

Apprentissage de structures linguistiques sous contraintes:
introduction
et application à la prédiction de relations temporelles

Philippe Muller
Equipe MELODI, IRIT, Univ. Paul Sabatier
`philippe.muller@irit.fr`

Séminaire CLLE
15 janvier 2011
En collaboration avec Pascal Denis (Alpage, INRIA)

- la plupart des problèmes de Traitement Automatique des langues considèrent des **structures** (séquences, arbres, graphes), et des transformations de ces structures
- la plupart des modèles employés en TAL sont fondés sur des approches statistiques
- l'application de ces modèles à des structures est difficile et restreint les facteurs pris en compte dans les modèles à une échelle très **locale**, ou très superficielle.
- ce qu'on vise ici est la meilleure prise en compte de contraintes **générales** dans les modèles prédictifs de structure
- schématiquement: remettre un peu de **symbolique** dans le **statistique**

- Structures et contraintes
- Apprentissage de structures
 - apprentissage local + décodage
 - exemple de prédictions structurées : coréférence, rôles sémantiques
- approche générale par ILP: retour sur les exemples
- le cas de la prédiction temporelle
 - le problème de l'ordonnancement temporel
 - état de l'art
 - solution par ILP
 - expérimentations

Structures linguistiques en TAL (à l'écrit)

une tâche de TAL \approx transformation de structures
résoudre la tâche \Rightarrow modèle prédictif de structures

séquence de caractères	\rightarrow	séquence de morphèmes
séquence de caractères	\rightarrow	séquence de tokens
séquence de tokens	\rightarrow	arbre de constituants ou graphe de dépendances syntaxiques ou séquence de chunks
séquence de chunks	\rightarrow	séquence d'arguments (semantic role labelling)
arbre syntaxique	\rightarrow	formule sémantique (FS, formule logique, etc)
phrase (tous niveaux)	\rightarrow	valeur simple (valence/opinion, spam, pertinence)

au-delà de la phrase :

séquence de clauses/DU	→	arbre/graphe de discours, séquence de parties
séquence de NPs/mentions	→	partition (coréférences)
séquence d'événements	→	graphe de relations entre événements
séquence de phrases	→	sous-ensemble (résumé extractif)
ensemble de textes	→	séquence de phrases (résumé multi-documents)
texte	→	classe (spam, opinion +/-, ...)

au-delà du texte

corpus+lexique → relations lexicales
(synonymes, (hyper,hyp)onymes, morphologie...)
corpus+lexique → ontologie(s)

dans presque tous les cas, **structures relationnelles**.

- les modèles à base de règles sont presque totalement remplacés par des modèles statistiques
- schématiquement :
 - un ensemble de facteurs calculés sur l'input,
"traits", "descripteurs"
 - choisir une famille de modèles paramétrables :
arbres de décision, SVM, maximum d'entropie, etc
 - ajuster les paramètres sur un ensemble d'exemples observés
entraînement
 - L'évaluation du modèle est faite sur un ensemble d'exemples
non vus pendant le paramétrage
test
- paradigme dominant : classification, i.e. associer une catégorie simple à un input

Dans le cas d'une tâche linguistique

- input = structure,
- le modèle de classification prédit une partie de l'output
- recombinaison des parties pour produire l'output structuré (“décodage”)

Exemple : séquence d'étiquettes morpho-syntaxiques

$$\begin{array}{ccc} x & \xrightarrow{f} & y \\ (x_1, x_2, x_3, x_4, x_5) & \rightarrow & (y_1, y_2, y_3, y_4, y_n) \\ (le, chat, brise, la, glace) & \rightarrow & (det, n, v, det, n) \end{array}$$

- prédictions isolées : $y_i = f(x_1, \dots, x_n)$
- mais interdépendance des y_i

Exemple : séquence d'étiquettes morpho-syntaxiques

x	\xrightarrow{f}	y
$(x_1, x_2, x_3, x_4, x_5)$	\rightarrow	$(y_1, y_2, y_3, y_4, y_n)$
$(le, chat, brise, la, glace)$	\rightarrow	$(det, n, v, det, ?)$

- prédictions isolées : $y_i = f(x_1, \dots, x_n)$
- mais interdépendance des y_i
- pour prédire le tag $f(x_i) = f(glace) = ?$

Exemple : séquence d'étiquettes morpho-syntaxiques

$$\begin{array}{ccc} x & \xrightarrow{f} & y \\ (x_1, x_2, x_3, x_4, x_5) & \rightarrow & (y_1, y_2, y_3, y_4, y_n) \\ (le, chat, brise, la, glace) & \rightarrow & (det, n, v, det, n) \end{array}$$

- prédictions isolées : $y_i = f(x_1, \dots, x_n)$
- mais interdépendance des y_i
- pour prédire le tag $f(x_i) = f(glace) = ?$
il vaut mieux savoir que $f(x_{i-1}) = det$ (\neq pronom)
plutôt que de juste savoir que $x_{i-1} = "la"$.

Exemple : séquence d'étiquettes morpho-syntaxiques

$$\begin{array}{ccc} x & \xrightarrow{f} & y \\ (x_1, x_2, x_3, x_4, x_5) & \rightarrow & (y_1, y_2, y_3, y_4, y_n) \\ (le, chat, brise, la, glace) & \rightarrow & (det, n, v, det, n) \end{array}$$

- prédictions isolées : $y_i = f(x_1, \dots, x_n)$
- mais interdépendance des y_i
- pour prédire le tag $f(x_i) = f(glace) = ?$
il vaut mieux savoir que $f(x_{i-1}) = \text{det}$ (\neq pronom)
plutôt que de juste savoir que $x_{i-1} = \text{"la"}$.
- \rightarrow apprentissage "joint" des y_i , avec restriction sur les dépendances
- solution usuelle: apprendre des contraintes (locales)
ex: $y_i = f(x, y_{i-1})$ (HMM, CRF).

- décodage complexe : optimiser l'ensemble, pas les décisions individuelles, c.à.d toutes les séquences possibles dans l'ensemble des dépendances
- → dépendances restent assez locales
 - pour limiter le nombre d'exemples d'entraînement nécessaires
 - pour limiter la combinatoire du décodage

- apprentissage : $f((x_1, \dots, x_n)) = (y_1, \dots, y_n)$
- structure : relations sur un ensemble de sorties :
séquence (tag,segmentation),
arbre (syntaxe),
graphe (discours)
- décodage
= reconstruction des structures possibles
= pb d'optimisation combinatoire
- contraintes plus globales sur les données ?
connaissances a priori ?
par ex: il doit y avoir (généralement) un verbe ?

une solution: mettre des pénalités sur le modèle s'il viole une contrainte

- soit au moment du décodage : Constraint-Driven Learning, (Chang et al. 2007),
- soit à l'entraînement : Generalized Expectation Constraints (Mann & Mc Callum, 2007),
- cf aussi : Posterior Regularization (Graça et al, 2007)

mais

- normalisation délicate
- que faire des contraintes "dures" ?
 $y_i = C \rightarrow y_j = C' ?$

Contraintes “dures” : contraintes sémantiques

résolution d'anaphores/coreférence

input:

un ensemble de mentions x_i ,

Alice, assise auprès de sa soeur sur le gazon, commençait à s'ennuyer de rester là à ne rien faire ; une ou deux fois elle avait jeté les yeux sur le livre que lisait sa soeur ; mais quoi ! pas d'images, pas de dialogues ! “La belle avance,” pensait Alice, “qu'un livre sans images, sans causeries !”

Contraintes “dures” : contraintes sémantiques

résolution d'anaphores/coreférence

input:

un ensemble de mentions x_i ,

prédiction:

un ensemble de coréférence (x_k, x_j)

contrainte “sémantique” évidente :

si (x, y) et (y, z) coréfère

alors (x, z) aussi

→ contrainte dure pour garder
cohérence sémantique

Alice, assise auprès de sa soeur sur le gazon, commençait à s'ennuyer de rester là à ne rien faire ; une ou deux fois elle avait jeté les yeux sur le livre que lisait sa soeur ; mais quoi ! pas d'images, pas de dialogues ! “La belle avance,” pensait Alice, “qu'un livre sans images, sans causeries !”

deux solutions :

- ① apprendre en ne considérant pas les sorties qui viole la contrainte
- ② apprendre sans contrainte, faire le décodage en respectant les contraintes

ici, on verra 2, plus simple

Décodage d'une structure en respectant les contraintes

- pour une séquence, eg: construire (y_1, y_2, \dots) puis ...
- pour la coréférence :
 - on peut interdire une séquence qui ne respecte pas la contrainte de transitivité
 - en résolvant dans l'ordre les paires possibles
 - x_2, x_1 ? oui/non
 - x_3/x_2 ? oui/non. si encore possible x_3/x_1 etc
- généralisation
 - problèmes joints, par ex. coréférence/typage des entités nommées
 - rôles sémantiques
- problème : on cherche à optimiser les prédictions \rightarrow combinatoire

Exemple : rôles sémantiques

ex:

- Bankers stole money from the government.
-

Exemple : rôles sémantiques

ex:

- Bankers stole money from the government.
ARG0 V ARG1 ARG2
-

Exemple : rôles sémantiques

ex:

- Bankers stole money from the government.
ARG0 V ARG1 ARG2
- They split it up on Treasure Island
ARG0 V ARG1 C-V1 A_LOC

attribution de rôles, optionalités, contraintes d'ordre etc
même problème de décodage, mais contraintes plus variées
tout refaire ?

Exemple : rôles sémantiques

ex:

- Bankers stole money from the government.
ARG0 V ARG1 ARG2
- They split it up on Treasure Island
ARG0 V ARG1 C-V1 A_LOC

attribution de rôles, optionalités, contraintes d'ordre etc
même problème de décodage, mais contraintes plus variées
tout refaire ?

cadre d'optimisation combinatoire sous contrainte: utiliser un outil générique

un cadre puissant et efficace: Integer Linear Programming

solveurs disponibles: on peut se concentrer sur la modélisation du problème

- contraintes = inégalité numériques
(programmation linéaire + entiers)
- coder le problème avec des variables entières x
- variable propositionnelle 0/1
- décision correspondant au problème
- contraintes exprimées sur les variables

- coréférence: $x_{i,j} = 1$ si i et j coréfèrent (0/1 = oui/non)
- $x_{ij} \geq 0, x_{ij} \leq 1$
- transitivité:
 $x_{ij} = 1$ et $x_{jk} = 1$ doit imposer $x_{ik} = 1$
- traduit en inégalités :
 $x_{ij} + x_{jk} - x_{ik} \leq 1$
- l'apprentissage doit donner un score pour chaque décision
si $f(x_{ij}=1)$ = proba de coréférence, notons le coût de se tromper $c_{ij} = -\log(f(x_{ij} = 1))$
alors on cherche à minimiser en prédiction
somme $c_{ij} * x_{ij}$
étant donné les c_{ij} (donné par un classifieur local)

(Denis & Baldrige, 2007)

en résumé, ILP =

- trouver le min/max d'une fonction coût
- étant donné un ensemble d'inégalités
- et un domaine de variables entières

Retour sur les rôles sémantiques

pour une phrase ensemble de segments (syntaxe), à étiquetter,
avec catégorie parmi ensemble de rôles

x_{ij} = segment i a le rôle j

contraintes:

- un seul rôle par segment

$$\sum_j x_{ij} = 1$$

- toujours au moins un arg0:

$$\sum_i x_i(\text{arg0}) \geq 1$$

- ordre des rôles: arg2 toujours après arg1:

$$x_i(\text{arg1}) + \sum_{j < i} x_j(\text{arg2}) \leq 1$$

- etc

cf (Koomen, Punyakanok, Roth, & Yih, 2005)

type de contraintes intégrables dans ce cadre

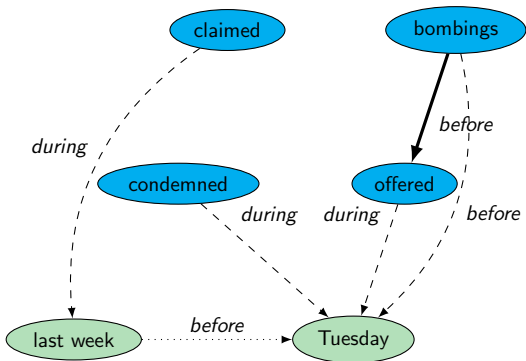
- ordre
- transitivité
- égalités de valeurs
- unicité, existence

aussi: expressivité mise dans les variables de décision.

on peut se contenter d'un classifieur simple pour apprendre les "morceaux" de structure

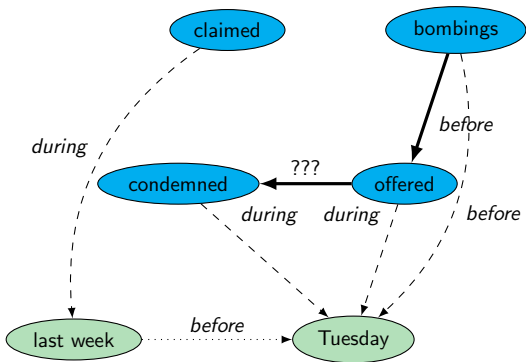
Ordonnement temporel

President Joseph Estrada on **Tuesday**_{t₄} **condemned**_{e₁} the **bombings**_{e₅} of the U.S. embassies in Kenya and Tanzania and **offered**_{e₁₂} condolences to the victims. [...] In all, the **bombings**_{e₁₀} **last week**_{t₅} **claimed**_{e₄} at least 217 lives.



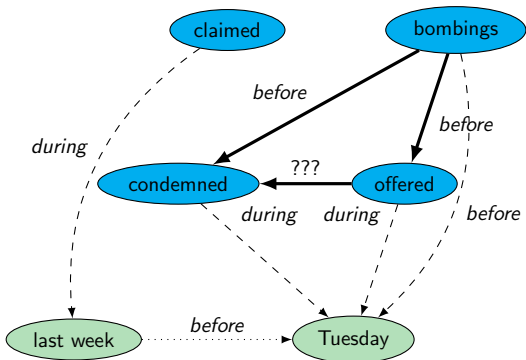
Ordonnement temporel

President Joseph Estrada on **Tuesday**_{t₄} **condemned**_{e₁} the **bombings**_{e₅} of the U.S. embassies in Kenya and Tanzania and **offered**_{e₁₂} condolences to the victims. [...] In all, the **bombings**_{e₁₀} **last week**_{t₅} **claimed**_{e₄} at least 217 lives.

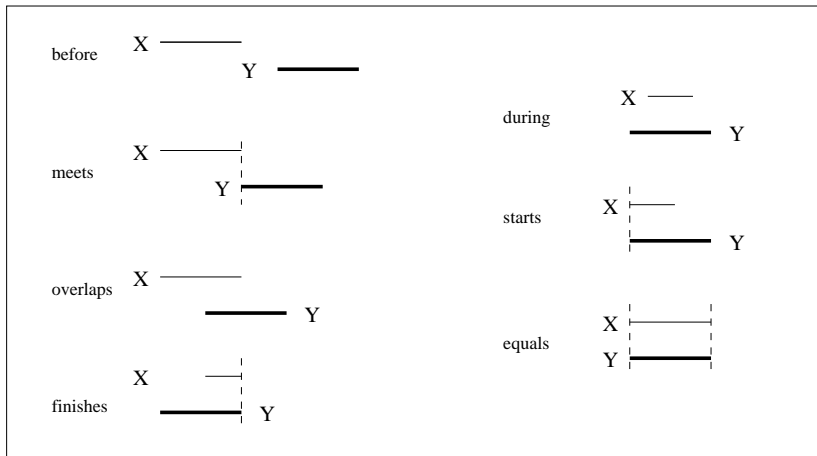


Ordonnement temporel \neq chronologie

President Joseph Estrada on **Tuesday**_{t₄} **condemned**_{e₁} the **bombings**_{e₅} of the U.S. embassies in Kenya and Tanzania and **offered**_{e₁₂} condolences to the victims. [...] In all, the **bombings**_{e₁₀} **last week**_{t₅} **claimed**_{e₄} at least 217 lives.



Relations entre intervalles de temps: Allen



Autres représentations : ISO-TimeML/Tempeval

- ISO-TimeML ∈ norme ISO/TC37/4 (annotations ling.)
- Tempeval: schéma ad hoc des campagnes Semeval (2007, 2010)

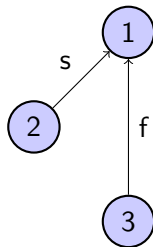
TimeML	Allen	Tempeval
BEFORE IBEFORE	before meet	before
(absent)	overlaps	overlaps
STARTS IS_INCLUDED FINISHES	starts during finishes	
(absent)	overlapsi	
IS_STARTED INCLUDES IS_FINISHED	startsi duringi	
IAFTER AFTER	meeti beforei	after
SIMULTANEOUS	equals	equals

Importance des aspects inférentiels

Les relations temporelles sont logiquement interdépendantes, par ex.

- $before(e_1, e_2), during(e_3, e_2) \models before(e_1, e_3)$
- $finishes(e_1, e_2), overlaps(e_3, e_2) \models overlaps(e_3, e_1) \vee meets(e_3, e_1) \vee before(e_3, e_1)$

et ces propriétés restreignent les ensembles cohérents de relations qui décrivent une situation temporelle décrite par un texte

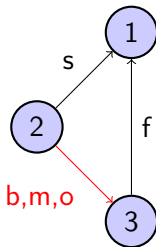


Importance des aspects inférentiels

Les relations temporelles sont logiquement interdépendantes, par ex.

- $before(e_1, e_2), during(e_3, e_2) \models before(e_1, e_3)$
- $finishes(e_1, e_2), overlaps(e_3, e_2) \models overlaps(e_3, e_1) \vee meets(e_3, e_1) \vee before(e_3, e_1)$

et ces propriétés restreignent les ensembles cohérents de relations qui décrivent une situation temporelle décrite par un texte

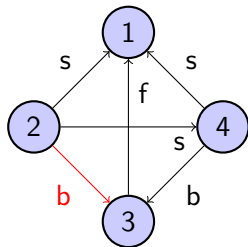


Importance des aspects inférentiels

Les relations temporelles sont logiquement interdépendantes, par ex.

- $before(e_1, e_2), during(e_3, e_2) \models before(e_1, e_3)$
- $finishes(e_1, e_2), overlaps(e_3, e_2) \models overlaps(e_3, e_1) \vee meets(e_3, e_1) \vee before(e_3, e_1)$

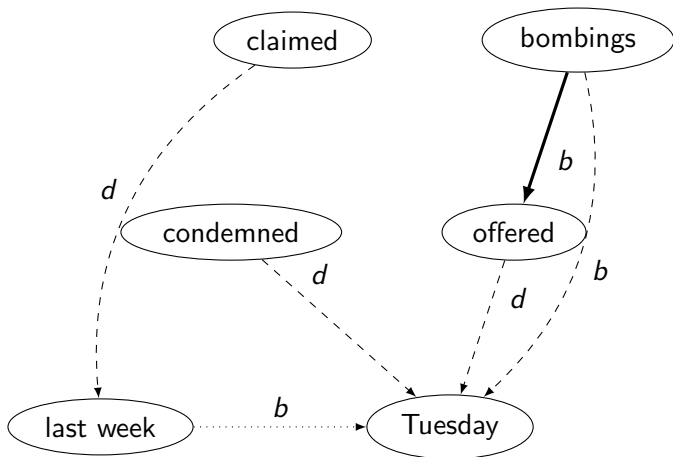
et ces propriétés restreignent les ensembles cohérents de relations qui décrivent une situation temporelle décrite par un texte



Représentation sous forme de graphe

annotation manuelle

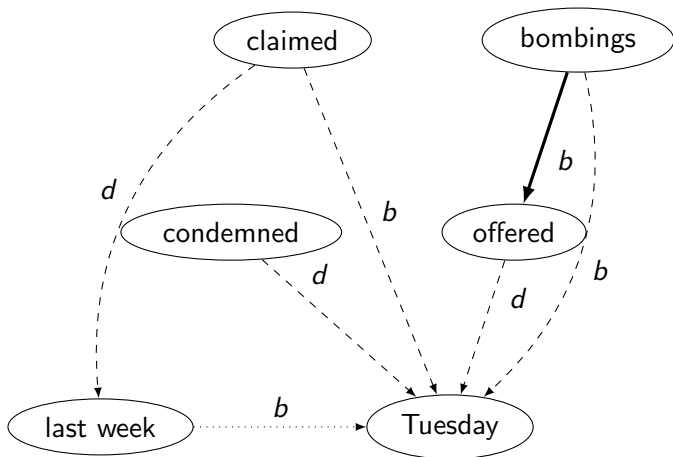
d=during, b=before



Temporal ordering: graph representation

après inférence: "saturation" des contraintes de dépendances entre 3 entités temporelles

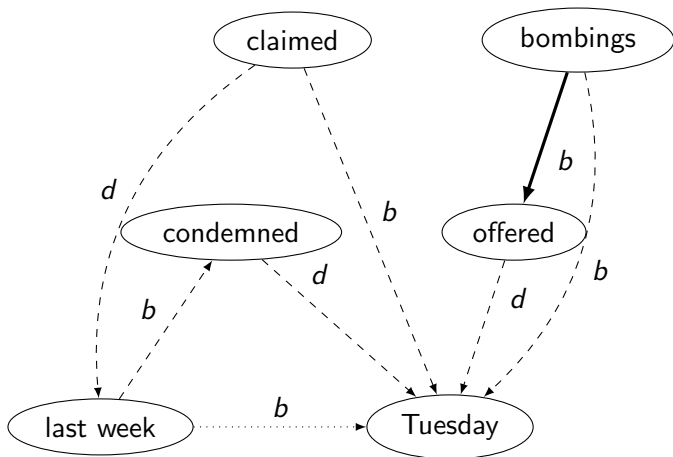
d=during, b=before



Temporal ordering: graph representation

(suite)

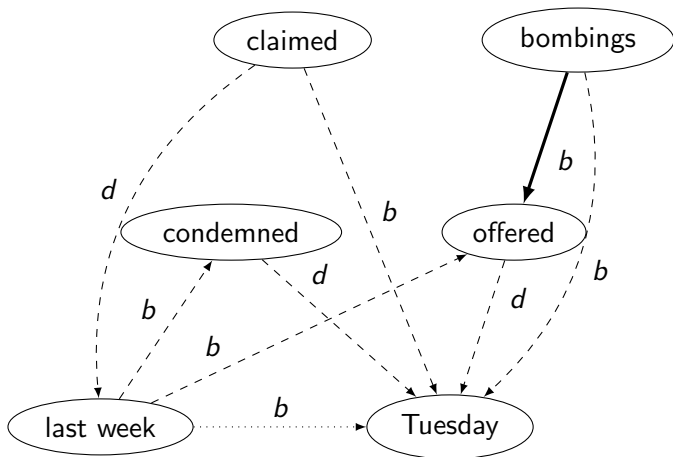
d=during, b=before



Temporal ordering: graph representation

suite

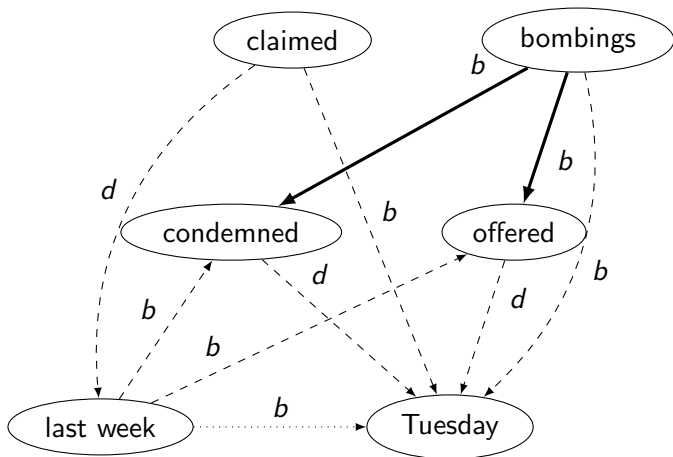
d=during, b=before



Temporal ordering: graph representation

suite

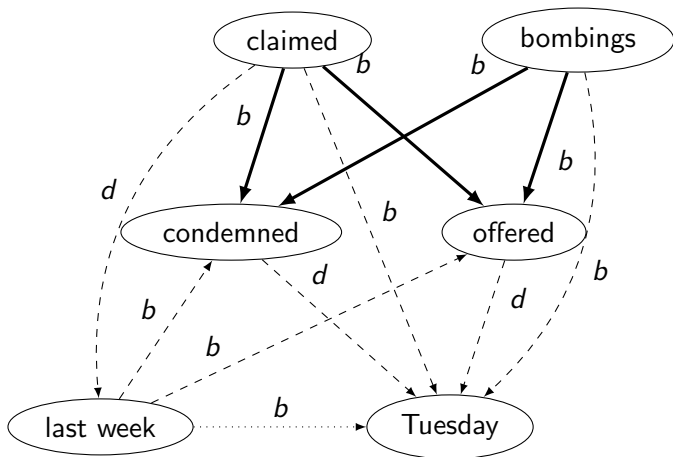
d=during, b=before



Temporal ordering: graph representation

suite

d=during, b=before



- il y a une longue tradition de modélisation du temps, en TAL et IA
- les modèles théoriques en linguistique se sont concentrées sur des cas idéalisés (Kamp Reyle), restreints à quelques phrases
- ces travaux et les premiers modèles en TAL se concentraient sur les sources d'information en jeu pour calculer l'ordre tempore: temps, aspect, lexique, structure discursive, connaissance du monde, pas toujours facilement accessibles
- plus récemment des approches à base d'apprentissage statistique ont été appliquées à des parties du problème, aussi grâce à la création de corpus annotés, par exemple TimeBank associé à la norme ISO-TimeML

Etat de l'art: classification de relation

Mani et al. (2006)

- relations considérées sont les 11 relations de TimeML où bien les 3 de TempEval
- classification très locale (i.e., paire donnée d'événements) plutôt que graphe complet
 - classification des paires présentes dans la référence TimeBank
 - l'ordre des paires est parfois présupposé, réduisant le problème à la prédiction d'une relation parmi 6.
- inférences temporelles seulement utilisées pour augmenter le pool d'exemple d'entraînement
- aucune vérification de la cohérence des décisions entre elles
- Problème: quel est le sens d'un ensemble incohérent de relations ?
de plus, il est difficile d'isoler des sous graphes cohérents.

Classification de relations, suite

TimeBank 1.2 après saturation

Allen	TimeML	Base (28%)	Sat. (37%)
b(efore)	<i>before</i>	785	12053
e(quals)	<i>simul., ident.</i>	1666	2462
d(uring)	<i>during, incl'ed</i>	370	1303
f(inish)	<i>ends</i>	43	82
s(tart)	<i>begins</i>	41	72
m(eet)	<i>ibefore</i>	39	78
o(verlap)	n/a	0	1

la saturation renforce l'importance des relations fréquentes (before, during)

Etat de l'art: forcer la cohérence

Bramsen et al. (2006), Tatu & Srikanth (2008), Chambers & Jurafsky (2008)

Principe général:

- 1 **apprentissage local**: classifieur entraîné sur les paires d'événements de la référence, fournit un score pour chaque paire+relation
- 2 **décodage global**: maximise les préférences locales sous contraintes de cohérence des paires entre elles.

Types d'inférences:

- Approximée/Algo. glouton: meilleur d'abord, NRO (natural reading order), ...
- Exacte: **Integer Linear Programming (ILP)**, Markov Logic Networks (MLN)

Etat de l'art: forcer la cohérence

formulation ILP

n événements, $|R|$ relations différentes

- 0 – 1 variables booléennes représentant (paire d'évt, relation)
⇒ $|R| \times n^2$ variables
- fonction objectif : maximiser les scores locaux des relations possibles pour chaque paire d'évt.
- contraintes: au moins une contrainte pour chaque triplet de variables pour exprimer la composition de relations, par ex. :

$$x_{(b,e_1,e_2)} + x_{(d,e_2,e_3)} \leq x_{(b,e_1,e_3)} + 1$$

⇒ $|R|^2 \times n^3$ contraintes

Etat de l'art: forcer la cohérence

Bramsen et al. (2006), Tatu & Srikanth (2008), Chambers & Jurafsky (2008)

Simplifications importantes dans ces travaux :

- jeu de relations réduit
- restriction de l'ensemble des entités considérées
- restriction aux paires (connues) de la référence
- oracle: fournit les liens evt-dates, liens entre dates

l'inférence ILP est NP-dure en général \rightarrow et devient impraticable avec le jeu complet de relation entre intervalles: $2^{13} = 8192$ disjonctions possibles entre 2 intervalles, même si 82 en pratique

- $2^{13} \times n^2$ (au mieux $82 \times n^2$) variables
- $2^{26} \times n^3$ (au mieux $82^2 \times n^3$) contraintes triangulaires

Contributions principales

Denis & Muller (2011)

- Inférence exacte rendue possible :
 - ① en réexprimant le problème d'optimisation par la prédiction des liens entre bornes temporelles des événements: même information, mais plus simple à optimiser
 - ② en forçant la cohérence sur des sous-parties significatives du graphe temporel complet

- prédiction exacte sur le graphe d'événements, sans présupposer les paires à relier, sans restriction sur les relations

Traduction intervalles/ppoints

$$R \in \text{Allen}, r_i \in \{<, >, =\}$$

$$R(e_1, e_2) \equiv r_1(e_1^-, e_2^-) \wedge r_2(e_1^+, e_2^-) \wedge r_3(e_1^-, e_2^+) \wedge r_4(e_1^+, e_2^+)$$

Conversion interval/pts
Allen ordre/bornes

b	(<, <, <, <)
m	(<, =, <, <)
o	(<, >, <, <)
s	(=, >, <, <)
d	(>, >, <, <)
f	(>, >, <, =)

Composition

o	<		>	>	
<	<	<			<
	<				
>			>	>	>
>			>	>	>
	<		>	>	

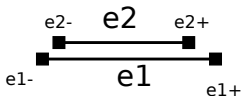
pour n événements

⇒ $5 \times (2n)^2$ rel./points au lieu de $82n^2$ rel. / intervalles

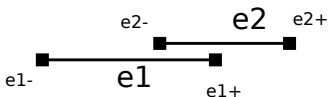
⇒ 17 compositions au lieu de 82^2

Probabilités des relations/points

during
 $p=0.1$



overlaps
 $p=0.2$



before
 $p=0.7$

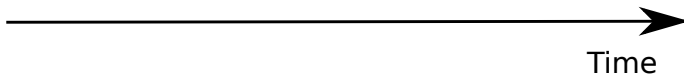


Probabilités des relations/points

$p=0.1$ e1- e2- e2+ e1+


$p=0.2$ e1- e2- e1+ e2+

$p=0.7$ e1- e1+ e2- e2+




Probabilités des relations/points

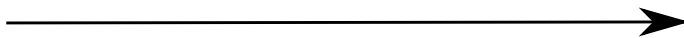

$p=0.1$ e1- e2- e2+ e1+



$p=0.2$ e1- e2- e1+ e2+



$p=0.7$ e1- e1+ e2- e2+



$p=0.7$ e1+ < e2-

$p=0.3$ e2- < e1+

$p=0.$ e2- = e1+

- TimeBank 1.2: 186 documents, $\sim 3000/15.000$ relations
- classifieur de relations à entropie maximum (Megam)
entraîné sur les graphes saturés
- traits standards de la littérature
- décodage ILP avec le solveur SCIP
avec limite de temps: 1 heure
- présuppose les liens evt-dates (E-T) et entre dates (T-T)
comme dans (Chambers & Jurafsky, 2008)
- Evaluation: précision/rappel des relations simples dans la
référence saturée
moyennées par texte

Traits standards

Mani et al. al (2006), Chambers & Jurafsky (2008)

Les traits incluent le prédicat verbal et les attributs TimeML :

- **aspect**: none, prog, perfect, prog perfect
- **class**: report, aspectual, state, l-state l-action, perception, occurrence
- **modality**: none, to, should, would, could can, might
- **polarity**: positive, negative
- **tense**: none, present, past, future

Traits additionnels :

- égalité des attributs des deux éventualités (par ex. même temps verbal)
- distance entre les deux événements (en nb de phrases ou en nb d'événements intercalés).

Experience 1 = état de l'art

- prédire les relations E-E **étant données** les paires de la référence
- étant données les liens T-T et E-T
- forçage de la cohérence globale au niveau de chaque texte
- moyenne des prédictions sur les **paires de la référence** **uniquement** \approx rappel

Expérience 1 = “baselines”

Méthodes de base pour comparaison :

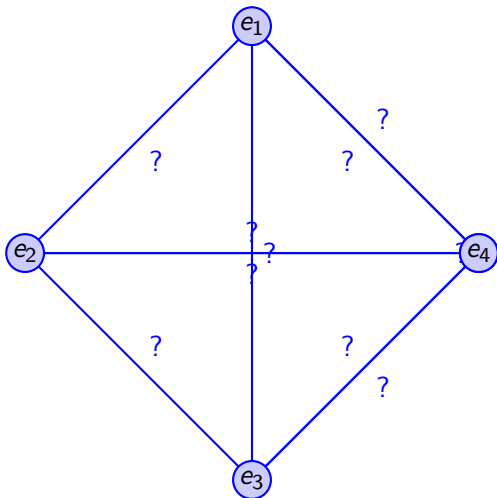
- **before**: prédit toujours la relation majoritaire (forcément cohérent)
- **zéro**: ne prédit rien, mais récupère les E-E inférés à partir des E-T et T-T seuls
- **NRO**: “Natural Reading Order”, relations prédites l’une après l’autre en vérifiant la cohérence à chaque étape

Systeme	F1-score
ILP	49.80
'before'	37.93
'zero'	26.01
natural reading order	20.08

NB: classifieur "local" sans vérification de la cohérence obtient **60.1%** mais produit **82%** de graphes incohérents

Prédire les relations sans connaissances a priori ?

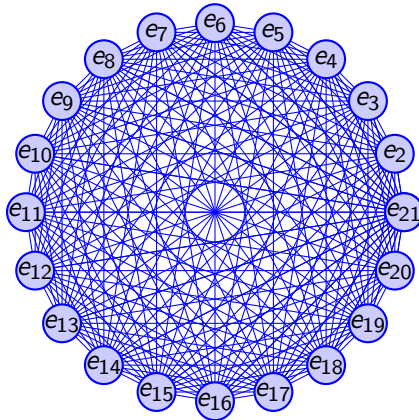
4 événements ? 6 relations potentielles!



Prédire les relations sans connaissances a priori ?

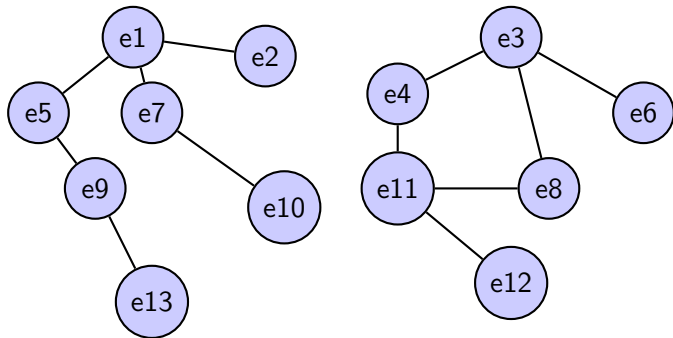
20 événements ? 190 relations potentielles

→ il faut diviser le problème



Décomposition 1

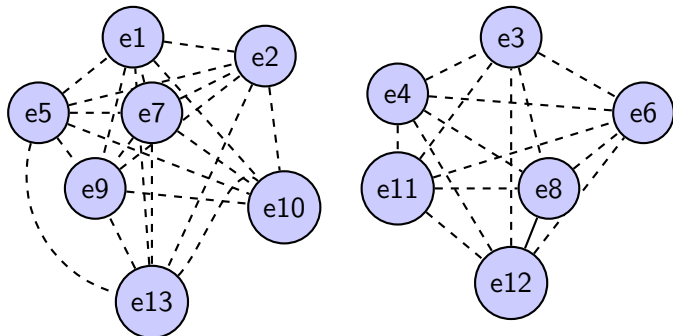
La référence



prédire sur les composantes connexes

Décomposition 1

Prédire sur les composantes connexes de la référence

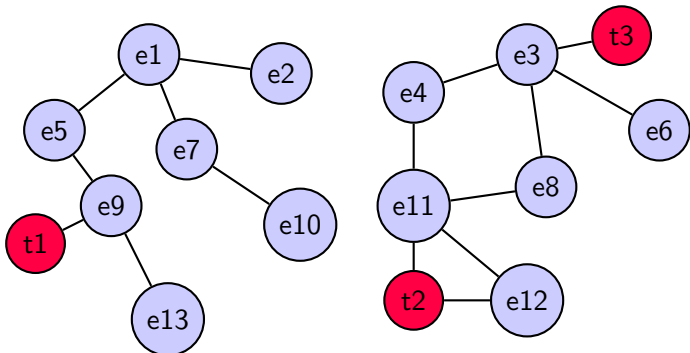


cohérence sur chaque composante → cohérence globale

Décomposition, épisode II

Regroupement "heuristique"

annotation de référence avec les dates :

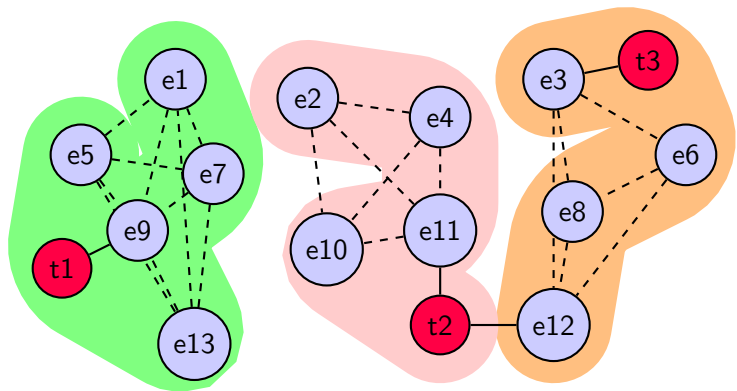


trouver des sous-groupes pertinents pour décomposer les prédictions

Décomposition, épisode II

Regroupement "heuristique"

regroupement des événements autour des dates voisines.



prédictions cohérentes sur chaque composante \nrightarrow cohérence globale

Résultats sans présupposer les paires E-E

la "vraie" tâche, mais sur les composantes connexes

Systeme	Précision	Rappel	F1-score	Incohérents
zero	100.00	26.01	41.28	0.00
ilp	33.02	54.07	41.00	5.93
nro	49.98	17.02	25.40	0.00
base locale	34.14	9.88	15.32	88.49
before	6.22	37.93	10.69	0.00
zero90%	82.01	5.51	10.34	66.19

NB

- la connaissance des liens E-T donne déjà beaucoup d'infos E-E (26%), mais résultat peu robuste : changer 10% des E-T fait chuter le score global. (zero90)
- avec la décomposition heuristique: (nro et ilp) $F1 \approx 15\%$

à emporter :

- il est possible de prédire de façon cohérente toutes les relations de TimeML, sans connaissance a priori des paires à relier

Perspectives sur l'ordonnancement temporel

- inférence jointe des prédictions entre événements, et des liens événements-dates
- utiliser les contraintes aussi pendant l'apprentissage ?
- peut-on apprendre directement les liens entre bornes des événements ? cela a-t-il un sens ?
- améliorer la décomposition ! = effort de modélisation (help wanted)

on a vu surtout :

- des contraintes de formes sur la structure
- des contraintes d'interdépendances entre tâches
- des contraintes "sémantiques" (inférences)

quid de contraintes + linguistiques ?