

Une méthode pour l'annotation de relations temporelles dans des textes et son évaluation

Philippe Muller, Xavier Tannier
IRIT-CNRS UMR 5505
Université Paul-Sabatier
118 Rte de Narbonne
31062 Toulouse Cedex 04
muller@irit.fr, tannier@emse.fr

Résumé - Abstract

Cet article traite de l'annotation automatique d'informations temporelles dans des textes et vise plus particulièrement les relations entre événements introduits par les verbes dans chaque clause. Si ce problème a mobilisé beaucoup de chercheurs sur le plan théorique, il reste en friche pour ce qui est de l'annotation automatique systématique (et son évaluation), même s'il existe des débuts de méthodologie pour faire réaliser la tâche par des humains. Nous proposons ici à la fois une méthode pour réaliser la tâche automatiquement et une manière de mesurer à quel degré l'objectif est atteint. Nous avons testé la faisabilité de ceci sur des dépêches d'agence avec des premiers résultats encourageants.

This paper focuses on the automated processing of temporal information in written texts, more specifically on relations between events introduced by verbs in every clause. While this latter problem has been largely studied from a theoretical point of view, it has very rarely been applied to real texts, if ever, with quantified results. The methodology required is still to be defined, even though there have been proposals in the human annotation case. We propose here both a procedure to achieve this task and a way of measuring the results. We have been testing the feasibility of this on newswire articles, with promising first results.

Mots-clefs – Keywords

Annotation, Temps, Discours
Text annotation, Time, Discourse

1 Introduction

Annoter l'information temporelle Cet article concerne l'annotation automatique d'informations temporelles dans des textes et vise plus particulièrement les relations entre événements introduits par des verbes dans chaque clause. Si la sémantique des marqueurs du temps et la structuration temporelle du discours sont des sujets d'étude très développés en linguistique formelle (Steedman, 1997)¹, l'annotation systématique (et quantifiée) de textes quelconques en est encore à ses débuts. Ce problème commence à générer des travaux en linguistique computationnelle (Harper *et al.*, 2001), liés aux enjeux que représente les informations temporelles entre autres pour la recherche d'information et les systèmes question-réponse.

On peut distinguer plusieurs tâches sous cette bannière :

- la détection de dates et de marqueurs temporels
- le repérage d'événements
- la datation d'événements
- la détermination de l'ordre des événements dans un texte

Le premier problème est relativement aisé à résoudre pour les dates à l'aide d'expressions régulières (Wilson *et al.*, 2001), mais nécessite une part d'analyse grammaticale dans des cadres plus larges (Vazov, 2001; Schilder & Habel, 2001). Le deuxième pose des problèmes ontologiques: qu'est-ce qui compte comme un événement (Setzer, 2001), comment considérer les verbes attitudinaux (croyances), les discours indirects, ... Le troisième rencontre une sous-détermination importante: beaucoup d'événements n'ont pas de date explicite associées dans le texte, et il n'est pas toujours possible d'en déterminer une contextuellement (Filatova & Hovy, 2001). Cela conduit à une approche plus adaptée à cette imprécision : annoter aussi les relations entre événements de manière symbolique (par exemple, e_1 est avant e_2), qui est la démarche suivie par (Katz & Arosio, 2001; Setzer, 2001) avec des annotateurs humains. Il se pose alors dans le cas de l'annotation humaine plusieurs niveaux de difficultés : le choix des relations à proposer et la comparaison entre annotateurs notamment. L'automatisation de telles annotations est de toutes façons loin d'être triviale, et nécessite une forme même simplifiée de modélisation discursive (cf. le travail préliminaire de (Grover *et al.*, 1995)). Nous nous proposons ici de montrer la faisabilité de cette tâche, et de proposer une façon de l'évaluer.

Le problème de l'évaluation Ce que l'on cherche à annoter est en fait une approximation du modèle que se fait le lecteur de l'information temporelle contenue dans un texte ; cela fait intervenir une part de raisonnement et l'intégration de diverses sources d'informations (lexicales et discursives). Comme l'a remarqué (Setzer, 2001), il est difficile d'accorder des annotateurs humains entre eux, car ils peuvent se contenter d'exprimer quelques relations entre événements (par exemple qu'un événement e_1 a lieu pendant un autre e_2 , et que e_2 a lieu avant e_3 , étant implicite qu'alors e_1 a lieu avant e_3). Une possibilité serait de demander explicitement d'établir un lien pour chaque paire d'événements, mais cela conduirait à une tâche très lourde pour l'humain, puisqu'il y a $O(n^2)$ paires pour n événements reconnus dans un texte.

¹Pour le français on peut se référer aussi à (Bras, 1991).

Pour comparer deux annotations, (Setzer, 2001) mais aussi de façon simplifiée (Katz & Arosio, 2001) propose alors de prendre les fermetures transitives des annotations par rapport à quelques règles d'inférence intuitives (comme celle que nous venons de prendre en exemple). On peut ensuite mesurer, pour chaque relation proposée aux annotateurs, que les événements reliés sont bien les mêmes avec les mesures classiques de précision et de rappel. Ceci améliore le protocole, mais (Setzer, 2001) note que cela n'empêche pas les humains d'oublier certaines informations, et qu'il faut alors repasser avec l'humain sur ses oublis. L'auteur estime alors à une heure environ le temps nécessaire par texte (en moyenne comprenant entre 15 et 40 événements). Cela rend l'automatisation d'autant plus cruciale. En fait on peut déceler deux problèmes dans cette méthode : la première est le choix des relations temporelles proposées aux annotateurs, à savoir "avant", "après", "pendant", "simultanément". La dernière est particulièrement difficile à estimer, puisqu'elle correspond d'après l'auteur à "à peu près au même moment" (*roughly at the same time*, p81). Le deuxième problème est lié au modèle inférentiel considéré. Celui-ci est en effet loin d'être "complet". S'il est difficile d'estimer dans quelle mesure un système de raisonnement se rapproche des capacités humaines, il existe néanmoins des systèmes formels plus précis pour modéliser l'information temporelle comme l'algèbre de relations d'Allen, très étudiée en raisonnement automatique. Nous pensons que la mesure de l'accord à partir de ce modèle sera plus précise, à condition d'appliquer pleinement le formalisme : à savoir que certaines informations déduites sont disjonctives (la relation entre deux événements est a OU b OU c).

Nous avons alors testé cette méthode et son évaluation sur une petite base de dépêches AFP, annotées par nos soins.

2 Méthode d'annotation automatique

Nous détaillons ici la méthode que nous avons utilisée pour annoter des textes de dépêches d'agence. Les étapes suivantes ont été réalisées sur les textes bruts, enrichis d'un tampon temporel précisant leur date d'émission :

- annotation en parties de discours (avec le logiciel Treetagger (Schmid, 1994), en gardant l'information du temps grammatical) ;
- un prétraitement est effectué (notamment la détection de certaines locutions prépositionnelles pour en faire des unités lexicales) ;
- une analyse grammaticale superficielle est effectuée avec des cascades d'analyseurs par expressions régulières² (cf. (Abney, 1996) dont nous avons utilisé le logiciel Cass) ; cela permet d'extraire les marqueurs de dates, ceux exprimant des relations temporelles (groupes prépositionnels, marqueurs temporels,...), ainsi que de fournir un découpage sommaire en clauses (un groupe verbal avec un temps fini par clause) et d'attacher à ces clauses les adjoints correspondants dans la plupart des cas (l'ambiguïté de certaines structures syntaxiques pouvant générer des erreurs à ce niveau). La figure 1 donne une idée des informations définies et extraites à ce niveau (pour l'instant le focus temporel est toujours l'événement précédemment trouvé), pour lesquelles une interprétation temporelle est donnée.

²Les règles que nous avons définies sont au nombre de 89 réparties en 29 niveaux.

Éléments temporels générés (reconnus) par <i>shallow parsing</i>	
date(1/2) : date non absolue ("le 25 mars", "en juin").	duree : durée quelconque ("pendant 3 ans").
dateabs : date absolue "le 14 juillet 1789".	duree2 : durée contenant deux dates (<i>du 11 février au 27 octobre...</i>).
daterelST(1/2) : date relative au moment d'élocution ("il y a 2 ans", "l'année dernière").	dureerelabs : durée absolue ("à partir du 14 juillet").
daterelTF(1/2/3) : date relative au focus temporel ("3 jours plus tard").	dureerelST : durée relative au moment d'élocution ("depuis un an").
datespecabs : date absolue, de forme particulière ("Au début des années 1980").	dureerelTF : durée relative au focus temporel ("depuis").
datespecrel : date relative, de forme particulière (mois, saisons).	tatome : atome de temps (<i>trois jours, 4 ans, ...</i>).

Figure 1: Lexique des éléments utilisés pour la gestion des adverbiaux temporels

- calcul de dates pour affiner les localisations temporelles des événements associés à des informations temporelles explicites (par exemple des dates relatives à la date de la dépêche comme *lundi dernier*).
- pour chaque événement reconnu associé à un adverbial temporel on établit une relation temporelle avec une date si cela est possible ;
- utilisation de règles d'enchaînement discursif pour le français, comparables à celle de (Grover *et al.*, 1995; Song & Cohen, 1991; Kameyama *et al.*, 1993) pour l'anglais, mais intégrant les relations temporelles d'Allen (Allen, 1984) au lieu d'un ensemble ad hoc (cf table 1 pour une idée de la forme de ces règles). Ces règles sont des règles par défaut en l'absence d'autres marqueurs, précisant l'ensemble des relations temporelles possibles entre deux événements consécutifs dans le texte³.
- calcul de relations par utilisation de procédures de calcul algébrique sur le graphe de relations liant les événements détectés et les entités temporelles autonomes (dates), en utilisant un algorithme de consistance de chemin (Allen, 1984).

Lors du découpage en clauses, nous déplaçons les relatives en fin de phrase, avec remplacement du pronom relatif par le groupe nominal associé par nos règles, de façon similaire à (Filatova & Hovy, 2001). Les relations d'Allen sont rappelées figure 2; on les abrègera par la suite en gardant la première lettre de la relation, et en ajoutant "i" pour la relation inverse. L'algorithme de calcul de relations est détaillé dans la section suivante.

3 Inférences des relations entre événements

Nous avons argumenté en faveur de l'utilisation des relations d'Allen pour l'annotation des relations temporelles non seulement parce qu'elles ont une sémantique claire, mais aussi parce

³Nous comptons enrichir plus tard ces contraintes en tenant compte d'un temps de référence et d'un focus temporel, suivant (Kameyama *et al.*, 1993; Song & Cohen, 1991).

Annotation de relations temporelles dans des textes

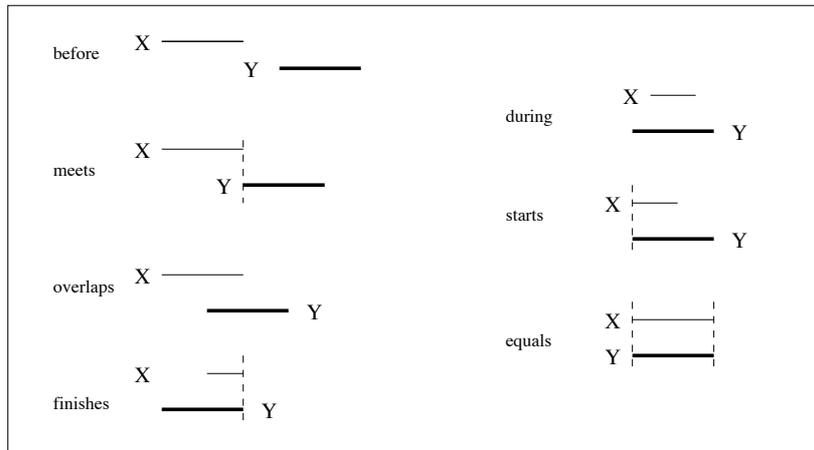


Figure 2: Les relations d'Allen entre deux événements X et Y (le temps s'écoulant de gauche à droite)

1/2	ps	pqp	imp	...
ps	e,b,m,s,d,f,o	e,s,d,f,bi,mi	b,m,o,e,s,d,f	
pqp	b,m,o	b,m,o,e,s,d,f,bi,mi	b,m,o,e,s,d,f	
imp	o,d,s,f,e,si,di,fi	bi,mi,oi	o,e,s,d,f,si,di,fi	
...				

Table 1: Contraintes temporelles discursive (extrait), ps=passé simple, imp=imparfait, pqp=plus que parfait, ...

que de nombreux travaux ont été faits sur l'inférence sur des contraintes exprimées avec ces relations. Nous pensons donc qu'un bon moyen de s'abstraire des aléas liés au choix des relations annotées par l'humain est de calculer toutes les relations possibles entre événements d'un texte (c'est-à-dire réaliser une saturation du graphe de contraintes), à partir d'une annotation automatique ou humaine, et de comparer les graphes résultants.

Cette saturation ne sera pas réalisée au moyen de règles spécifiques (relation entre e1 et e2) + (relation entre e2 et e3) donne (une relation simple entre e1 et e3) (à la manière de (Setzer, 2001; Katz & Arosio, 2001)), mais en utilisant l'algèbre complète de relations d'Allen, pour fournir un modèle que nous pensons plus complet. Cela permet notamment de détecter plus sûrement une incohérence dans l'annotation. Une algèbre de relation peut se définir sur un ensemble de relations mutuellement exclusives (deux relations ne peuvent être valides en même temps entre deux objets) et exhaustives (au moins une relation relie deux objets quelconques). L'algèbre est construite en considérant des ensembles de relations de base $\{r_1, r_2, \dots\}$ que l'on peut interpréter comme de l'information disjonctive entre deux entités. On peut alors parler d'intersection et d'union de relations. On peut alors généraliser les règles de transitivité de la forme: $r(x, y) \wedge r'(y, z) \rightarrow r''(x, z)$ à tous les couples de relations r et r' de l'algèbre complète. L'algorithme suivant, appelé "propagation par consistance de chemin" (Allen, 1984), assure que l'on arrive à un point fixe de saturation du graphe :

Soient $\left\{ \begin{array}{l} A = \text{l'ensemble des arcs du graphe} \\ N = \text{l'ensemble des nœuds du graphe} \\ U = \text{la disjonction des treize relations d'Allen} \\ R_{m,n} = \text{la relation entre deux nœuds } m \text{ et } n \end{array} \right.$

1. *modifié* = 0
2. Pour chaque couple de noeuds $(i, j) \in N \times N$ et pour tout $k \in N$ tel que $((i, k) \in A \wedge (k, j) \in A)$
 - (a) $R1_{i,j} = \text{compose}(R_{i,k}, R_{k,j})$ (voir plus loin)
 - (b) Si aucun arc (une relation $R2_{i,j}$) n'existait déjà entre i et j , alors $R2_{i,j} = U$
 - (c) faire l'intersection : $R_{i,j} = R1_{i,j} \cap R2_{i,j}$
 - (d) Si $R_{i,j} = \emptyset$ (inconsistance détectée)
alors : afficher un message d'erreur
 - (e) Si $R_{i,j} = U$ (aucune information) ne rien faire
sinon mettre l'arc à jour
modifié = 1
3. Si *modifié* = 1, recommencer en 1.

La fonction $\text{compose}(R_{i,k}, R_{k,j})$ utilise les 169 relations de transitivité entre les relations d'Allen. Pour chaque couple (r_1, r_2) de relations binaires d'Allen tel que $r_1 \in R_{i,k}$ et $r_2 \in R_{k,j}$, on applique la transition adaptée. Le résultat est la disjonction de toutes les relations ainsi obtenues.

Il faut noter que cet algorithme est correct (s'il détecte une incohérence alors le graphe est bien incohérent), mais incomplet dans le cas général (il ne détecte pas nécessairement un scénario incohérent)⁴.

4 Mesures de l'annotation

Afin de valider la méthode employée pour annoter, nous avons comparé les résultats avec une annotation produite "manuellement". Il n'est pas envisageable de faire annoter par l'humain (expert ou non) des relations par l'intermédiaire des relations d'Allen, trop nombreuses et trop précises (ce qui obligerait à des notations disjonctives difficiles à gérer pour l'annotateur moyen). En opposition à (Setzer, 2001), nous voulons cependant des relations dont la sémantique est précise, et que nous pouvons alors relier à l'algèbre d'Allen pour effectuer les inférences qui permettent de comparer deux "modèles" donnés par un ensemble de contraintes. Nous avons alors choisi (suivant (Li *et al.*, 2001)) des relations similaires aux relations de (Bruce, 1972) (qui a d'ailleurs inspiré Allen), dont la définition en termes de relation d'Allen est donnée table 2. Pour s'abstraire des particularités d'une annotation et pour se concentrer sur le modèle temporel sous-jacent nous allons donc mesurer une distance entre l'annotation du système et

⁴Il existe des sous-algèbres pour lesquelles cette procédure est également complète, et il serait intéressant de voir dans quelle mesure les annotations sont restreintes ou non à ces ensembles de relations; nous n'avons pas effectué cette étude pour l'instant.

BEFORE	$\forall i \forall j (i \text{ before } j \Leftrightarrow ((i < j) \vee (i \text{ m } j)))$
AFTER	$\forall i \forall j (i \text{ after } j \Leftrightarrow ((i > j) \vee (i \text{ mi } j)))$
OVERLAPS	$\forall i \forall j (i \text{ overlaps } j \Leftrightarrow ((i \text{ o } j)))$
IS_OVERLAPED	$\forall i \forall j (i \text{ is_overlaped } j \Leftrightarrow ((i \text{ oi } j)))$
INCLUDES	$\forall i \forall j (i \text{ includes } j \Leftrightarrow ((i \text{ di } j) \vee (i \text{ si } j) \vee (i \text{ fi } j) \vee (i = j)))$
IS_INCLUDED	$\forall i \forall j (i \text{ is_included } j \Leftrightarrow ((i \text{ d } j) \vee (i \text{ s } j) \vee (i \text{ f } j) \vee (i = j)))$

Table 2: Relations proposées à l'annotation

celle de l'humain à partir du graphe saturé des relations temporelles sur les événements détectés. Nous souhaitons alors ne pas seulement mesurer l'exactitude des relations "simples" (non disjonctives) entre événements, comme (Setzer, 2001), car cela rend le résultat très dépendant du choix des relations proposées à l'humain, et nous voulons de plus avoir une mesure graduelle de l'imprécision du résultat automatique: trouver une relation "avant OU pendant" entre deux événements est mieux que proposer "après" quand l'humain a proposé "avant". Nous voulons alors mesurer deux choses: la cohérence de l'annotation automatique avec l'humain (l'information contenue dans le texte implique celle trouvée automatiquement) et la précision de cette information (une information trop disjonctive est moins précise qu'une information bien ciblée, par exemple (a ou b ou c) est moins bon que (a ou b) si la relation à trouver est (a)). Nous proposons donc les mesures élémentaires suivantes entre deux relations attribuées entre deux événements, où S désigne l'ensemble des relations proposées par le système (a ou b ou c ou ...) et H celle proposée par l'humain (après propagation des contraintes):

$$\text{finesse} = \frac{|S \cap H|}{|S|} \quad \text{cohérence} = \frac{|S \cap H|}{|H|}$$

La mesure globale de l'annotation est alors la moyenne de ces mesures sur tous les arcs comportant de l'information selon l'humain (on exclut les arcs où toutes les relations sont possibles), après saturation du graphe de relations entre événements.

D'une certaine façon la finesse mesure la quantité d'information récupérée par le système et la cohérence donne une mesure des erreurs que commet le système vis à vis de l'information présente dans le texte. En ce sens finesse et cohérence sont les pendants des mesures de rappel et précision quand on évalue des annotations simples⁵. Evidemment si S=H sur tous les arcs, les mesures valent 1. Si le système ne donne aucune information, S est une disjonction de toutes les relations, donc $H \subseteq S$, $H \cap S = H$ et la cohérence vaut trivialement 1, mais la finesse sera alors très faible.

Ces mesures peuvent bien sûr aussi être utilisées pour estimer l'accord entre deux annotateurs humains.

5 Résultats

Pour voir si la mesure a un sens, on s'en sert tout d'abord pour comparer un texte annoté aléatoirement de la façon suivante: à chaque paire d'événements successifs du texte on associe

⁵Précision est un terme un peu ambigu puisqu'en français il s'appliquerait mieux à notre mesure de finesse, alors que celle-ci correspond plutôt au "rappel" d'informations ; pour éviter les confusions, nous avons choisi de nouveaux termes.

une relation d'annotation au hasard, on sature le graphe, et on évalue les résultats par rapport à l'annotation humaine. Le résultat est typiquement très faible, comme on peut le voir table (3), sur une dépêche prise comme exemple.

Ici on fait trois séries de mesure: une sur les relations simples (non disjonctives) annotées par l'humain (et après saturation), une sur toutes les relations d'annotation (simples ou disjonctions de simples), et une sur l'équivalence en relation d'Allen (qui reflète peut-être mieux le calcul sous-jacent, au détriment de la mesure de la performance de la tâche réelle visée).

	Finesse	Cohérence
relations d'annotations (toutes)	0.114	0.011
relations d'annotations (simples seulement)	0.125	0.095
relations d'Allen	0.083	0.094

Table 3: Exemple d'évaluation sur une annotation "au hasard"

Nous avons ensuite utilisé ces mesures pour estimer le rapport entre notre algorithme et l'annotation humaine. Un ensemble, pour l'instant assez limité⁶, de 8 textes a été utilisé, constitué de dépêches prises au hasard sur le site de l'AFP, pour un total d'environ 2300 mots et 160 événements⁷. Ces textes avaient chacun entre 10 et 40 événements, pour lesquels les taux de rappel et de précision sont supérieurs à 97%. Nous avons alors comparé les résultats en ne gardant que les événements correctement reconnus. Cela ne propage pas les erreurs sur les annotations au niveau de la cohérence, mais handicape bien sûr le niveau de finesse, puisque les inférences sont faites avec moins d'informations. Nous avons fait la moyenne des résultats par texte.

	Finesse	Ecart-type	Cohérence	E.-T.
relations d'annotations (toutes)	0.477499	0.286781	0.449899	0.175922
relations d'Allen	0.448222	0.289856	0.495755	0.204974

Table 4: Evaluation des résultats

Ces résultats semblent encourageants au vu des simplifications que nous avons opérées sur toute la chaîne de traitement, même si les variations importantes entre textes (et leur nombre encore limité) doivent inciter à la prudence. L'objectif était surtout ici de montrer la faisabilité de cette tâche de façon automatique, en fournissant également des outils pertinents pour mesurer le résultat, et pour étalonner chaque étape par la suite. Il faut noter que les points de comparaison manquent, puisqu'à notre connaissance, seuls (Li *et al.*, 2001) et (Mani & Wilson, 2000) mentionne avoir tenté cette annotation, en marge de leurs systèmes d'annotation d'expressions temporelles. Les premiers ne considèrent que les relations entre deux événements présents dans une même phrase ; les seconds n'ont pas évalué leur méthode.

Enfin, remarquons que l'annotation humaine elle-même est une tâche difficile. Pour l'instant les textes ont été annotés par les deux auteurs, avec résolution après coup des différences. Nous n'avons donc pas de mesure inter-annotateurs, qui donnerait une borne maximum aux performances que l'on peut atteindre d'un système⁸, mais nous prévoyons d'étudier ce problème par la suite.

⁶Nous continuons le difficile processus d'annotation manuelle pour obtenir une base plus importante de dépêches.

⁷Par comparaison, le corpus anglais de (Setzer, 2001) comprend 6 dépêches pour moins de 2000 mots et 160 événements.

⁸Nous remercions le lecteur pour sa remarque à ce propos.

6 Conclusion

L'objectif de ce travail était de montrer la faisabilité de l'annotation automatique de relations entre événements dans un texte. Nous avons proposé une méthode d'évaluation de cette tâche et donné des résultats préliminaires. Il est bien évident que ceux-ci doivent être affinés, et ce sur des données plus nombreuses. Nous pouvons maintenant tenter d'améliorer certaines composantes de la méthode : étiquetage, analyse syntaxique, règles d'enchaînements des temps de discours, calcul des dates. Nous comptons en particulier évaluer la précision de notre reconnaissance d'adverbiaux temporels, qui sont la source majeure d'information de l'algorithme. La mesure de proximité entre réseaux temporels devrait nous permettre d'ajuster ces divers éléments. Elle devrait permettre aussi de préparer le terrain pour une combinaison de l'approche symbolique avec des méthodes d'apprentissage (il serait sans doute assez facile par exemple d'améliorer les règles d'enchaînement temporels avec un simple algorithme local opérant sur l'espace des règles possibles, en partant de notre premier modèle).

Références

- ABNEY S. (1996). *Corpus-Based Methods in Language and Speech*, chapter Part-of-Speech Tagging and Partial Parsing. Kluwer Academic Publisher.
- ALLEN J. (1984). Towards a general theory of action and time. *Artificial Intelligence*, **23**, 123–154.
- BRAS M. (1991). *Calcul des structures temporelles du discours*. PhD thesis, Université Paul Sabatier, Toulouse.
- BRUCE B. (1972). A model for temporal references and its application in a question answering program. *Artificial Intelligence*, **3**(1-3), 1–25.
- FILATOVA E. & HOVY E. (2001). Assigning time-stamps to event-clauses. In (Harper *et al.*, 2001).
- GAGNON M. & LAPALME G. (1996). From conceptual time to linguistic time. *Computational linguistics*, **22**(1), 91–127.
- GROVER C., HITZEMAN J. & MOENS M. (1995). Algorithms for analysing the temporal structure of discourse. In *Sixth International Conference of the European Chapter of the Association for Computational Linguistics: ACL*.
- L. HARPER, I. MANI & B. SUNDHEIM, Eds. (2001). *ACL Workshop on Temporal and Spatial Information Processing*, 39th Annual Meeting and 10th Conference of the European Chapter. Association for Computational Linguistics.
- KAMEYAMA M., PASSONNEAU R. & POESIO M. (1993). Temporal centering. In *Proceedings ACL 1993*, p. 70–77.
- KATZ G. & AROSIO F. (2001). The annotation of temporal information in natural language sentences. In (Harper *et al.*, 2001), p. 104–111.
- LI W., WONG K.-F. & YUAN C. (2001). A model for processing temporal reference in chinese. In (Harper *et al.*, 2001).
- MANI I. & WILSON G. (2000). Robust temporal processing of news. In *Proceedings ACL 2000*.
- SCHMID H. (1994). Probabilistic part-of-speech tagging using decision trees. In *International Conference on New Methods in Language Processing*.
- SETZER A. (2001). *Temporal Information in Newswire Articles: an Annotation Scheme and Corpus Study*. PhD thesis, University of Sheffield, UK.

- SHILDER F. & HABEL C. (2001). From temporal expressions to temporal information: Semantic tagging of news messages. In (Harper *et al.*, 2001), p. 65–72.
- SONG F. & COHEN R. (1991). Tense interpretation in the context of narrative. In *Proceedings of AAAI'91*, p. 131–136.
- STEEDMAN M. (1997). Temporality. In J. V. BENTHEM & A. TER MEULEN, Eds., *Handbook of Logic and Language*. Elsevier Science B.V.
- VAZOV N. (2001). A system for extraction of temporal expressions from french texts based on syntactic and semantic constraints. In (Harper *et al.*, 2001).
- WILSON G., MANI I., SUNDHEIM B. & FERRO L. (2001). A multilingual approach to annotating and extracting temporal information. In (Harper *et al.*, 2001).