

Grivaz, Cecile: Université de Genève

Mesurer et améliorer l'accord inter annotateurs pour les relations causales entre propositions en français

Dans le cadre du développement d'un programme informatique capable de reconnaître automatiquement les expressions de relations causales dans des textes, nous voulons annoter un corpus en repérant les relations causales qui y sont exprimées.

Les travaux de Inui (2005) qui a réalisé un corpus similaire en japonais, montrent que l'annotation de relations causales est difficile et peut résulter en un accord inter annotateurs faible. Nous devons donc construire un ensemble de consignes d'annotation qui servira de guide aux annotateurs afin d'augmenter leur accord. Pour obtenir un bon ensemble de consignes, nous les testerons sur des échantillons, et nous les raffinerons jusqu'à ce que nous obtenions un jeu de consignes qui permette un accord inter annotateurs suffisamment haut. Cette méthodologie se base sur le travail de Hovy et de ses collègues (2006) qui ont décrit un système similaire pour la tâche qui consiste à relier les mots d'un corpus aux concepts correspondants dans une ontologie.

Nous avons réalisé une première expérience dans laquelle nous demandons à des sujets de répondre à un certain nombre de questions liées à la causalité à propos d'énoncés dont il est difficile de décider si ils expriment une relation causale ou non. En nous basant sur des résultats théorique et sur une expérience précédente, nous avons identifié des caractéristiques de la causalité (asymétrie temporelle, propriété contrefactuelle, etc.). Pour chaque énoncé, on demande aux sujets de déterminer si ces caractéristiques sont présentes. On leur demande ensuite de se baser sur ces résultats pour déterminer si l'énoncé exprime une relation causale. Il nous sera ainsi possible de calculer un accord inter annotateur, non seulement sur la présence de relation causale, mais également sur les propriétés associées à la causalité. Cette expérience nous permettra ainsi d'isoler les propriétés les plus difficiles à identifier et d'affiner nos consignes d'annotations en fonction de ces résultats.

Références:

- Eduard H. Hovy/Mitchell P. Marcus/Martha Palmer/Lance A. Ramshaw/Ralph M. Weischedel (2006):
Ontonotes: The 90. In Robert C. Moore/Jeff A. Bilmes/Jennifer Chu-Carroll/Mark Sanderson (eds.),
HLT-NAACL. The Association for Computational Linguistics.
- Takashi Inui (2005): Creating an Annotated Corpus for the Analysis of Causal Relations. COE-LKR2005.