

# BACANAL

## Balades Aléatoires Courtes pour ANALyses Lexicales

*Applications à la substitution lexicale (SemDis 2014)*

Yann Desalle<sup>†</sup>, Emmanuel Navarro<sup>\*\*</sup>, Yannick Chudy<sup>\*\*</sup>  
Pierre Magistry<sup>‡</sup>, Bruno Gaume<sup>\*</sup>

<sup>†</sup> : ATILF, CNRS & Université de Lorraine

<sup>\*\*</sup> : IRIT, CNRS & Université de Toulouse

<sup>‡</sup> : Graduate Institute of Linguistics, National Taiwan University

<sup>\*</sup> : CLLE, CNRS & Université de Toulouse

Séminaire de l'axe CARTEL :

08 décembre 2014

# Plan de l'exposé

- 1 La substitution lexicale
- 2 Balades Aléatoires
- 3 Application à la substitution lexicale
  - Réseaux lexicaux
  - Méthodes
  - Résultats Sem Dis2014
- 4 Conclusion

- 1 La substitution lexicale
- 2 Balades Aléatoires
- 3 Application à la substitution lexicale
- 4 Conclusion

# La substitution lexicale

## Exemple

Et cette confiance **fonde** la responsabilité des praticiens.

Et cette confiance **établit** la responsabilité des praticiens.

Et cette confiance **crée** la responsabilité des praticiens.

...

# La substitution lexicale

## Exemple

Et cette confiance **fonde** la responsabilité des praticiens.

Et cette confiance **établit** la responsabilité des praticiens.

Et cette confiance **créé** la responsabilité des praticiens.

...

## Tâche

**Construire une liste ordonnées de  $n$  substituts lexicaux qui n'altèrent pas le sens de l'énoncé.**

ex: **établir**, **constituer**, **créer**, former, instituer, **assurer**, mettre, **instaurer**, **poser**, construire

# La substitution lexicale

## Exemple

Et cette confiance **fonde** la responsabilité des praticiens.

Et cette confiance **établit** la responsabilité des praticiens.

Et cette confiance **créé** la responsabilité des praticiens.

...

## Tâche

**Construire une liste ordonnées de  $n$  substituts lexicaux qui n'altèrent pas le sens de l'énoncé.**

ex: **établir**, **constituer**, **créer**, former, instituer, **assurer**, mettre, **instaurer**, **poser**, construire

## Les méthodes de l'état de l'art :

### ▶ Méthodes supervisées [12, 4, 3, 11, 2, 8, 5, 7, 4] :

- 1 Ressources : ressources lexicales (thésauri, dictionnaires) + corpus
- 2 Filtre à partir de ressources lexicales  $\Rightarrow$  candidats-substituts
- 3 Ordonnancement des candidats-substituts
- 4  $\Rightarrow$  Liste ordonnée de candidats-substituts

### ▶ Méthodes non supervisées [10] :

- 1 Ressources : corpus seulement
- 2 analyse distributionnelle de corpus + calculs matriciels
- 3  $\Rightarrow$  Liste ordonnée de candidats-substituts

## Les méthodes de l'état de l'art :

### ▶ Méthodes supervisées [12, 4, 3, 11, 2, 8, 5, 7, 4] :

- 1 Ressources : ressources lexicales (thésauri, dictionnaires) + corpus
- 2 Filtre à partir de ressources lexicales  $\Rightarrow$  candidats-substituts
- 3 Ordonnancement des candidats-substituts
- 4  $\Rightarrow$  Liste ordonnée de candidats-substituts

### ▶ Méthodes non supervisées [10] :

- 1 Ressources : corpus seulement
- 2 analyse distributionnelle de corpus + calculs matriciels
- 3  $\Rightarrow$  Liste ordonnée de candidats-substituts

## Nos méthodes « simples »:

### ► Méthodes supervisées:

- 1 Ressources : ressources lexicales (thésauri, dictionnaires)
- 2 **Aucun filtre**
- 3 Ordonnancement de **l'ensemble de la ressource**
- 4  $\Rightarrow$  Liste ordonnée de candidats-substituts

### ► Méthodes non supervisées:

- 1 Ressources : corpus seulement
- 2 **Analyse de corpus  $\Rightarrow$  Graphes**
- 3 **Balades aléatoires courtes sur les graphes**
- 4  $\Rightarrow$  Liste ordonnée de candidats-substituts

## Agrégation de méthodes

- $\omega$  le mot-cible à remplacer
- $\mathcal{V}_i = M_i(G_i, P, \omega)$  et  $\mathcal{V}_j = M_j(G_j, P, \omega)$  : deux méthodes simples
- $\mathcal{V} = \text{Agreg}(\mathcal{V}_i, \mathcal{V}_j)$

## Nos méthodes « simples »:

### ► Méthodes supervisées:

- 1 Ressources : ressources lexicales (thésauri, dictionnaires)
- 2 **Aucun filtre**
- 3 Ordonnancement de **l'ensemble de la ressource**
- 4  $\Rightarrow$  Liste ordonnée de candidats-substituts

### ► Méthodes non supervisées:

- 1 Ressources : corpus seulement
- 2 **Analyse de corpus  $\Rightarrow$  Graphes**
- 3 **Balades aléatoires courtes sur les graphes**
- 4  $\Rightarrow$  Liste ordonnée de candidats-substituts

## Agrégation de méthodes

- $\omega$  le mot-cible à remplacer
- $\mathcal{V}_i = M_i(G_i, P, \omega)$  et  $\mathcal{V}_j = M_j(G_j, P, \omega)$  : deux méthodes simples
- $\mathcal{V} = \text{Agreg}(\mathcal{V}_i, \mathcal{V}_j)$

## Nos méthodes « simples »:

### ▶ Méthodes supervisées:

- 1 Ressources : ressources lexicales (thésauri, dictionnaires)
- 2 **Aucun filtre**
- 3 Ordonnancement de **l'ensemble de la ressource**
- 4  $\Rightarrow$  Liste ordonnée de candidats-substituts

### ▶ Méthodes non supervisées:

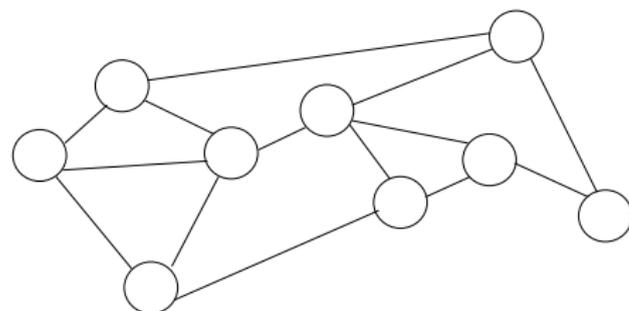
- 1 Ressources : corpus seulement
- 2 **Analyse de corpus**  $\Rightarrow$  **Graphes**
- 3 **Balades aléatoires courtes sur les graphes**
- 4  $\Rightarrow$  Liste ordonnée de candidats-substituts

## Agrégation de méthodes

- ▶  $\omega$  le mot-cible à remplacer
- ▶  $\mathcal{V}_i = M_i(G_i, P, \omega)$  et  $\mathcal{V}_j = M_j(G_j, P, \omega)$  : deux méthodes simples
- ▶  $\mathcal{V} = \text{Agreg}(\mathcal{V}_i, \mathcal{V}_j)$

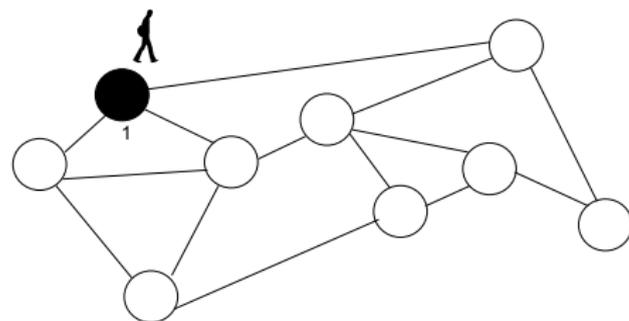
- 1 La substitution lexicale
- 2 Balades Aléatoires**
- 3 Application à la substitution lexicale
- 4 Conclusion

# Balades Aléatoires



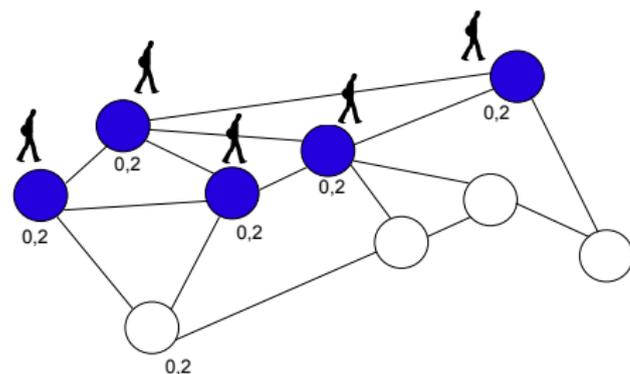
- 1 État initial ( $t=0$ ) :
- 2  $t = 1$
- 3  $t = 2$
- 4  $t = 3$
- 5  $t = 4$  pour des graphes de symétrie

# Balades Aléatoires



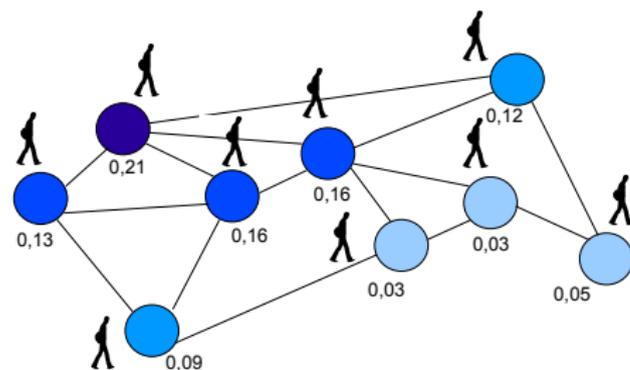
- 1 État initial ( $t=0$ ) :
- 2  $t = 1$
- 3  $t = 2$
- 4  $t = 3$
- 5  $t = 4$  pour des graphes de symétrie

# Balades Aléatoires



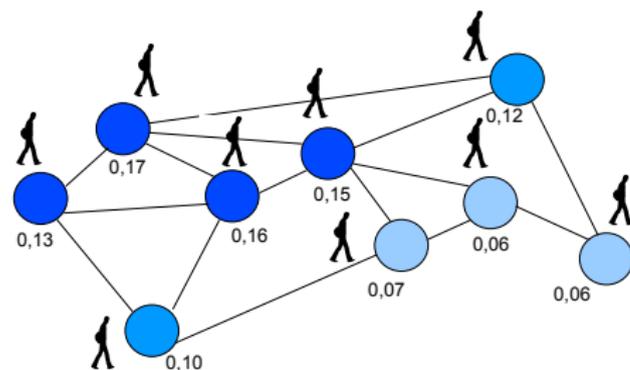
- 1 État initial ( $t=0$ ) :
- 2  $t = 1$
- 3  $t = 2$
- 4  $t = 3$
- 5  $t = 4$  pour des graphes de symétrie

# Balades Aléatoires



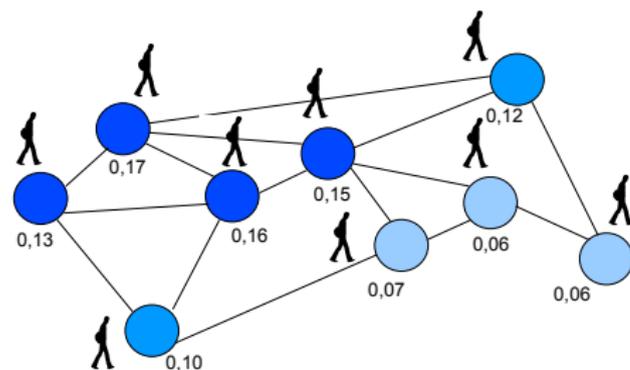
- 1 État initial ( $t=0$ ) :
- 2  $t = 1$
- 3  $t = 2$
- 4  $t = 3$
- 5  $t = 4$  pour des graphes de symétrie

# Balades Aléatoires



- 1 État initial ( $t=0$ ) :
- 2  $t = 1$
- 3  $t = 2$
- 4  $t = 3$
- 5  $t = 4$  pour des graphes de symétrie

# Balades Aléatoires

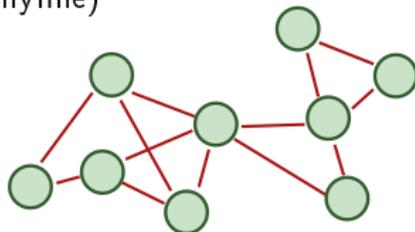


- 1 État initial ( $t=0$ ) :
- 2  $t = 1$
- 3  $t = 2$
- 4  $t = 3$
- 5  $t = 4$  pour des graphes de symétrie

- 1 La substitution lexicale
- 2 Balades Aléatoires
- 3 Application à la substitution lexicale**
  - Réseaux lexicaux
  - Méthodes
  - Résultats SemDis2014
- 4 Conclusion

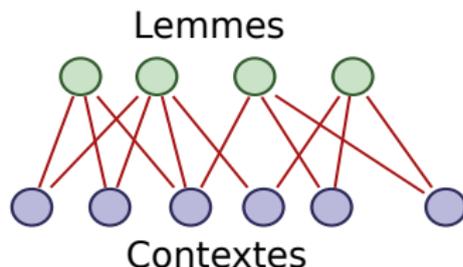
## Réseaux lexicaux construits à partir de ressources lexicales :

- ▶ Réseaux lexicaux construits à partir de DicoSyn :
  - Union de 7 dictionnaires classiques (Baillly, Benac, Du Chazaud, Guizot, Lafaye, Larousse et Robert)
  - Un lien entre deux unités lexicales synonymes
  - $\Rightarrow$  **Gdsyn<sub>A</sub>**, **Gdsyn<sub>N</sub>**, **Gdsyn<sub>V</sub>**
- ▶ Réseaux lexicaux construits à partir de Jeux De Mots (JDM)
  - JDM : construit par les foules en utilisant un jeu décrit dans [6]
  - Tâche du jeu : association de mots à une cible: synonymie, association libre ...
  - Un lien entre deux mots si associés par les participants
  - $\Rightarrow$  **GjdmS<sub>A</sub>**, **GjdmS<sub>N</sub>**, **GjdmS<sub>V</sub>** (synonymie)
  - $\Rightarrow$  **GjdmA** (association libre)



## Réseaux lexicaux construits à partir de corpus :

- ▶ Réseaux lexicaux construits à partir du corpus LM10 :  
corpus journalistique, 200 millions de mots  
⇒ **Glm10<sub>N</sub>**, **Glm10<sub>A</sub>**, **Glm10<sub>V</sub>**
- ▶ Réseaux lexicaux construits à partir du corpus frWaC [1] :  
corpus du web français, 1,6 milliard de mots  
⇒ **Gfrwac<sub>A</sub>** **Gfrwac<sub>N</sub>** **Gfrwac<sub>V</sub>**



## Réseaux lexicaux construits à partir de corpus :

- ▶ Réseaux lexicaux construits à partir du corpus LM10 :  
corpus journalistique, 200 millions de mots  
⇒ **Glm10<sub>N</sub>**, **Glm10<sub>A</sub>**, **Glm10<sub>V</sub>**
- ▶ Réseaux lexicaux construits à partir du corpus frWaC [1] :  
corpus du web français, 1,6 milliard de mots  
⇒ **Gfrwac<sub>A</sub>** **Gfrwac<sub>N</sub>** **Gfrwac<sub>V</sub>**

## Etapas de construction des graphes

- 1 Analyse syntaxique par Talismane [9] au laboratoire CLLE
- 2 Module de déduction de relations syntaxiques logiques :  
coordination, attribut du sujet, complément prépositionnels ...
- 3  $\forall$  item lexical  $l$  et contexte syntaxique  $c = (rel, l_c)$  :  
création d'un lien entre  $l$  et  $c$  si  $c$  est un contexte syntaxique de  $l$
- 4 Pondération par une pseudo information mutuelle

$$IM = \frac{freq((*, *)) \times freq((l, c))}{freq((l, *)) \times freq((*, c))} \quad (1)$$

# Méthodes par vision simple

$P = \ll \text{Et cette confiance } \mathbf{fonde} \text{ la responsabilité du praticien. } \gg$

$\omega = \mathbf{fonde}$  le mot-cible de  $P$

$C_P^\omega = \{(NC.\text{confiance}, \text{Dep.suj}), (NC.\text{responsabilité}, \text{Dep.obj}), (CC.\text{et}, \text{Gov.dep\_coord})\}$

## Voisins de *fonder*

### ► Dans $Gdsyn_V$ : 32 voisins

affermir, appuyer, asseoir, assurer, baser, bâtir, commencer, compter, constituer, construire, créer, engendrer, enter, forger, former, instaurer, instituer, justifier, lancer, mettre, motiver, organiser, ouvrir, placer, poser, reposer, table, échafauder, édifier, élever, ériger, établir

### ► Dans $GjdmS_V$ : 15 voisins

affermir, appuyer, asseoir, assoir, baser, bâtir, constituer, créer, former, instaurer, instituer, justifier, édifier, élever, ériger

### ► Dans $GjdmA$ : 47 voisins

acte fondateur, amorcer, aménager, assurer, attaquer, commencer, composer, concevoir, construction, construire, disposer, débiter, démarrer, enfanter, engendrer, engrener, entamer, entreprendre, entreprise, esquisser, fixer, fondateur, fondation, fondement, foyer, imaginer, implanter, installer, inventer, maison, mettre, montrer, partir, placer, poser, presser, production, produire, préluder, réaliser, se fonder, ébaucher, échafauder, élaborer, équilibrer, établir, étrenner

# Méthodes par vision simple

$P = \ll \text{Et cette confiance } \mathbf{fonde} \text{ la responsabilité du praticien. } \gg$

$\omega = \mathbf{fonde}$  le mot-cible de  $P$

$C_p^\omega = \{(\text{NC.confiance, Dep.suj}), (\text{NC.responsabilité, Dep.obj}), (\text{CC.et, Gov.dep\_coord})\}$

## Voisins de *fonder*

### ► Dans **Glm10<sub>v</sub>** : 583 voisins

NC.espoir.Dep.obj (freq=144, IM=120.347), NC.société.Dep.obj (freq=138, IM=59.6531), NC.famille.Dep.obj (freq=98, IM=77.4457), NC.revue.Dep.obj (freq=85, IM=317.996), V.venir.Gov.Prep/de (freq=82, IM=7.29996), NC.principe.Dep.suj (freq=82, IM=84.03), NC.compagnie.Dep.obj (freq=79, IM=120.454), NC.parti.Dep.obj (freq=78, IM=46.1816), NC.association.Dep.obj (freq=78, IM=100.757), NC.valeur.Dep.suj (freq=76, IM=73.5628)

### ► Dans **Gfrwacy** : 824 voisins

NC.famille.Dep.obj (freq=1144, IM=387.762), NC.monastère.Dep.obj (freq=517, IM=4889.38), NC.société.Dep.obj (freq=472, IM=137.059), NC.groupe.Dep.obj (freq=363, IM=71.3103), NC.école.Dep.obj (freq=283, IM=126.836), NC.action.Dep.obj (freq=270, IM=26.9095), V.être.Gov.Prep/de (freq=264, IM=1.91398), V.permettre.Gov.Prep/de (freq=253, IM=2.75423), NC.association.Dep.obj (freq=202, IM=77.0306), NC.compagnie.Dep.obj (freq=200, IM=383.908)

## 7 méthodes par vision simple sur différents graphes lexicaux

Chaque méthode construit une liste ordonnée de lemmes:

- ▶  $T_1$ , liste ordonnée sur l'axe paradigmatique de  $\omega$ 
  - par rapport à  $\omega$
  - indépendamment du contexte  $C_P^\omega$  de la phrase  $P$
- ▶  $T_2$ , liste ordonnée sur l'axe syntagmatique de  $C_P^\omega$ 
  - par rapport  $C_P^\omega$
  - indépendamment de  $\omega$
- ▶  $T_3$ , liste ordonnée
  - par rapport à  $\omega$
  - axes paradigmatiques ou syntagmatiques
  - indépendamment de  $C_P^\omega$

## 7 méthodes par vision simple sur différents graphes lexicaux

Chaque méthode construit une liste ordonnée de lemmes:

- ▶  $T_1$ , liste ordonnée sur l'axe paradigmatique de  $\omega$ 
  - par rapport à  $\omega$
  - indépendamment du contexte  $C_P^\omega$  de la phrase  $P$
- ▶  $T_2$ , liste ordonnée sur l'axe syntagmatique de  $C_P^\omega$ 
  - par rapport  $C_P^\omega$
  - indépendamment de  $\omega$
- ▶  $T_3$ , liste ordonnée
  - par rapport à  $\omega$
  - axes paradigmatiques ou syntagmatiques
  - indépendamment de  $C_P^\omega$

## 7 méthodes par vision simple sur différents graphes lexicaux

Chaque méthode construit une liste ordonnée de lemmes:

- ▶ **T<sub>1</sub>**, liste ordonnée sur l'axe paradigmatique de  $\omega$ 
  - par rapport à  $\omega$
  - indépendamment du contexte  $C_P^\omega$  de la phrase  $P$
- ▶ **T<sub>2</sub>**, liste ordonnée sur l'axe syntagmatique de  $C_P^\omega$ 
  - par rapport  $C_P^\omega$
  - indépendamment de  $\omega$
- ▶ **T<sub>3</sub>**, liste ordonnée
  - par rapport à  $\omega$
  - axes paradigmatiques ou syntagmatiques
  - indépendamment de  $C_P^\omega$

## 7 méthodes par vision simple sur différents graphes lexicaux

Chaque méthode construit une liste ordonnée de lemmes:

- ▶  $T_1$ , liste ordonnée sur l'axe paradigmatique de  $\omega$ 
  - par rapport à  $\omega$
  - indépendamment du contexte  $C_P^\omega$  de la phrase  $P$
- ▶  $T_2$ , liste ordonnée sur l'axe syntagmatique de  $C_P^\omega$ 
  - par rapport  $C_P^\omega$
  - indépendamment de  $\omega$
- ▶  $T_3$ , liste ordonnée
  - par rapport à  $\omega$
  - axes paradigmatiques ou syntagmatiques
  - indépendamment de  $C_P^\omega$

## 7 méthodes par vision simple sur différents graphes lexicaux

Chaque méthode construit une liste ordonnée de lemmes:

- ▶ **T<sub>1</sub>**, liste ordonnée sur l'axe paradigmatique de  $\omega$ 
  - par rapport à  $\omega$
  - indépendamment du contexte  $C_P^\omega$  de la phrase  $P$
- ▶ **T<sub>2</sub>**, liste ordonnée sur l'axe syntagmatique de  $C_P^\omega$ 
  - par rapport  $C_P^\omega$
  - indépendamment de  $\omega$
- ▶ **T<sub>3</sub>**, liste ordonnée
  - par rapport à  $\omega$
  - axes paradigmatiques ou syntagmatiques
  - indépendamment de  $C_P^\omega$

- ▶ les méthodes **T1** et **T3** tiennent compte de la polysémie de la cible  $\omega$
- ▶ les méthodes **T2** tiennent compte du contexte  $C_P^\omega$
- ▶ **T2** renforce les listes **T1** et **T3** sur le sens que prend  $\omega$  dans le contexte  $C_P^\omega$

Méthode	Type de liste
$\mathcal{V}_1 = \vartheta(Gdsyn, \{\omega\}, 3)$	$T_1$
$\mathcal{V}_2 = \vartheta(GjdmS, \{\omega\}, 3)$	$T_1$
$\mathcal{V}_3 = \vartheta(GjdmA, \{\omega\}, 3)$	$T_3$
$\mathcal{V}_4 = \vartheta(Glm10, \{\omega\}, 2)$	$T_1$
$\mathcal{V}_5 = \vartheta(Gfrwac, \{\omega\}, 2)$	$T_1$
$\mathcal{V}_6 = \vartheta(Glm10, C_P^\omega, 3)$	$T_2$
$\mathcal{V}_7 = \vartheta(Gfrwac, C_P^\omega, 3)$	$T_2$

Table: Sept visions simples

# Agregation de vision simples

Soit  $A$  et  $B$ , les vecteurs retournés par deux visions simples de  $\omega$ .

**3 méthodes d'agregation sont définies :**

$$\text{Agreg}_1(\mathbf{A}, \mathbf{B}) = [C]_i = \begin{cases} [A]_i \cdot [B]_i & \text{si } [A]_i \neq 0 \\ & \text{et } [B]_i \neq 0 \\ [A]_i & \text{sinon} \end{cases} \quad (2)$$

$$\text{Agreg}_2(\mathbf{A}, \mathbf{B}) = [C]_i = \begin{cases} [B]_i & \text{si } [A]_i = 0 \\ [A]_i & \text{sinon} \end{cases} \quad (3)$$

$$\text{Agreg}_3(\mathbf{A}, \mathbf{B}) = [C]_i = \begin{cases} [B]_i & \text{si } [A]_i \neq 0 \\ 0 & \text{sinon} \end{cases} \quad (4)$$

## Méthodes soumises à SemDis2014 :

► **Méthode non supervisée :**

$$\vartheta_8 = \text{Agreg}_1(\text{Agreg}_1(\vartheta_4, \vartheta_5), \text{Agreg}_1(\vartheta_6, \vartheta_7))$$

► **Méthode supervisée :**

$$\vartheta_9 = \text{Agreg}_2(\text{Agreg}_1(\text{Agreg}_1(\vartheta_2, \vartheta_3), \vartheta_6), \text{Agreg}_3(\text{Agreg}_2(\vartheta_1, \vartheta_2), \text{Agreg}_1(\text{Agreg}_1(\vartheta_2, \vartheta_3), \vartheta_6)))$$

# Résultats

$P = \ll Et \text{ cette confiance } \mathbf{fonde} \text{ la responsabilité du praticien. } \gg$

$\omega = \mathbf{fonde}$  le mot-cible de  $P$

$C_P^\omega = \{(NC.\text{confiance}, Dep.\text{suj}), (NC.\text{responsabilité}, Dep.\text{obj}), (CC.\text{et}, Gov.\text{dep\_coord})\}$

Gold	créer, forger, constituer, justifier, être à la base, entraîner, assurer, impliquer, baser, instaurer, induire, définir, être à l'origine, établir, installer, poser, supporter
$\vartheta(Gdsyn, \{\omega\}, 3)$	établir, bâtir, créer, construire, faire, organiser, former, constituer, élever, placer
$\vartheta(GjdmS, \{\omega\}, 3)$	bâtir, constituer, créer, élever, établir, édifier, ériger, instaurer, instituer, appuyer
$\vartheta(GjdmA, \{\omega\}, 3)$	responsabilité, confiance, charge, meilleur ami, sureté, affect, condamnation, devoir, poids, dette
$\vartheta(Glm10, \{\omega\}, 2)$	fondre, diriger, présider, animer, créer, rejoindre, perpétuer, abriter, érier, racheter
$\vartheta(Gfrwac, \{\omega\}, 2)$	se marier, ème, rejoindre, pondre, échanger, arranger, engendrer, dater, rivaliser, diriger
$\vartheta(Glm10, C_P^\omega, 3)$	se décréter, se mériter, se rétablir, endosser, se rejeter, se démentir, se renvoyer, saisissant, se évanouir, imputer
$\vartheta(Gfrwac, C_P^\omega, 3)$	généraliser, se mériter, se acquérir, aveugler, se rejeter, endosser, décliner, régner, se décréter assumer

$\mathcal{V}_8$  reposer, se installer, se fonder, assumer, diriger, régner, rejoindre, quitter, animer, se appuyer

$\mathcal{V}_9$  établir, constituer, créer, former, instituer, assurer, mettre, instaurer, poser, construire

Table: Résultats sur la phrase : « Et cette confiance <fonde> la responsabilité du praticien. »

Méthodes	Type	best	oot
$\mathcal{V}_1 = \vartheta(Gdsyn, \{\omega\}, 3)$	supervisée	.0453	.3245
$\mathcal{V}_2 = \vartheta(GjdmS, \{\omega\}, 3)$	supervisée	.0645	.3519
$\mathcal{V}_3 = \vartheta(GjdmA, \{\omega\}, 3)$	supervisée	.0022	.0736
$\mathcal{V}_4 = \vartheta(Glm10, \{\omega\}, 2)$	non-supervisée	.0259	.1347
$\mathcal{V}_5 = \vartheta(Gfrwac, \{\omega\}, 2)$	non-supervisée	.0319	.0799
$\mathcal{V}_6 = \vartheta(Glm10, C_P^\omega, 3)$	non-supervisée	.0061	.0368
$\mathcal{V}_7 = \vartheta(Gfrwac, C_P^\omega, 3)$	non-supervisée	.0024	.0228
<hr/>			
$\mathcal{V}_8$	non-supervisée	.0511	.2129
$\mathcal{V}_9$	supervisée	.0970	.4017

Table: Résultats des méthodes BACANAL

# Résultats

Méthodes	Type	best	oot
$\mathcal{V}_1 = \vartheta(Gdsyn, \{\omega\}, 3)$	supervisée	.0453	.3245
$\mathcal{V}_2 = \vartheta(GjdmS, \{\omega\}, 3)$	supervisée	.0645	.3519
$\mathcal{V}_3 = \vartheta(GjdmA, \{\omega\}, 3)$	supervisée	.0022	.0736
$\mathcal{V}_4 = \vartheta(Gjdm10, \{\omega\}, 3)$	supervisée	.0050	.1047
<b>▶ <math>\mathcal{V}_8 / \mathcal{V}_5</math>: best: +60%; oot: +166%</b>			
<b>▶ <math>\mathcal{V}_9 / \mathcal{V}_2</math>: best: +50%; oot: +14%</b>			
$\mathcal{V}_5 = \vartheta(Gjdm10, \{\omega\}, 3)$	supervisée	.0318	.2287
$\mathcal{V}_6 = \vartheta(Gjdm10, \{\omega\}, 3)$	supervisée	.0318	.2287
$\mathcal{V}_7 = \vartheta(Gjdm10, \{\omega\}, 3)$	supervisée	.0318	.2287
$\mathcal{V}_8$	non-supervisée	.0511	.2129
$\mathcal{V}_9$	supervisée	.0970	.4017

Table: Résultats des méthodes BACANAL

- 1 La substitution lexicale
- 2 Balades Aléatoires
- 3 Application à la substitution lexicale
- 4 Conclusion

# Conclusion

- ▶ Les méthodes BACANAL sont au niveau de l'état de l'art
  - en supervisé ( $oot = .40$ )
  - en non supervisé ( $oot = .21$ )
- ▶ Les méthodes par agrégation améliorent les méthodes « simples »
- ▶ Mais,  $oot < .50 \Rightarrow$  rappel élevé au rang 10 difficile à obtenir
- ▶ Amélioration des résultats : la qualité des graphes prime sur leur taille
  - $\mathcal{V}_4 = \vartheta(Glm10, \{\omega\}, 2)$  ( $best = .0259$  &  $oot = .1347$ )
  - $\mathcal{V}_5 = \vartheta(Gfrwac, \{\omega\}, 2)$  ( $best = .0319$  &  $oot = .0799$ )
- ▶  $\Rightarrow$  **Besoin de ressources de qualité**

MERCI

 Marco Baroni, Silvia Bernardini, Adrian Ferraresi, and Eros Zanchetta.  
The wacky wide web: a collection of very large linguistically processed  
web-crawled corpora.

In Proceedings of the Seventh International Language Resources and  
Evaluation (LREC'09), volume 43(3), pages 209–226, 2009.

 G. Dahl, A.-M. Frassica, and R. Wicentowski.  
SW-AG: Local context matching for english lexical substitution.

In  
Proceedings of the 4th workshop on Semantic Evaluations (SemEval-2007),  
pages 304–307, Prague, Czech Republic, 2007.

 C. Giuliano, A. Gliozzo, and C. Strapparava.  
FBK-irst: Lexical substitution task exploiting domain and syntagmatic  
coherence.

In  
Proceedings of the 4th workshop on Semantic Evaluations (SemEval-2007),  
pages 145–148, Prague, Czech Republic, 2007.

 S. Hassan, A. Csomai, C. Banea, R. Sinha, and R. Mihalcea.

LINT: Subfinder: Combining knowledge resources for automatic linguistic