« *Un haut responsable américain chargé des questions de désarmement aura des échanges avec le gouvernement britannique sur les moyens de vérifier les promesses faites par la Libye sur l'abandon de ses programmes d'armes de destruction massive , a annoncé MERCREDI le département d'État* »

- Si une expression temporelle modifie un verbe qui apparaît dans une proposition introduite par un verbe de discours rapporté, alors on ajoute l'attribut REPORTED="YES" à l'annotation de l'expression temporelle.

Par exemple, dans la phrase suivante extraite du même corpus, l'annotation de l'expression temporelle *en 2004* comporte l'attribut REPORTED="YES" car elle fait partie du discours rapporté introduit par le verbe *demander*.

« *Dans ses vœux , le Premier ministre John Howard a demandé de ne pas baisser les bras EN 2004 [...]* »

Il faut noter que les différentes annotations peuvent se combiner (par exemple, la modalité et le discours rapporté peuvent être exprimés pour un même événement).

Ainsi, l'expression temporelle *vendredi* dans « *La réunion devrait avoir lieu vendredi, a déclaré M. Hong* » est annotée avec les deux attributs MODALITY et REPORTED.

4.5. Cas particuliers liés au corpus

Lors du développement des annotateurs linguistique et temporel, nous avons pris en compte certaines spécificités de notre corpus. Nous avons ainsi décidé que les expressions temporelles *today* et *now* pour l'anglais, *aujourd'hui* pour le français, n'étaient pas pertinentes pour la détection des dates saillantes. En effet, dans le corpus de dépêches de l'AFP, ces expressions sont principalement synonymes de *de nos jours* et ne datent pas réellement un événement par rapport à la DCT. Une autre spécificité du corpus réside dans le fait que si la DCT d'un document correspond à un lundi et si un événement au passé est décrit avec l'expression temporelle *le lundi* ou *lundi*, l'ancrage temporel de cet événement est celui de la DCT et non le lundi précédent, contrairement à ce que l'emploi du verbe au passé aurait pu laisser supposer. Nous avons adapté la normalisation des dates pour prendre en compte ces cas particuliers.

4.6. Mise en œuvre et exemple

Comme dit précédemment, un module de reconnaissance d'entités nommées est intégré à XIP. La normalisation et l'annotation des expressions temporelles a été adaptée à partir de (Hagège et Tannier, 2008). Nous avons utilisé l'API Java fournie avec XIP pour réaliser la normalisation et l'annotation des expressions temporelles. Les sorties sont des fichiers XML où seules les informations désirées sont conservées (informations structurelles pour différencier les titres du contenu des dépêches, la DCT) et enrichies avec les annotations linguistiques décrites précédemment. La figure 4 montre un exemple d'analyse d'un extrait de dépêche en anglais.

Dans cet extrait de dépêche, une seule expression temporelle (*Wednesday*) est normalisée alors que *The year 2004* et *in a decade* ne sont pas considérées comme pertinentes. La première réfère plutôt à un événement, car elle apparaît comme sujet. Cette dépendance grammaticale est calculée par XIP. La seconde expression n'est pas normalisée puisqu'elle est d'une granularité supérieure à l'année.

L'annotation de *Wednesday* possède l'attribut DECLARATION indiquant que cette expression date un événement réalisé par un verbe de discours rapporté. La valeur normalisée de l'expression correspond au 5 janvier 2005 qui est un mercredi. Les entités nommées sont annotées également.

Dans le corpus AFP en anglais, 11,5 millions d'expressions temporelles ont été détectées parmi lesquelles 845 000 sont des dates absolues (7 %) et 4,6 millions sont des dates relatives (40 %). En français, 9,4 millions d'expressions temporelles comprenaient 1,1 million de dates absolues (12 %) et 3,8 millions de dates relatives (40 %).

```

<DCT value="20050105"/>
<EC TYPE="TIMEX" value="unknown">The year 2004</EC>
was the deadliest
<EC TYPE="TIMEX" value="unknown">in a decade</EC>
for journalists around the world, mainly because of the number of reporters killed in
<EC TYPE="LOCORG">Iraq</EC>,
the media rights group
<EN TYPE="ORG">Reporters Sans Frontieres</EN>
(Reporters Without Borders) said
<EC TYPE="DATE" SUBTYPE="REL" REF="ST" DECLARATION="YES"
value="20050105">Wednesday</EC>.

```

Figure 4. Exemple d'une sortie XIP pour une dépêche. *TYPE* est le type de l'expression, *SUBTYPE* est le sous-type d'une date (absolue ou relative), *REF* la référence de cette date, par rapport au moment d'énonciation (*ST*) ou par rapport à un moment de référence (*RT*).

4.7. Évaluation de l'annotation temporelle

Nous avons souhaité évaluer les performances de nos outils d'annotation en les comparant aux corpus TimeBank du français et de l'anglais. Nous rappelons que cette évaluation concerne les dates absolues et les dates relatives à la DCT. Le corpus de référence ayant servi à l'annotation correspond donc aux ressources TimeBank dans lesquelles toutes les expressions temporelles annotées qui ne correspondent pas à une date absolue ou à une date relative au moment de l'énonciation sont ignorées. Nous évaluons les trois points suivants :

- 1) la reconnaissance de l'expression ;
- 2) le typage de l'expression annotée ;
- 3) la normalisation de l'expression annotée.

Dans le cas où une annotation du système n'a pas exactement la même délimitation que celle présente dans le corpus de référence ou dans le cas où nous avons une absence totale de normalisation pour une expression dont l'étendue est correctement délimitée, nous considérons une erreur pour l'annotateur. Pour évaluer le typage et la normalisation des expressions, nous ne considérons que les annotations des expressions qui sont correctement reconnues et délimitées. Cette mesure est plus rigoureuse qu'une mesure qui prendrait en compte séparément la normalisation et la détection, mais nous avons choisi de procéder de cette façon car la détection des dates saillantes ne peut se faire, selon notre méthode, sans une normalisation des dates repérées.

4.7.1. Pour le français

Pour le français, nous avons comparé les sorties de l'annotateur avec le TimeBank du français (Bittar *et al.*, 2011) adapté en filtrant les expressions temporelles que nous ne considérons pas dans notre travail. Après ce filtrage le corpus contenait

299 expressions temporelles annotées avec la balise <TIMEX3> (sur les 608 du corpus original). Nous avons mesuré le rappel, la précision et la F-mesure, mesures standard dans l'évaluation de systèmes d'extraction d'information ainsi que l'accord (coefficient Kappa (Cohen, 1960)). Contrairement aux autres mesures utilisées, le calcul du Kappa prend en compte la notion d'accord par chance entre deux annotations. Les résultats figurent dans le tableau 1. Pour la détection des expressions temporelles, les performances sont acceptables, mais peuvent encore être améliorées. L'analyse plus fine des résultats montre que les principales erreurs sont dues à des manques de couverture dans la grammaire, qui est encore en cours de développement. Par ailleurs, pour cette expérience, nous avons utilisé la grammaire générale du français et n'avons procédé à aucune adaptation en fonction du corpus. Pour l'évaluation du typage et de la normalisation, nous avons considéré le sous-ensemble de 239 expressions qui étaient correctement reconnues et délimitées. Aucune erreur n'a été commise sur le typage des expressions (dates absolues et dates relatives par rapport à la DCT). De très bons résultats ont été obtenus pour la normalisation des expressions. La principale source d'erreurs pour la normalisation provient des cas où une expression de date relative par rapport à la DCT apparaît dans un contexte dans lequel elle n'est pas reliée syntaxiquement (directement ou indirectement) à un prédicat verbal. En effet, pour certaines de ces expressions, le temps du verbe associé est nécessaire pour procéder correctement à la normalisation. Dans certains cas, l'absence du lien syntaxique entre expression temporelle et prédicat verbal est dû à une erreur de l'analyseur syntaxique. Dans d'autres cas, cependant, il s'agit d'expressions apparaissant dans des phrases ne contenant pas de verbes.

	Précision	Rappel	F-mesure	Kappa
Détection	0,9	0,8	0,85	0,66
Attribut type	1,0	1,0	1,0	1,0
Attribut value	0,96	1,0	0,97	0,95

Tableau 1. Performances du système d'annotation pour le français sur l'ensemble du corpus d'évaluation

4.7.2. Pour l'anglais

Nous avons procédé de manière similaire pour évaluer les annotations de l'anglais. Comme pour le français, nous avons effectué un filtrage sur le TimeBank 1.2 de l'anglais pour enlever les expressions qui ne rentraient pas dans le cadre de ce travail. De plus, nous avons choisi d'effectuer l'évaluation sur un sous-ensemble du corpus total, prenant un document sur trois. Ce choix a été motivé par deux facteurs. Premièrement, nous avons voulu utiliser un corpus de taille semblable à celui utilisé pour le français en termes de nombre d'expressions temporelles, et cela afin de faciliter la comparaison des résultats sur les deux langues. Deuxièmement, l'analyse des erreurs devant s'effectuer en partie manuellement, nous avons estimé judicieux de réduire la taille du corpus afin de faciliter et d'accélérer cette partie de la tâche. Le corpus d'évaluation que nous avons utilisé contient 320 expressions temporelles annotées avec la balise <TIMEX3> (sur les 1 414 du corpus original).

Les résultats sont présentés dans le tableau 2. Comme pour le français, nous avons évalué le typage et la normalisation sur le sous-ensemble des expressions qui étaient correctement reconnues et délimitées (un total de 198 pour l’anglais). Aucune erreur n’a été commise sur le typage des expressions. Les résultats pour la normalisation des expressions sont également assez bons, même si légèrement inférieurs à ceux obtenus pour le français. Le rappel obtenu pour le repérage des expressions temporelles est bien en dessous de ce que l’on espérait obtenir pour le type d’expressions que l’on traite ici. Comme pour le français, la grammaire de l’anglais n’a pas été développée avec la perspective de traiter un corpus particulier et, par conséquent, certains types d’expressions temporelles, qui sont récurrentes dans le TimeBank de l’anglais, ne sont pas encore traités par la grammaire, par exemple, les dates numériques telles que « 19980108 » et « 11/02/89 » et les dates en toutes lettres « *the year two thousand* » (« *l’an deux mille* ») et « *in nineteen seventy-nine* » (« *en mille neuf cent soixante-dix-neuf* »). Dans le premier cas, les patrons qui serviraient à détecter ces dates (sans contexte), pourraient reconnaître d’autres types de chiffres. Comme la grammaire a été développée avec un but d’application générale, ces expressions ne sont pas reconnues, ce qui a généré du silence. Pour le deuxième type d’expressions, s’il est relativement plus commun d’écrire ce type de dates au format numérique (« *the year 2000* » et « *in 1979* »), ces expressions méritent d’être intégrées dans la grammaire. Comme pour le français, un certain nombre d’erreurs proviennent d’une erreur d’analyse en amont par le parseur. Ainsi on trouve des expressions telles que « *Friday* » (« *vendredi* ») et « *November* » (« *novembre* ») qui ne sont pas repérées, ou sont repérées mais pas normalisées, ce qui constitue une erreur dans notre mesure.

	Précision	Rappel	F-mesure	Kappa
Détection	0,9	0,62	0,74	0,41
Attribut type	1,0	1,0	1,0	1,0
Attribut value	0,91	1,0	0,94	0,88

Tableau 2. Performances du système d’annotation pour l’anglais sur l’ensemble du corpus d’évaluation

En conclusion, les résultats de ces deux évaluations montrent que des améliorations sont possibles et ont permis de mettre en lumière une série d’erreurs qui pourront être corrigées. Les mesures utilisées pour cette évaluation ne prennent en compte que les expressions correctement repérées et normalisées. Dans nos travaux futurs, nous comptons étendre cette évaluation sur l’ensemble des expressions repérées par le système. Nous souhaitons également étendre la normalisation pour prendre en compte les expressions autres que les dates absolues et relatives au moment d’énonciation.

5. Expériences sur les dates saillantes et résultats

Dans la section 5.1, nous proposons deux approches « *baseline* » afin de montrer la difficulté de la tâche (la section 5.3 discute aussi ce point). Dans la section 5.2.1, nous

présentons nos expériences qui utilisent un filtrage simple et des statistiques sur les dates calculées par Lucene. Enfin, la section 5.2.2 détaille nos expériences avec une approche par apprentissage.

Dans toutes ces expériences, nous utilisons trois valeurs différentes pour classer les dates par ordre d'importance :

- $occ(d)$ est le nombre d'unités textuelles (phrases) contenant la date d ;
- Lucene fournit des documents classés associés à leur score de pertinence. $luc(d)$ est la somme des scores Lucene pour les unités textuelles contenant la date d ;
- une adaptation du $tf.idf$ pour les dates :

$$tf.idf(d) = luc(d) \cdot \log \frac{N}{df(d)}$$

où N est le nombre de paragraphes indexés et $df(d)$ est le nombre de paragraphes contenant la date d . Les $tf.idf$ sont sommés pour donner le score de la date.

Dans tous les cas, comme expliqué plus haut, la sortie du système est une liste de dates, ordonnées de la plus saillante à la moins saillante, accompagnées par des phrases associées à ces dates et au thème de la recherche. Un extrait simplifié d'une telle sortie est présenté à la figure 5.

Pour évaluer nos résultats, nous les avons comparés aux chronologies manuelles de l'AFP (voir section 3.2) en utilisant la *Mean Average Precision* (MAP), qui est une métrique répandue et adaptée à l'évaluation de classement. La MAP donne un poids plus fort aux éléments bien classés. Il s'agit de calculer la précision à chaque position de la liste ordonnée de résultats (ici, de dates) et d'en faire la somme tant que toutes les dates n'ont pas été trouvées. Pour une requête, la précision moyenne *AP* (*Average Precision*) est donc :

$$AP = \frac{\sum_{k=1}^n (P(k) \times rel(k))}{r}$$

où k est le rang de la date dans la liste, n le nombre de dates retournées, r le nombre de dates de la chronologie de référence, $P(k)$ la précision à ce rang et $rel(k)$ la valeur de pertinence (0 ou 1) de cette date. La MAP est la moyenne des précisions moyennes sur l'ensemble des requêtes. La signification statistique des résultats est indiquée par la *p-value*, résultat du t-test de Student ($t(90) = 1,9867$).

5.1. Baselines

Nous avons implémenté les deux approches suivantes qui font office de *baseline*. Pour celles-ci, ainsi que pour les expériences suivantes, l'unité textuelle choisie est le paragraphe. Chaque date est ainsi associée à une partie du texte et non à l'ensemble de la dépêche, celle-ci pouvant relater d'autres événements sans relation directe avec la date spécifiée.

```

<data title="La crise politique à Madagascar" keywords="madagascar"
datemin="2009-01-16" datemax="2010-08-13">
  <event date="2009-03-17" rank="1">
    <doc date="2009-08-08">Lâché par l'armée, Ravalomanana avait remis le 17 mars
    ses pouvoirs à un directoire militaire qui les avait immédiatement transférés à Andry Ra-
    joelina, devenu depuis le nouvel homme fort de Madagascar.</doc>
    <doc date="2010-10-11">Madagascar est plongée dans une grave crise politique
    depuis fin 2008 et l'éviction de M. Ravalomanana le 17 mars 2009, au profit de M. Ra-
    joelina, ex-maire d'Antananarivo</doc>
    <doc date="2009-03-19">Elle avait auparavant validé l'ordonnance par laquelle le
    président Ravalomanana, lâché par l'armée, a démissionné mardi en transférant les pleins
    pouvoirs à un directoire militaire, qui les a ensuite remis au chef de l'opposition.</doc>
    ...
  </event>
  <event date="2009-08-09" rank="2">
    <doc date="2009-08-09">Le sommet sur Madagascar organisé à Maputo s'est ter-
    miné dimanche, et la médiation a donné rendez-vous aux leaders malgaches dans une
    dizaine de jours au Mozambique, selon des médiateurs.</doc>
    <doc date="2009-08-12">La présidence de l'Union européenne s'est félicitée dans
    un communiqué d'un accord signé dimanche à Maputo, qui prévoit la mise en place à
    Madagascar d'un gouvernement de transition avant des élections d'ici fin 2010.</doc>
    ...
  </event>
  ...
</data>

```

Figure 5. Exemple de sortie (simplifiée) du système de sélection des dates saillantes

BL_{abs} . L'indexation et la recherche de document sont effectuées au niveau du para-
 graphe par le moteur de recherche Lucene (le titre et les mots-clés du document
 sont ajoutés à chaque paragraphe). Pour une requête donnée, les 10 000 premiers
 documents sont renvoyés⁶. Pour cette *baseline*, seules les **dates absolues** sont
 prises en compte. Les dates sont classées selon une des trois valeurs décrites
 précédemment (*occ*, *luc* or *tf.idf*). Nous avons ainsi obtenu les *runs* BL_{abs}^{occ} ,
 BL_{abs}^{luc} et BL_{abs}^{tfidf} .

$Chieu_{burst}$. Nous avons implémenté et adapté à notre tâche le système décrit dans
 (Chieu et Lee, 2004) qui extrait d'une collection de documents les événements
 pertinents par rapport à une requête, et les classe en fonction d'une combinaison
 de mesures appelées *Burstiness* et *Interest*. La première étape de l'algorithme
 consiste à associer chacune des phrases renvoyées par le système à une date, soit
 une date absolue si elle existe dans la phrase, soit la DCT. Ensuite, les phrases

6. Ce chiffre a été défini de façon empirique.

	s	\bar{s}
K	a	b
\bar{K}	c	d

Tableau 3. Table de contingence pour la mesure *Burstiness*. La ligne K représente des documents temporellement proches ($\|date(s') - date(s)\| \leq k$), la ligne \bar{K} des documents temporellement éloignés ($\|date(s') - date(s)\| > k$). La colonne s mesure la similarité, la colonne \bar{s} la dissimilarité (voir les formules pour a, b, c, d ci-dessous).

sont classées selon la mesure *Burstiness* : chaque paire phrase/date s est comparée à la requête avec la mesure *Interest*. Cette dernière consiste à calculer le cosinus entre chaque vecteur lexical en tenant compte d'une période de k jours autour de la date associée à s afin de filtrer les phrases avec une date trop éloignée de la date initiale. Ceci permet de remplir la table de contingence (voir tableau 3) telle que définie en 1, 2, 3, 4 :

$$a = \sum_{\|date(s') - date(s)\| \leq k} \text{cosine}(\vec{s}', \vec{s}) \quad [1]$$

$$b = \sum_{\|date(s') - date(s)\| \leq k} (1 - \text{cosine}(\vec{s}', \vec{s})) \quad [2]$$

$$c = \sum_{\|date(s') - date(s)\| > k} \text{cosine}(\vec{s}', \vec{s}) \quad [3]$$

$$d = \sum_{\|date(s') - date(s)\| > k} (1 - \text{cosine}(\vec{s}', \vec{s})) \quad [4]$$

Le score final de la mesure *Burstiness* pour une paire phrase/date s est obtenu en faisant la somme des *log-likelihood ratio* avec k variant de 1 à n tel que défini en 5.

$$Burstiness(s) = \sum_{k=1}^n LL(s)$$

où [5]

$LL(s) = \log$ likelihood ratio de la table de contingence 3.

La mesure *Interest* est par la suite réutilisée afin de retirer les phrases doublons en tenant compte du classement obtenu préalablement. Le système décrit par Chieu et Lee (2004) a pour but de produire des événements (texte), et non des dates. Pour l'adapter à notre problème, nous choisissons d'ordonner les dates selon la valeur de *Burstiness* de l'événement correspondant.

Modèles	Scores MAP	
	Corpus anglais	Corpus français
<i>Baselines avec uniquement les dates absolues</i>		
BL_{abs}^{occ}	0,2627	0,3303
BL_{abs}^{luc}	0,2782	0,3516
$BL_{abs}^{tf.idf}$	0,2778	0,3528
<i>Baseline système Chieu</i>		
$Chieu_{burst}$	0,4895	0,4823

Tableau 4. Scores MAP pour les deux systèmes baselines

Lors des tests préliminaires, nous avons fait varier k entre 1 et 20. Les meilleurs résultats pour la MAP ont été obtenus avec k variant de 1 à 10 ($n = 10$). Cependant, le système reste extrêmement lent lorsque le nombre de phrases renvoyées pour la requête est important malgré l'optimisation de la mesure *Interest* décrite dans l'article.

Le tableau 4 présente les résultats pour ces deux systèmes. La *baseline* qui utilise uniquement des dates absolues présente des résultats extrêmement faibles quel que soit le type de classement. Le système *Chieu* présente des résultats intéressants avec un score MAP autour de 0,5.

5.2. Extraction de dates saillantes

5.2.1. Avec filtrage simple

Dans ces expériences, l'index Lucene a été construit de la façon suivante : chaque **paragraphe** contenant une date normalisée par XIP est considéré comme un document. Le titre ainsi que les mots-clés de la dépêche AFP sont ajoutés à chaque paragraphe du document lors de l'indexation. Pour une requête, les 10 000 premiers documents sont renvoyés. Différentes combinaisons de filtrage ont été mises en place afin de ne pas prendre en compte les dates associées à un verbe de discours rapporté (R), un verbe de déclaration (D), un verbe modal (M) ou encore un verbe au futur (F). L'objectif de ces filtres est de pouvoir retirer les références à des événements incertains qui pourraient bruyé les résultats.

Cette série d'expériences est appelée par la suite SD^7 , avec différents indices en fonction du type de filtrage utilisé. Par exemple, une série de tests avec un filtrage de des verbes modaux et des verbes au futur sera appelée $SD_{M,F}$. Pour toutes ces combinai-

7. Pour Salient Dates en anglais

Modèles	Scores MAP	
	Anglais	Français
<i>Tests avec toutes les dates</i>		
SD^{occ}	0,6502	0,6279
SD^{luc}	0,6962	0,6722
$SD^{tf.idf}$	0,6982	0,6756
<i>Tests avec filtrage</i>		
SD_R^{luc}	0,6975	0,6719
SD_F^{luc}	0,6967	0,6712
SD_M^{luc}	0,6978	0,6723
SD_D^{luc}	0,7066 **	0,6687
SD_{RFMD}^{luc}	0,7127 **	0,6658
$SD_R^{tf.idf}$	0,6996	0,6752
$SD_F^{tf.idf}$	0,6993 **	0,6744
$SD_M^{tf.idf}$	0,7005 *	0,6754
$SD_D^{tf.idf}$	0,7091 **	0,6713
$SD_{RFMD}^{tf.idf}$	0,7146 **	0,6684

Tableau 5. Scores MAP obtenus en combinant extraction de dates saillantes avec XIP et filtrage simple. La signification statistique des résultats avec filtrage comparés aux résultats sans filtrage est indiquée par le t-test de Student (* : $p < 0,05$ (significatif) ; ** : $p < 0,01$ (très significatif)). L'utilisation du $tf.idf(d)$ par rapport au $occ(d)$ présente également une progression très significative dans les deux langues.

sons, les dates obtenues sont classées en faisant la somme des scores Lucene (*luc*) des paragraphes associés ou par le $tf.idf$ ⁸.

Le tableau 5 présente les résultats obtenus pour chacune de ces expériences. Les scores MAP obtenus sont meilleurs que pour les *baselines*. L'utilisation du $tf.idf(d)$ apporte une très légère amélioration par rapport aux scores *luc*. Les différents filtrages permettent une amélioration globale mais nécessitent d'être approfondis afin de pouvoir affiner les résultats.

8. Nous ne présentons plus par la suite les classements obtenus en fonction de la fréquence d'apparition de chacune des dates (*occ*), ceux-ci étant, comme pour les approches de base, systématiquement plus faibles.

5.2.2. Avec apprentissage

Nous avons utilisé l'ensemble de chronologies manuelles comme corpus d'apprentissage afin d'entraîner nos modèles.

Nous utilisons IcsiBoost⁹, une version libre du classifieur BoosTexter, fondé sur Ada-Boost (Freund et Schapire, 1997), un algorithme de *boosting* de classifieurs simples (des arbres de décision à 1 niveau de profondeur). Concernant le paramètre définissant le nombre d'itérations de l'algorithme, les expériences ont montré qu'un optimum était atteint aux alentours des 1 000 tours.

Dans notre approche, nous considérons deux classes : les dates saillantes sont les dates contenues dans les chronologies manuelles tandis que toutes les autres dates sont considérées comme des dates non saillantes. Cette approche présente cependant un biais important. Les choix des journalistes sont en effet très subjectifs et les chronologies ne devant pas dépasser une certaine longueur, des dates pertinentes peuvent être exclues. Ces questions seront discutées dans la section 5.3.

Nous avons choisi de ne pas représenter chaque paragraphe comme une instance du classifieur mais plutôt de regrouper tous les paragraphes correspondant à la même date avant la phase d'apprentissage. Ainsi, chaque instance correspond à une date unique, et les traits associés concernent l'ensemble des paragraphes contenant cette date.

Les traits utilisés pour ces expériences sont les suivants :

1) traits représentant le fait que plus une date est mentionnée, plus celle-ci doit être importante :

- somme des scores Lucene des paragraphes contenant cette date,
- nombre de paragraphes contenant cette date,
- ratio entre les scores Lucene de cette date et les scores Lucene de toutes les dates,
- ratio entre le nombre de phrases contenant cette date et le nombre de phrases de toutes les dates ;

2) traits représentant le fait qu'un événement important est toujours mentionné longtemps après qu'il se soit produit :

- distance en nombre de jours entre la date et l'évocation la plus récente de cette date (la plus tardive dans le corpus),
- distance en nombre de jours entre cette date et la DCT ;

3) autres traits :

- meilleur classement Lucene de la date,

9. <http://code.google.com/p/icsiboost/>

- nombre de fois où la date est absolue dans les paragraphes (*i.e.*, nombre de paragraphes renvoyés, dans lesquels la date était une date absolue, normalisée sans recours à une connaissance externe ou contextuelle, par opposition aux dates relatives),
- nombre de fois où la date est relative et normalisée dans les paragraphes,
- nombre total de mots-clés de la requête présents dans le titre, le texte ou les entités nommées des documents renvoyés,
- nombre de fois où la date est extraite d'un discours rapporté ou au futur.

Nous n'avons pas pour but de classifier de manière binaire les dates, mais plutôt de les classer. Pour cette raison, nous avons utilisé la probabilité $P(d)$ renvoyée par le classifieur combinée au score Lucene des paragraphes ou au $tf.idf$ de la date d :

$$score(d) = P(d) \times val(d)$$

où $val(d)$ est soit $luc(d)$ ou $tf.idf(d)$.

Étant donné le biais signalé ci-dessus, et surtout la faible quantité de données d'apprentissage, les approches communément appelées *Learning to Rank*, qui visent à apprendre un modèle de tri à partir de listes partiellement ordonnées et d'un ensemble de traits, ne semblent pas adaptées à notre problème. En effet, ces méthodes demandent un gros volume de données d'apprentissage, ce que nous ne pouvons pas avoir dans notre cas.

Nous avons évalué cette approche avec une validation croisée classique sur quatre partitions du corpus d'apprentissage. Chaque chronologie manuelle est ainsi répartie aléatoirement dans une des partitions. Nous utilisons trois sous-corpus en tant que données d'apprentissage (deux sont utilisés en apprentissage tandis que le dernier permet de calibrer le classifieur) et le quatrième corpus est réservé au test : quatre jeux de test tournants sont ainsi créés. Le score final d'un test est ensuite calculé en faisant la moyenne de tous les scores obtenus sur chacun des quatre corpus de test.

Le tableau 6 présente les scores MAP obtenus avec apprentissage. ML_{base}^{luc} utilise le score Lucene et uniquement le premier type de traits décrit précédemment. ML^{luc} et $ML^{tf.idf}$ utilisent les trois types de traits. Nous observons ainsi une progression du score MAP d'environ 0,08 point sur le corpus anglais et de 0,02 sur le corpus français.

5.2.3. Analyse des erreurs

Une analyse plus détaillée des erreurs montre que celles-ci peuvent être de plusieurs types. Un manque de précision dans les requêtes définies manuellement peut conduire le système à proposer des dates non valides pour le thème. Ainsi la requête « Niger » utilisée pour retracer les différents événements de la crise au Niger en 2009, conduit le système à proposer les dates des matchs de la coupe du Niger qui s'est déroulée durant l'été 2009. De la même façon, la requête « Georges Tron », associée à la chronologie sur les accusations d'agression sexuelle, ne permet pas de faire la distinction entre les dates des interventions du secrétaire d'État et celles de l'affaire proprement dite.

Modèles	Scores MAP	
	Anglais	Français
<i>Tests avec apprentissage</i>		
ML_{base}^{luc}	0,7033	0,6684
ML^{luc}	0,7905 **	0,6988
$ML^{tf.idf}$	0,7918 **	0,7008

Tableau 6. Scores MAP obtenus pour l'extraction de dates saillantes avec apprentissage. Les résultats sur l'anglais présentent une amélioration très significative ($p \approx 6.10^{-4}$) comparés respectivement à SD^{luc} et $SD^{tf.idf}$.

D'autre part, même si la requête semble correcte, des documents moins pertinents sélectionnés par le moteur de recherche peuvent introduire du bruit dans les résultats. Par exemple, pour la chronologie des événements survenus lors de la crise dans le monde de la formule 1 en 2009, l'ensemble des dates des événements associés à cette crise est retrouvé mais l'ordre de celles-ci est bruité par les dates de Grands Prix qui ont également une fréquence importante dans le corpus. De même, la chronologie retraçant l'ensemble des crimes commis à Marseille au cours de l'année 2009, avec comme requête « Marseille crime », obtient une très mauvaise précision moyenne (0,02). Une recherche dans le corpus montre que l'ensemble des dépêches relatant ces événements font généralement référence à des lieux plus précis (tels que les quartiers nord, la cité des Bleuets ou l'hôpital de la Timone). Ces documents pertinents ne sont donc pas renvoyés par Lucene, qui donne la priorité à d'autres documents relatifs aux crimes ayant eu lieu pendant cette période, sans lien avec la ville de Marseille. C'est, dans ce cas, la taille de l'unité textuelle choisie qui est en cause.

De plus, certains événements choisis par les journalistes pour figurer dans les chronologies manuelles n'ont pas de lien direct avec la thématique. Ainsi la date du 21 février 2011 dans la chronologie « Decline and fall of Hosni Mubarak » est associée aux violences en Libye, pays frontalier. Ce dernier exemple montre la difficulté d'évaluer les résultats de notre système par rapport à des chronologies de référence construites manuellement de manière subjective (voir section 5.3).

Enfin, nous pouvons remarquer que les performances sont nettement meilleures pour l'anglais que pour le français. Une analyse plus détaillée des résultats montre que pour l'anglais, 0,6 % seulement des dates présentes dans les chronologies de référence ne sont pas du tout retrouvées par le système, alors qu'elles sont 3,6 % pour le français. Nous avons constaté qu'il arrive que le corpus de dépêches en français ne contienne pas du tout (ou parfois très peu) de références à certains événements décrits dans les chronologies, ainsi que quelques incohérences entre la date de la chronologie et la date fournie dans le corpus. Ce nombre plus grand de dates non retrouvées en français peut expliquer en partie les performances moindres du système pour le français.

Topic	AP_C	AP_E
<i>Maroc</i>	0,5847	0,5718
<i>Kirghizistan</i>	0,6125	0,9989
<i>Libye</i>	0,7856	1
<i>Liban</i>	0,4673	0,7652

Tableau 7. Précision moyenne pour quatre thèmes, en comparant avec la chronologie originale (AP_C , évaluation automatique) et avec l'avis de l'expert (AP_E , évaluation manuelle)

5.3. Qu'est-ce qu'un bon résultat ?

Les chronologies existantes, écrites manuellement par les journalistes de l'AFP, constituent une ressource très utile pour une évaluation objective de notre système. En effet, elles sont totalement dissociées de notre sujet de recherche et sont une représentation exacte du type de sortie que nous souhaitons.

Pourtant, la composition d'une telle chronologie est une tâche très subjective, et aucune méthode simple ne permet d'estimer ce que serait l'accord entre deux journalistes écrivant la même chronologie. En effet, seuls des experts d'un domaine peuvent rédiger une chronologie sur ce domaine, et calculer le traditionnel « accord interannotateur » nécessiterait la participation d'au moins deux experts de chaque domaine concerné, ce qui semble impossible à obtenir.

Nous proposons deux expérimentations spécifiques à ce point délicat.

5.3.1. Évaluation manuelle

Dans la première expérimentation, nous avons choisi quatre sujets en anglais, concernant le Printemps arabe¹⁰, et nous avons proposé la chronologie produite par notre système pour chacun de ces sujets à un expert de l'AFP. Nous lui avons demandé d'estimer la pertinence des 30 premières dates de la chronologie.

Le tableau 7 présente les résultats de cette évaluation. Il offre une comparaison de la précision moyenne par rapport à la chronologie originale composée manuellement (AP_C) d'une part, et au jugement de l'expert humain (AP_E) d'autre part. Ces valeurs montrent que, pour trois *runs* sur les quatre, de nombreuses dates suggérées par le système sont considérées comme valides par l'expert, alors qu'elles n'étaient pas présentes dans la chronologie originale.

Malgré le faible nombre de sujets concernés, ceci montre que notre système peut être une aide efficace pour le journaliste, puisque celui-ci considère la plupart des dates proposées comme pertinentes.

10. « Arab revolt timeline for Morocco », « Kyrgyzstan unrest timeline », « Lebanon's new government : a timeline », « Libya timeline ».

Thème	Nombre de dates			Rappel	Précision	F-mesure
	EN (réf.)	FR	communes			
Mine chilienne	17	18	15	0,88	0,83	0,86
Lybie (avr. 2011)	14	20	14	1	0,7	0,82
Lybie (juil. 2011)	17	26	11	0,65	0,42	0,51
Maroc (juin 2011)	10	17	7	0,7	0,41	0,52
Syria (juin 2011)	20	26	15	0,75	0,58	0,65
Yémen (mars 2011)	15	17	13	0,87	0,76	0,81
Total	93	124	71	0,81	0,62	0,70

Tableau 8. *Accords entre les chronologies françaises et anglaises sur le même thème (même jour de parution)*

5.3.2. Accord interannotateur sur deux langues

Dans le but d'estimer un accord interannotateur entre deux journalistes écrivant la même chronologie, nous avons extrait de notre collection de chronologies manuelles des paires de chronologies répondant aux contraintes suivantes :

- l'une est écrite en français, l'autre en anglais ;
- le thème des deux chronologies donné par le titre est le même ;
- les deux chronologies ont été écrites le même jour ;
- les auteurs des deux chronologies sont différents.

Six paires de chronologies répondent à ces critères. Nous avons considéré arbitrairement les chronologies anglaises comme la référence et donnons les scores de précision, rappel et F-mesure obtenus par la chronologie française correspondante¹¹. On voit dans le tableau 8 que l'accord entre deux journalistes créant une chronologie sur le même thème est donc relativement faible.

Pour une tâche donnée, un système n'a pas nécessairement pour mission d'approcher 100 % de résultats en commun avec la référence, mais plutôt de s'approcher de celle-ci avec un accord comparable à l'accord possible entre deux humains. Ce résultat permet donc de mettre en perspective les valeurs de MAP obtenues par notre système et données plus haut.

6. Conclusion et perspectives

En présentant un système d'extraction des dates saillantes à partir de la requête d'un utilisateur, cet article montre l'importance de la prise en compte de l'information temporelle, et comment une analyse temporelle relativement simple permet de pointer

11. La MAP n'ayant ici pas de sens puisque les dates des chronologies ne sont pas ordonnées.

sur les événements importants d'un thème, par l'intermédiaire des dates associées à ces événements.

Notre système obtient des résultats satisfaisants puisque proches des chronologies produites manuellement par les journalistes. Ainsi, pour l'anglais notre système produit 52,8 % de chronologies ayant un score de précision moyenne supérieur à 0,8 par rapport aux chronologies de référence alors qu'il ne produit que 6,6 % de chronologies ayant un score inférieur à 0,5. Pour le français, en revanche, le système ne propose que 34 % de chronologies ayant une précision moyenne supérieure à 0,8, et 22,4 % de chronologies ayant un score inférieur à 0,5. Comme nous l'avons noté, cette différence semble s'expliquer par l'absence de nombreux événements du corpus en français, ainsi qu'une redondance moins importante des informations.

Le but final du projet et la suite logique de notre travail sont de repérer les événements correspondant à ces dates saillantes. Ceci passe donc par la prise en compte du contenu textuel des documents et des passages associés aux dates. Les travaux existants sur les scores de similarité entre les phrases et de « centralité » thématique devraient permettre d'apporter de bonnes indications sur les phrases les plus représentatives d'un événement parmi celles associées à une date. Une autre direction intéressante dans laquelle nous comptons nous engager est l'utilisation de techniques de catégorisation pour déterminer si les passages associés à une date correspondent vraiment à un seul et même événement. Ceci devrait permettre d'affiner la détection des dates saillantes.

Concernant l'amélioration de l'extraction des dates saillantes elles-mêmes, elle pourrait bénéficier d'approches hybrides en utilisant à la fois l'analyse temporelle efficace décrite dans cet article, et les techniques fondées sur le contenu, plus proches du domaine du résumé automatique multidocument.

Enfin, si nous avons présenté ici des résultats séparés pour l'anglais et le français, il sera intéressant d'étudier la possibilité de combiner les informations en provenance des corpus des deux langues pour une extraction multilingue des dates saillantes.

Remerciements

Ce travail a été partiellement financé par l'Agence Nationale pour la Recherche (ANR) dans le cadre du projet Chronolines (ANR-10-CORD-010). Nous remercions l'Agence France Presse (AFP) pour la mise à disposition du corpus et son concours pour l'évaluation.

7. Bibliographie

- Abney S., « Parsing By Chunks », *Principle-Based Parsing*, Kluwer Academic Publishers, p. 257-278, 1991.
- Aït-Mokhtar S., Chanod J.-P., Roux C., « Robustness beyond Shallowness : Incremental Deep Parsing », *Natural Language Engineering*, vol. 8, p. 121-144, 2002.
- Allen J. (ed.), *Topic Detection and Tracking*, Springer, 2002.
- Allen J., Gupta R., Khandelwal V., « Temporal summaries of new topics », *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '01, p. 10-18, 2001.
- Alonso O., Baeza-Yates R., Gertz M., « Exploratory Search Using Timelines », *SIGCHI 2007 Workshop on Exploratory Search and HCI Workshop*, 2007.
- Alonso O. R., Temporal information retrieval, PhD thesis, University of California at Davis, Davis, CA, USA, 2008. Adviser-Gertz, Michael.
- Barzilay R., Elhadad N., « Inferring Strategies for Sentence Ordering in Multidocument News Summarization », *Journal of Artificial Intelligence Research*, vol. 17, p. 35-55, 2002.
- Battistelli D., Couto J., Minel J.-L., Schwer S., « Representing and visualizing calendar expressions in texts », *Proceedings of STEP 08*, Venice, Italy, sept, 2008.
- Bittar A., Amsili P., Denis P., Danlos L., « French TimeBank : An ISO-TimeML Annotated Reference Corpus », *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011)*, Portland, Oregon, July, 2011. Short paper.
- Bittar A., Hagège C., Moriceau V., Tannier X., Tesseidre C., « Temporal Annotation : A Proposal for Guidelines and an Experiment with Inter-annotator Agreement », *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, Istanbul, Turkey, May, 2012.
- Brants T., Chen F., Farahat A., « A System for new event detection », *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '03, ACM, New York, NY, USA, p. 330-337, 2003.
- Chieu H. L., Lee Y. K., « Query based event extraction along a timeline », *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '04, p. 425-432, 2004.
- Cohen J., « A Coefficient of Agreement for Nominal Scales », *Educational and Psychological Measurement*, vol. 43, n° 6, p. 551-558, 1960.
- Freund Y., Schapire R. E., « A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting », *Journal of Computer and System Sciences*, vol. 55, n° 1, p. 119-139, 1997.
- Fung G. P. C., Yu J. X., Yu P. S., Lu H., « Parameter free bursty events detection in text streams », *VLDB '05 : Proceedings of the 31st international conference on Very large data bases*, p. 181-192, 2005.
- Hagège C., Tannier X., « XTM : A Robust Temporal Text Processor », *Computational Linguistics and Intelligent Text Processing, proceedings of 9th International Conference CICLing 2008*, Springer Berlin / Heidelberg, Haifa, Israel, p. 231-240, February, 2008.

- Harabagiu S., Bejan C. A., « Question Answering Based on Temporal Inference », *Proceedings of the Workshop on Inference for Textual Question Answering*, Pittsburg, Pennsylvania, USA, July, 2005.
- Jung H., Allen J., Blaylock N., de Beaumont W., Galescu L., Swift M., « Building timelines from narrative clinical records : initial results based-on deep natural language understanding », *Proceedings of BioNLP 2011 Workshop*, BioNLP '11, Association for Computational Linguistics, Stroudsburg, PA, USA, p. 146-154, 2011.
- Kanhabua N., « Exploiting temporal information in retrieval of archived documents », *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2009, Boston, MA, USA, July 19-23, 2009*, p. 848, 2009.
- Kim Y., Choi J., « Recognizing Temporal Information in Korean Clinical Narratives through Text Normalization », *Health Inform Res*, vol. 17, n° 3, p. 150-5, 2011.
- Kumaran G., Allen J., « Text classification and named entities for new event detection », *SIGIR '04 : Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, p. 297-304, 2004.
- Li W., Li W., Lu Q., Wong K.-F., « A Preliminary Work on Classifying Time Granularities of Temporal Questions », *Proceedings of Second international joint conference in NLP (IJCNLP 2005)*, Jeju Island, Korea, oct, 2005a.
- Li Z., Wang B., Li M., Ma W.-Y., « A Probabilistic Model for Restrospective News Event Detection », *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM Press, New York City, NY, USA, Salvador, Brazil, 2005b.
- Llorens H., Saquete E., Navarro-Colorado B., Gaizauskas R., « Time-Surfer : Time-based Graphical Access to Document Content », in P. Clough, C. Foley, C. Gurrin, G. Jones, W. Kraaij, H. Lee, V. Mudoch (eds), *Advances in Information Retrieval*, vol. 6611 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, p. 767-771, 2011.
- Mestl T., Cerrato O., Øines J., Myrseth P., Gustavsen I.-M., « Time Challenges - Challenging Times for Future Information Search », *D-Lib Magazine*, 2009.
- Parent G., Gagnon M., Muller P., « Annotation d'expressions temporelles et d'événements en français », in F. Bechet, P. Bellot, J.-F. Bonastre, T. Jimenez (eds), *Actes de TALN 2008*, ATALA, LIA, Avignon, June, 2008.
- Pustejovsky J., Ingria R., Saurí R., Littman J., Gaizauskas R., Setzer A., Katz G., *The Specification Language TimeML*, Oxford University Press, chapter 27, p. 545-558, 2005.
- Pustejovsky J., Lee K., Bunt H., Romary L., « ISO-TimeML : An International Standard for Semantic Annotation », in N. C. C. Chair, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, D. Tapias (eds), *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, European Language Resources Association (ELRA), Valletta, Malta, may, 2010.
- Roux C., « Annoter les documents XML avec un outil d'analyse syntaxique », *11^e Conférence annuelle de Traitement Automatique des Langues Naturelles*, ATALA, Fès, Maroc, April, 2004.
- Saquete E., Vicedo J. L., Martínez-Barco P., Muñoz R., Llorens H., « Enhancing QA Systems with Complex Temporal Question Processing Capabilities », *Journal of Artificial Intelligence Research*, vol. 35, p. 775-811, 2009.

- Smith D. A., « Detecting events with date and place information in unstructured text », *JCDL'02 : Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, ACM, New York, NY, USA, p. 191-196, 2002.
- Swan R., Allen J., « Automatic generation of overview timelines », *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '00, ACM, New York, NY, USA, p. 49-56, 2000.
- Uzzaman N., Allen J., « Event and temporal expression extraction from raw text : first step towards a temporally aware system », *International Journal of Semantic Computing*, 2011.
- Verhagen M., Gaizauskas R., Schilder F., Hepple M., Katz G., Pustejovsky J., « SemEval-2007 - 15 : TempEval Temporal Relation Identification », *Proceedings of SemEval workshop at ACL 2007*, Association for Computational Linguistics, Morristown, NJ, USA, Prague, Czech Republic, June, 2007.
- Verhagen M., Sauri R., Caselli T., Pustejovsky J., « SemEval-2010 - Task 13 : Evaluation events, time expressions and temporal relations - TempEval-2 », *Proceedings of SemEval workshop at ACL 2010*, Association for Computational Linguistics, Morristown, NJ, USA, Uppsala, Sweden, July, 2010.
- Wang Y., Zhu M., Qu L., Spaniol M., Weikum G., « Timely YAGO : Harvesting, Querying, and Visualizing Temporal Knowledge from Wikipedia », *Proceedings of the 13th International Conference on Extending Database Technology (EDBT), Lausanne, Switzerland, March 22-26*, p. 697-700, 2010.
- Yan R., Kong L., Huang C., Wan X., Li X., Zhang Y., « Timeline Generation through Evolutionary Trans-Temporal Summarization », *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, EMNLP 2011, 27-31 July 2011, Edinburgh, UK*, p. 433-443, 2011a.
- Yan R., Wan X., Otterbacher J., Kong L., Li X., Zhang Y., « Evolutionary timeline summarization : a balanced optimization framework via iterative substitution », *Proceeding of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2011, Beijing, China, July 25-29, 2011*, p. 745-754, 2011b.
- Yang Y., Pierce T., Carbonell J. G., « A study on retrospective and on-line event detection », *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM Press, New York City, NY, USA, Melbourne, Australia, August, 1998.