Dialogue: structures et similarités

Nicholas avec l'aide de Tim, Antoine V., Cedric, Soumya

Melodi/axe discours

24 février 2017

What we're after

- what would a notion of semantic similarity for conversations ideally look like, assuming perfect information ?
- how do different notions of meaning lead to different notions of semantic similarity ?
- what can we do practically ?

What do we bring?

- theories of conversation
- familiarity with supervised methods for discourse structure learning,
- theories of lexical meaning
- compositionality at the sentential level
- unsupervised methods for computing lexical meanings

Discourse

- the interpretation of a text is dynamic and depends on context.
- example : *John fell. Max pushed him.*
- each discourse constituent has one or more rhetorical functions relative to other DUs
- several DUs can work together to convey the same rhetorical function.
- discourse structures as directed acyclic graphs with 2 edges.
- such structures have been verified both for text and multi-party dialogue.

Formalization : a reminder

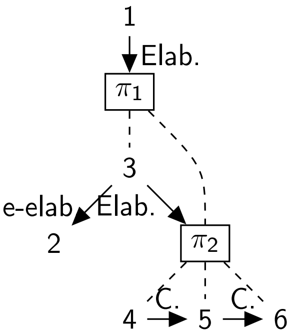
A directed acyclic graph :

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

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

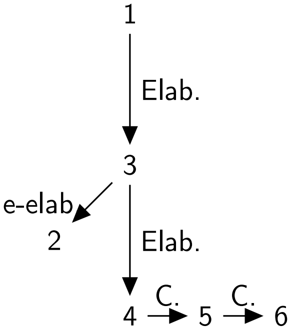
An example from Monologue

[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



Simplifying to structures without CDUs

A graph without E_2 edges and only EDUs as nodes.



Dialogue Structure

- Conversation as a game of sequential message exchanges
- speakers may speak in turns, each turn is finite but no limit on how long
- speakers make different commitments in the game.
- The conversational game is a sequence of commitment structures.

Constraints from the sequential game structure

- turn constraint : people react to prior messages
- there are no backwards attachments (e.g., subordinate clause prior to main clause) across speaker turns
- except in very rare cases (Sacks) (Perret et al. NAACL 2016).

STAC corpus annotations, a good place to start thinking about similarity

- $\sim$ 1100 negotiation dialogues, 11k EDUs and relations, 45 games annotated so far.
- annotation tool Glozz used by 3 annotators, annotations revised by experts.
- 5 years of work...
- but we have confidence in the quality of the annotations.

Comparisons

- Client turns are often much longer than a typical turn in *Settlers*
- intra turn structure will be more involved in the longer turns.

Annotation scheme

- segmentation of dialog turns into discourse units, with domain level act annotations
- labelling with domain-related speech acts (negotiation moves)
- relational rhetorical annotation from SDRT but with relations for dialogue (QAP, Q-elab, Acknowledgment, Correction).

Lexical semantics

- How do we understand the meaning of words ?
- closed class terms and morphemes vs. open class terms
- a hybrid account of composition (Asher et al. Computational Linguistics 2016, Van de Cruys 2011, V.de C. et al. 2013)
- But we have a long way to go to integrate the two. structured prediction ?

Syntactic similarities

- Parse Eval : $d(t, t') = 0$ iff for each constituent in t there is a constituent in t' with the same span over terminals (and same label strong version) and vice versa.
- Leaf Ancestor : compare the paths from root nodes of t and t' to terminals according to the Levenshtein distance function. This distance is defined as the minimum number of edits needed to transform one string into the other, with the allowable edit operations being insertion, deletion, or substitution of a single character. Various operations or cost functions furnish variants of this string distance.
- various edit distances
- The leaf ancestor metric penalizes less the boundaries of constituents in a parse than Parseval does. But doesn't give an idea of the structure as a whole.

Moving to semantic similarity

- structure $\neq$ content.
- What metrics respect the semantic structure (say the standard interpretation of Boolean connectives) ?
- What metrics respect the structure of first order structures ?

Starting very simply

- a conversation is a sequence of turns composed of sequences of sentences/words.
- Let **L** be a list of formulas composed together by .
- an interpretation of **L** : : $||.|| : \mathbf{L} \longrightarrow X$, X a non-empty set.
- The "target" of $||.||$ typically has structure—e.g.,
  - a BA with $\bot, \top$ a 2 element lattice (meets and joins everywhere defined).
  - an intuitionistic or S4 Kripke modal structure $(W, \leq)$
  - a possible worlds structure $(W, \subseteq)$
  - a metric space (X, d) with d a metric over X like euclidean distance or cosine distance over $\mathbb{R}^n$

Structural constraints on interpretations

Suppose that $||.||$ has as a target a partial order ; we can read $\leq$ as $\models$.

Axiom name	Meaning
contraction	$\|\vec{\alpha}, \phi, \phi, \vec{\beta}\| \leq \|\vec{\alpha}, \phi, \vec{\beta}\|$
expansion	$\|\vec{\alpha}, \phi, \vec{\beta}\| \leq \|\vec{\alpha}, \phi, \phi, \vec{\beta}\|$
exchange	$\|\vec{\alpha}, \phi, \psi, \vec{\beta}\| = \|\vec{\alpha}, \psi, \phi, \vec{\beta}\|$
right monotonicity	$\|\vec{\alpha}, \phi\| \leq \|\vec{\alpha}\|$
left monotonicity	$\|\phi, \vec{\alpha}\| \leq \|\vec{\alpha}\|$
$\vec{\epsilon}$–$\top$	$\|\vec{\alpha}\| \leq \|\vec{\epsilon}\|$
adjunction	If $\|\vec{\alpha}\| \leq \|\vec{\beta}\|$ and $\|\vec{\alpha}\| \leq \|\vec{\gamma}\|$
	then $\|\vec{\alpha}\| \leq \|\vec{\beta}\vec{\gamma}\|$
mix	If $\|\vec{\alpha}_1\| \leq \|\vec{\beta}_1\|$ and $\|\vec{\alpha}_2\| \leq \|\vec{\beta}_2\|$
	then $\|\vec{\alpha}_1\vec{\alpha}_2\| \leq \|\vec{\beta}_1\vec{\beta}_2\|$

Stronger properties

Definition (Conjunctive, intersective intepretation)

- $\| \cdot \|$ is **conjunctive iff it is lattice-valued and** $\forall \vec{\phi}, \vec{\psi}$, $\|\vec{\phi}\vec{\psi}\| = \|\vec{\phi}\| \wedge \|\vec{\psi}\|$.
- $\| \cdot \|$ **is intersective iff it is set-valued and** $\forall \vec{\phi}, \vec{\psi}$, $\|\vec{\phi}\vec{\psi}\| = \|\vec{\phi}\| \cap \|\vec{\psi}\|$.

Definition

We let $\| \cdot \|^t$ be the interpretation function for $\mathbf{L}^*_{\mathbf{prop}}$ defined by $\|\vec{\phi}\|^t := \| \bigwedge \{\phi \in Rng(\vec{\phi})\} \|^t$.

Example

$\| \cdot \|^t$ is **intersective. If** $\| \cdot \|$ **is intersective, then it is finite** $\subseteq$**–lattice-valued.**

What is semantic similarity ?

- we need a function $d\colon (\mathbf{L}^* \times \mathbf{L}^*) \to \mathbb{R}$ such that :
- $d(\vec{\phi}, \vec{\phi}) = 0$
- $d(\vec{\phi}, \vec{\psi}) = d(\vec{\psi}, \vec{\phi})$
- $d(\vec{\phi}, \vec{\chi}) \leq d(\vec{\phi}, \vec{\psi}) + d(\vec{\psi}, \vec{\chi})$

Semantic constraints

Definition

min sem separation : If for every $\vec{\chi_1}, \vec{\chi_2}$ with $\|\vec{\chi_1}\| = \|\vec{\chi_2}\|$, we have $d(\vec{\phi}, \vec{\chi_1}) = d(\vec{\psi}, \vec{\chi_2}),$ then$\|\vec{\phi}\| = \|\vec{\psi}\|$

Definition

sem separation :
If for every $\vec{\chi}$ we have $d(\vec{\phi}, \vec{\chi}) = d(\vec{\psi}, \vec{\chi})$ then $\|\vec{\phi}\| = \|\vec{\psi}\|$

Definition

zero $\Rightarrow$ sem$\equiv$: If $d(\vec{\phi}, \vec{\psi}) = 0$, then $\|\vec{\phi}\| = \|\vec{\psi}\|$

Definition

zero $\Rightarrow$ sem$\equiv$: If $\|\vec{\phi}\| = \|\vec{\psi}\|$, then $d(\vec{\phi}, \vec{\psi}) = 0$.

Definition

sem preservation : If $\|\vec{\phi}\| = \|\vec{\psi}\|$, then for every $\vec{\chi}$ we have $d(\vec{\phi}, \vec{\chi}) = d(\vec{\psi}, \vec{\chi})$

A few facts

- sem separation implies min sem separation
- zero $\Rightarrow$ sem$\equiv$ implies sem separation
- with the triangle inequality, zero $\Rightarrow$ sem$\equiv$ iff sem separation
- Sem preservation implies zero $\Rightarrow$ sem$\equiv$
- syntactic based measures (e.g., edit distances over propositional letters, or δ_{count}, $\delta_{synt,count}$) do not in general support zero $\Rightarrow$ sem$\equiv$.
  $\delta_{count}(\vec{\phi}, \vec{\psi}) := card(Rng\vec{\phi} \ominus Rng\vec{\psi})$
  e.g., for δ_{count} $\|p\neg p\| = \|q\neg q\|$ but $\delta_{count}(p\neg p, q\neg q) = 4$.

Semantic Definitions

Definition (Semantic Symmetric Difference Metric)

The semantic symmetric difference metric is the cardinality of the symmetric difference between the respective interpretations.
$d_{\ominus}(\vec{\phi}, \vec{\psi}) := card(\|\vec{\phi}\| \ominus \|\vec{\psi}\|)$

Definition (Proportional metric)

The proportional metric decreases from 1 to 0 as the ratio between the intersection and the union of the respective interpretations increases.

$$d_{\alpha}(\vec{\phi}, \vec{\psi}) = \begin{cases} 0 & \text{if } \|\vec{\phi}\| = \|\vec{\psi}\| = \emptyset \\ 1 - \frac{card(\|\vec{\phi}\| \cap \|\vec{\psi}\|)}{card(\|\vec{\phi}\| \cup \|\vec{\psi}\|)} & \text{otherwise} \end{cases}$$

Observe that, with this measure, pairs of sequences with non-empty disjoint interpretations will always be at distance 1 of each other.

Continuation based metrics

Definition (Contination based Metric)

using the idea that an admissible continuation is a *consistent* one :

$$d_{\mathcal{C}}(\vec{\phi}, \vec{\psi}) = \frac{2^{card(\|\vec{\phi})\|} + 2^{card(\|\vec{\psi}\|)} - 2 \cdot 2^{card(\|\vec{\phi}\| \cap \|\vec{\psi}\|)}}{2^{card(\|\vec{\phi}\|) + card(\|\vec{\psi}\|)}}.$$

More constraints

Definition (rebar property)

$$d(\vec{\phi}, \vec{\phi}\vec{\psi}) \leq d(\vec{\phi}, \vec{\psi})$$

Definition (antitonicity)

$$\text{If } \|\vec{\phi}\| \preceq \|\vec{\psi}\| \text{ then } d(\vec{\epsilon}, \vec{\psi}) \leq d(\vec{\epsilon}, \vec{\phi})$$

Definition (Signature Invariance)

states that the relative proximity of conversations should not depend on irrelevant aspects pertaining to the choice of signature. If $\vec{\phi}, \vec{\psi}, \vec{\chi} \in \mathbf{L}'^*$ and $\mathbf{L}' \subseteq \mathbf{L}$, then we have $d_L(\vec{\phi}, \vec{\chi}) \leq d_L(\vec{\psi}, \vec{\chi})$ iff $d_{L'}(\vec{\phi}, \vec{\chi}) \leq d_{L'}(\vec{\psi}, \vec{\chi})$

All our semantic metrics satisfy Rebar, Antitonicity and Signature Invariance.

More substantive constraints

Definition (Uniform preservation)

Extending conversations with a given piece of information should not change the relative proximity of conversations. Formally :

$$\text{If } d(\vec{\phi}, \vec{\chi}) \leq d(\vec{\psi}, \vec{\chi}) \text{ then } d(\vec{\phi}\vec{\phi}_0, \vec{\chi}\vec{\phi}_0) \leq d(\vec{\psi}\vec{\phi}_0, \vec{\chi}\vec{\phi}_0)$$

Definition (Anti-preservation)

$$\text{If } d(\vec{\phi}\vec{\phi}_0, \vec{\chi}\vec{\phi}_0) \leq d(\vec{\psi}\vec{\phi}_0, \vec{\chi}\vec{\phi}_0) \text{ then } d(\vec{\phi}, \vec{\chi}) \leq d(\vec{\psi}, \vec{\chi})$$

Triviality facts

Fact

Let $\|\cdot\|$ *satisfy* **exchange**, **contraction** *and* **expansion** *and let* d *be a semantic metric. If* d *satisfies* **Strong Preservation** *and* **Sem** $\equiv\Rightarrow 0$*, then whenever* $d(\phi, \chi) \leq d(\psi, \chi)$ *then* $d(\psi\chi, \phi\psi\chi) = 0$*.*

Fact

if the interpretation satisfies **contraction**, **expansion** *and* **exchange**, *then the only semantic metric satisfying* **Anti-Preservation** *and* **Sem** $\equiv\Rightarrow 0$ *is the trivial metric.*

Extending sequences with more or less similar continuations

Definition (Action Pref)

$$\text{If } d(\vec{\phi}, \psi_1) < d(\vec{\phi}, \psi_2) \text{ then } d(\vec{\phi}, \vec{\phi}\psi_1) \leq d(\vec{\phi}, \vec{\phi}\psi_2)$$

Definition (Coherent Deviation)

$$\text{If } d(\vec{\phi}, \chi) < d(\vec{\psi}, \chi) \text{ then } d(\vec{\phi}, \vec{\phi}\chi) < d(\vec{\psi}, \vec{\psi}\chi)$$

d satisfies **Coherent Deviation** and the **triangle inequality** iff d is the *trivial metric*.
The converses of these two conditions are only satisfied by the trivial metric.

Constraints on conjunctions and disjunctions

Definition ($\wedge$ rule)

$$\text{If } \|\vec{\phi_1}\| \wedge \|\vec{\phi_2}\| \leq \|\vec{\phi_1}\| \wedge \|\vec{\phi_3}\| \text{ then } \delta(\vec{\phi_1}, \vec{\phi_2}) \geq \delta(\vec{\phi_1}, \vec{\phi_3})$$

Fact

d satisfies the **triangle inequality** *and the* $\wedge$ **rule** *iff d is the trivial metric.*

Remark : we can weaken the $\wedge$ rule by replacing $\leq$ with $<$. The continuation metric satisfies this and is not trivial.

Outlook

- We can extend some distance metrics (e.g. a Hamming distance) to simple first order structures (and hence simple discourse structures which are in FO(∃),
- We don't know how to do this for the continuation style metrics.
- But in principle we have a theoretical basis on which to do structural similarity over dialogue structures.