

Annotation d’expressions temporelles et d’événements en français

Gabriel Parent¹ Michel Gagnon¹ Philippe Muller²

(1) École Polytechnique de Montréal

(2) Université Paul-Sabatier, Toulouse

Résumé. Dans cet article, nous proposons une méthode pour identifier, dans un texte en français, l’ensemble des expressions adverbiales de localisation temporelle, ainsi que tous les verbes, noms et adjectifs dénotant une éventualité (événement ou état). Cette méthode, en plus d’identifier ces expressions, extrait certaines informations sémantiques : la valeur de la localisation temporelle selon la norme TimeML et le type des éventualités. Pour les expressions adverbiales de localisation temporelle, nous utilisons une cascade d’automates, alors que pour l’identification des événements et états nous avons recours à une analyse complète de la phrase. Nos résultats sont proches de travaux comparables sur l’anglais, en l’absence d’évaluation quantitative similaire sur le français.

Abstract. We propose a method to extract expressions of temporal location in French texts, and all verbs, nouns and adjectives that denote an event or a state. This method also computes some semantic information : the value of the temporal location according to the TimeML standard and the types of eventualities. For temporal location expression, we use a cascade of transducers, whereas event identification is based on full syntactic parsing. Our results are compared to similar work on English, as no other empirical evaluation has been done on French before.

Mots-clés : Extraction d’informations temporelle, TimeML.

Keywords: Temporal information extraction, TimeML.

1 Introduction

Le problème de l’extraction d’informations à partir de données textuelles reçoit depuis quelques années une très grande attention. Il inclut généralement plusieurs étapes, dont la reconnaissance d’entités et d’événements dans lesquelles elles sont impliquées, ainsi que les informations temporelles, spatiales et autres qui peuvent leur être rattachées. Mais bien que le problème de l’identification des entités nommées soit relativement bien compris (Nadeau & Sekine, 2007), l’identification d’événements et de leurs relations temporelles demeure un problème qui est loin d’être résolu. Or, que ce soit pour les applications de réponses à des questions, ou la simple recherche d’information, on est souvent appelé à rechercher ou annoter dans les documents les éléments qui nous permettent de répondre à des questions de la forme suivante : Quand un certain événement a-t-il eu lieu ? Que s’est-il passé pendant une certaine période de temps donnée ? Dans quels événements une personne a-t-elle été impliquée ?

Pour donner une idée, environ 15% des questions posées dans le cadre des campagnes CLEF portent sur des informations temporelles. Ce nombre est à peu près similaire dans les instances des campagnes sur l'inférence textuelle (RTE). Dans les campagnes d'évaluation de résumé automatique (DUC), nous avons dénombré 30% environ de phrases contenant un adverbial temporel, à la fois dans les textes et les résumés de référence, ce qui montre l'importance de cette information dans la structuration et la cohérence d'un texte.

Pour arriver à capturer cette information, il faut évidemment être capable d'analyser les documents afin d'en extraire la structure temporelle. Cette tâche exige une analyse aux niveaux syntaxique, sémantique et discursif. Au niveau syntaxique, il faut reconnaître les différentes formes d'expressions adverbiales temporelles que l'on peut rencontrer dans un texte : *demain, hier, la semaine prochaine, deux jours après son départ, le 8 janvier à 8h00, depuis le 4 avril, etc.* Il faut aussi reconnaître les différentes occurrences d'éventualités¹, qui peuvent être exprimés sous forme verbale, nominale ou adjectivale :

- (1) a. Secoués par une violente secousse d'une magnitude de 5,8 sur l'échelle de Richter jeudi à 03 h 15 (02 h 15 GMT), les Algérois sont descendus nombreux jeudi soir dans les rues d'Alger, a constaté l'AFP.
- b. Le conflit a commencé le mois dernier.

En plus de repérer les expressions pertinentes, il faut aussi distinguer les formes qui ne représentent pas une localisation temporelle ou une éventualité à extraire :

- (2) a. Jean est un chauffeur du *dimanche*.
- b. Jean, *20 ans*, est le plus jeune joueur de l'équipe.
- c. Il est déconseillé de *regarder* la télévision durant de longues heures.
- d. Jean est *intelligent*.
- e. Jean a retiré son *annonce*.

Dans les deux premiers exemples, on retrouve des expressions qui dans d'autres contextes pourraient dénoter une localisation temporelle ou une durée, mais qui dans ce cas jouent plutôt le rôle d'un épithète. Dans l'exemple suivant, le verbe *regarder* réfère à une activité générique plutôt qu'à un événement spécifique. La phrase 2d, quant à elle, nous révèle une caractéristique permanente de Jean, plutôt que d'exprimer un état temporaire. Finalement, dans le dernier exemple, on retrouve un nom qui, dans certains contextes, pourrait être vu comme un événement (comme dans *L'annonce de sa victoire a été faite tard hier soir*), mais qui dans ce cas est plutôt un objet.

Pour l'identification des expressions adverbiales temporelles, nous verrons qu'une méthode basée sur une cascade d'automates est suffisante, étant donné la nature locale des informations nécessaires pour les identifier. L'identification des événements implique par contre de faire appel à une analyse syntaxique, combinée à des règles lexico-sémantiques. Dans l'exemple 2c, on doit savoir que le verbe *regarder*, complément du verbe *déconseillé*, ne doit pas être annoté puisque la sémantique du verbe principal n'implique que l'action a été réalisée. Dans l'exemple 1b, on détecte que *conflit* réfère à un événement, parce qu'il est le sujet du verbe *a commencé*, soit un verbe qui, par son aspect, implique que son sujet soit un événement non ponctuel. On voit aussi que certaines informations lexicales doivent être utilisées.

¹Nous regroupons sous le vocable *éventualité* tout événement ou état.

Nous verrons, dans la section qui suit, que quelques solutions pour résoudre ce problème ont été proposées pour les textes en anglais. Pour le français, à notre connaissance, nous sommes les seuls à avoir abordé ce problème en suivant une méthodologie comparable. Suivent une présentation de nos méthodes pour l'identification des expressions de localisation temporelle, dans un premier temps, puis l'identification des termes qui réfèrent à des événements ou des états. Pour chacune, nous présentons une démarche d'évaluation et des résultats. Finalement, nous évaluons le potentiel des approches proposées.

2 Travaux reliés

La détection d'expressions temporelles et d'événements a suscité un certain nombre de travaux, pour l'essentiel restreints à l'anglais dans le cadre des campagnes MUC ou ACE (avec récemment l'introduction de la tâche sur le chinois et l'espagnol)², et à l'intérieur du groupe de travail préparant la norme TimeML (Pustejovsky *et al.*, 2005) (avec des exceptions rares sur l'allemand et l'italien)³. Si en français on peut citer les travaux de (Bernard & Maurel, 1993; Molinès, 1989) sur l'interprétation des adverbiaux temporels, il n'existe pas de travaux récents ayant la couverture des travaux sur l'anglais ou avec des validations empiriques comparables. Le travail initié dans (Muller & Tannier, 2004) avait l'ambition de se placer dans un cadre complet similaire à TimeML, mais la quantité de données concernées était alors assez faible. Nous présentons ici la suite de ce travail, restreint au repérage des adverbiaux temporels (et au calcul de leur référence selon la norme TimeML) et aux événements.

A l'intérieur des travaux similaires sur l'anglais on peut distinguer deux types d'approches : des approches à base de règles de pattern-matching de complexité variable, écrites "manuellement" (Mani & Wilson, 2000; Mazur & Dale, 2007) ou bien des approches à base d'apprentissage automatique, qui intègrent presque toujours une analyse syntaxique partielle (du simple étiquetage morpho-syntaxique à du chunking plus ou moins élaboré), cf notamment (Boguraev & Ando, 2005). Dans ce cadre le problème de détection est ramené à une classification de type IOB (In-Out-Boundary) où chaque item lexical est classifié comme à l'intérieur (I), à l'extérieur (O) ou délimitant une frontière (B) d'une expression temporelle (événement ou adverbial), éventuellement avec un niveau d'imbrication interne (Hacioglu *et al.*, 2005). Le calcul des références temporelles est alors fait par des règles lexicales ad hoc, à l'exception de (Baldwin, 2002) qui tente de généraliser aussi l'interprétation à base d'apprentissage, avec des résultats plus mitigés.

La reconnaissance d'adverbiaux temporels obtient généralement des scores comparables avec les deux méthodes, avec des F-scores voisins de 80-85% pour la détection et le calcul de référence, selon la manière de compter (cf section 4) et les normes (Timex2 pour Ace, Timex3 pour TimeML). En particulier, les travaux de (Negri & Marseglia, 2004), (Mani & Wilson, 2000) et (Mazur & Dale, 2007) montrent bien qu'on peut espérer de bons résultats avec un système basé sur un ensemble de règles, semblable à ce que nous proposons ici.

Le système EVITA (Saurí *et al.*, 2005), qui utilise une analyse partielle des phrases, un système de patrons lexicaux-syntaxiques et Wordnet, obtient un taux de précision de 74% pour le classement des événements et un rappel de 87%, mais il est à noter que ces résultats ne sont pas obtenus avec un corpus d'évaluation indépendant du corpus d'entraînement. Les résultats d'évaluations ne sont pas très détaillés. Notamment, on ne connaît pas le nombre total d'occurrences

²<http://www.nist.gov/speech/tests/ace/ace07/>.

³<http://www.timeml.org>

et les performances pour chaque type d'expression reconnue (noms, verbes, adjectifs). Il est intéressant de comparer ce travail avec le système STEP (Bethard & Martin, 2006), entièrement basé sur de l'apprentissage automatique, utilisant le même corpus d'entraînement et des traits syntaxiques incluant ceux d'EVITA. En termes de précision et de rappel, les résultats obtenus avec l'ensemble total de caractéristiques est légèrement supérieur à ceux obtenus en n'utilisant que le sous-ensemble des traits pris par EVITA. Dans la littérature, on retrouve d'autres approches qui visent à identifier les événements cités dans un texte, mais aucune de celles-ci n'aborde le problème de manière aussi complète que les deux systèmes que nous venons de citer.

L'idée d'utiliser une analyse syntaxique pour l'identification d'événements n'est pas nouvelle. Dans un contexte d'extraction d'informations, (Madhyastha *et al.*, 2003) ont utilisé une grammaire de dépendance (Sleator & Temperley, 1991) pour identifier les contextes syntaxiques des événements recherchés. Similairement, (Yakushiji *et al.*, 2001) se servent d'une analyse syntaxique complète de la phrase pour ramener la recherche d'information à une recherche dans un ensemble limité de formes syntaxiques canoniques. Mais dans les deux cas, il s'agit d'une méthode appliquée à un domaine restreint et dont le but n'est pas d'identifier les événements d'un texte, mais plutôt l'extraction d'informations relatives à un ensemble d'événements pré-définis.

3 La norme TimeML

La norme TimeML est le fruit d'un groupe de travail dirigé par James Pustejovsky (Pustejovsky *et al.*, 2005) qui vise à standardiser les annotations sémantiques reliées à la temporalité dans un texte en langage naturel. Le guide d'annotation est tournée vers l'anglais, mais les éléments du standard sont a priori indépendants de la langue considérée. La norme traite de trois sortes d'entités annotables : les adverbiaux (objet TIMEX3), les éventualités (EVENT), et les signaux (SIGNAL) ; elle introduit aussi plusieurs sortes de relations entre ces entités : liens d'ordonnement (TLINK), de modalité (SLINK), ou bien aspectuels (ALINK), couvrant ainsi l'essentiel des informations temporelles que l'on peut associer à un texte.

Les TIMEX ont plusieurs attributs, comme le type (date, heure, durée), la valeur absolue correspondant à la localisation temporelle, le temps d'ancrage, pour les localisations qui sont calculées relativement à une autre (deux jours après *son départ*) et la quantité ou la fréquence, pour des localisations qui correspondent à plus d'un moment dans le temps (tous les jeudi d'avril). Les éventualités ont pour attribut une classe parmi : occurrence (la plupart des événements), état, action ou état intensionnel (ceux-ci, comme dans *je voudrais qu'il vienne*, supposent un second événement qui ne se réalise pas nécessairement), action d'énonciation ou de perception (dire, révéler, entendre, voir, etc.) et finalement les événements aspectuels (commencer, finir de, continuer, etc.). Les autres attributs sont principalement : le temps grammatical, l'aspect progressif ou non (s'il est déterminable), la forme (nom, verbe, adjectif). Finalement, la balise <SIGNAL> sert à identifier un item (généralement une préposition) marquant une relation temporelle entre événements et localisations temporelles : *Il viendra après le 3 avril.*

Nous nous concentrons ici sur les entités à repérer, les relations qu'elles entretiennent étant un problème à part.

4 L'identification des expressions de localisation temporelle

La détection d'expressions de localisation temporelle est ici focalisée sur l'identification des expressions correspondant à la norme TimeML, qui en distingue trois types : les dates, les "temps" (comme dans 'à trois heures') et les durées. Cette étape inclut également le calcul de la référence temporelle suivant la norme ISO modifiée par la norme TimeML⁴, ainsi que certains attributs supplémentaires (par exemple des classes de modificateurs, comme "environ", etc). Nous avons ajouté quelques attributs pertinents pour les tâches en aval de cette reconnaissance (calcul des référents des dates et liens avec les événements), qui permettent notamment d'introduire la notion de perspective temporelle, nous inspirant des travaux de (Molinès, 1989). Nous faisons donc la distinction entre une perspective absolue, déictique car relative au temps d'énonciation (TE), ou anaphorique car relative à un point de perspective temporel ou focus temporel (FT). Cela permet de séparer les cas suivants :

- date relative au TE : *il y a 2 ans, l'année dernière.*
- dates relatives au FT : *3 jours plus tard, le mois suivant.*
- il reste des dates ambiguës TE/FT : *le 25 mars, en juin.*
- durées avec ancrage relatif au TE : *depuis un an.*
- durées avec ancrage relatif au FT : *depuis.*

Nous avons également constitué un lexique des éléments pertinents pour le temps (environ 120), associant un ensemble de traits qui instancie directement une ou des valeurs définies par la norme TimeML ou qui préparent leur calcul. Par exemple :

- matin : [unité:h valeur:MO] (unité ISO concernée : heures, valeur : morning)
- environ : [mod:approx]
- janvier : [unité:m val:1] (unité ISO concernée : mois)

Le traitement de cette étape est constitué d'un pipe-line utilisant un tokeniseur, un étiqueteur morpho-syntaxique (le Treetagger de (Schmid, 1994)), certains post-traitements pour isoler les éléments lexicaux pertinents ou corriger des erreurs de l'étiqueteur, puis une cascade d'automates effectuant une analyse syntaxique superficielle (en utilisant le Cass de (Abney, 1996)) à base d'environ 90 règles séparées en une cascade de 20 niveaux⁵. La structure syntaxique superficielle fournissait auparavant l'essentiel des informations syntaxiques nécessaires en aval (Muller & Tannier, 2004) mais est maintenant corrigée par le niveau suivant avec un analyseur plus complet, qui reprend uniquement les structures adverbiales détectées ici. Nous n'avons pour l'instant gardé cette analyse plus locale que pour le calcul de la référence temporelle (date exacte) quand celle-ci est ambiguë ou nécessite le contexte pour être interprétée. La dernière étape du traitement consiste à calculer l'attribut "val" de TimeML qui est une normalisation de la référence temporelle pour les dates, heures et durées, à partir de l'analyse précédente. Les dates pouvant poser problème se répartissent dans les deux classes TE/FT :

- par rapport à la date de la dépêche : "Les deux présidents se rencontrent mardi"
- par rapport à un point de perspective : "Les deux hommes se sont déjà vus deux jours avant"

Nous présentons ci-dessous les résultats obtenus à la fois sur le corpus d'étude (une vingtaine de dépêches AFP), et sur le corpus de test, obtenu par mise à jour du corpus de (Baldwin, 2002), qui utilisait une norme précédente (Timex2, issu de la campagne ACE). Celui-ci était constitué de 45 dépêches AFP. La colonne "app" indique de façon similaire à la campagne ACE le nombre

⁴Cf aussi le groupe ISO TC37/4 sur la gestion de ressources linguistiques.

⁵Tous ces outils sont disponibles sur demande.

d'adverbiaux détectés mais dont l'étendue est incomplète (par exemple "vingt-quatre heures" pour "les dernières vingt-quatre heures" ou "un mois" pour "un mois et demi"). La colonne "corrects" indique les valeurs exactement trouvées.

	Humain	Trouvés	App.	Corrects	Valeurs correctes
nombre	233	222	216	207	119
précision			0,97	0,93	0,77
rappel			0,93	0,93	0,72
f_score			0,95	0,93	0,74
	Humain	Trouvés	Apps	Corrects	Valeurs correctes
nombre	544	518	485	432	244
precision			0,93	0,83	0,56
rappel			0,89	0,79	0,45
f_score			0,91	0,81	0,5

TAB. 1 – Résultats pour les entités Timex3 sur corpus d'étude (haut) et de test (bas).

Nous voyons ici qu'avec un jeu de règles assez simples et une étude sur un corpus réduit, nous arrivons à une couverture déjà assez large en généralisation (du moins sur ce genre particulier), avec des résultats aussi bons que les travaux sur l'anglais. La plupart des expressions manquées sont des constructions non prévues, les erreurs des confusions entre âge et période (*M. Rimbaud, 17 ans vs. 100 ans de solitude*), des nombres pris pour des années, et des erreurs d'annotation manuelle. Les valeurs "erronées" sont la plupart du temps juste sous-spécifiées. Nous présentons aussi parmi les valeurs le cas des dates incomplètes (cf ci-dessus) qui nécessitent une interprétation contextuelle (le cas des autres dates étant plus facile) :

corpus	d'étude	de test	Score total
/TE	42/50	103/140	0.76
/FT	11/19	39/77	0.52
Score total	0.77	0.65	0.68

TAB. 2 – Proportion des dates incomplètes bien reconnues.

A partir de là nous avons étendu et continuerons à étendre le corpus de référence en intégrant d'autres genres : biographies (une vingtaine pour l'instant), textes historiques, juridiques. Nous avons pour l'instant 96 textes annotés pour 40.000 tokens et 1220 expressions Timex3. En comparaison, TimeBank1.2, le corpus de référence de TimeML pour l'anglais comprend 183 textes pour 60.000 tokens et 1400 expressions environ. Nous ne présentons pas les résultats (meilleurs) de la détection sur ce corpus en cours de construction car il mélange les genres de façon déséquilibrée pour l'instant. De plus les biographies dopent les scores un peu artificiellement car les dates sont beaucoup plus contraintes et plus faciles à calculer.

5 Identification des événements

Contrairement à la plupart des approches déjà existantes, notre module de reconnaissance des événements ne se limite pas à extraire le contexte local pour l'annotation des événements. L'analyse syntaxique complète de la phrase est utilisée pour l'élaboration de règles d'annotation. Nous utilisons un analyseur syntaxique robuste, soit celui qui est au coeur du Correcteur 101TM,

un correcteur gramatical très performant développé par Machina Sapiens Inc. Cet analyseur produit un arbre de dépendance qui est fourni en entrée à notre système. Dans un deuxième temps, les étiquettes TIMEX3 qui ont été créées par le module décrit précédemment sont ajoutées à l'arbre de dépendance : chaque noeud de l'arbre correspondant à un mot se trouvant dans la portée d'une balise TIMEX3 se voit associer une référence à cette annotation.

La tâche d'annotation temporelle est naturellement divisée en trois étapes qui correspondent aux différentes catégories lexicales : verbes, noms et adjectifs. Des stratégies différentes ont été établies pour chaque catégorie tout en s'appuyant sur le même principe : les *patrons-actions*. Ces patrons-actions sont les constituants de base de notre système d'annotation et sont au nombre de 80. Parmi ce nombre, 31 utilisent l'information syntaxique, soit le nom de la relation entre le noeud et son parent (e.g. *Épithète*, *Attribut*, *RelativeQui*). L'annotation des arbres syntaxiques s'effectue donc par une cascade de 6 ensembles de patrons-actions, dont 3 ensembles spécifiques pour les trois catégories lexicales d'événements.

Un patron-action (PA) est constitué avant tout d'un segment d'arbre syntaxique, potentiellement sous-défini, qui sera utilisé pour trouver les occurrences de ce patron dans le texte analysé. Dans un patron-action, les attributs d'un noeud syntaxique sont entre crochets ([' , ']); on explicite ses enfants en les mettant entre parenthèses. On mentionne la relation entre les noeuds tout juste après une paranthèse ouvrante. Par exemple, si on souhaite construire un patron qui couvre tous les noeuds de catégorie lexicale «verbe» possédant un fils avec lequel il est lié par la relation «Sujet» et de catégorie lexicale «pronom», le patron suivant est utilisé :

([catégorieLexicale = verbe] (SUJET [catégorieLexicale = pronom]))

La liste des attributs pouvant être utilisés est facilement extensible : pour notre part, nous utilisons 10 attributs tels que «lemme», «catégorieLexicale», «tempsVerbe», «estSingulier», etc. Pour chaque PA, nous cherchons les instances de ce segment d'arbre dans l'ensemble des arbres de dépendance que nous sommes en train d'annoter. Pour chaque instance ainsi trouvée, nous appliquons l'«action» associée au patron-action. Une action agit sur un noeud en particulier en modifiant ses attributs ; dans notre cas, les actions ajoutent ou modifient une étiquette de TimeML associée à un noeud. Parmi les 8 actions que nous avons définies, il y a «créerÉtiquetteEvent», «rallongerEtendueÉtiquette» et «supprimerÉtiquette». L'ordre dans lequel on recherche et applique les PA a une importance, puisque la modification d'un attribut à un noeud pourrait influencer la détection d'un PA recherché par la suite.

Pour identifier les verbes dénotant un événement, nous avons défini 29 patrons-action. La première règle à être exécutée consiste à ajouter une étiquette «EVENT» à tous les verbes présents dans nos arbres syntaxiques. Par la suite, 28 PA servent à filtrer les verbes n'exprimant pas d'événement ou d'état. Certains de ces PA font exclusivement référence à l'information lexicale des noeuds. Par exemple, un des PA mentionne qu'il faut «supprimer l'étiquette EVENT s'il est accroché à un noeud dont le lemme est *falloir*». D'autres règles utilisent l'information syntaxique, par exemple pour supprimer l'étiquette EVENT si elle est accrochée à un noeud dont le lemme est *venir* et possédant un fils «Objet Indirect» dont la forme est *de + verbe infinitif*. On peut ainsi distinguer la forme *il vient à Montréal*, qui exprime un événement, de la forme *il vient de finir son travail*, où le verbe *venir* exprime plutôt un aspect inchoatif.

Pour détecter les événements exprimés par des nominalisations, nous avons utilisé une ressource linguistique nommée VerbaCTION⁶. Ce lexique est le fruit du travail d'une équipe de l'INALF, enrichi par la suite à l'ERSS, en utilisant WebAffix (Tanguy & Hathout, 2002) ; VerbaCTION

⁶Cf. <http://w3.univ-tlse2.fr/erss/ressources/verbaaction/main.html>

visé à regrouper les couples «verbe» :«nom» où le nom peut être utilisé pour dénoter l'action ou l'activité exprimée par le verbe. Un travail manuel a été effectué pour éliminer les items indésirables (les déverbaux qui ne sont pas des événements au sens véhiculé par TimeML). En plus de l'apport de ce lexique, nous utilisons l'information syntaxique pour annoter tous les noms dépendants d'une préposition annotée comme SIGNAL (e.g. *pendant ses études, après la guerre, etc.*) et filtrer ceux qui sont compléments de nom (une bourse d'*admission*).

Pour les adjectifs dénotant des éventualités, nous notons qu'il est très difficile de trouver une règle qui augmente le rappel sans que l'ajout de cette règle ne cause une diminution importante de la précision. La raison est simple : dans notre corpus, nous avons 284 noeuds de catégorie lexicale «adjectif» et seulement 18 (6%) de ceux-ci sont annotés comme événement ou état. Ce taux passe à près de 75% pour les événements verbaux. La stratégie utilisée pour les adjectifs est donc à l'opposé de celle utilisée pour les verbes : on étiquette seulement les adjectifs pour lesquels on est à peu près certain qu'ils désignent véritablement un événement ou un état. Tel que mentionné précédemment, Evita utilise une stratégie semblable, c'est-à-dire qu'il annoté seulement les adjectifs annotés dans leur corpus de développement. Cette approche n'est certes pas durable pour une application où le domaine n'est pas restreint. C'est pourquoi nous avons tenté d'identifier des PA assurant un bon rappel et une bonne précision de manière à ne pas dépendre de l'existence d'un corpus de développement. Entres autres, on annoté seulement les adjectifs qui sont liés à un verbe copule par la relation «Attribut».

	Précision %	Rappel %	F-Measure %
<i>Corpus développement</i>	86,8	80,6	83,5
<i>Corpus évaluation</i>	62,5	77,7	69,3
Evita <i>TimeBank 1.2</i>	74.0	87.3	80.1

TAB. 3 – Évaluation comparative de notre approche

Comme pour la plupart des systèmes d'annotation, l'utilisation d'un corpus est essentielle pour l'édification et l'évaluation des performances du système. Nous aurions pu adopter le même corpus que celui utilisé pour l'évaluation de l'annotation des marqueurs temporels. Cependant, ce corpus est incomplet du point de vue des étiquettes EVENT : ce ne sont pas tous les attributs qui sont inclus et surtout, l'annotation des EVENT ne correspond pas tout à fait à la spécification 1.2.1 de TimeML. Nous avons donc repris quelques textes, et construit un autre corpus de développement à l'aide de 13 biographies issues du corpus initial et de 22 rubriques de sports issues de www.rds.ca. L'outil d'annotation graphique *Callisto*⁷ a été utilisé pour faciliter la tâche. Ces 35 fichiers contiennent 9234 mots répartis dans 400 phrases. Nous avons ainsi ajouté 840 étiquettes EVENT, 419 étiquettes TIMEX3 et 210 étiquettes SIGNAL. Bien que ce corpus apparaisse dans la présentation des résultats, il n'a été utilisé que pour développer l'ensemble de PA d'annotation ; l'évaluation est effectuée sur un autre corpus. Ce dernier a été construit de manière indépendante, à partir des nouvelles du site de l'École Polytechnique de Montréal. Il est donc normal de s'attendre à ce que l'évaluation soit influencée par la nature différente du corpus d'évaluation. Le corpus d'évaluation contient 2673 mots dans 90 phrases, 210 étiquettes EVENT, 87 étiquettes TIMEX3 et 47 étiquettes SIGNAL. Les résultats globaux sont illustrés au tableau 3.

Le tableau 4 présente les résultats par catégorie lexicale, de manière à pouvoir discuter des avan-

⁷<http://callisto.mitre.org>

Catégorie	Corpus développement			Corpus évaluation		
	Préc. %	Rappel %	F-Score %	Préc. %	Rappel %	F-Score %
Nominaux	61,5	40,0	48,4	54,7	53,7	54,2
Verbaux	94,1	97,3	95,7	65,6	90,9	76,2
Adjectivaux	66,7	77,8	71,8	NIL	NIL	NIL

TAB. 4 – Évaluation comparative de notre approche, segmenté en trois catégories

tages de notre méthode. En segmentant les résultats de cette manière, il apparaît que le problème principal de notre approche est notre faible capacité à extraire efficacement les nominalisations. C'est aussi la tâche qui semble avoir le moins bon accord inter-annotateur et dont la spécification nécessite donc un approfondissement. Le guide d'annotation TimeML pour l'anglais ne fournit pas suffisamment de détails sur l'annotation des noms.

Les résultats obtenus sur le corpus de développement au niveau des adjectivaux montrent bien l'avantage de notre approche par rapport aux méthodes syntaxiques plus locales : alors qu'il est impossible pour un système avec chunker seul de déterminer correctement si un adjectif exprime une éventualité ou non, l'analyse syntaxique nous permet de le faire avec un f-score de 71,8%. C'est principalement parce que nous connaissons la relation qui relie l'adjectif avec son père syntaxique. On remarquera que dans notre corpus d'évaluation, aucun des 161 adjectifs qu'il contient n'a été annoté, ni par l'humain, ni par notre système.

6 Conclusion

Cet article présente des résultats sur deux tâches reliées : la détection d'expressions adverbiales temporelles (et le calcul de leur référence suivant la norme TimeML) et la détection et la classification d'événements, suivant cette même norme. Nous prôtons l'utilisation de règles explicites et d'une analyse grammaticale fine des phrases, qui sont pour nous les conditions d'une généralisation efficace, dans un domaine où la pauvreté des données et la complexité de leur collecte rend délicat l'apprentissage automatique de la tâche en question. Ces informations sont de plus indispensables aux tâches qui en découlent, à savoir les liens temporels que l'on peut établir entre les entités temporelles d'un texte et plus généralement la structure temporelle globale du texte, ce qui est en fait notre objectif réel à moyen terme. Nous ne connaissons aucun travail sur le français qui aurait été évalué empiriquement comme nous l'avons fait ici, et les résultats sur d'autres langues sont comparables. Il faut également noter que certaines des informations que nous extrayons au passage dépassent un peu le cadre strict des évaluations ACE ou de la norme TimeML, puisque nous avons repris des classes d'expressions plus détaillées, et que nous extrayons des constructions syntaxiques complètes réservées pour les analyses en aval des liens entre entités temporels, notamment la structure des groupes prépositionnels ayant plusieurs composantes temporelles, là où TimeML ne préconise que l'enregistrement séparé des expressions et où l'évaluation est confiée à des fonctions externes non-spécifiées. Par ailleurs nous n'extrayons pas toutes les informations, certaines classes d'événements au sens de TimeML étant ignorées pour l'instant.

Références

- ABNEY S. (1996). *Corpus-Based Methods in Language and Speech*, chapter Part-of-Speech Tagging and Partial Parsing. Kluwer Academic Publisher.
- BALDWIN J. A. (2002). Learning temporal annotation of French news. Master's thesis, Georgetown University, Washington, DC.
- BERNARD P. & MAUREL D. (1993). Interrogation en langage naturel d'une base de données : interprétation des adverbiaux de localisation temporelle. *Mathématiques, Informatiques et Sciences humaines*, (123), 45–52.
- BETHARD S. & MARTIN J. H. (2006). Identification of event mentions and their semantic class. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP 2006)*, p. 146–154, Sydney.
- BOGURAEV B. & ANDO R. (2005). TimeML-compliant text analysis for temporal reasoning. In L. P. KAEHLING & F. GIUNCHIGLIA, Eds., *Proceedings of IJCAI05*, p. 997–1003.
- HACIOGLU K., CHEN Y. & DOUGLAS B. (2005). Automatic time expression labeling for english and chinese text. In *CICLing*, p. 548–559.
- MADHYASTHA H. V., BALAKRISHNAN N. & RAMAKRISHNAN K. (2003). Event information extraction using link grammar. In *Proceedings of the 13th International Workshop on Research Issues in Data Engineering : Multi-lingual Information Management (RIDE'03)*, p. 16–23, Los Alamitos, CA, USA.
- MANI I. & WILSON G. (2000). Robust temporal processing of news. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL 2000)*, p. 69–76, New Brunswick, New Jersey.
- MAZUR P. & DALE R. (2007). The DANTE temporal expression tagger. In *Proceedings of the 3rd Language & Technology Conference (LTC)*, Poznan, Poland.
- MOLINÈS F. (1989). *Acceptabilité et interprétation des adverbiaux de localisation temporelle : grammaire ou dictionnaire*. Mémoire de DES, Université Toulouse Le Mirail.
- MULLER P. & TANNIER X. (2004). Annotating and measuring temporal relations in texts. In *Proceedings of Coling 2004*, volume I, p. 50–56, Genève.
- NADEAU D. & SEKINE S. (2007). A survey of named entity recognition and classification. *Journal of Linguisticae Investigationes*, **30**(1), 3–26.
- NEGRI M. & MARSEGLIA L. (2004). Recognition and normalization of time expressions : Ictirst at tern 2004. In *TERN 2004 Evaluation Workshop*, Alexandria, VA : Technical Report.
- PUSTEJOVSKY J., INGRIA R., SAURI R., CASTANO J., LITTMAN J., GAIZAUSKAS R., SETZER A., KATZ G. & MANI I. (2005). The specification language TimeML. In I. MANI, J. PUSTEJOVSKY & R. GAIZAUSKAS, Eds., *The Language of Time : A Reader*. Oxford University Press.
- SAURÍ R., KNIPPEN R., VERHAGEN M. & PUSTEJOVSKY J. (2005). Evita : A robust event recognizer for QA systems. In *Proceedings of HLT/EMNLP*, p. 700–707, Vancouver.
- SCHMID H. (1994). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the International Conference on New Methods in Language Processing*.
- SLEATOR D. & TEMPERLEY D. (1991). Parsing English with a link grammar. In *Technical report CMU-CS-91-196* : Carnegie Mellon University.
- TANGUY L. & HATHOUT N. (2002). Webaffix : un outil d'acquisition morphologique dérivationnelle à partir du web. In J.-M. PIERREL, Ed., *Actes de la 9^e Conférence Annuelle sur le Traitement Automatique des Langues Naturelles (TALN-2002)*, p. 245–254, Nancy : ATALA.
- YAKUSHIJI A., TATEISI Y. & MIYAO Y. (2001). Event extraction from biomedical papers using a full parser. In *Proceedings of the Pacific Symposium on Biocomputing*, p. 408–419.