Human judgements on causation in French texts

Cécile Grivaz

Département de Linguistique
University of Geneva
Switzerland

Computational Linguistics and Phonetics
Saarland University
Germany

18 November 2009

* This work was financed by a Swiss National Foundation project (n° 100012-113382) and by a scholarship from the Ernest Boninchi Foundation
What is causal?

- John’s ear fell; he took a shower.
- He broke his arm while skiing.
- It is triangle; it has three sides.
Traditional test of causation: intuition

- John’s ear fell; he took a shower.
- Does it seem causal to you?
- Can you reword it as “Taking a shower caused John’s ear to fall.”?
- Need an *a priori* knowledge.
Linguistic tests

- Can you say *It is triangle because it has three sides*?
- What about *John went home because his car is not in the parking any longer*?
- Linguistic tests tend to be *ambiguous*. 
Need for complex guidelines

- These tests are **not sufficient** to identify causation.
- We need to study causation in human judgement and its features.
Overview

- What tests do people use when they think about causation?
- Are there simpler clues that are statistically associated with causation?
- Do our guidelines work?

All of this in a French framework.
Intuitive features of causation

- What tests do people use when they think about causation?
- Does this text contain a causal relation? Justify your answer.
Example (translated from French)

Example

Context: In a murder trial, X confessed the murder of Hughes.

The fact that similarities exist between the murder of Worms and the murder of Hughes convinced the policemen that X was also responsible for the death of Worms which happened sooner. They continued their questioning of X.
Setting of the experiment

- Short segments of text from the BAF corpus.
- The texts contain *parce que*, *donc* or *mais*.
- In most of the texts the markers are removed.
- 9 subjects, 10 segments per subject.
Types of answer

- **Rewording:**
  
  Une explication est donnée.

  Je ne vois pas de relation causale.

- **Linguistic tests:**
  
  parce que, donc, c’est parce que, est le fruit de, entraîne, permet, est la cause de . . .

- **Presence of an explicit marker:**
  
  ’donc’ apparaît. Donc introduit une conséquence.

  Le ’mais’ exprime une nuance, une restriction dans ce cas précis.
Types of answer

- Presence of a non-causal relation:
  
  *C'est une description. La 2ème phrase apporte seulement une précision.*

- Other:
  
  ??
**Number of justifications of each type**

**Figure:** No real-world justification. Many rewordings.
Conclusions on elicitation of intuitive tests of causation

- Difficulty to obtain systematic justifications.
- Lack of variety in justifications.
- High amount of rewording

Human reasoning does not consciously make use of intuitive tests of causation.
Association of the judgement of causation and its features

• Which features are statistically associated to judgements of causation?
• These features can be used in annotation guidelines.
Experiment

- Features from the previous experiment and from theoretical work.
- **Judgements** on causation and its features on 24 artificial sentences that are ambiguous to causation.
- 4 subjects.

**Example**
I put the chicken in the oven, I’m cooking.
Features

- Temporal order.
- Ability to build a causal chain.
- Linguistics test by using *parce que* (because) and *donc* (so).
- Counterfactuality.
- Paraphrases.
**Results**

![Bar chart showing P-values of Fisher's exact test of association for each feature and the presence of causation.](image)

**Figure:** $P$-value of Fisher’s exact test of association for each feature and the presence of causation.
Guidelines

• There is no easily usable necessary and sufficient condition for causation.
• Annotators must rely on intuition.
• Features of causation are necessary but not sufficient conditions.
• Some features can rule out non-causal cases
• Some can help annotators to clarify their intuition.
Tests of causation

- From *divergences* between our predictions and annotators answers.
- From divergences between experts.
Annotation experiment

• 4 Annotators.
• 15 short texts from *De la terre à la lune*.
• No training, only the guidelines.
• Low pairwise agreement.
• $\kappa = 0.84$ between the majority of the annotators answers and our judgements.

Example

Bref, il ne fut plus permis, même au moins lettré des Yankees, d’ignorer un seul des faits relatifs à son satellite, ni à la plus bornée des vieilles mistress d’admettre encore de superstitieuses erreurs à son endroit. *La science leur arrivait sous toutes les formes; elle les pénétrait par les yeux et les oreilles;* impossible d’être un âne...en astronomie.
Features that help to rule out non-causal occurrences

- Temporal order
- Counterfactuality.
- Ontological asymmetry.
Features that help to clarify intuition

- Ability to build causal chains.
- Linguistic tests.
Conclusion

- If there exist features of causation that are used in human reasoning, they are not conscious.
- There exist features of causation that are associated with causation, e.g. causal chains, linguistic tests and counterfactuality.
- Our understanding of causation can be effectively transmitted through our instructions.