

Intégration des constructions à verbe support dans TimeML

André Bittar, Laurence Danlos

► **To cite this version:**

André Bittar, Laurence Danlos. Intégration des constructions à verbe support dans TimeML. TALN 2009, Jun 2009, Senlis, France. 2009. <inria-00522340>

HAL Id: inria-00522340

<https://hal.inria.fr/inria-00522340>

Submitted on 30 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Intégration des constructions à verbe support dans TimeML

André Bittar et Laurence Danlos

ALPAGE, INRIA Rocquencourt, 78150 Le Chesnay

Université Paris Diderot, 75013 Paris

andre.bittar@linguist.jussieu.fr, laurence.danlos@linguist.jussieu.fr

Résumé. Le langage TimeML a été conçu pour l’annotation des informations temporelles dans les textes, notamment les événements, les expressions de temps et les relations entre les deux. Des consignes d’annotation générales ont été élaborées afin de guider l’annotateur dans cette tâche, mais certains phénomènes linguistiques restent à traiter en détail. Un problème commun dans les tâches de TAL, que ce soit en traduction, en génération ou en compréhension, est celui de l’encodage des constructions à verbe support. Relativement peu d’attention a été portée, jusqu’à maintenant, sur ce problème dans le cadre du langage TimeML. Dans cet article, nous proposons des consignes d’annotation pour les constructions à verbe support.

Abstract. TimeML is a markup language developed for the annotation of temporal information in texts, in particular events, temporal expressions and the relations which hold between the two. General annotation guidelines have been developed to guide the annotator in this task, but certain linguistic phenomena have yet to be dealt with in detail. A common problem in NLP tasks, whether in translation, generation or understanding, is that of the encoding of light verb constructions. Relatively little attention has been paid to this problem, until now, in the TimeML framework. In this article, we propose annotation guidelines for light verb constructions.

Mots-clés : TimeML, verbes support, discours, sémantique.

Keywords: TimeML, light verbs, discourse, semantics.

1 Introduction

TimeML¹ est un langage d’annotation qui a pour objectif l’explicitation des informations temporelles contenues dans des textes, en se focalisant sur l’annotation de critères de surface, c’est-à-dire ce qui est lexicalisé dans le texte. Ce langage permet d’annoter les unités linguistiques réalisant des événements, les expressions temporelles, les marqueurs de relation (typiquement des prépositions comme *avant* ou *après*) ainsi que les relations aspectuelles, temporelles ou de subordination qui peuvent exister entre événements et expressions temporelles. Sont considérés basiquement comme des événements introduisant une balise <EVENT> des verbes "ordinaires" comme *écrire* ou *commettre* dans (1).

- (1)a. Luc a écrit ce livre.
- b. Luc a commis ce film.

Sont aussi considérés basiquement comme des événements introduisant une balise <EVENT> des noms comme *destruction* ou *agression* dans (2). Pour de telles phrases, TimeML codera à raison deux événements : l’événement décrit par le verbe "ordinaire" (*décrire* ou *filmer*) et l’événement décrit par l’objet de ces verbes (*destruction* ou *agression*).

- (2)a. Luc a décrit la destruction de Rome (par Brutus).
- b. Max a filmé l’agression de Luc contre Marie.

Mais qu’en est-il pour les constructions à “verbe support” illustrées dans les exemples (3) ?

- (3)a. Luc a commis une agression contre Marie.
- b. Marie a subi une agression de la part de Luc.

Pour (3a), un codage où le verbe *commettre* introduit une balise <EVENT> comme en (1b) et où le nom *agression* introduit indépendamment une autre balise <EVENT> comme en (2b) n’est pas, en soi, totalement correct : un seul événement est en jeu dans une construction à verbe support comme (3a), qui peut être paraphrasée par une phrase à verbe "ordinaire" *Luc a agressé Marie* qui ne met en jeu qu’un seul événement. De même pour (3b) qui peut être paraphrasé par *Marie a été agressée par Luc*.

Cet article se propose donc d’examiner le codage des constructions à verbe support en français dans TimeML, ce qui n’a pas été fait de façon détaillée pour les autres langues couvertes par TimeML (l’anglais, l’italien, le coréen et le chinois) à notre connaissance. La Section 2 présentera une introduction à TimeML, ses enjeux et sa méthodologie. La Section 3 présentera les principales données linguistiques concernant les constructions à verbe support. La Section 4 proposera un codage linguistiquement justifié des constructions à verbe support dans TimeML. Enfin, la Section 5 portera sur des questions de faisabilité et de méthodologie.

2 Présentation de TimeML

TimeML (Pustejovsky *et al.*, 2003) est un langage de spécification conçu pour l’annotation et la normalisation d’informations temporelles dans des textes. Il a principalement été élaboré

¹La version de TimeML que l’on prend en compte ici est celle élaborée dans les consignes de l’ISO (Lee *et al.*, 2007), effectivement ISO-TimeML.

dans le cadre de l'amélioration de systèmes questions-réponses en leur permettant de traiter des questions d'ordre temporel sur les entités et les événements. Le schéma d'annotation prévoit les fonctionnalités suivantes : l'annotation des événements, l'étiquetage des expressions temporelles et la normalisation de leur valeur², ainsi que la mise en évidence des relations aspectuelles, temporelles ou de subordination qui peuvent exister entre ces deux types d'entités temporelles. Les marqueurs de relations peuvent également être annotés.

TimeML adopte une conception large des événements qui regroupe les événements et certains états, ce qui correspond plus à la notion d'éventualité (Bach, 1986). La balise <EVENT> permet de classer un événement annoté selon une ontologie prédéfinie. L'ontologie TimeML compte 7 classes d'événements différents : ASPECTUAL (*commencer, continuer, terminer...*), I_ACTION (*nier, permettre, empêcher...*), I_STATE (*vouloir, croire, aimer...*), PERCEPTION (*voir, entendre, percevoir...*), REPORTING (*dire, citer, affirmer...*), STATE (*être + ADJ*), et OCCURRENCE (les événements qui ne rentrent pas dans l'une des autres classes). Nous orientons le lecteur vers (Saurí *et al.*, 2005) pour une description de ces classes. La balise <EVENT> comporte des attributs permettant de représenter les informations grammaticales comme la partie du discours, d'éventuels temps et modes verbaux, la présence d'une négation, ainsi que les propriétés aspectuelles d'un événement donné. Les consignes d'annotation précisent que l'on marque de cette balise la tête lexicale du constituant qui dénote l'événement.

Les expressions temporelles sont marquées par la balise <TIME3>. Elles se divisent en quatre classes : les dates (type DATE, *le 15 janvier, 15.01.2008*), les heures (type TIME, *15h20, l'après-midi*), les durées (type DURATION, *5 jours, deux ans*) et les ensembles (type SET, *tous les jours, hebdomadaire, annuellement*). TimeML permet aussi de raisonner avec des expressions temporelles sous-spécifiées, comme *lundi prochain* et *l'année dernière*, dont la valeur doit être déterminée par rapport à un point temporel de référence.

Les événements et les expressions temporelles sont mis en relation par trois sortes de liens (links) : liens temporels (TLINK), aspectuels (ALINK) et de subordination (SLINK). La première sorte capture des relations temporelles entre deux entités (EVENT-EVENT, TIME3-TIME3, ou EVENT-TIME3), la deuxième capture les phases dans le déroulement d'un événement, enfin la troisième est essentielle pour les raisonnements qui dépendent de la véracité ou la certitude des propositions qui dénotent des événements. Les mots fonctionnels qui signalent explicitement un de ces liens sont annotés avec la balise SIGNAL. Le plus souvent ce sont des prépositions comme *avant, après, pendant* ou *lors de*.

3 Présentation des constructions à verbe support

La notion de verbe support est traditionnellement utilisée (Gross, 1981) pour un verbe (noté V_{sup}) qui a comme objet (direct ou indirect) un nom prédicatif (N_{pred}) dénotant une éventualité, comme sujet un participant à cette éventualité – en gros, le participant qui est le sujet du verbe morphologiquement associé au N_{pred} s'il existe – et comme objet oblique éventuel l'autre participant s'il y en a un, voir (4a). Cette notion a été élargie pour couvrir les verbes support converses (V_{conv}) où, à l'instar des constructions passives, les rôles syntaxiques des participants sont inversés, voir (4b) (Gross, 1982). Enfin, elle peut être élargie aux verbes supports inverses (V_{inv}) où le N_{pred} se trouve en position sujet, voir (4c) (Danlos, 2009).

²La valeur étant la représentation codifiée de l'intervalle ou point temporel précis désigné par l'expression.

- (4)a. Ce politicien a mené une attaque contre le libéralisme [V_{sup}]
 (Ce politicien a attaqué le libéralisme)
- b. Le libéralisme a fait l'objet d'une attaque de la part de ce politicien [V_{conv}]
 (Le libéralisme a été attaqué par ce politicien)
- c. Un cri a échappé à Luc [V_{inv}]
 (Luc a crié)

Dans les exemples (4), un seul événement est en jeu, comme en témoigne la (pseudo-)paraphrase avec une phrase à verbe ordinaire donnée entre parenthèses. Nous verrons à la Section 4.1 comment prendre en compte ce phénomène en TimeML.

Les verbes support de (4) ont un aspect neutre dans la mesure où aucune phase de l'éventualité dénotée par le N_{pred} n'est mise en évidence. Ceci n'est pas le cas dans l'exemple (5) où *lancer*, verbe support à aspect non neutre, indique qu'on s'intéresse au début de l'attaque.

- (5) Ce politicien a lancé une attaque contre le libéralisme [V_{asp}]

Dans une construction à verbe support, le N_{pred} est la tête d'un GN et en tant que tel il peut être quantifié, (6a), la quantification portant le nombre d'occurrences de l'éventualité concernée, ou modifié, par exemple, par des adjectifs temporels, (6b), ou de fréquence, (6c).

- (6)a. Ce politicien a lancé (quatre + plusieurs) attaques contre le libéralisme.
 b. Luc a fait un voyage nocturne.
 c. Luc faisait une visite hebdomadaire à sa mère.

4 Intégration des verbes support dans TimeML

Nous présentons maintenant notre solution pour le traitement des constructions à verbe support dans le cadre de l'annotation TimeML.

4.1 Phénomènes de base

Avant de présenter notre solution, nous devons expliquer comment TimeML permet de capturer les relations aspectuelles. D'abord celles qui existent entre un verbe matrice aspectuel et son complément événementiel sous forme de verbe infinitif, voir (7a), ou de groupe nominal, voir (7b). Il s'agit ici de verbes tels que *commencer*, *entamer*, *continuer*, *poursuivre*, *terminer* ou *arrêter* qui indiquent la phase initiale, médiane ou terminale dans le déroulement d'un événement introduit comme complément. Pour l'exemple (7a), l'annotation est donnée ci-dessous.

- (7)a. Luc a commencé à préparer une choucroute.
 b. Luc a commencé la préparation d'une choucroute.

```
Luc a <EVENT eid="e1" eiid="ei1" class="ASPECTUAL" pos="VERB"
tense="PAST">commencé</EVENT> à <EVENT eid="e2" eiid="ei2"
class="OCCURRENCE" pos="VERB" vform="INFINITIVE">préparer</EVENT> une
choucroute.
<ALINK lid="l1" eventInstanceID="ei1" relatedToEventInstance="ei2"
relType="INITIATES"/>
```

Intégration des constructions à verbe support dans TimeML

Le verbe *commencer* est balisé `<EVENT>`. Il reçoit la classe `ASPECTUAL` et les valeurs `VERB` et `PAST` pour ses traits grammaticaux de catégorie (`pos` pour *part-of-speech*) et de temps verbal (attribut `tense`) respectivement. Le verbe *préparer* est lui aussi marqué par la balise `<EVENT>`. Il est de classe `OCCURRENCE`, et sa balise contient un attribut `vform` qui capture le fait qu'il s'agit d'un infinitif. Les deux balises `<EVENT>` contiennent des attributs `eid` et `eiid`, qui servent à identifier les événements et leurs instances. La dernière balise est un `<ALINK>` — identifié par l'attribut `lid` — qui désigne une relation aspectuelle. Ses attributs `eventInstanceID` et `relatedToEventInstance` sélectionnent les instances des événements concernés : `ei1` pour le verbe *commencer* et `ei2` pour le verbe *préparer*. L'attribut `relType` précise le type de relation aspectuelle en jeu, ici `INITIATES`, qui indique que `ei1` se situe au début de `ei2`.

Pour (7b), la seconde balise `<EVENT>` est placée autour du nom et seule la valeur de l'attribut `pos` change (cette balise ne contient pas l'attribut `vform`).

```
Luc a <EVENT eid="e1" eiid="ei1" class="ASPECTUAL" pos="VERB"
tense="PAST">commencé</EVENT> la <EVENT eid="e2" eiid="ei2"
class="OCCURRENCE" pos="NOUN">préparation</EVENT> d'une choucroute.
<ALINK lid="l1" eventInstanceID="ei1" relatedToEventInstance="ei2"
relType="INITIATES"/>
```

Considérons maintenant les phrases à verbe support données en (8), (8a) étant construite autour de *lancer* à valeur aspectuelle inchoative (de classe TimeML `ASPECTUAL`), (8b) autour de *mener* à valeur aspectuelle neutre (de classe TimeML `OCCURRENCE`).

- (8)a. Une mère de famille a lancé une procédure contre une institutrice à Bourges.
- b. Une mère de famille a mené une procédure contre une institutrice à Bourges.

Étant donné leur similitude formelle et sémantique, (7b) et (8a) doivent recevoir des annotations similaires. En conséquence, l'annotation de (8a) doit comprendre deux balises `<EVENT>` qui sont reliées par l'intermédiaire d'un `<ALINK>` indiquant une relation aspectuelle de type `INITIATES` entre le verbe support *lancer*, de classe `ASPECTUAL`, et son complément nominal.

Maintenant, les annotations de (8a) et (8b) doivent être identiques sauf pour la relation aspectuo-temporelle entre le verbe support et son complément nominal et la classe du verbe. Pour un verbe support à valeur aspectuelle neutre, c'est-à-dire de classe `OCCURRENCE`, nous proposons que cette relation aspectuelle soit de type `IDENTITY`, comme cela a été fait dans (Pustejovsky *et al.*, 2005) pour l'anglais et dans (Caselli, 2008) pour l'italien — cependant, dans les deux cas sans argumentation poussée. Nous obtenons donc l'annotation suivante pour (8b).

```
Une mère de famille a <EVENT eid="e1" eiid="ei1" class="OCCURRENCE"
pos="VERB" tense="PAST">mené</EVENT> une <EVENT eid="e2" eiid="ei2"
pos="NOUN">procédure</EVENT> contre une institutrice à Bourges.
<TLINK lid="l1" eventInstanceID="ei1" relatedToEventInstance="ei2"
relType="IDENTITY"/>
```

En conclusion, nous proposons que les phrases à verbe support reçoivent des annotations comportant deux balises `<EVENT>`, l'une pour le V_{sup} , l'autre pour le N_{pred} . Pour un verbe support à valeur aspectuelle non neutre, ces deux balises sont reliées par un `<ALINK>` précisant le type de relation aspectuelle entre le verbe support et son complément nominal. Pour un verbe support à valeur aspectuelle neutre, ces deux balises sont reliées par un `<TLINK>` de type `IDENTITY`. Nous allons maintenant examiner d'autres phénomènes concernant les phrases à verbe support.

4.2 Quantification d'occurrences

Tout groupe nominal événementiel ordinaire porte une quantification et c'est bien le cas du GN construit autour du N_{pred} dans les constructions à verbe support, cf Section 3. Dans les versions antérieures de TimeML (Saurí *et al.*, 2006), l'attribut `cardinality` servait à capturer la valeur des quantifications adverbiales, comme dans (9a). Ces consignes ne prenaient pas en compte la quantification des occurrences d'événements nominaux (9b).

- (9)a. Max a voyagé en Tanzanie deux fois.
- b. Deux violents séismes ont fait des centaines de morts au Japon.

D'un point de vue temporel, il peut être pertinent d'annoter la quantification d'événements nominaux. C'est le cas dans les exemples de (10), qui précisent qu'un certain nombre d'accidents ont eu lieu sur une période donnée. Il est également important de capturer la quantification d'occurrences dans les constructions à verbe support, voir (12).

- (10)a. 934 accidents ont endeuillé le pays au mois de mars.
- b. Plusieurs accidents ont endeuillé le pays au mois de mars.
- c. Aucun accident n'a endeuillé le pays au mois de mars.

L'emploi de l'attribut `cardinality` n'est pas très détaillé dans les consignes d'annotation TimeML actuelles. Nous proposons d'employer cet attribut pour capturer la quantification d'occurrences (différente d'une valeur de 1) sur tout nom événementiel marquant par une balise `<EVENT>`. Les événements de cardinalité 1 n'auront pas cet attribut. Nous proposons que le domaine des valeurs soit des chiffres pour des expressions numériques précises (*934*, *aucun*) comme dans (10a) et (10c) et des expressions à opérateur unaire pour les valeurs vagues, comme *>1* pour *plusieurs* dans (10b).

L'annotation du nom événementiel dans (10a) serait alors :

```
934 <EVENT eid="e1" eiid="ei1" class="OCCURRENCE" pos="NOUN"
cardinality="934">accidents</EVENT> ont endeuillé le pays au mois de mars.
```

Prenons maintenant les deux phrases à verbe ordinaire (11a) et (11b). (11a) comporte une quantification adverbiale d'occurrence et (11b) comporte une quantification adverbiale de fréquence d'occurrence.

- (11)a. Max a voyagé en Tanzanie deux fois.
- b. Max a voyagé en Tanzanie deux fois par an.

Dans (11a), la quantification adverbiale *deux fois* porte sur l'événement décrit par le verbe *voyager*. La version actuelle de TimeML traite ce genre de phrase de la façon suivante. La quantification adverbiale est capturée dans une balise `<TIMEX3>` exprimant la fréquence (`freq="2X"` pour *deux fois*). Un `<TLINK>` explicite la relation entre l'événement et l'expression temporelle.

```
Max a <EVENT eid="e1" eiid="ei1" class="OCCURRENCE" pos="VERB"
tense="PAST">voyagé</EVENT> en Tanzanie <TIMEX3 tid="t1" type="SET"
freq="2X">deux fois</TIMEX3>.
<TLINK lid="l1" eventInstanceID="ei1" relatedToTime="t1"
relType="IS_INCLUDED"/>
```

Selon TimeML l'annotation de (11b) sera la suivante. La quantification adverbiale de fréquence *deux fois par an* est capturée dans une seule balise `<TIMEX3>` de type `SET` qui désigne un ensemble de valeurs. L'intervalle temporel est capturé dans l'attribut `value` de valeur `P1Y` (period of one year). L'attribut `freq` capture la fréquence.

Intégration des constructions à verbe support dans TimeML

```
Max a <EVENT eid="e1" eiid="ei1" class="OCCURRENCE" pos="VERB"
tense="PAST">voyagé</EVENT> en Tanzanie <TIMEX3 tid="t1" type="SET"
value="P1Y" freq="2X">deux fois par an</TIMEX3>. <TLINK lid="l1"
eventInstanceID="ei1" relatedToTime="t1" relType="IS_INCLUDED"/>
```

Venons-en maintenant au cas des constructions à verbe support (12a-d).

- (12)a. Max a fait un voyage en Tanzanie.
- b. Max a fait deux voyages en Tanzanie.
- c. Max a fait un voyage en Tanzanie deux fois par an.
- d. Max a fait deux voyages en Tanzanie par an.

(12a) est annoté selon les consignes que nous avons présentées dans la Section 4.1 pour les verbes support à valeur aspectuelle neutre. Pour (12b), qui comporte une quantification sur le N_{pred} , la balise <EVENT> du N_{pred} a l'attribut *cardinality* de valeur 2.

```
Max a <EVENT eid="e1" eiid="ei1" class="OCCURRENCE" pos="VERB"
tense="PAST">fait</EVENT> deux <EVENT eid="e2" eiid="ei2" pos="NOUN"
cardinality="2">voyages</EVENT> en Tanzanie.
<TLINK lid="l1" eventInstanceID="ei1" relatedToEventInstance="ei2"
relType="IDENTITY"/>
```

L'annotation de (12c), avec une quantification adverbiale de fréquence d'occurrence, peut être obtenue de façon "compositionnelle" à partir de (11b) et (12a). L'annotation est analogue à l'annotation pour (11b) sauf qu'il s'agit d'une construction à verbe support. Le N_{pred} est alors annoté par une balise <EVENT> et un <TLINK> explicite la relation entre le V_{sup} et le N_{pred} .

```
Max a <EVENT eid="e1" eiid="ei1" class="OCCURRENCE" pos="VERB"
tense="PAST">fait</EVENT> un <EVENT eid="e2" eiid="ei2"
pos="NOUN">voyage</EVENT> en Tanzanie <TIMEX3 tid="t1" type="SET"
value="P1Y" freq="2">deux fois par an</TIMEX3>.
<TLINK lid="l1" eventInstanceID="ei1" relatedToEventInstance="ei2"
relType="IDENTITY"/>
<TLINK lid="l2" eventInstanceID="ei1" relatedToTime="t1"
relType="IS_INCLUDED"/>
```

Enfin passons à l'annotation de (12d), où il y a une quantification sur le N_{pred} ainsi qu'une expression de fréquence *par an* rattachée au groupe nominal. Le verbe support est marqué par une balise <EVENT>. Le N_{pred} comporte également une balise <EVENT>, avec un attribut *cardinality* pour la quantification d'occurrences. Maintenant une difficulté se présente lorsque l'on doit annoter l'expression de fréquence. Les consignes TimeML (Lee *et al.*, 2007) précisent que l'étendue de la balise <TIMEX3> doit correspondre soit à un groupe nominal (ex. *deux fois par an*), soit un groupe adjectival (ex. *annuel*), soit un groupe adverbial (ex. *annuellement*), ce qui n'est pas le cas de l'expression de fréquence ici en forme d'un groupe prépositionnel. De plus, comme nous montre l'agrammaticalité dans (13a), l'expression de fréquence seule ne peut apparaître dans les constructions à verbe ordinaire à la place d'un adverbe comme *annuellement*. On ne saurait non plus l'annoter au même titre qu'un adjectif, comme *annuel* (13b) et (13c), qui représente à la fois la quantification et la fréquence d'occurrence (par synonymie avec *une fois par an*). Effectivement, le groupe prépositionnel qui exprime la fréquence ne peut être séparé du groupe nominal quantifié qu'il modifie.

- (13)a. Max a voyagé annuellement / *par an.
- b. Le voyage annuel / *par an de Max.
- c. Le voyage de Max était annuel / *par an.

Il est alors difficile de savoir comment arriver à une annotation adéquate dans le cadre du langage TimeML actuel. La solution qui nous semble la mieux adaptée à cette situation est celle d'étendre le langage TimeML pour capturer non seulement la quantification mais aussi la fréquence d'occurrence des événements nominaux dans la balise `<EVENT>`. Nous proposons l'attribut `periodicity` (pour ne pas reprendre `freq` qui sert dans la balise `<TIMEX3>`), qui prend la valeur normalisée de l'intervalle temporel en question. Ainsi, pour la phrase (12d) nous proposons une annotation similaire à celle pour (12b) mais cette fois avec l'attribut `periodicity` de valeur `P1Y` (period - one year).

```
Max a <EVENT eid="e1" eiid="ei1" class="OCCURRENCE" pos="VERB"
tense="PAST">fait</EVENT> deux <EVENT eid="e2" eiid="ei2" pos="NOUN"
cardinality="2" periodicity="P1Y">voyages</EVENT> en Tanzanie par an.
<TLINK lid="l1" eventInstanceID="ei1" relatedToEventInstance="ei2"
relType="IDENTITY"/>
```

L'introduction de l'attribut `periodicity` permet une annotation linguistiquement cohérente des phrases à verbe support avec quantification et fréquence d'occurrences sur le N_{pred} .

5 Faisabilité et méthodologie

Afin de pouvoir effectuer une annotation automatique des phénomènes présentés ici, un système doit disposer de certaines ressources lexicales, notamment pour déterminer quelle balise (`<ALINK>` ou `<TLINK>`) il faut utiliser pour expliciter la relation entre un V_{sup} et son N_{pred} , voir Section 4.1. Un lexique spécifiant les V_{sup} possibles pour un N_{pred} donné, ainsi que la variante aspectuelle (inchoatif ou neutre) de la construction, semble indispensable. Des travaux sur la reconnaissance des constructions à verbe support sont décrits dans (Laporte *et al.*, 2008), mais à notre connaissance les lexiques résultants ne contiennent pas d'information sur la valeur aspectuelle du verbe dans ces constructions.

Une possibilité pour palier ce manque de ressources lexicales est de recourir à des approximations, par exemple en choisissant la relation uniquement en fonction du verbe. Ainsi, les verbes ayant habituellement une valeur aspectuelle non neutre (*commencer, lancer, terminer, ...*) lorsqu'ils ont un objet direct événementiel, imposeraient l'utilisation de la balise `<ALINK>`. Les autres verbes imposeraient l'utilisation d'une balise `<TLINK>` pour expliciter la relation avec leur argument événementiel. La faiblesse de cette approximation réside dans le fait que les verbes de cette deuxième classe peuvent aussi avoir une valeur aspectuelle lorsqu'ils se combinent avec certains noms événementiels. Nous citons le cas de *prendre*, qui a une valeur aspectuelle inchoative dans (14a) et une valeur neutre dans (14b).

- (14)a. L'économie a pris son essor.
- b. Luc a pris une revanche sur Marie.

Malgré cette faiblesse, l'implémentation de ces simples heuristiques permettrait de constituer automatiquement un lexique de paires $V_{sup}—N_{pred}$ corrélées avec leur classe aspectuelle en effectuant une correction manuelle pour rendre compte des cas comme 14. Nous entreprenons actuellement une implémentation de cette approximation dans le cadre des modules que nous développons pour l'annotation TimeML automatique. Nous orientons le lecteur vers (Bittar, 2008) pour les détails sur ces modules. Nous nous servirons des résultats pour évaluer à la fois la fréquence du phénomène linguistique en question et la couverture des heuristiques proposées pour son traitement.

6 Conclusion

Une extension des consignes d'annotation du langage TimeML s'est avérée nécessaire pour l'annotation correcte des constructions à verbe support. Premièrement, pour rendre compte des variantes aspectuelles des V_{sup} et, deuxièmement, pour capturer la quantification et la fréquence d'occurrences des N_{pred} . De fait, nous avons élaboré des consignes linguistiquement motivées pour la bonne annotation des phénomènes en question. Cependant, pour arriver à une implémentation exacte de ces consignes, le développement de ressources lexicales demeure une nécessité. Dans un premier temps, une implémentation approximative est envisagée. Une première évaluation de ce système permettra de jauger le potentiel apport des ressources lexicales proposées.

Références

- BACH E. (1986). The Algebra of Events. *Linguistics and Philosophy*, **9**(1), 5–16.
- BITTAR A. (2008). Annotation des informations temporelles dans des textes en français. In *Actes de RECITAL 2008*, Avignon, France.
- CASELLI T. (2008). *It-TimeML : TimeML Annotation Guidelines for Italian, Version 1.0*. Rapport interne, Istituto di Linguistica Computazionale, C.N.R.
- DANLOS L. (2009). Extension de la notion de verbe support. In *Supports et prédicats non verbaux dans les langues du monde*. Paris.
- GROSS G. (1982). Un cas de constructions inverses : *donner* et *recevoir*. *Langages*, **136**, 41–52.
- GROSS M. (1981). Les bases empiriques de la notion de prédicat sémantique. *Langages*, **63**, 7–52.
- LAPORTE E., RANCHHOD E. & YANNAKOPOULOU A. (2008). Syntactic variation of support verb constructions. *Linguisticae Investigationes*, **31**(2).
- LEE K., PUSTEJOVSKY J., BUNT H., BOGURAEV B. & IDE N. (2007). *Language resource management - Semantic annotation framework (SemAF) - Part 1 : Time and events*. International Organization for Standardization, Geneva, Switzerland. http://lirics.loria.fr/doc_pub/SemAFCD24617-1Rev12.pdf.
- PUSTEJOVSKY J., MEYERS A., PALMER M. & POESIO M. (2005). Merging PropBank, NomBank, Timebank, Penn Discourse Treebank and Coreference. In *ACL 2005 Workshop on Frontiers in Corpus Annotations II : Pie in the Sky*, Ann Arbor, Michigan, USA.
- PUSTEJOVSKY J., NO J. C., INGRIA R., SAURÍ R., GAIZAUSKAS R., SETZER A. & KATZ G. (2003). TimeML : Robust Specification of Event and Temporal Expressions in Text. In *Proceedings of IWCS-5, Fifth International Workshop on Computational Semantics*.
- SAURÍ R., KNIPPEN R., VERHAGEN M. & PUSTEJOVSKY J. (2005). Evita : A Robust Event Recognizer for QA Systems. In *Proceedings of HLT/EMNLP 2005*, p. 700–707.
- SAURÍ R., LITTMAN J., KNIPPEN B., GAIZAUSKAS R., SETZER A. & PUSTEJOVSKY J. (2006). *TimeML Annotation Guidelines Version 1.2.1*.