Contextual Phenomena in NLP: Time, Space and Discourse

Phénomènes contextuels en TAL : temps, espace et discours

Philippe Muller

Université de Toulouse

Jury:

- Isabelle Tellier
- Philippe Blache
- James Pustejovsky
- Violaine Prince
- Claudette Cayrol
- Nicholas Asher

soutenance HDR, 6 mai 2014

Introduction: natural language understanding

- from textual data to formal representation
- **coherence** of a textual object: a document is not just a set of sentences
- interaction between elements : clauses, sentences, document parts
- coherence results from the use of linguistic devices and shared competence, so that implicit information is recovered
  - from linguistic conventions
  - from external knowledge

between Natural Language Processing & Artificial Intelligence

Introduction: research topics and methods

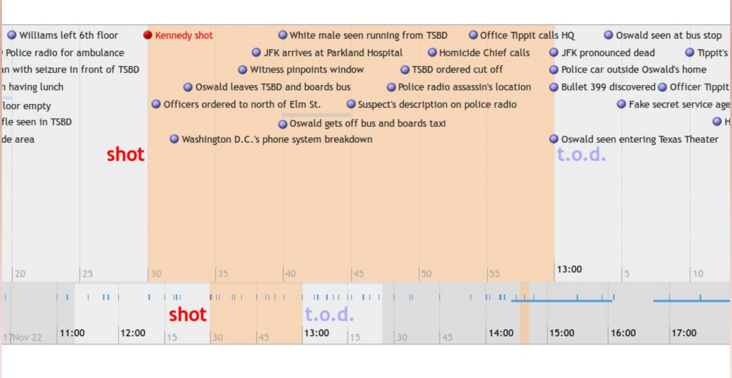
- a partial view of semantic content:
  - time and space localizations
  - structure of a document as a reflection of a writer's intentions and commitments
    and structure of a conversation as a reflection of speakers information exchange
  - lexical semantics as a clue to discourse organisation
- analysing and predicting semantic structures
  - collecting data
  - models of interpretation:
    - cues (inductive classification)
    - processes (decoding with constraints)
  - experimental validation

- coreference:
  George bought a car. **It** *belonged to Jon Voigt.*
- impliciteness:
  George does not want to sell his car. It belonged to Jon Voigt.
  (explanation)
- deictics:
  Last week*, I wasn't* **here***.*
- lexical cohesion:
  I needed some sort of **vehicle***. I bought a* **bike**
- extra linguistic knowledge to resolve ambiguity
  One morning I shot an elephant in my pajamas.

(Groucho Marx)

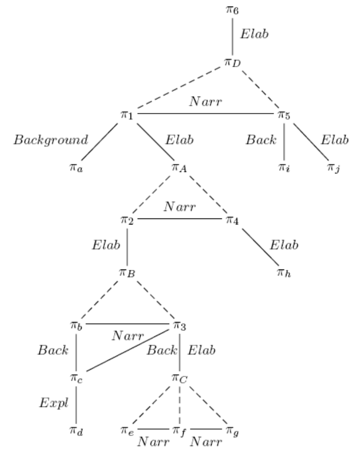
- coreference:
  George bought a car. **It** *belonged to Jon Voigt.*
- impliciteness:
  George does not want to sell his car. It belonged to Jon Voigt.
  (explanation)
- deictics:
  Last week, *I wasn't* **here**.
- lexical cohesion:
  I needed some sort of **vehicle**. *I bought a* **bike**
- extra linguistic knowledge to resolve ambiguity ... or not
  One morning I shot an elephant in my pajamas. How he got into my pajamas I'll never know. (Groucho Marx)

At 12:29 p.m. CST, as President Kennedy's uncovered limousine entered Dealey Plaza, Nellie Connally, then the First Lady of Texas, turned around to President Kennedy, who was sitting behind her, and commented, "Mr. President, you can't say Dallas doesn't love you," which President Kennedy acknowledged by saying "No, you certainly can't." Those were the last words ever spoken by John F. Kennedy. He gave his reply just after the Main-to-Houston Street turn (with photos and films even showing him leaning in towards Mrs. Connally on Houston Street to reply to her).
From Houston Street, the presidential limousine made the planned left turn onto Elm Street, allowing it access to the Stemmons Freeway exit. As it turned on Elm, the motorcade passed the Texas School Book Depository. Shots were fired at President Kennedy as they continued down Elm Street. About 80% of the witnesses recalled hearing three shots.
A minority of the witnesses recognized the first gunshot blast they heard as a weapon blast, but there was hardly any reaction to the first shot from a majority of the people in the crowd or those riding in the motorcade. Many later said they heard what they first thought to be a firecracker, or the exhaust backfire of a vehicle, just after the President started waving.



Example: discourse structure

At 12:29 p.m. CST, as President Kennedy's uncovered limousine entered Dealey Plaza, Nellie Connally, then the First Lady of Texas, turned around to President Kennedy, who was sitting behind her, and commented, "Mr. President, you can't say Dallas doesn't love you," which President Kennedy acknowledged by saying "No, you certainly can't." Those were the last words ever spoken by John F. Kennedy. He gave his reply just after the Main-to-Houston Street turn (with photos and films even showing him leaning in towards Mrs. Connally on Houston Street to reply to her).
From Houston Street, the presidential limousine made the planned left turn onto Elm Street, allowing it access to the Stemmons Freeway exit. As it turned on Elm, the motorcade passed the Texas School Book Depository. Shots were fired at President Kennedy as they continued down Elm Street. About 80% of the witnesses recalled hearing three shots.
A minority of the witnesses recognized the first gunshot blast they heard as a weapon blast, but there was hardly any reaction to the first shot from a majority of the people in the crowd or those riding in the motorcade. Many later said they heard what they first thought to be a firecracker, or the exhaust backfire of a vehicle, just after the President started waving.



Example: Temporal structure

Temporal situation

Philippine President Joseph Estrada on Tuesday$_{t_4}$ condemned$_{e_1}$ the bombings$_{e_5}$ of the U.S. embassies in Kenya and Tanzania

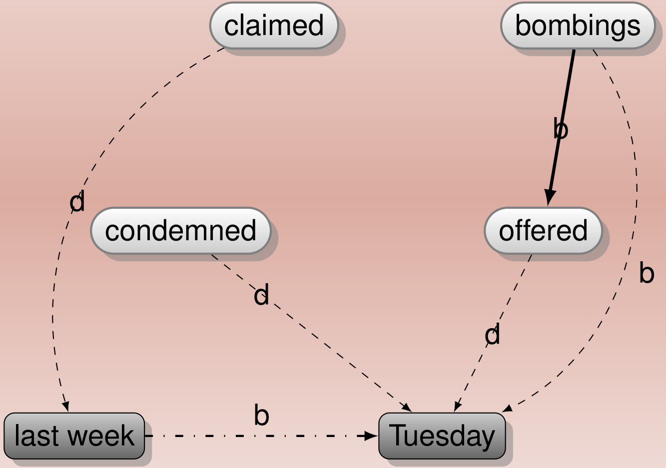
and offered$_{e_{12}}$ condolences to the victims. [...]

In all, the bombings$_{e_{10}}$ last week$_{t_5}$ claimed$_{e_4}$ at least 217 lives.

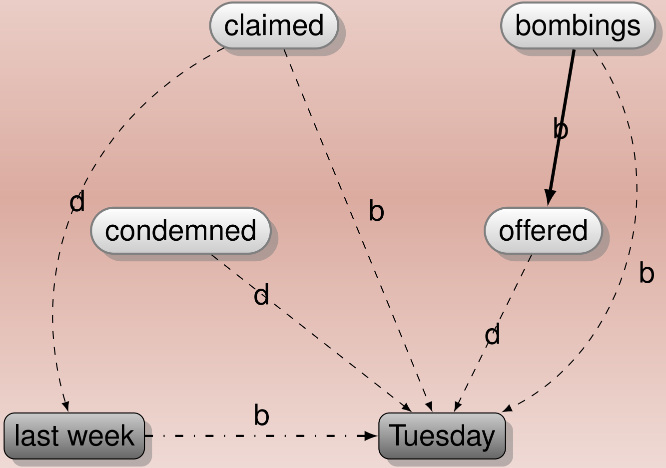
- Sentence-level information : condemn / on Tuesday, ...
- Inter-sentential information: “bombings” coreference, ...
- Contextual information: date of publication / “last week”, Tuesday
- Inference: knowledge of temporal relations
  (x before y) and (z during y) implies (x before z)

(slightly simplified from Aquaint/TimeBank)

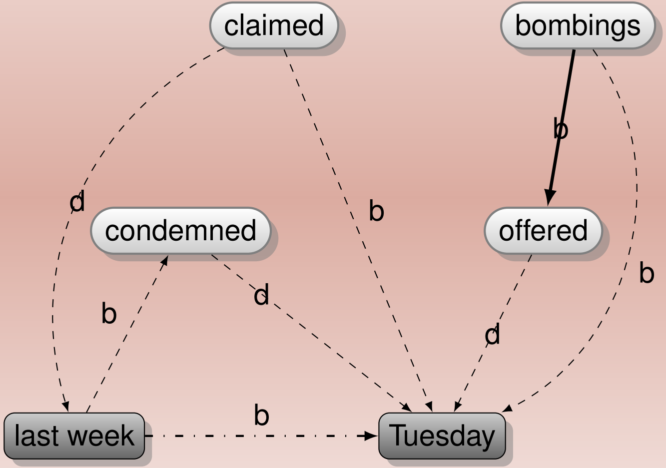
d=during, b=before



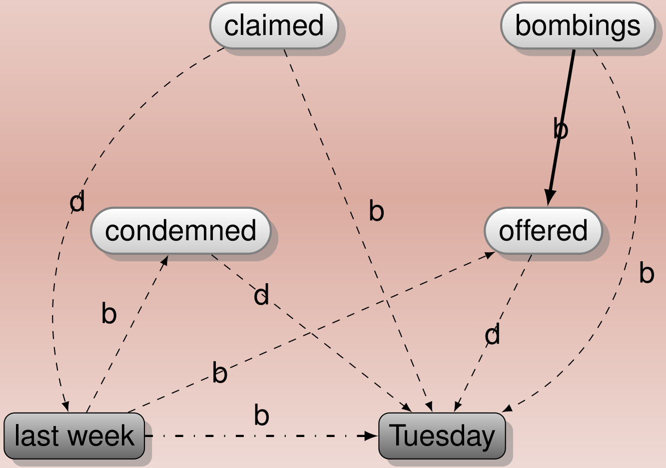
d=during, b=before



d=during, b=before

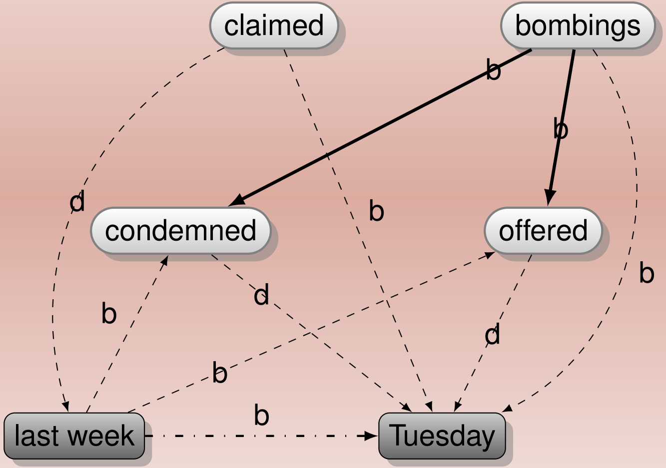


d=during, b=before

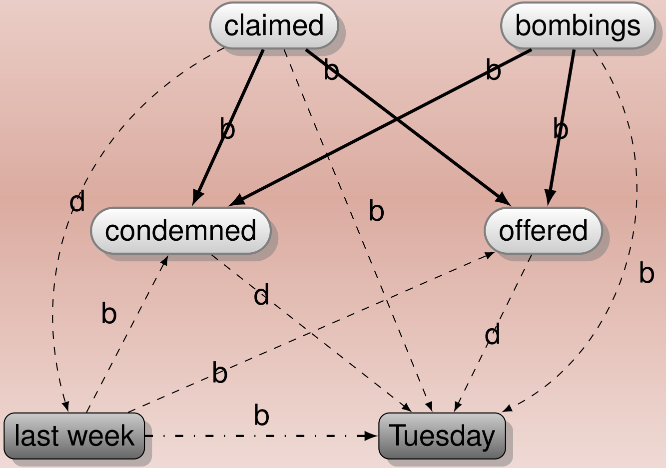


Reference annotation

d=during, b=before



d=during, b=before



Standard NLP methodology

- reference data = human annotated data
    - annotation scheme
    - reproducible annotation
- inductive methods : design a model based on a training dataset
    - Machine learning methods: classification, sequence learning, tree learning, graph learning
    - for structure: restrict the hypothesis space
- evaluate on separate test data set

for semantic / pragmatic problems ?

- reference data = human annotated data
  - annotation scheme semantics of scheme !
  - reproducible annotation difficult for human (even experts)
- inductive methods : design a model based on a training dataset
  - Machine learning methods: classification, sequence learning, tree learning, graph learning
    usually not enough data
  - for structure: restrict the hypothesis space
    usually not enough data
    + constraints not well known
- evaluate on separate test data set
  not just identifying labels, but comparing structure,
  and respecting contraints (eg coherence)

Issues for time

representation of temporal relations : which relations ? what inferences do they allow for ?

how to extract them ?

how to use inferential properties for structure prediction ?

how to evaluate success with respect to reference annotation ?

what is coherence ?

Personal contributions

representation of temporal relations : which relations ? what inferences do they allow for ?
comparison of various representation schemes: (Coling 2010)

how to extract them ?

how to use inferential properties for structure prediction ?

how to evaluate success with respect to reference annotation ?

what is coherence ?

Personal contributions

representation of temporal relations : which relations ? what inferences do they allow for ?
comparison of various representation schemes: (Coling 2010)

how to extract them ?
result on French (Coling 2004a), English (IJCAI 2011)

how to use inferential properties for structure prediction ?

how to evaluate success with respect to reference annotation ?

what is coherence ?

Personal contributions

representation of temporal relations : which relations ? what inferences do they allow for ?
comparison of various representation schemes: (Coling 2010)

how to extract them ?
result on French (Coling 2004a), English (IJCAI 2011)

how to use inferential properties for structure prediction ?
constrained decoding (IJCAI 2011)

how to evaluate success with respect to reference annotation ?

what is coherence ?

Personal contributions

representation of temporal relations : which relations ? what inferences do they allow for ?
comparison of various representation schemes: (Coling 2010)

how to extract them ?
result on French (Coling 2004a), English (IJCAI 2011)

how to use inferential properties for structure prediction ?
constrained decoding (IJCAI 2011)

how to evaluate success with respect to reference annotation ?
non biased measures (LREC 2008, JAIR 2011)

what is coherence ?

joint work with Pascal Denis (IJCAI, Coling 2010), Xavier Tannier (JAIR, Coling 2004a, LREC), Michel Gagnon, Gabriel Parent

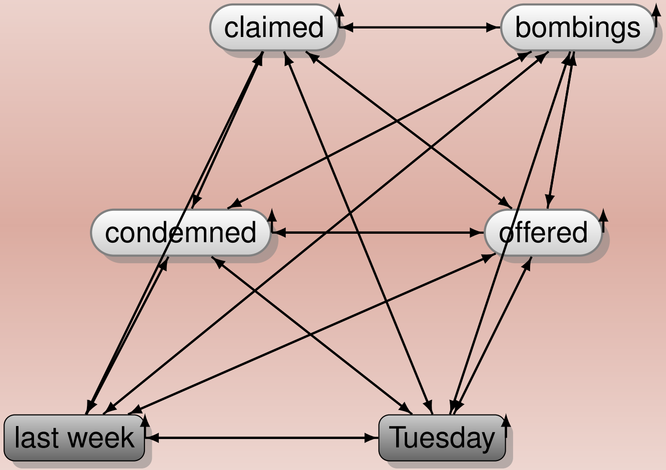
Structure predictions with constraints: temporal case

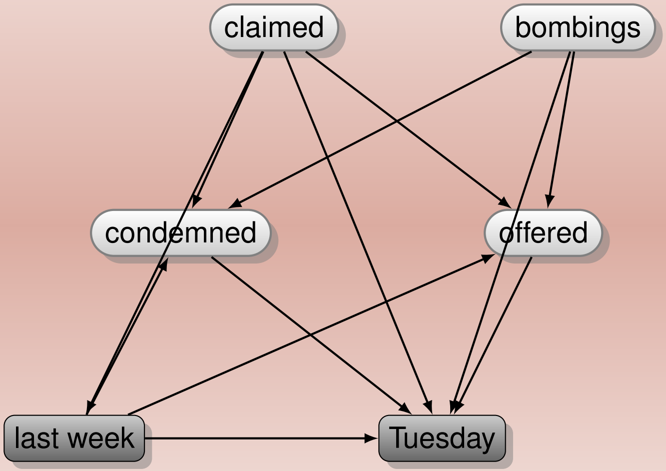
- hard constraints : derive from semantic of relations
- definition of coherence: composition of relations and non contradiction in “closure” of temporal relations wrt composition (Allen algebra)

Procedure:

- local machine learning model for event/event relations
- structure prediction = best coherent subset of relations
  $\rightarrow$ optimization problem (solved with Integer Linear Programming)

(IJCAI 2011): beat existing coherence-preserving methods, do not need to assume the related event pairs


claimed
bombings
condemned
offered
last week
Tuesday

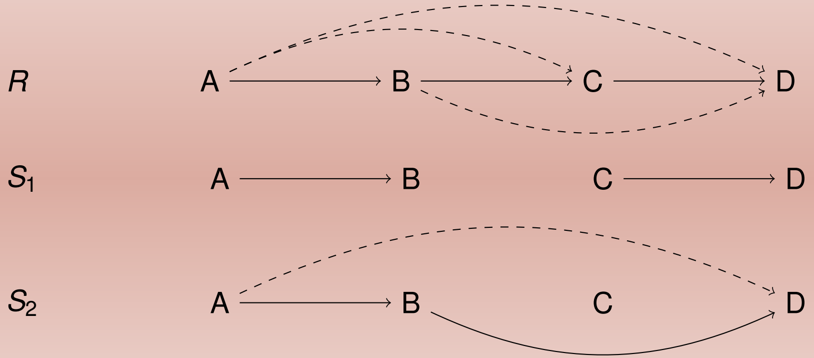

claimed
bombings
condemned
offered
last week
Tuesday

The problem of evaluating semantic structures

- mismatch annotation syntax / semantics (implicit information) solved with closure ?
- closure introduces an even greater bias towards majority relations
- instability of evaluations

$$
\begin{tikzcd}
R & A \arrow[r] & B \arrow[r] & C \arrow[r] & D \\
S_1 & A \arrow[r] & B & C \arrow[r] & D \\
S_2 & A \arrow[r] & B \arrow[rr, bend right] & C & D
\end{tikzcd}
$$

Inference and its bias



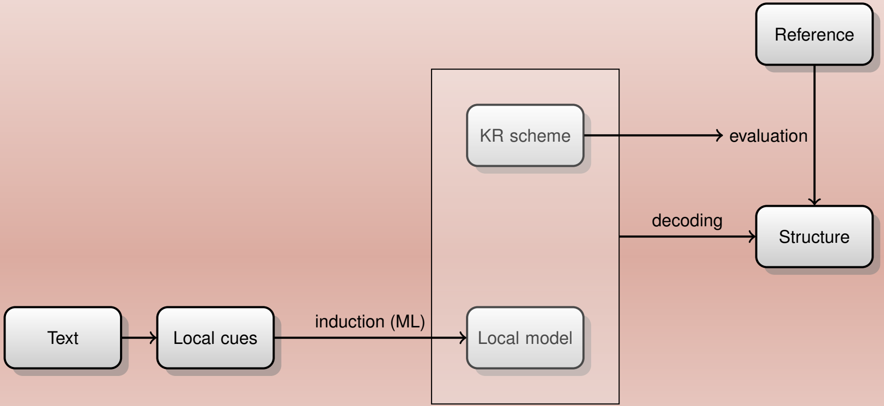
Solution and evaluation methodology

find a stable subset of the information

- that can generate the closure
- that is minimal
- so that information should be related to number of events ($\approx$ human annotation)
- experiment: degrade the reference (remove information) and compare to it

(JAIR 2011)
idea can be extended to semantic structures with equivalences

Temporary summary: semantic structure prediction



– KR scheme: test coherence and enrich representation
– decoding : optimise prediction under constraints
– evaluation: comparison also uses domain knowledge

similar issues

- representation : localization relations, motion
- annotation (ISO)
- extraction
- global structure / interaction with discourse/dialogue
- inference: role less clear for NLP

Space

similar issues

- representation : localization relations, motion
  KR 98, Comp. Intell. 2002
- annotation (ISO)
- extraction
  Itipy Project
- global structure / interaction with discourse/dialogue
  Laurent Prévot's Ph.D.
- inference: role less clear for NLP

Standard NLP methodology

- reference data = human annotated data
  - annotation scheme
  - reproducible annotation
- inductive methods : design a model based on a training dataset
  - Machine learning methods: classification, sequence learning, tree learning, graph learning
  - for structure: restrict the hypothesis space
- evaluate on separate test data set

Standard NLP methodology

for semantic / pragmatic problems ?

- reference data = human annotated data
  - annotation scheme semantics of scheme !
  - reproducible annotation difficult for human (even experts)
- inductive methods : design a model based on a training dataset
  - Machine learning methods: classification, sequence learning, tree learning, graph learning
    usually not enough data
  - for structure: restrict the hypothesis space
    usually not enough data
    + constraints not well known
- evaluate on separate test data set
  not just identifying labels, but comparing structure,
  and respecting contraints (eg coherence)

- representation of discourse structure :
  - what discourse units ? how to find them ?
  - which relations ? what semantics ? how to extract them ?
- how to use discourse properties and constraints for structure prediction ?
- how to evaluate success with respect to reference annotation ?

Discourse parsing: segmentation

Jerry had a wonderful evening yesterday. He had a fantastic meal. He had a soup. He ate lobster. He saw a great movie.

Jerry had a wonderful evening yesterday.

He had a fantastic meal.

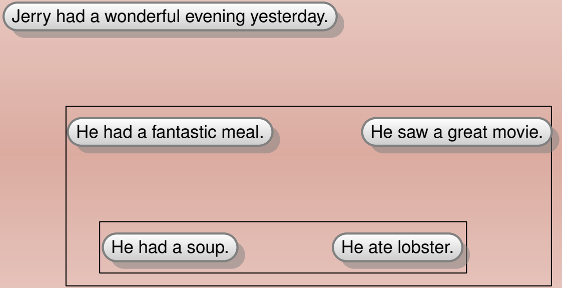
He saw a great movie.

He had a soup.

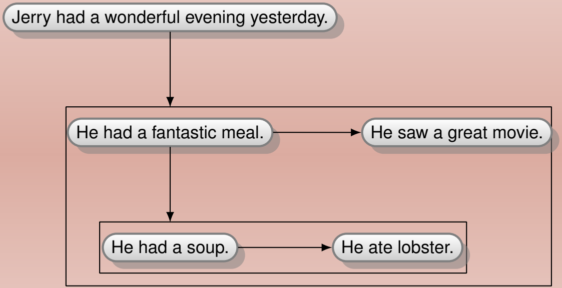
He ate lobster.

Discourse parsing: complex discourse units

Jerry had a wonderful evening yesterday. He had a fantastic meal. He had a soup. He ate lobster. He saw a great movie.

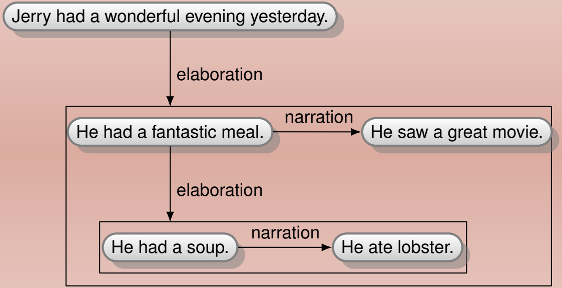


Jerry had a wonderful evening yesterday. He had a fantastic meal. He had a soup. He ate lobster. He saw a great movie.



Discourse parsing: labelling

Jerry had a wonderful evening yesterday. He had a fantastic meal. He had a soup. He ate lobster. He saw a great movie.



representation of discourse structure : which relations ? what semantics ?

how to find discourse units ?

how to extract discourse links ?

how to use properties for structure prediction ?

how to evaluate success with respect to reference annotation ?

representation of discourse structure : which relations ? what semantics ?
ANR annodis: annotated corpus
comparison of expressivity of theories (Sigdial 2013)
(Afantenos, Venant, Denis, Asher)

how to find discourse units ?

how to extract discourse links ?

how to use properties for structure prediction ?

how to evaluate success with respect to reference annotation ?

Contributions

representation of discourse structure : which relations ? what semantics ?
ANR annodis: annotated corpus
comparison of expressivity of theories (Sigdial 2013)
(Afantenos, Venant, Denis, Asher)

how to find discourse units ?
segmentation (Lrec 2010), topic segmentation (Taln 2010)
(Afantenos, Denis) (Adam, Fabre)

how to extract discourse links ?

how to use properties for structure prediction ?

how to evaluate success with respect to reference annotation ?

Contributions

representation of discourse structure : which relations ? what semantics ?
ANR annodis: annotated corpus
comparison of expressivity of theories (Sigdial 2013)
(Afantenos, Venant, Denis, Asher)

how to find discourse units ?
segmentation (Lrec 2010), topic segmentation (Taln 2010)
(Afantenos, Denis) (Adam, Fabre)

how to extract discourse links ?
markers (Charlotte Roze PhD), lexical cohesion (Clementine Adam PhD), lexical regularities (Juliette Conrath PhD)

how to use properties for structure prediction ?

how to evaluate success with respect to reference annotation ?

representation of discourse structure : which relations ? what semantics ?
ANR annodis: annotated corpus
comparison of expressivity of theories (Sigdial 2013)
(Afantenos, Venant, Denis, Asher)

how to find discourse units ?
segmentation (Lrec 2010), topic segmentation (Taln 2010)
(Afantenos, Denis) (Adam, Fabre)

how to extract discourse links ?
markers (Charlotte Roze PhD), lexical cohesion (Clementine Adam PhD), lexical regularities (Juliette Conrath PhD)

how to use properties for structure prediction ?
structure decoding with constraints : (Coling 2012)
(Afantenos, Denis, Asher)

how to evaluate success with respect to reference annotation ?

Contributions

representation of discourse structure : which relations ? what semantics ?
ANR annodis: annotated corpus
comparison of expressivity of theories (Sigdial 2013)
(Afantenos, Venant, Denis, Asher)

how to find discourse units ?
segmentation (Lrec 2010), topic segmentation (Taln 2010)
(Afantenos, Denis) (Adam, Fabre)

how to extract discourse links ?
markers (Charlotte Roze PhD), lexical cohesion (Clementine Adam PhD), lexical regularities (Juliette Conrath PhD)

how to use properties for structure prediction ?
structure decoding with constraints : (Coling 2012)
(Afantenos, Denis, Asher)

how to evaluate success with respect to reference annotation ?
inferential equivalent for discourse relations (Charlotte Roze PhD)

Constraints for prediction

Experimenting on the Annodis French discourse corpus

- Machine learning on standard cues (markers, position, intra-phrastic relations,...)
  - **local** model for attachment
  - **local** model for relation labelling given attachment
- joint decoding as optimal subset of the complete graph with edge probability combining local models, with various constraints
  - force a tree (MST decoding)
  - force an incremental tree w or w/ right frontier constraint (custom A* decoding)

A* here for ease of customisation in incremental construction.
(Coling 2012): decoding beats existing greedy or heuristic decodings on Annodis corpus.

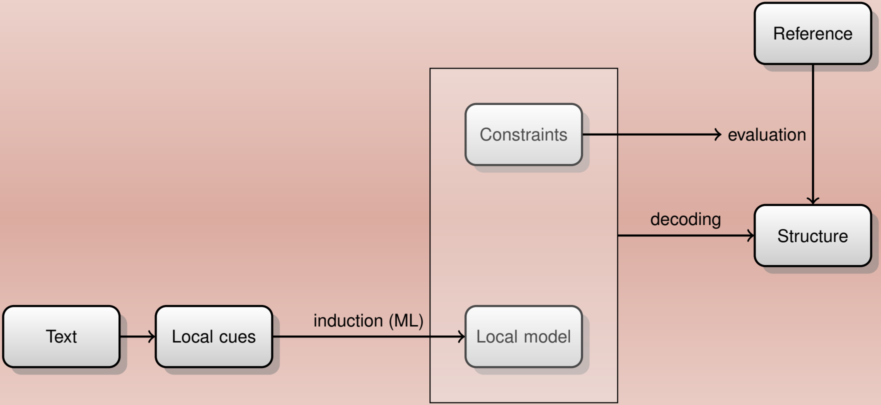
Discourse Evaluation Issues

some problems similar to the temporal case

- some annotator disagreement on seemingly equivalent structures
  - confusion to semantic proximity between relations
  - different forms because of equivalent formulations ?
- explore the idea of **inference on relations**
- composition of relations' semantic consequences at least locally
- protocol to investigate potential inferences from introspective work and corpus exploration

(Charlotte Roze PhD, article in prep.)

Discours parsing summary



Constraints: semantic constraints, interpretation constraints

Dialogue

- dialogue as another kind of structured documents
- focus on interaction modelling
  - information sharing (common ground, agreement structures)
  - negociation
  - argumentation
- mostly linguistic analysis and formalisation (Laurent Prevot PhD) (Semdial 2003, 2005; TAL 2002; 2 book chapters)
- prediction of structures w.i.p. within the ERC Stac Project

Lexical semantics

two dimensions of discourse analysis

- rhetorical : structural coherence (eg markers, organisation)
- topical : lexical coherence (continuity and shift)

semantic link between lexical items in coherent texts for

- high level topical segmentation
- local interpretation (discourse relations)

within this : semantic relatedness

Le gorille est après le bonobo et le chimpanzé , du point de vue génétique , l' animal le plus proche de l' humain .

Cette parenté a été confirmée par les similitudes entre les chromosomes et les groupes sanguins . Notre génome ne diffère que de 2 % de celui du gorille .

Redressés , les gorilles atteignent une taille de 1,75 mètre , mais ils sont en fait un peu plus grands car ils ont les genoux fléchis . L' envergure des bras dépasse la longueur du corps et peut atteindre 2,75 mètres .
Il existe une grande différence de masse entre les sexes : les femelles pèsent de 90 à 150 kilogrammes et les mâles jusqu' à 275. En captivité , particulièrement bien nourris , ils atteignent 350 kilogrammes .

Le pelage dépend du sexe et de l' âge . Chez les mâles les plus âgés se développe sur le dos une fourrure gris argenté , d' où leur nom de "dos argentés" . Le pelage des gorilles de montagne est particulièrement long et soyeux .

Comme tous les anthropoïdes , les gorilles sont dépourvus de queue . Leur anatomie est puissante , le visage et les oreilles sont glabres et ils présentent des torus supra-orbitaires marqués .

- cohesion: lexical items in text are semantically related

Le gorille est après le bonobo et le chimpanzé , du point de vue génétique , l' animal le plus proche de l' humain .

Cette parenté a été confirmée par les similitudes entre les chromosomes et les groupes sanguins . Notre génome ne diffère que de 2 % de celui du gorille .

Redressés , les gorilles atteignent une taille de 1,75 mètre , mais ils sont en fait un peu plus grands car ils ont les genoux fléchis . L' envergure des bras dépasse la longueur du corps et peut atteindre 2,75 mètres .
Il existe une grande différence de masse entre les sexes : les femelles pèsent de 90 à 150 kilogrammes et les mâles jusqu' à 275. En captivité , particulièrement bien nourris , ils atteignent 350 kilogrammes .

Le pelage dépend du sexe et de l' âge . Chez les mâles les plus âgés se développe sur le dos une fourrure gris argenté , d' où leur nom de “dos argentés” . Le pelage des gorilles de montagne est particulièrement long et soyeux .

Comme tous les anthropoïdes , les gorilles sont dépourvus de queue . Leur anatomie est puissante , le visage et les oreilles sont glabres et ils présentent des torus supra-orbitaires marqués .

- cohesion: lexical items in text are semantically related
- semantically related items reflect topical structure

Contributions

use of distributional semantic relatedness (resource by CLLE-ERSS)

1. topic segmentation
2. cohesion filtering in context

Clementine Adam PhD: TALN 2010, TAL 2013, ACL 2014

(C. Fabre)

study of different semantic relatedness:

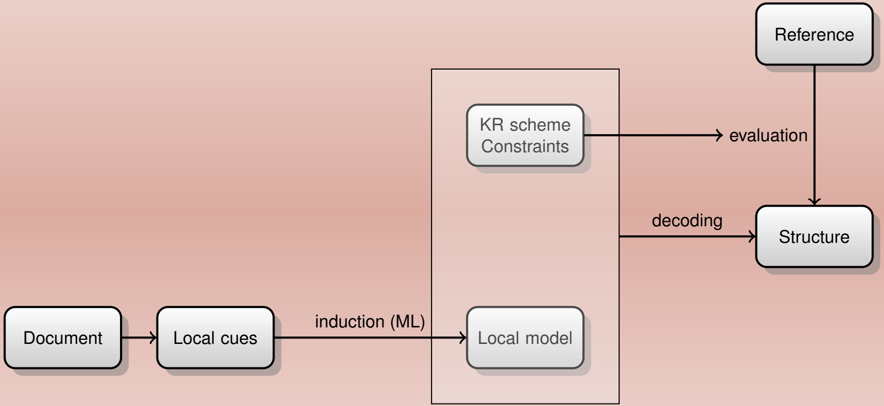
- in dictionaries (Coling 2004b, Textgraph 2006) (B. Gaume, N. Hathout)
- from multilingual corpora (CAI 2011) (P. Langlais)

unsupervised lexical relation extraction for discourse analysis

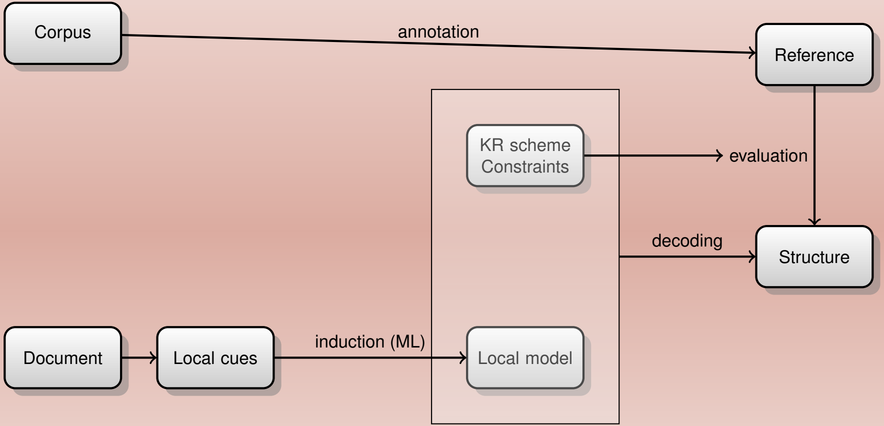
The witness **demonstrated** his good faith during the cross-examination.
The jury was **convinced**.

w.i.p. Juliette Conrath's Ph.D (S. Afantenos, N. Asher)

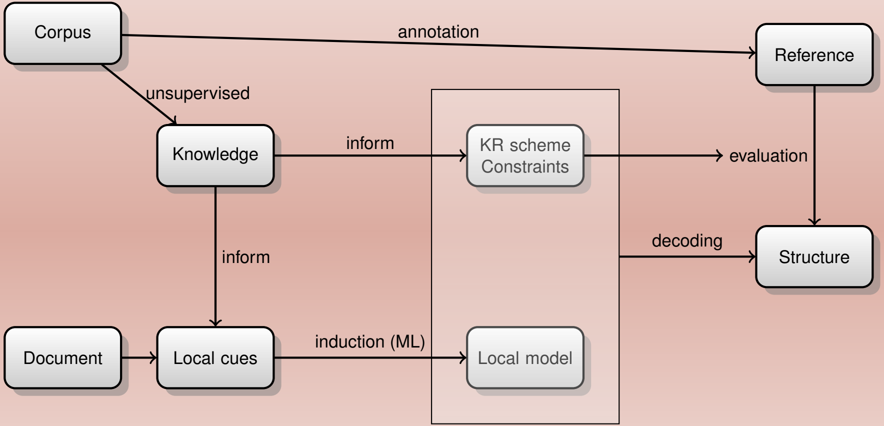
Synthesis: pragmatics and NLP



Synthesis: pragmatics and NLP



Synthesis: pragmatics and NLP



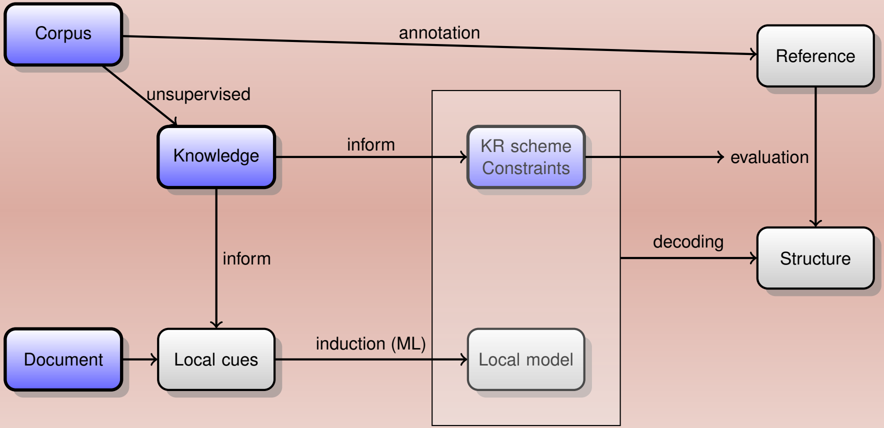
Perspectives

more general “collaboration” between Knowledge Representation and Machine Learning for semantic NLP

- to manage data sparseness, and to improve current models
- leverage unsupervised approaches

go beyond the textual document in isolation,

- exploit explicit structures in recent communication forms: forums, microblogging
- take into account the dynamics of communication: correspondances, on-going discussions


Corpus
annotation
Reference
unsupervised
Knowledge
inform
KR scheme
Constraints
evaluation
inform
decoding
Structure
Document
Local cues
induction (ML)
Local model