

# Une proposition pour l'extraction de relations non prédicatives

Mouna Kamel

IRIT-Université Paul Sabatier-31069 Toulouse Cedex  
kamel@irit.fr

**Résumé.** Les relations sémantiques généralement reconnues par les méthodes d'extraction sont portées par des structures de type prédicats-arguments. Or, l'information recherchée est souvent répartie sur plusieurs phrases. Pour détecter ces relations dites complexes, nous proposons un modèle de représentation des connaissances basé sur les graphes conceptuels.

## 1 Problématique

L'Extraction d'Information consiste à identifier de l'information bien précise d'un texte en langue naturelle et à la représenter sous forme structurée. Les composantes de l'information recherchée sont généralement prédéfinies et circonscrites à un domaine spécifique, et les principaux travaux réalisés en matière d'identification de relations sémantiques ont essentiellement concerné les relations portées par une structure de type prédicats-arguments. Les principales approches d'identification de ces relations ont été basées sur l'analyse syntaxique (identification du verbe et ses arguments) (Khélif, 2006), ou sur la définition de patrons lexi-co-syntaxiques (Aussenac et al., 2000).

L'étude de corpus de domaines différents montre que bonne partie de l'information pertinente peut aussi être distribuée sur plusieurs phrases, par le biais de relations exprimées à l'aide de variations linguistiques, comme la coréférence, l'anaphore ou l'ellipse. Les méthodes classiques d'extraction de relations ne sont alors plus adaptées.

La résolution de relations non prédicatives, et plus particulièrement d'une certaine forme elliptique (formes passives où l'argument agent est effacé) utilisée fréquemment, nous a conduits à proposer une représentation des connaissances du domaine considéré, à l'aide du modèle des graphes conceptuels, car ce modèle est doté d'opérations et offre des procédures de raisonnement (Salvat, 1997).

## 2 Identification de relations non prédicatives

Nous avons étudié un corpus de résumés d'articles scientifiques décrivant des expériences génétiques menées par des chercheurs sur un ensemble de patients porteurs d'une même maladie génétique, le but étant de localiser les régions chromosomiques affectées. Une des relations pertinentes identifiées est *Conditions Expérimentales* qui relie l'ensemble des patients observés au type d'analyse subie. Dans l'exemple "*A study was conducted on 22 MM patients. The authors used G-banding*", la relation *Conditions Expérimentales* ne peut être détectée par une approche classique. La mise en œuvre d'une procédure de raisonnement qui établit un lien entre les auteurs (*authors*) et l'étude (*study*) menée par les auteurs (*authors*)