

# Degré de relation sémantique dans une ontologie pour la commande en langue naturelle

Laurent Mazuel, Nicolas Sabouret

Laboratoire Informatique de Paris 6 (LIP6)  
{laurent.mazuel, nicolas.sabouret}@lip6.fr

IC, 04 Juillet 2007

# Plan

- 1 Contexte et motivation
- 2 Distance/Similarité/Degré de relation sémantique
- 3 Notre mesure de degré de relation sémantique
- 4 Conclusion & Perspectives

# Plan

- 1 Contexte et motivation
- 2 Distance/Similarité/Degré de relation sémantique
- 3 Notre mesure de degré de relation sémantique
- 4 Conclusion & Perspectives

# Contexte

## Besoin

- 1 Système d'IHM en TAL [Corby, al04], [Milward, al03]...
- 2 Besoin d'ontologies,
- 3 Besoin d'approximation sémantique

## Exemples

- Équivalence/pont « terme utilisateur », « terme système »
- « cube »  $\Rightarrow$  « carré »
- « Ethernet »  $\Rightarrow$  « wifi »

# Travaux existants

## Autres travaux du domaine

- 1 Essentiellement basé sur WordNet [Mohammad, al06], [Patwardhan, al06], [Cho, al03], [Yang, al05]
- 2 Non conforme aux évaluations psychologiques

## Exemples

### Modélisation incomplète

- « téléphone » et « communication » : 7/10
- *Modélisé ni par WordNet, ni par les mesures connues*

## Conclusion

*Besoin d'une mesure de distance de degré de relation sur les ontologies en général (et non WordNet)*

# Plan

- 1 Contexte et motivation
- 2 Distance/Similarité/Degré de relation sémantique
- 3 Notre mesure de degré de relation sémantique
- 4 Conclusion & Perspectives

# Distance/Similarité sémantique

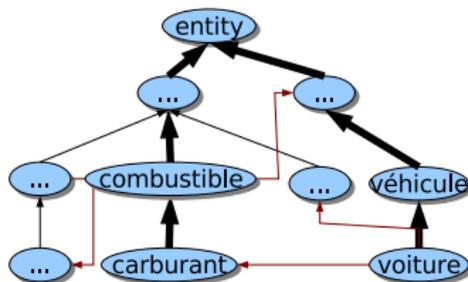
## Definition

Distance sémantique [Resnik 95] :

- Lié à la hiérarchie de concept, inverse de la **similarité**

## Exemples

- « voiture » très proche de « automobile »
- « carburant » loin de « voiture »



# Degré de relation sémantique

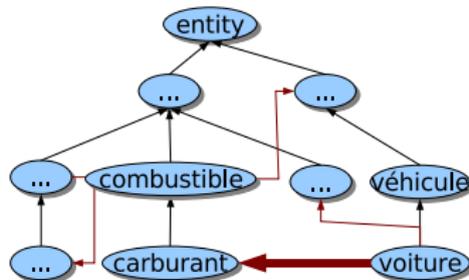
## Definition

Degré de relation sémantique :

- Lié à aux différentes relations, apparenté à la **similarité**

## Exemples

- « voiture » très proche de « automobile »
- « carburant » très proche de « voiture »



# Travaux de Jiang & Conrath

## Pourquoi utiliser ces travaux ?

- 1 Mesure courante et efficace ([Budanitsky06])
- 2 Mesure *uniquement* hiérarchique, mais *pondération des arêtes*

## Mesure pondérée

*Proposition d'une mesure efficace prenant en compte la pondération des arêtes*

# Vocabulaire formel

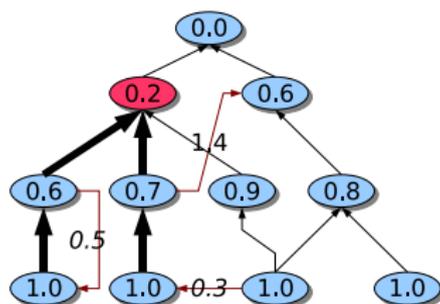
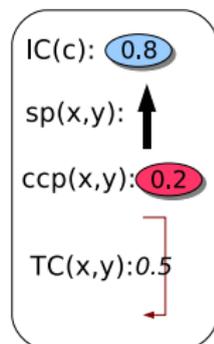
## Definitions

$IC(c_i)$  La quantité d'information d'un concept ( $\in [0, 1]$ )

$sp(c_i, c_j)$  Le plus court chemin dans l'ontologie/hierarchie

$ccp(c_i, c_j)$  Plus proche commun parent dans la hiérarchie

$TC(c_i, c_j)$  Poids associé à l'arête  $\{x, y\}$



# Distance de Jiang & Conrath [Jiang&Conrath 97]

## Definitions

Complète :  $dist_{JC}(c_1, c_2) =$

$$\sum_{c \in sp(c_1, c_2)} [IC(c) - IC(pere(c))] \times TC(c, pere(c))$$

Simple ( $TC = 1$ ) :  $dist_{JCsimple}(c_1, c_2) =$

$$(IC(c_1) + IC(c_2)) - 2 \times IC(ccp(c_1, c_2))$$

## Exemples

- « car » et « automobile »  $\longrightarrow dist = 0.0, sim = 1.0$
- « car » et « gasoline »  $\longrightarrow dist = 1.46, sim = 0.27,$   
 $ccp = \text{« entity »}$

# Travaux d'Aleksovski

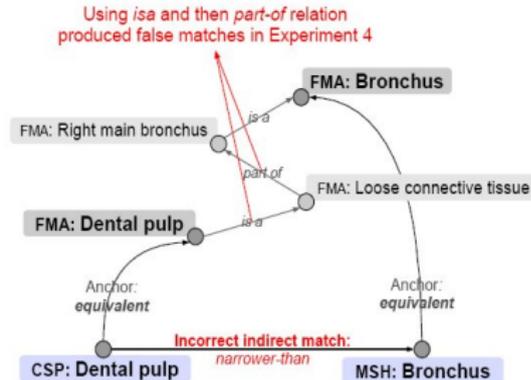
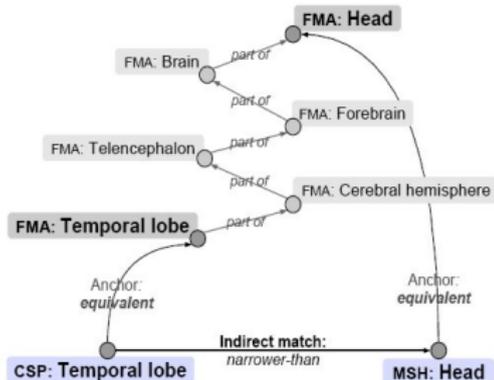
## Pourquoi utiliser ces travaux ?

- 1 Relation  $\Rightarrow$  graphe au lieu d'une hiérarchie
- 2 Création d'une multitude de chemins pas forcément sémantique

## Etude des chemins corrects

*Étude des conditions pour qu'un chemin soit correct sémantiquement*

# Exemple



## Conclusion

❶ Pas de mixité relation/hiérarchie :

- ❶ part-of d'abord ;
- ❷ is-a ensuite.

# Conclusion pour notre mesure

## Hypothèses sur notre mesure de degré de relation sémantique :

- 1 Relation d'abord *puis* Hiérarchie (Aleksovski) :
  - 1 Composante relationnelle +
  - 2 Composante hiérarchique
- 2 Homogène à la distance de Jiang & Conrath
  - 1 Utilisation des pondérations *par type d'arêtes*
  - 2 Linéarité de la somme  $\Rightarrow$  somme des 2 composantes

# Plan

- 1 Contexte et motivation
- 2 Distance/Similarité/Degré de relation sémantique
- 3 Notre mesure de degré de relation sémantique**
- 4 Conclusion & Perspectives

# Composante hiérarchique

## Hypothèse

- Homogène à Jiang & Conrath si hiérarchique uniquement  
→ Utilisation de  $sim_{JCsimple}$

## Theorem

$$dist_{is-a}(c_1, c_2) = dist_{JCsimple}(c_1, c_2)$$

# Composante relationnelle

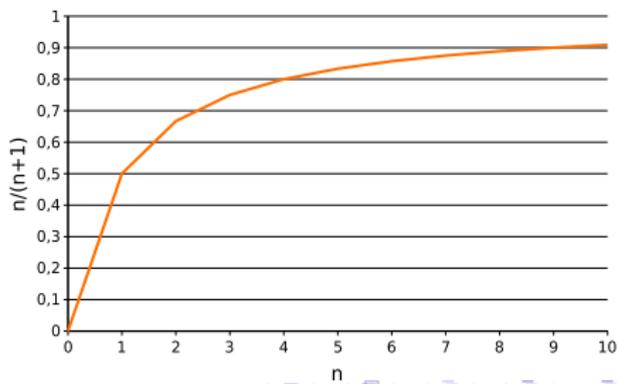
## Hypothèses

- Bornée dans l'intervalle  $[0, 1]$
  - Dépendante de la distance du chemin (logarithme)
- Proche de la progression  $IC(c_i)$

## Theorem

$$dist_X(c_1, c_2) = TC_X \times \frac{|sp_X(c_1, c_2)|}{(|sp_X(c_1, c_2)| + 1)}$$

Evolution de la fonction  $n/(n+1)$



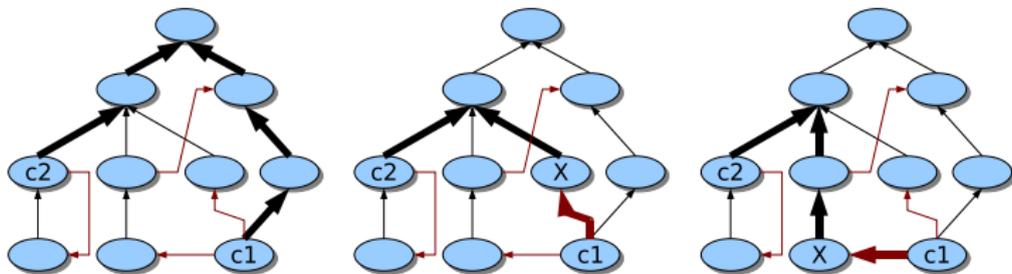
# Degré de non-relation sémantique

## Hypothèses

Recherche du « point d'articulation X »  
→ transition relation/hiérarchie

## Theorem

$$dist_{ONT}(c_1, c_2) = \min_{t \in C, X \in R} \left\{ TC_X \times \left( \frac{|sp_X(c_1, t)|}{|sp_X(c_1, t)| + 1} \right) + dist_{simpleJC}(t, c_2) \right\}$$



# Degré de relation sémantique

## Conversion

- Conversion classique [Resnik 95], [Budanitsky 06]
- La distance doit être **bornée**

## Theorem

$$sim_{ONT}(c_1, c_2) = 1 - \min \left\{ \frac{dist_{ONT}(c_1, c_2)}{2 + \max_{X \in \mathcal{R}} TC_X}, \frac{dist_{simpleJC}(c_1, c_2)}{2} \right\}$$

# Plan

- 1 Contexte et motivation
- 2 Distance/Similarité/Degré de relation sémantique
- 3 Notre mesure de degré de relation sémantique
- 4 Conclusion & Perspectives

# Conclusion

## Une mesure de degré de relation sémantique

- Généralisation de la formule classique de Jiang & Conrath
- Prise en compte des hypothèses sémantique d'Aleksovski
- Formule paramétrable ( $TC_X$ )

# Perspectives

## Evaluation

- 1 Différentes types d'application :
  - 1 Choix des agents
  - 2 Choix des ontologies (Swoogle ?)
- 2 Test psychologique ([Miller & Charles 91], [Finkelstein 01])

## Evolution

- 1 Apprentissage par renforcement des  $TC_X$
- 2 Composante relationnelle
- 3 Étude d'une spécialisation la formule générale de [Lin 98] (formule « en compétition » avec celle de Jiang & Conrath)

# Merci !

Merci pour votre attention !  
Des questions ?