Some observations on annotation

Nicholas Asher

Melodi/axe discours

Utrecht, June 2017

Why do annotations ?

- A discovery process for linguistically interesting phenomena, especially beyond the sentence.
- Building an annotation model encodes linguistic generalizations and then tests them, useful for linguists.
- Testing on real data necessary as there are too many variables to control for a full semantics beyond the sentence.
- also means there are compromises...
- a dialectical process in annotation : devise a model, annotate, revise the model, annotate, ...
- the model gets better !

Some things you want to know about your model

- Is the structure or semantics we develop in the annotation model replicable ?
- How complex are the features on which the decisions about structure are made ?
  e.g., are the features easily recoverable from the raw data, or do they require complex processing ?
- Test this with formal machine learning methods and computable features

Discourse, a case in point

- the interpretation of a text is dynamic and depends on context.
- example : *John fell. Max pushed him.* (Lascarides & Asher 1993)
- each discourse constituent has one or more rhetorical functions relative to other DUs
- several DUs can have the same rhetorical function. Recursive Structure and CDUs
- formalize annotations of discourse structure as directed acyclic graphs.
- How does this picture carry over to dialogue ?

Formalization : a reminder

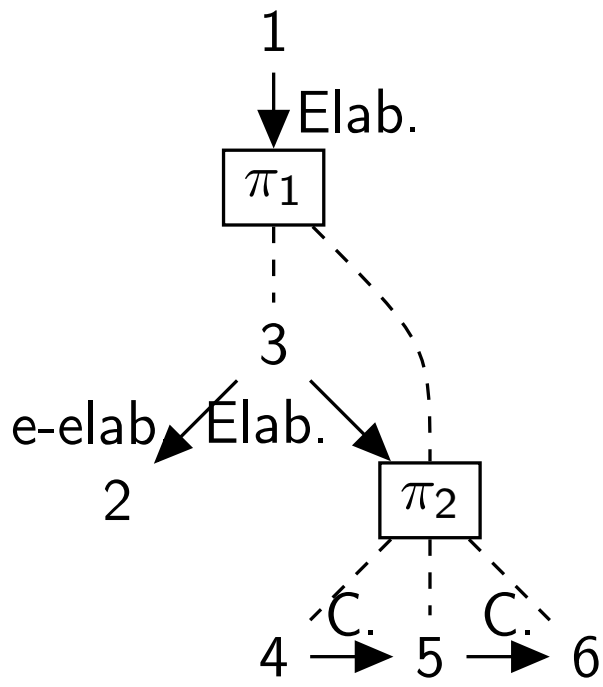
A directed acyclic graph :

$$(V, E_1, E_2, \ell)$$

- V a set of discourse constituents,
- $E_1 \cup E_2 \subseteq V \times V$,
- $\ell : E_1 \rightarrow$ Relation-Types a labelling.

An example from Monologue

[Principes de la sélection naturelle.]_1 [La théorie de la sélection naturelle [telle qu'elle a été initialement décrite par Charles Darwin,]_2 repose sur trois principes :]_3 [1. le principe de variation]_4 [2. le principe d'adaptation]_5 [3. le principe d'hérédité]_6

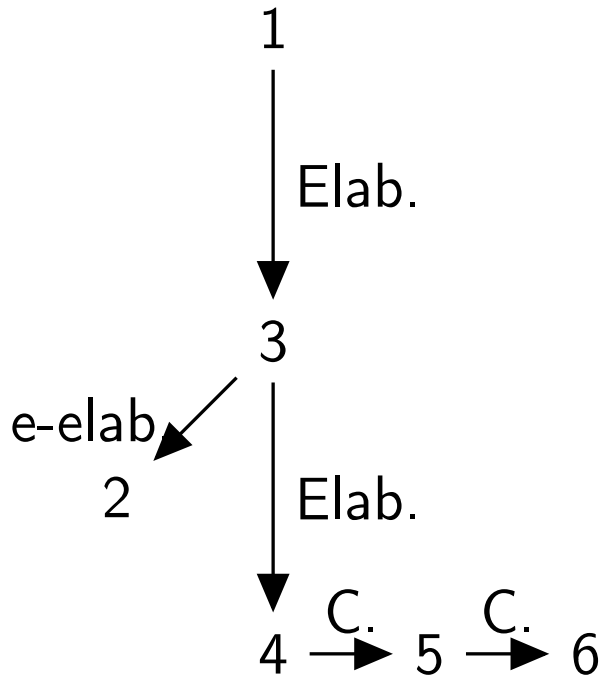


some observations

- notice the nested or recursive structure
- this affects content, in particular temporal relations but also anaphora

Simplifying to structures without CDUs

A graph without E_2 edges and only EDUs as nodes.



Dialogue Structure

Conversation as a game of message exchange involving a kind of signaling game :

- X plays ϕ
- Y decodes a message in strategic equilibrium (safety, credibility)
- Y decides what signal to send in return
- X decodes a message.
- ...

Choice of corpus

- cooperativity at various levels
- strategic goals behind the conversation
- multi-party
- we wanted to look at structural properties of such conversations.

The *Settlers* corpus

A Do you have rock ?

B I've got lots of wheat

[in fact, B has a rock]

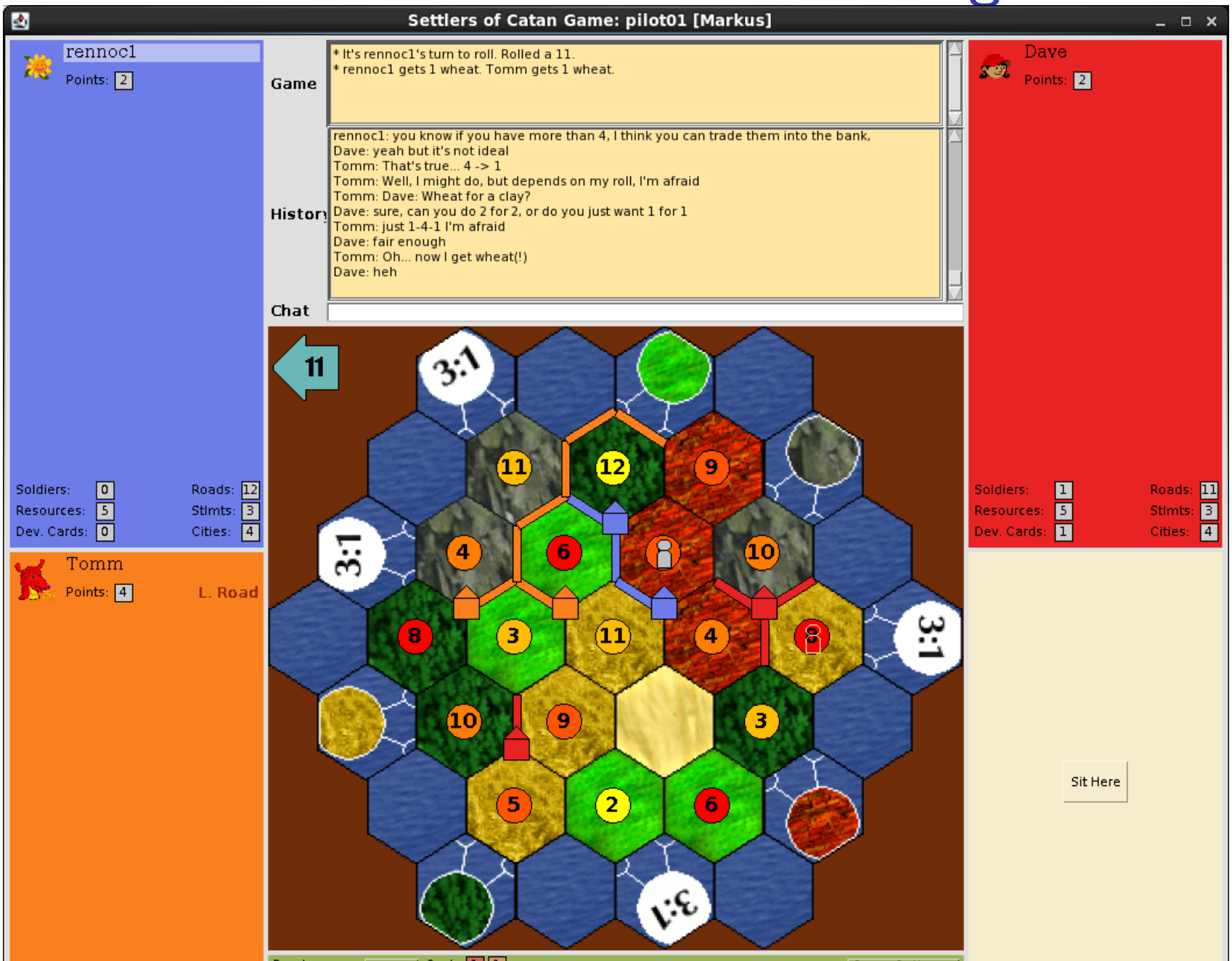
A I'll give you 2 clay for a rock

B How about 2 clay for a wheat ?

A I'll give 1 clay for 3 wheat

B OK, it's a deal.

The game

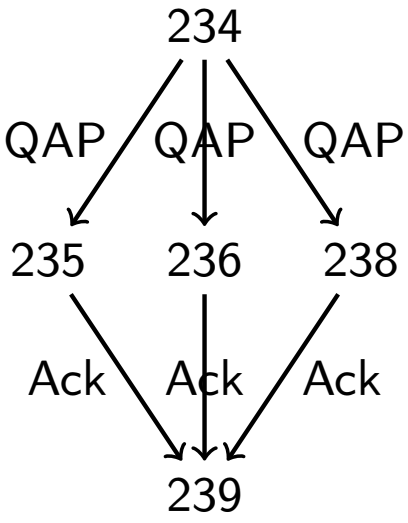

Settlers of Catan Game: pilot01 [Markus]
rennoc1
Points: 2
Soldiers: 0
Roads: 12
Resources: 5
Stlmts: 3
Dev. Cards: 0
Cities: 4
Tomm
Points: 4
L. Road
Game
* It's rennoc1's turn to roll. Rolled a 11.
* rennoc1 gets 1 wheat. Tomm gets 1 wheat.
History
rennoc1: you know if you have more than 4, I think you can trade them into the bank,
Dave: yeah but it's not ideal
Tomm: That's true... 4 -> 1
Tomm: Well, I might do, but depends on my roll, I'm afraid
Tomm: Dave: Wheat for a clay?
Dave: sure, can you do 2 for 2, or do you just want 1 for 1
Tomm: just 1-4-1 I'm afraid
Dave: fair enough
Tomm: Oh... now I get wheat(!)
Dave: heh
Chat
11
3:1
3:1
3:1
3:1
Dave
Points: 2
Soldiers: 1
Roads: 11
Resources: 5
Stlmts: 3
Dev. Cards: 1
Cities: 4
Sit Here

Some specificities of the Settlers corpus

- conversation involving strategic reasoning, multi party, clearly opposing general goals
- also allowed us to study situated communication, because the visual environment of the game constituted a controlled nonlinguistic, virtual environment.

Multi party dialogue structures

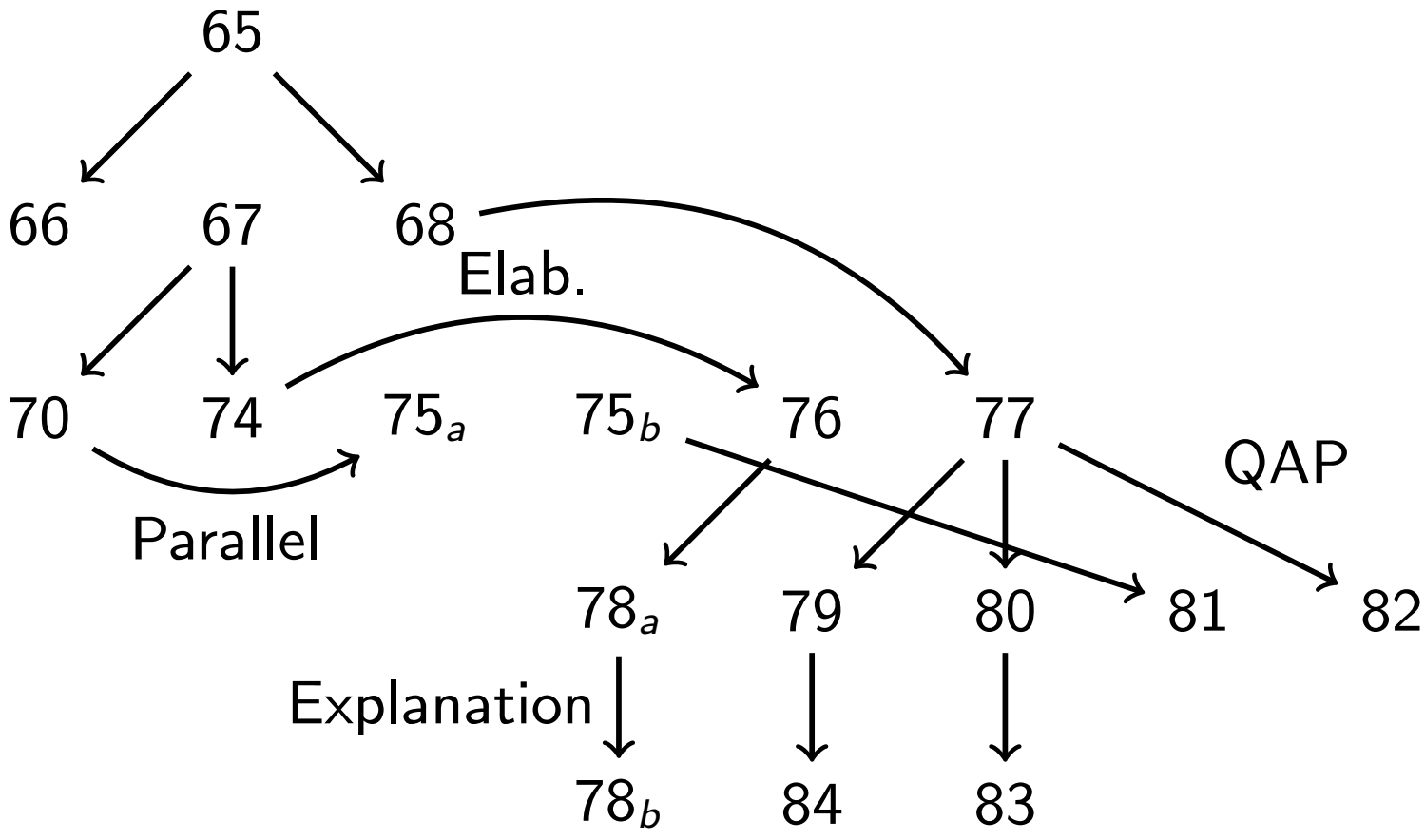
234	gotwood4sheep	anyone got wheat for a sheep ?
235	inca	sorry, not me
236	CheshireCatGrin	nope. you seem to have lots of sheep !
237	gotwood4sheep	yup baaa
238	dmm	i think i'd rather hang on to my wheat i'm afraid
239	gotwood4sheep	kk I'll take my chances then...



Multi party dialogue and threads

65	lj	anyone want sheep for clay ?
66	gw	got none, sorry :(
67	gw	so how do people know about the league ?
68	wm	no
70	lj	i did the trials
74	tk	i know about it from my gf
75	gw	[yeah me too,]$_a$
		[are you an Informatics student then, lj ?]$_b$
76	tk	did not do the trials
77	wm	has anyone got wood for me ?
78	gw	[I did them]$_a$ [because a friend did]$_b$
79	gw	lol william, you cad
80	gw	afraid not :(
81	lj	no, I'm about to start math
82	tk	sry no
83	gw	my single wood is precious
84	wm	what's a cad ?

A dependency graph



Not so much recursive structure, but multiple interleaved threads.

In fact two Stac corpora

The Linguistic Corpus :

- 39 *Settlers* games
- SDRT-annotated
- 1091 dialogues
- 10677 EDUs
- 1284 CDUs
- 10191 relations

Two Stac corpora

The Situated Corpus :

- 36 *Settlers* games (so far)
- SDRT-annotated
- $\approx$ dialogues
- $\approx$ 30,000 EDUs
- $\approx$ 5000 CDUs
- $\approx$ 27000 relations

Annotation scheme : several levels

- segmentation of dialog turns into discourse units
- labelling with domain-related speech acts (negotiation moves)
- relational rhetorical annotation from SDRT but with relations for dialogue (QAP, Q-elab, Acknowledgment, Correction).
- integrating nonlinguistic events into the relational structure (extending the relational annotation)

Domain level acts

- offer : *I'll give you 2 clay for a rock*
- counteroffer : *How about 2 clay for a wheat ?*
- accept : *OK, it's a deal.*
- refusal : *I don't think so.*
- has-resource : *I have wheat*
- strategic comment : *joel fancies a bit of your clay*
- other (non relevant for negotiation)

Example annotation

Speaker	Id	Turn	Dom. function	Rhet. function
Euan	47	[And I alt tab back from the tutorial.]_1	other	
		[What's up ?]_2	other	Result*(47_1,47_
Joel	48	[do you want to trade ?]	offer ⟨Joel, ?, ?,Euan⟩	Q-elab(47_2, 48)
Card.	49	[joel fancies a bit of your clay]	strat.-comment	Expl*(48, 49)
Joel	50	[yes]	other	Ackn(49, 50)
Joel	51	[!]	other	Comment(50, 51)
Euan	52	[Whatcha got ?]	counteroffer ⟨Euan, ?, ?,Joel⟩	Q-elab([48-50], 52)
Joel	53	[wheat]	has-resources ⟨Joel,wheat⟩	QAP(52, 53)
Euan	54	[I can wheat for clay.]	counteroffer ⟨Euan,wheat,clay,Joel⟩	Elab([52,53], 54)
Joel	55	[awesome]	accept(54)	Ackn(54, 55)

Discourse specificities

- Most frequent relations : QAP, Comment, Acknowledge, Q-elab, (Result and Sequence for the situated corpus)
- traditional discourse markers not relevant for most of these.
- nonlinguistic actions often had

Structures in multiparty dialogue

- non-treelike structures exist, whereas most theories that investigate full discourse structures .
- long distance, crossing dependencies,
- recursively structured discourse representations (CDUs).
- So tree spanning algorithms like MST are not conceptually right for multiparty dialogue
- Move to DAGs and various algorithms for getting the right sort (ILP)
- Lots of interesting investigations to do about situated discourse structures.

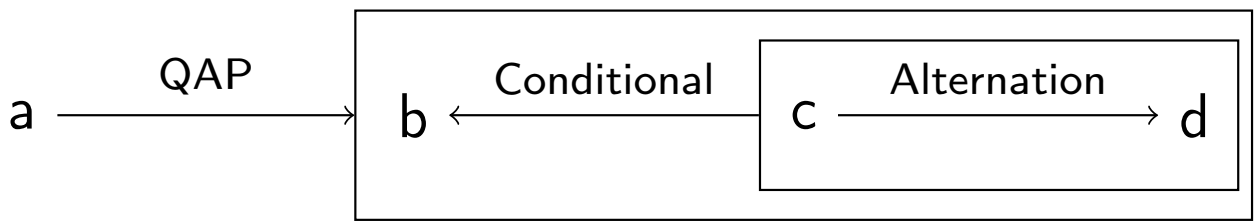
Constraints from conversational structure

- reactivity : turn constraint : people react to prior messages
- there are no backwards attachments (e.g., n. *I can trade* n+k. *if you want*) across turns from different speakers
- except in very rare cases (Sacks)

Complex Discourse Units

Example

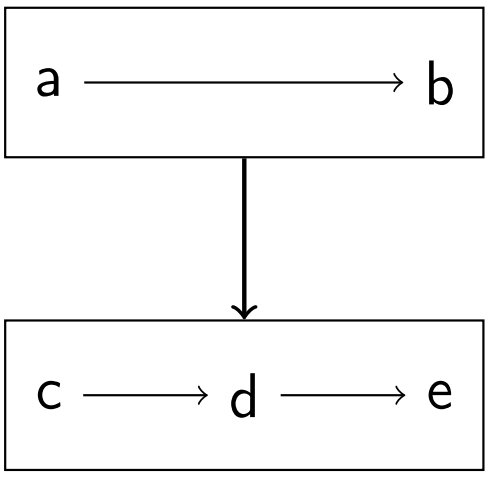
Alice [Do you have a sheep ?]$_a$
Bob [I do,]$_b$ [if you give me clay]$_c$
Bob [or wood.]$_d$

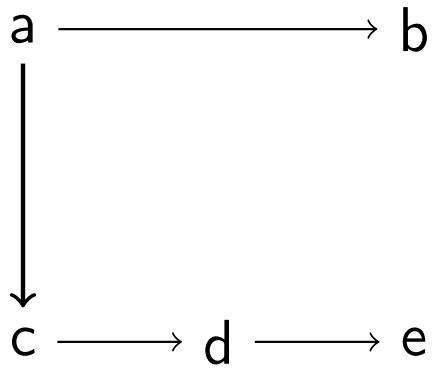


No reliable method has been identified in the literature for identifying CDUs.

We approximate CDUs in the SDRT hypergraph by relations between EDUs only, thus creating a dependency graph.

Distributing relations

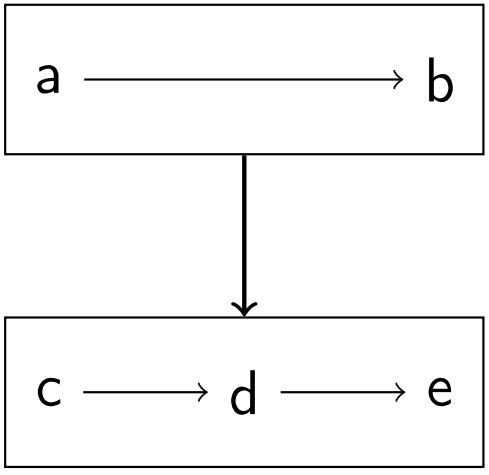




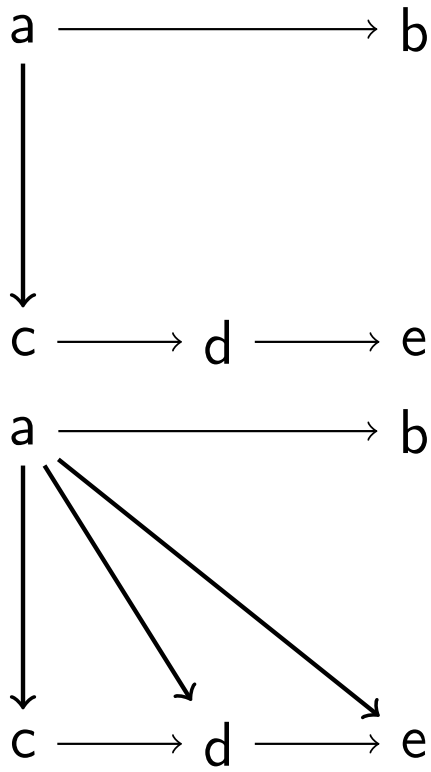
No distribution

Head points to head

Distributing relations



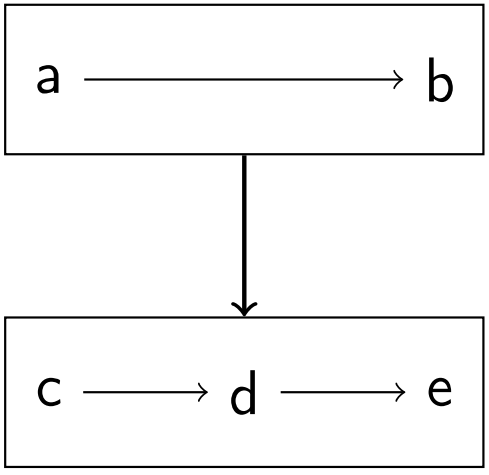
[I'll buy a card]$_a$
[and not a road]$_b$
[because I have
sheep]$_c$ [and wheat]$_d$
[and ore]$_e$



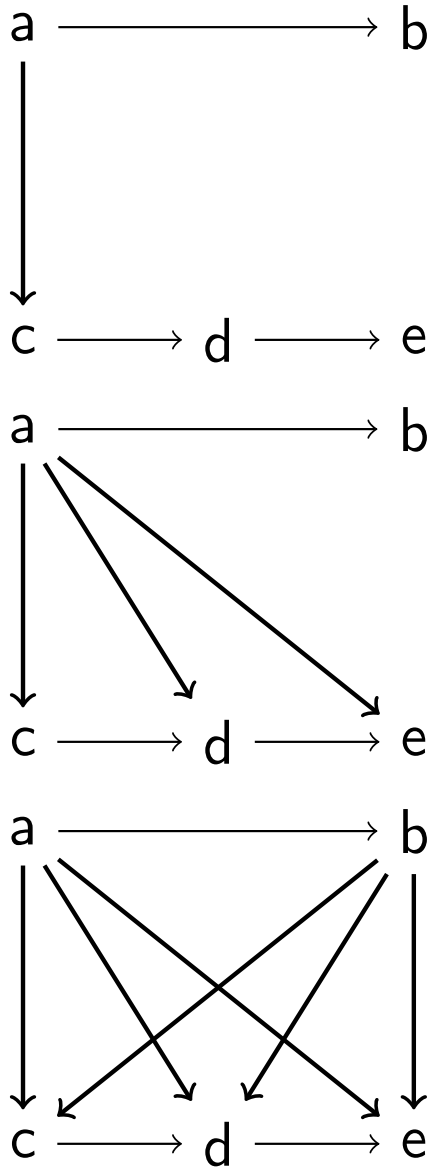
No distribution
Head points to head

Partial distribution
Relation semantics determine distribution to the source/target CDU components

Distributing relations



[I'll buy a card]$_a$
[and not a road]$_b$
[because I have sheep]$_c$ [and wheat]$_d$
[and ore]$_e$



No distribution
Head points to head

Partial distribution
Relation semantics determine distribution to the source/target CDU components

Full distribution
All relations distribute to every component

Evaluation setup

Pair modelization : Maximum Entropy model

Graph building methods

- Last : all EDUs are linked to the previous one, forming a single chain.
- Local : pair attachment and labelling models are treated as classifiers.
- MST : Maximum Spanning Tree *(Afantenos et al., 2015)*
- ILP : Integer Linear Programming

Three CDU replacement strategies

Each one creates its own version of the corpus

Evaluation F1 scores on test corpus

Method	Unlabelled	Labelled	Edge count
	No distribution		10191
Last	0.584	0.391	
Local	0.483	0.429	
MST	0.671	0.516	
ILP	**0.689**	**0.531**	
	Partial distribution		11734
Last	0.593	0.426	
Local	0.471	0.396	
MST	0.647	0.488	
ILP	**0.668**	**0.519**	
	Full distribution		13675
Last	0.582	0.420	
Local	0.541	0.443	
MST	0.613	0.466	
ILP	**0.675**	**0.527**	

Comparisons to the best text parsers

parser	U	R
HHN16 HILDA	64.7	49.1
SHV15 D *	64.2	49.5
JCN15 1S-1S	65.8	48.9
FH14 gCRF *	67.6	50.1
LLC16	62.9	42.1
JE14 DPLP **	**69.2**	**53.7**
BCS17	68.2	51.6
our method (MST)	69.0	50.8

Table – Dependency evaluation. U = unlabelled dependencies, R = relation labelled dependencies.

Analysis of the Results

- ILP outperforms MST even on trees (no-distribution strategy)
- Score gap between ILP and MST widens as edge count increases
- Partial distribution produces the hardest parsing task overall

Analysis of the Results

- ILP outperforms MST even on trees (no-distribution strategy)
- Score gap between ILP and MST widens as edge count increases
- Partial distribution produces the hardest parsing task overall

What the F1 scores don't tell you

- ILP consistently obtains higher recall
- Inside turns, clusters of edges are frequent
- Long-distance relations incur mixed results

Conclusions

- A small set of features
- A simple local model
- Richer well-formed structures
- Linguistic constraints and generalizations are important for the computational implementation

Future work

- Detection of CDUs from clusters of predicted edges
- Extended constraint sets for specific domains
- Structured prediction influencing local models
- Enhanced representations of discourse units