



Annotation de la temporalité en corpus : contribution à l'amélioration de la norme TimeML

Anais Lefeuvre, Jean-Yves Antoine, Agata Savary, Emmanuel Schang, Lotfi Abouda, Denis Maurel, Iris Eshkol

► To cite this version:

Anais Lefeuvre, Jean-Yves Antoine, Agata Savary, Emmanuel Schang, Lotfi Abouda, et al.. Annotation de la temporalité en corpus : contribution à l'amélioration de la norme TimeML. ATALA. TALN'2014, Jul 2014, Marseille, France. Conférence TALN'2014, 2, pp.F14-2029, 2014, <<http://www.taln2014.org/site/>>. <hal-01075207>

HAL Id: hal-01075207

<https://hal.archives-ouvertes.fr/hal-01075207>

Submitted on 16 Oct 2014

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.

Annotation de la temporalité en corpus : contribution à l'amélioration de la norme TimeML

Anaïs Lefeuvre¹, Jean-Yves Antoine¹, Agata Savary¹, Emmanuel Schang², Lotfi Abouda², Denis Maurel¹,
Iris Eshkol²

(1) Université François Rabelais de Tours, laboratoire LI

(2) Université d'Orléans, laboratoire LLL, UMR 7270

Résumé. Cet article propose une analyse critique de la norme TimeML à la lumière de l'expérience d'annotation temporelle d'un corpus de français parlé. Il montre que certaines adaptations de la norme seraient conseillées pour répondre aux besoins du TAL et des sciences du langage. Sont étudiées ici les questions de séparation des niveaux d'annotation, de délimitation des éventualités dans le texte et de l'ajout d'une relation temporelle de type associative.

Abstract. This paper reports a critical analysis of the TimeML standard, in the light of a temporal annotation that was conducted on spoken French. It shows that the norm suffers from weaknesses that must be corrected to fit the needs of NLP and corpus linguistics. These limitations concern mainly 1) the separation of different levels of linguistic annotation, 2) the delimitation in the text of the events, and 3) the absence of a bridging temporal relation in the norm.

Mots-clés : annotation temporelle, TimeML, éventualités, relations temporelles, expressions temporelles

Keywords: temporal annotation, TimeML, eventualities, temporal relations, time expressions

1 Introduction : normalisation des ressources linguistique en TAL

La lecture des actes de la première édition de la conférence TALN, organisée il y a 20 ans à Marseille, nous renseigne sur l'évolution de la discipline, marquée par le développement des approches par apprentissage : seules 13% des communications de l'époque évoquaient des approches centrées sur les données. Cette évolution fut celle du passage de la linguistique computationnelle à l'ingénierie des langues. D'un point de vue méthodologique, elle s'est traduite par le recours de plus en plus systématique aux corpus. Avec l'émergence d'Internet et du *Big Data*, une masse sans cesse accrue d'information textuelle est désormais disponible. La question qui se pose n'est plus l'accès au matériau linguistique mais à son annotation, description enrichie essentielle à de nombreuses applications. L'annotation manuelle de corpus étant coûteuse, la création de telles ressources reste limitée. Pour palier cette difficulté, certains s'en remettent à une annotation automatique dont les erreurs sont compensées par la masse de données enrichies (Clark et Lappin 2010). La diffusion d'annotations manuelles d'envergure et de qualité reste néanmoins une question cruciale. Afin de favoriser la réutilisabilité des ressources, il est nécessaire de définir des normes d'annotation acceptées de tous. Cet effort de normalisation reste toutefois insuffisant. La reconnaissance des entités nommées est symptomatique de cette situation. La notion d'entité nommée n'a cessé d'évoluer au fil des campagnes d'évaluation sans arriver à un réel standard consensuel. D'où l'existence de corpus reposant sur des typologies peu compatibles.

L'annotation en temporalité semble avoir évité cet écueil. Elle a fait l'objet de la publication d'une norme, TimeML proposée par le comité TC37 de l'ISO (2008). Elle a été adoptée par de multiples corpus, avec parfois des adaptations comme ce fut le cas pour le French TimeBank (Bittar and al., 2011). Cette norme présente toutefois à nos yeux comme à d'autres (Battistelli, 2009) des insuffisances qu'il convient de circonscrire. Après un état de l'art sur l'annotation en temporalité, cet article présente une analyse critique de certaines faiblesses de la norme ISO TimeML et tente d'y apporter des réponses. Cette réflexion est portée par un consortium dont l'ambition est de réaliser un corpus de grande envergure annoté en temporalité, comme il l'a déjà fait pour l'annotation en coréférence avec le corpus ANCOR_Centre (Muzerelle et al. 2013, Muzerelle et al. 2014). Ce texte ne constitue pas la proposition d'une nouvelle norme, mais le fruit de réflexions issues d'ateliers d'annotation que nous portons auprès de la communauté TALN pour alimenter le début d'une nécessaire évolution de la norme TimeML.

2 Annotation en temporalité

L'annotation de la temporalité linguistique demande dans un premier lieu de définir ce qu'est un évènement, élément constitutif de la temporalité du discours. On retiendra deux définitions pertinentes, la première de Mani et al. (2005) :

“We consider “events” a cover term for situations that happen or occur. Events can be punctual or last for a period of time. We also consider as events those predicates describing states or circumstances in which something obtains or holds true.” (Mani et al., 2005)

“Nous utiliserons (...) le terme d'éventualité (...) comme terme générique pour désigner quelque chose qui se produit, qui a lieu ou qui est vrai sur une période donnée de temps.” (De Saussure, 1997)

Ces deux définitions rendent compte de la diversité des évènements évoqués dans le discours. Diversité que nous chercherons à représenter par le terme générique d'éventualité (Bach, 1981), et non d'évènement. Pour analyser temporellement un énoncé, il est également nécessaire de détecter les relations entre ces entités du discours. (Allen, 1983) en propose un modèle standard reposant sur 13 relations minimales entre intervalles.

(1) *Tu restes la semaine à Paris, j'y serai à partir de mercredi.*

(2) *Je reste la semaine à Paris d'où je pars pour Londres.*

L'exemple (1) illustre la relation d'*overlap*, qui suppose l'existence d'une intersection entre les intervalles mobilisés par “*tu restes la semaine à Paris*” et “*j'y serai à partir de mercredi*”. L'exemple (2) illustre la relation *meet*. Celle-ci décrit le fait qu'une éventualité, (“*je pars pour Londres*”) succède immédiatement à une autre (“*je reste la semaine à Paris*”).

L'annotation de la temporalité a été initialement motivée par des visées applicatives, notamment en extraction d'information. Des campagnes d'évaluation ont défini des tâches bien délimitées, de complexité croissante, chacune étant associée à un schéma d'annotation spécifique. Ainsi, MUC-7 et CoNLL 2002-2003 ont proposé les tâches de repérage d'entités nommées incluant des expressions temporelles. ACE 2005-2007 a complété cette tâche par une phase de normalisation d'expressions temporelles et a introduit la tâche de repérage et de caractérisation d'évènements. TempEval à son tour a proposé une tâche d'identification automatique des relations temporelles à partir des expressions temporelles et évènements déjà identifiés dans le texte. Un des résultats de ces initiatives est la construction progressive d'un standard d'annotation qui a donné lieu à la norme TimeML, dont nous donnons un aperçu à la section suivante.

Un des challenges récents dans la reconnaissance d'entités nommées, y compris temporelles, est celui du repérage des entités imbriquées (Savary et al. 2010, Gravier et al., 2012), comme par exemple dans l'entité nommée [*conseil général [du Morbihan]*]. L'imbrication permet de mettre directement en évidence certaines relations inter-entités, mais complexifie la tâche des systèmes de reconnaissance automatique. Certains travaux tentent de contourner la difficulté du repérage des frontières de ces entités/mentions en identifiant seulement leurs têtes lexicales. Comme nous l'expliquerons dans la section 3.2, nous nous plaçons du côté des approches qui délimitent la portée textuelle complète des entités/mentions (Muzerelle et al. 2013, Ogrodniczuk et al. 2013).

Notons enfin, que l'intérêt de la prise en compte de l'information temporelle est partagée par d'autres domaines. Ainsi, les spécialistes de représentation de connaissances, et notamment des Linked Open Data, ont identifié le besoin de raisonnement automatique de nature spatio-temporelle, et ont pour cela proposé des extensions aux ontologies existantes, telles que YAGO2 (Hoffart et al., 2011), et aux formalismes associés (RDF, SPARQL, etc.), en y incluant notamment des dimensions temporelles (par addition d'intervalles temporels pour caractériser l'existence des entités).

3 TimeML

Dans le domaine du TALN, la norme ISO TimeML s'est imposée comme un standard de facto pour l'annotation de la temporalité. Des corpus annotés suivant le format ISO TimeML existent ainsi pour l'anglais, le français, l'italien, le portugais, le coréen, le roumain et le chinois. TimeML distingue tout d'abord trois types d'unités:

- EVENT, reprenant globalement la notion d'éventualité typée (*state, occurrence, etc.*) et attachée à la tête du syntagme (verbal, nominal, adjectival ou prépositionnel) correspondant. Un EVENT est assorti d'un ou plusieurs MAKEINSTANCE qui permettent d'annoter les différentes instances du même EVENT en donnant plusieurs traits à ceux-ci (part-of-speech, tense, aspect, etc.). Par exemple, l'énoncé suivant sera annoté avec un EVENT attaché à “*enseigne*” et deux objets MAKEINSTANCE, un pour chacune des instances de cette éventualité.

(3) *Jean a enseigné deux fois lundi*

- TIMEX : les expressions temporelles expriment une date calendaire, un horaire, une durée, un ensemble d’horaires (*date, time, duration, set*). Pour l’exemple (3), “*lundi*” sera annoté TIMEX.
- SIGNAL : termes linguistiques qui marquent l’existence d’une relation temporelle entre deux éventualités.

Enfin, 3 relations sont proposées :

- TLINK : relation entre deux EVENTS, un EVENT et un TIMEX, ou encore entre deux TIMEXs, suivant la typologie de Allen. TimeML ajoute une relation *identity* pour la coréférence temporelle.
- ALINK : relation aspectuelle entre EVENTS, permet d’annoter l’opération d’un EVENT sur un second telle que la sélection de sa phase initiale, finale, son apogée etc. Dans l’exemple suivant, “*démarré*” et “*cours*” sont deux EVENTS reliés par une relation ALINK dont le type est *initiates* :

(4) *Jean a démarré son cours à 14h.*

- SLINK : relation de subordination entre deux EVENTS, marquant l’influence du premier sur la factivité du second. Dans l’exemple suivant “*oublié*” et “*était à Paris*” sont liés par un SLINK de type *factive*.

(5) *Jean a oublié qu’il était à Paris ce jour là.*

4 Limites de TimeML

Dans le cadre du projet TEMPORAL, financé par la MSH Val de Loire, nous comptons ajouter une couche d’annotation temporelle au corpus ANCOR_Centre (Muzerelle et al. 2013, 2014). Couvrant un large éventail de relations anaphoriques ou de coréférence, ce grand corpus oral ne recense toutefois pas les anaphores abstraites (Dipper & Zinsmeister, 2010) telles que :

- (6) L1 : *Pierre a encore cassé sa voiture.*
L2 : *Venant de lui, ça ne m’étonne pas.*

Ici, le pronom anaphorique “*ça*” a pour antécédent l’éventualité décrite par l’ensemble de l’énoncé de L1. TEMPORAL avait pour objectif initial de rendre compte de ce genre de situations avec la norme TimeML. Une démarche méthodologique consistant à réaliser à intervalles réguliers des ateliers d’annotation entre experts nous a toutefois révélé certaines limitations pénalisantes de la norme, qui sont l’objet du présent article.

4.1 Confusion entre les niveaux d’annotation

ISO TimeML cherche visiblement à intégrer dans une seule couche d’annotation tous les éléments nécessaires à la résolution des références temporelles. Ce choix conduit à des corpora auto-suffisants mais induit des incohérences du fait de la juxtaposition d’annotations de niveaux linguistiques différents. C’est ainsi le cas de la distinction entre événement et instance (EVENT/MAKEINSTANCE). EVENT est en effet directement lié à un observable linguistique, marqueur de l’éventualité dans l’énoncé, tandis que l’instance réfère à la réalisation effective de cette éventualité. Paradoxalement, TimeML confère à l’instance l’ensemble des traits qui décrivent l’éventualité, alors que ceux-ci sont clairement liés à sa réalisation linguistique en contexte (*tense, aspect, etc.*).

(7) *Jean a enseigné deux fois lundi.*

Ainsi, l’exemple (7), transposé du guide d’annotation TimeML, montre une répétition d’éventualités. La norme propose de définir un EVENT pour décrire une éventualité traduite par le verbe “*a enseigné*”, puis deux instances correspondant aux deux réalisations successives de cette éventualité. Il serait préférable de définir des niveaux d’annotation séparés, ce que permet le format déporté d’ISO TimeML. Nous conseillons même de ne conserver que les objets EVENT en leur intégrant les attributs de MAKEINSTANCE comme dans le French TimeBank (Bittar et al. 2011), nous limitant à une annotation objective purement linguistique. L’annotation combinée des deux niveaux pose en effet des questions en termes de fiabilité des données. Considérons maintenant l’énoncé (8) :

(8) *Jean a enseigné aux L1 le lundi et Marc aux L2 le mardi.*

Doit-on encore considérer comme TimeML qu’il n’y a qu’une seule éventualité et plusieurs instances, alors que les actions diffèrent, puisque les arguments du verbe varient ? Dès lors, où s’arrête la distinction entre éventualités et instances ? Dans le cadre du projet TEMPORAL qui réunit nos deux laboratoires, nous proposons de nous limiter à une annotation purement linguistique qui limite ce genre de questionnement. L’annotation pourrait ensuite être complétée par d’autres couches pour les communautés travaillant sur la recherche d’information ou les systèmes question/réponse.

4.2 Délimitation des éventualités

La remarque précédente pose la question de la délimitation des éventualités. ISO TimeML a choisi de caractériser les éventualités uniquement par la tête lexicale de leur observable linguistique, en ignorant leurs arguments. Ce choix est motivé par le souci de simplifier la tâche des annotateurs (Bittar, 2008). Cette question a préoccupé Pustejovsky et ses collègues, qui ont proposé d'annoter à part les arguments des têtes lexicales portant une éventualité (Pustejovsky et al. 2006). Cette proposition n'a toutefois pas été retenue dans la norme TimeML.

Nos expériences d'annotation n'ont pas démontré de difficulté à délimiter des éventualités à large empan. Il nous semble donc préférable d'annoter l'ensemble de l'unité linguistique décrivant l'éventualité. En effet, le calcul de la temporalité d'un énoncé se fait à partir de tous ses éléments pertinents, ce dont doit rendre compte le balisage dû à l'annotation. Cette approche permet de se dispenser des relations artefactuelles auxquelles TimeML est condamné dans le cas d'encapsulation d'éventualités comme dans l'exemple (9) :

(9) [vous voulez [venir avec votre chien]]

Ce choix est cohérent avec les pratiques observées pour l'annotation des entités nommées. Il permet d'établir une procédure homogène et syntaxiquement cohérente avec le traitement des éventualités nominales. L'exemple ci-dessous fait ainsi un parallèle entre les délimitations d'entités imbriquées proposées pour notre annotation temporelle (10a) et pour une annotation en entités nommées (10b) telle qu'envisagée par exemple dans la campagne ETAPE.

(10a) [le début de [la seconde phase de [colonisation de l'archipel]]]

(10b) [le président de [la première réunion de [la nouvelle assemblée constituante]]]

L'intérêt de ce choix est manifeste dans le cas des anaphores abstraites ou de l'encapsulation d'éventualités. Dans l'exemple (11), l'éventualité ancrée par *vouloir* a une portée sur les deux autres éventualités, ce qui est aisément capturé par l'encapsulation. TimeML peut représenter ce type d'énoncé, mais d'une manière moins élégante avec des relations.

(11) [vous voulez [venir avec votre chien] et [faire l'intégralité de la promenade]]

Il n'en reste pas moins qu'une annotation à large empan peut conduire à de nouvelles difficultés que nous avons-nous-même expérimentées. Tout d'abord, les relations entre les têtes lexicales et leurs arguments peuvent s'étendre sur de grandes longueurs. Nous proposons de nous limiter à l'inclusion des éléments sous-catégorisés (arguments de valence pour un verbe, par exemple). D'autre part, les éventualités ainsi délimitées peuvent être discontinues. C'est le cas à l'oral où des disfluences telles que les incises peuvent découper l'éventualité. On peut formellement décrire ces discontinuités (notion de schéma sous la plateforme Glozz, par exemple (Mathet & Widlöcher, 2009)) mais cela complique la tâche d'annotation. C'est pourquoi notre réflexion s'oriente à l'heure actuelle vers une annotation sur treebank, inspirée de l'annotation des expressions polylexicales dans le Prague Dependency Treebank (Bejček & Straňák, 2010): puisqu'une éventualité est représentée dans un treebank par le nœud de l'arbre syntaxique qui correspond à la tête lexicale tout en couvrant la partie de l'énoncé concerné, identifier cette éventualité se réduirait à pointer ce nœud dans l'arbre tout en lui assignant des attributs spécifiques à la temporalité.

Des incohérences plus anecdotiques de délimitation peuvent également se rencontrer. TimeML distingue ainsi deux types d'entités qui modifient ou précisent la portée temporelle des éventualités ou des expressions temporelles :

- les signaux, qui indiquent la présence d'une relation temporelle (TLINK) entre deux éventualités ou expressions temporelles. Ainsi, la préposition *avant* sera annotée comme un signal dans l'énoncé *je suis parti avant midi*.
- les modificateurs (MOD), qui altèrent ou précisent une portée temporelle. Par exemple, la locution *en fin de* sera annoté comme un modificateur dans l'énoncé *je suis parti en fin de matinée*.

Ces entités jouent des rôles relativement proches mais donnent lieu à des conventions d'annotation radicalement différentes. Les signaux sont des objets d'annotation propres, délimités par un balisage spécifique (<SIGNAL> avant </SIGNAL>) alors que les modificateurs conduisent simplement à l'ajout d'un attribut MOD dans l'élément qu'ils caractérisent. ISO TimeML donnera ainsi sur notre exemple l'annotation (simplifiée) suivante : en fin de <EVENT ... MOD='END'> matinée </EVENT>. Une description homogène rendrait plus fiable le processus d'annotation.

4.3 Expressions temporelles

L'annotation des expressions temporelles dans TimeML dérive du tag TIMEX2 utilisé dans TIDES (Ferro 2005). Cette norme répond plus aux besoins de la RI que du TAL. Les expressions temporelles y sont en effet intégralement évaluées temporellement en contexte. Ainsi, "*le mois dernier*" sera par exemple annotée 2014-05. Cette convention pose des

problèmes de cohérence : le contexte d'évaluation ne sera en effet pas toujours connu pour permettre cette résolution temporelle. Cela renforce par ailleurs la part de subjectivité laissée à l'annotateur, qui ne cherchera pas toujours l'information nécessaire à cette évaluation. Ici encore, TimeML mélange deux niveaux d'annotation : celui du signifiant linguistique et celui de son interprétation qui relève d'un raisonnement purement temporel. Nous proposons donc d'annoter "*le mois dernier*" sous la forme -1M (mois antérieur) et de laisser à une autre couche d'annotation la question de la résolution complète de l'expression temporelle.

4.4 Attributs liés aux éventualités

Enfin, certaines limitations concernent les attributs liés aux éventualités. L'une d'entre elles est spécifique l'adaptation au français de TimeML (Bittar et al. 2011). Le French TimeBank ajoute, pour le français, des temps dans le système de valeurs de l'attribut TimeML *tense*. On retrouve un temps *past* mais également un temps *imperfect*. Ce choix manque de pertinence linguistique : le terme *imperfect* étant traditionnellement attaché à une notion d'aspect, les valeurs du trait ici croisent d'un côté une notion de temps linguistique (*tense* en anglais et qui pourrait justifier *imperfect*) et de l'autre une représentation du temps extralinguistique en 3 parties *past*, *present*, *future*.

D'autres limitations correspondent à des manques de la norme. Par exemple ISO TimeML prévoit un attribut *neg* qui permet de représenter la négation d'une éventualité (exemple : *il ne pleut pas*). Dans cet esprit, il nous semble manquer un attribut pour marquer l'occurrence de l'éventualité dans une question. Cette information peut par exemple avoir une utilité pour l'identification des actes de langage liés à l'éventualité. Trois valeurs peuvent être portées par cet attribut :

- YEST si la question porte sur la temporalité de l'évènement. Par exemple : *quand viendras-tu ?*
- YES si la question porte sur l'occurrence ou non de l'éventualité. Par exemple : *viendras-tu ?*
- NO lorsqu'il n'y a pas d'une question ou qu'elle porte sur une autre dimension. Par exemple : *Où vas-tu ?*

4.5 Relations temporelles

Le dernier élément important de la norme concerne les relations temporelles. ISO TimeML intègre un type de relation (TLINK) qui peut concerner de manière indifférente les éventualités ou les expressions temporelles, alors que les types ALINK et SLINK ne concernent que les éventualités. Nous regrettons là encore le groupement sous la même annotation d'objets décrivent des réalités différentes (cf § 3.1.). Il serait plus cohérent de réserver le type TLINK aux seules relations entre éventualités, et d'ajouter un type spécifique pour les relations d'ancrage liant une expression temporelle à une éventualité. Pustejovsky suit d'ailleurs cette logique lorsqu'il propose (séminaire invité, Paris 2014) de créer une relation MLINK pour représenter une relation entre un TIMEX et un EVENT dont le premier donne la durée du second.

On relève en outre que le type TLINK se limite aux relations d'ordonnement (égalité, précédence, post-occurrence...) de Allen (1983). Cette convention ignore l'existence de correspondances plus complexes, comme par exemple des relations d'inclusion temporelle ontologiquement décrites. Considérons l'exemple suivant :

(12) *Je suis né à Noël, précisément, le 25 décembre 1966. Cette année-là il a fait très froid durant tout l'hiver*

TimeML représente *25 décembre 1966* comme une date mais ne peut exprimer la relation d'inclusion qu'elle partage ontologiquement avec l'expression englobante *cette année-là* ? On se retrouve ici avec une relation de type meronymique (*partie_de*) au niveau temporel. Celle-ci s'apparente par analogie avec les anaphores associatives que l'on retrouve lorsque l'on s'intéresse à la référence (Webber 1988). L'attribut *reltype* de TLINK doit décrire ce type de relation. Par analogie, on propose de rajouter la valeur *bridging* au système de valeur associé à cet attribut.

5 Conclusion

Cet article a cherché à caractériser certaines limitations de la norme ISO TimeML qui demandent à être corrigées pour atteindre une annotation qui réponde réellement aux besoins du TALN, des sciences du langage mais également de l'ingénierie des connaissances. Certaines réponses à ces observations ne requièrent qu'une modification à la marge de la norme. La question de la délimitation des éventualités étant plus problématique de notre point de vue. Nous proposons de ne pas se limiter à la tête lexicale mais bien de l'étendre à tout le signifiant linguistique correspondant. La taille des ressources annotées avec la norme TimeML reste encore limitée. Il nous semble donc plus qu'urgent de la remettre partiellement en question, sur des points qui, comme nous l'avons rappelé, ont souvent été déjà questionnés.

Références

- ALLEN J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11), 832–843.
- BACH E. W. (1981). Time, Tense, and Aspect : An Essay in English Metaphysics . In *Radical Pragmatics*, 63–81.
- BATTISTELLI D. (2009). *La temporalité linguistique: circonscrire un objet d'analyse ainsi que des finalités à cette analyse* Habilitation à diriger des Recherches, Université de Nanterre-Paris X.
- BEJČEK E. & STRAŇÁK P. (2010). Annotation of multiword expressions in the Prague Dependency Treebank. *Language Resources and Evaluation*, 44(1–2). 7–21.
- BITTAR A. (2008) Annotation des informations temporelles dans des textes en français. Actes *RECITAL'2008*.
- BITTAR A., AMSILI P., DENIS P., DANLOS L. (2011). French TimeBank: An ISO-TimeML Annotated Reference Corpus. Proc. *ACL'2011*, Portland, Oregon : États-Unis.
- CLARK A. & LAPPIN S. (2010). Unsupervised learning and grammar induction. *The Handbook of Computational Linguistics and Natural Language Processing*, 57.
- DE SAUSSURE L. (1997). Le temps chez Beauzée : algorithmes de repérage et comparaison avec Reichenbach. *Cahiers Ferdinand de Saussure*, (49):171–195.
- DIPPER S., ZINSMEISTER H. (2010). Towards a standard for annotating abstract anaphora. In *LREC 2010 Workshop on Language Resources and Language Technology Standards*, 54-59.
- FERRO L. (2005). TIDES 2005 standard for the annotation of temporal expressions. *MITRE Corp. Tech. Report*.
- GRAVIER G., ADDA G., PAULSON N., CARRÉ M., GIRAUDEL A. & GALIBERT O.(2012). The ETAPE corpus for the evaluation of speech-based TV content processing in the French language. In *International Conference on Language Resources, Evaluation and Corpora*, na. Turquie.
- ISO (2008). ISO DIS 24617-1: 2008 Language resource management - Semantic annotation framework - Part 1: Time and Events. International Organization for Standardization, ISO Central Secretariat, Genève, Suisse.
- HOFFART J., SUCHANEK F. M., BERBERICH K., LEWIS-KELHAM E., DE MELO G. & WEIKUM G. (2011). YAGO2: exploring and querying world knowledge in time, space, context, and many languages, in Proc. *20th International Conference on World Wide Web, WWW 2011*, Hyderabad, India (Companion Volume), 229–232.
- MANI I., PUSTEJOVSKY J., GAIZAUSKAS R., SAURI R., CASTANO J., LITTMAN J., SETZER A., & KATZ G. (2005). The specification language TimeML. In Mani I., Pustejovsky J., Gaizauskas R. (Eds.), *The language of time : a reader*, chapter 27. Oxford.
- MATHET, Y., WIDLÖCHER, A. (2009). La plate-forme GLOZZ : environnement d'annotation et d'exploration de corpus. *Actes de TALN-2009*, pages 1–10
- MUZERELLE J., LEFEUVRE A., ANTOINE J-Y, SCHANG E, MAUREL D., VILLANEAU J., ESHKOL I. (2013). ANCOR, premier corpus de français parlé d'envergure annoté en coréférence et distribué librement. Actes *TALN'2013*, Les Sables d'Olonnes.
- MUZERELLE J., LEFEUVRE A., SCHANG E., ANTOINE J-Y., PELLETIER A., MAUREL D., ESHKOL I., VILLANEAU J. (2014). ANCOR_Centre, a large free spoken french coreference corpus. Proc. *LREC'2014*, Reykjavik.
- OGRODNICZUK M., ZAWISŁAWSKA M., SAVARY A., GŁOWŃSKA K. (2013). Coreference_Annotation Schema for an Inflectional Language, Proc. *CICLING'2013*, Samos, Greece, In. *LNCS 7816*, Springer Verlag, 394-407.
- PUSTEJOVSKY J., LITTMAN J., SAURI R. (2006). Argument Structure in TimeML. In Katz G., Pustejovsky J., Schilder F. (Eds.) *Annotating, Extracting and Reasoning about Time and Events*, Dagstuhl Seminar Proc., Schloss Dagstuhl, Allemagne.
- SAVARY A., WASZCZUK J., PRZEPIÓRKOWSKI A. (2010). Towards the Annotation of Named Entities in the National Corpus of Polish. in Proc. *LREC'10*, Valetta, Malta.
- WEBBER B. L. (1988). Tense as discourse anaphor. *Computational Linguistics*, 14(2), 61-73.