

# Skillex

Exploiter la structure des réseaux lexicaux  
extraits de ressources lexicales  
pour la modélisation en sciences cognitives

Bruno Gaume

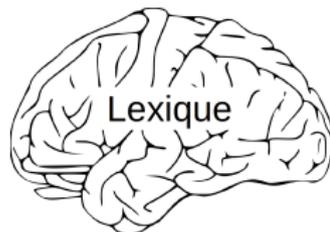
Axe CARTEL

Travaux réalisés en collaboration avec Karine Duvignau (DidAPs)

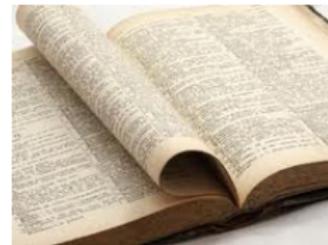
12 Octobre 2015



# Motivation



Artefact Cognitif



## Question

Les dictionnaires peuvent-ils être utilisés pour modéliser une partie du fonctionnement cognitif du lexique ?

## Réponse du conseil scientifique d'un appel à projet (2000) :

Le projet repose sur un pari : penser que la théorie des graphes permettra de faire émerger des propriétés sémantiques « cachées » dans les dictionnaires, pour lequel il a semblé au Conseil que toutes les précautions n'avaient pas été prises. Les définitions des dictionnaires sont des textes extrêmement variables d'une entrée à l'autre et d'un dictionnaire à l'autre. Les auteurs n'indiquent pas quel dictionnaire ils vont exploiter, et espèrent gérer la variabilité au moyen de méthodes qu'ils ont expérimentées dans des contextes bien différents. Les chances sont très faibles de trouver des relations lexicales fines en partant des données grossières décrites (cooccurrence dans une définition – fréquences). **Quand bien même des relations seraient trouvées ainsi, la pertinence de cette méthodologie vis-à-vis des sciences cognitives resterait à démontrer.**

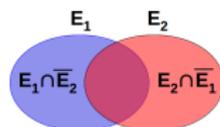
- 1 Structure des réseaux lexicaux
- 2 Un score d'habilité lexicale: *Skillex*
- 3 Perspectives

- 1 Structure des réseaux lexicaux 
- 2 Un score d'habilité lexicale: *Skillex*
- 3 Perspectives

# 1 Structure des réseaux lexicaux : Problème

**Constat** : Il existe de nombreux désaccords entre les membres d'une même communauté linguistique :

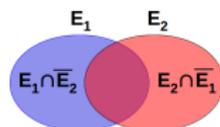
- ▶ **Beaucoup** de paires de mots sont synonymes pour l'individu  $I_1$  et pas pour l'individu  $I_2$  et vice et versa



# 1 Structure des réseaux lexicaux : Problème

**Constat** : Il existe de nombreux désaccords entre les membres d'une même communauté linguistique :

- ▶ **Beaucoup** de paires de mots sont synonymes pour l'individu  $I_1$  et pas pour l'individu  $I_2$  et vice et versa



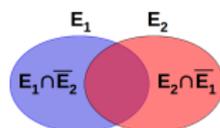
**Incompatible avec :**

**Hypothèse** : La synonymie reflète la structure sémantique du lexique commune aux membres d'une même communauté linguistique

# 1 Structure des réseaux lexicaux : Problème

**Constat** : Il existe de nombreux désaccords entre les membres d'une même communauté linguistique :

- ▶ **Beaucoup** de paires de mots sont synonymes pour l'individu  $I_1$  et pas pour l'individu  $I_2$  et vice et versa



**Résoudre cette contradiction apparente**

**Hypothèse** : La synonymie reflète la structure sémantique du lexique commune aux membres d'une même communauté linguistique



**bavarder** : parler, discourir, causer

**causer** : parler, discuter, bavarder, provoquer, exciter

**discourir** : parler, palabrer, pérorer, bavarder

**discuter** : parler, palabrer, causer, pérorer

**exciter** : causer, provoquer

**palabrer** : parler, discourir, pérorer, discuter

**parler** : palabrer, discourir, causer, pérorer, discuter, bavarder

**pérorer** : parler, palabrer, discourir, discuter

**provoquer** : causer, exciter

**bavarder** : parler, discourir, causer

**causer** : parler, discuter, bavarder, provoquer, exciter

**discourir** : parler, palabrer, pérorer, bavarder

**discuter** : parler, palabrer, causer, pérorer

**exciter** : causer, provoquer

**palabrer** : parler, discourir, pérorer, discuter

**parler** : palabrer, discourir, causer, pérorer, discuter, bavarder

**pérorer** : parler, palabrer, discourir, discuter

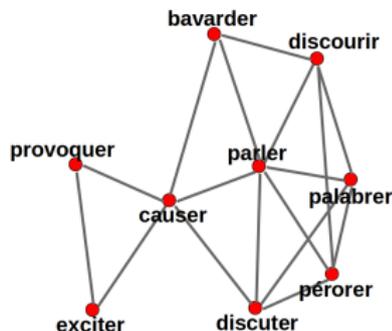
**provoquer** : causer, exciter

$G = (V, E)$  où  $E \subseteq P_2^V$

$V = \{\text{bavarder, causer, discourir, discuter, exciter, palabrer, parler, pérorer, provoquer}\}$

$E = \{\{\text{bavarder, parler}\}, \{\text{bavarder, discourir}\}, \{\text{bavarder, causer}\}, \{\text{causer, discuter}\}, \{\text{causer, parler}\}, \{\text{causer, provoquer}\}, \{\text{causer, exciter}\}, \{\text{discourir, parler}\}, \{\text{discourir, palabrer}\}, \{\text{discourir, pérorer}\}, \{\text{discuter, parler}\}, \{\text{discuter, palabrer}\}, \{\text{discuter, pérorer}\}, \{\text{exciter, provoquer}\}, \{\text{palabrer, parler}\}, \{\text{palabrer, pérorer}\}, \{\text{parler, pérorer}\}\}$

Un graphe de  $n = |V|$  sommets et  $m = |E|$  arêtes



$G = (V, E)$  où  $E \subseteq P_2^V$

$V = \{\text{bavarder, causer, discourir, discuter, exciter, palabrer, parler, pérorer, provoquer}\}$

$E = \{\{\text{bavarder, parler}\}, \{\text{bavarder, discourir}\}, \{\text{bavarder, causer}\}, \{\text{causer, discuter}\}, \{\text{causer, parler}\}, \{\text{causer, provoquer}\}, \{\text{causer, exciter}\}, \{\text{discourir, parler}\}, \{\text{discourir, palabrer}\}, \{\text{discourir, pérorer}\}, \{\text{discuter, parler}\}, \{\text{discuter, palabrer}\}, \{\text{discuter, pérorer}\}, \{\text{exciter, provoquer}\}, \{\text{palabrer, parler}\}, \{\text{palabrer, pérorer}\}, \{\text{parler, pérorer}\}\}$

Un graphe de  $n = |V|$  sommets et  $m = |E|$  arêtes

## Structures des graphes de terrain

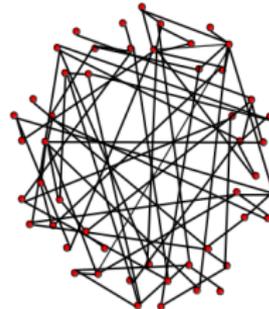
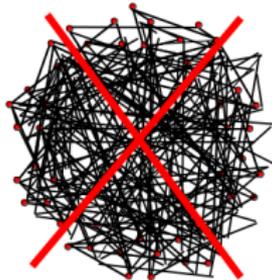
## Structures des graphes de terrain

- ▶ Les graphes Lexicaux
- ▶ Les graphes d'accointance d'un groupe d'humains
- ▶ Le graphe des appels téléphoniques aux USA
- ▶ Le graphe du World Wide Web
- ▶ Le graphe des collaborations cinématographiques
- ▶ Le graphe de *Caenorhabditis elegans*

## Structures des graphes de terrain

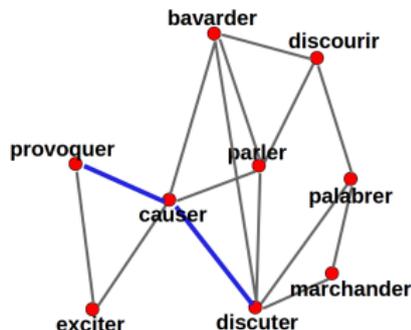
- ▶ Faible densité
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde (loi de puissance)
- ▶ Fort coefficient de clustering : zones denses en arêtes

## Structures des graphes de terrain



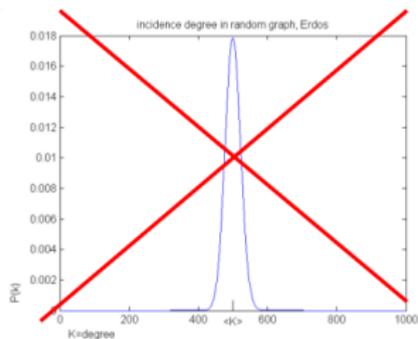
- ▶ **Faible densité**
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde (loi de puissance)
- ▶ Fort coefficient de clustering : zones denses en arêtes

## Structures des graphes de terrain



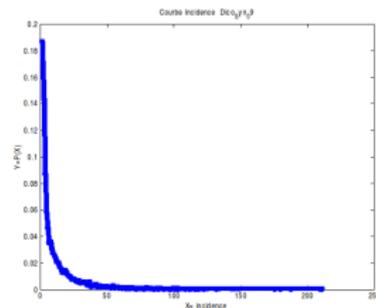
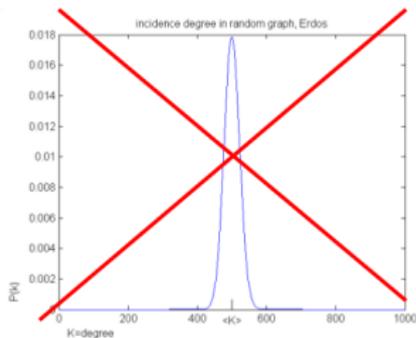
- ▶ Faible densité
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde (loi de puissance)
- ▶ Fort coefficient de clustering : zones denses en arêtes

## Structures des graphes de terrain



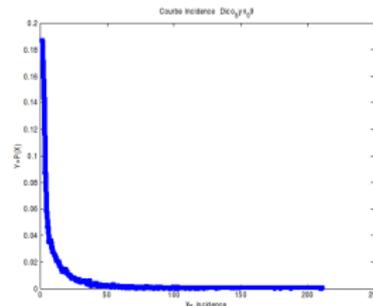
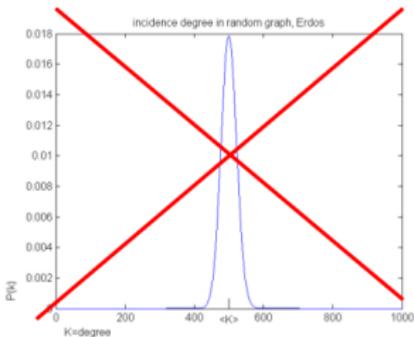
- ▶ Faible densité
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde (loi de puissance)
- ▶ Fort coefficient de clustering : zones denses en arêtes

## Structures des graphes de terrain



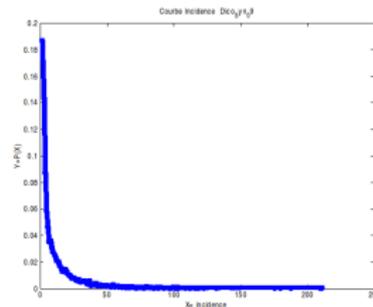
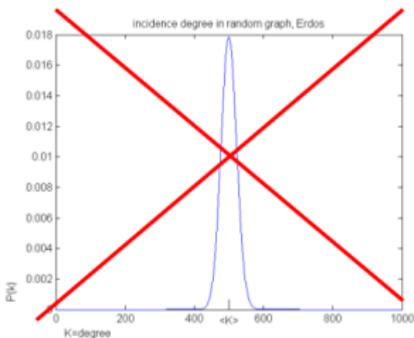
- ▶ Faible densité
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde (loi de puissance)
- ▶ Fort coefficient de clustering : zones denses en arêtes

## Structures des graphes de terrain



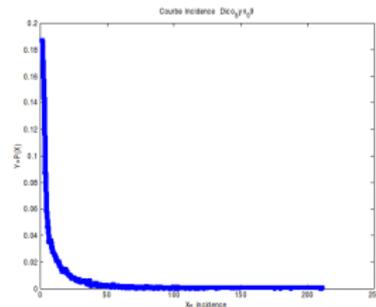
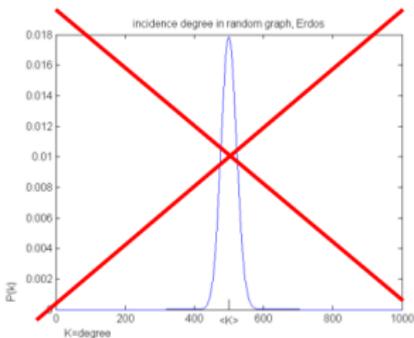
- ▶ Faible densité
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde (loi de puissance)
- ▶ Fort coefficient de clustering : zones denses en arêtes

## Structures des graphes de terrain



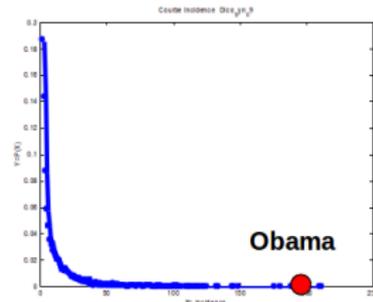
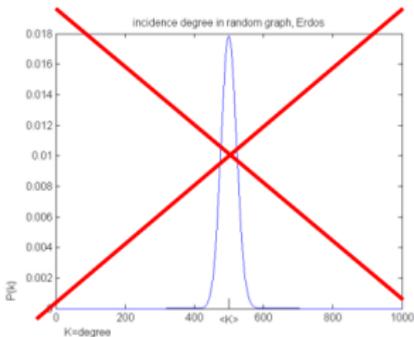
- ▶ Faible densité
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde (loi de puissance)
- ▶ Fort coefficient de clustering : zones denses en arêtes

## Structures des graphes de terrain



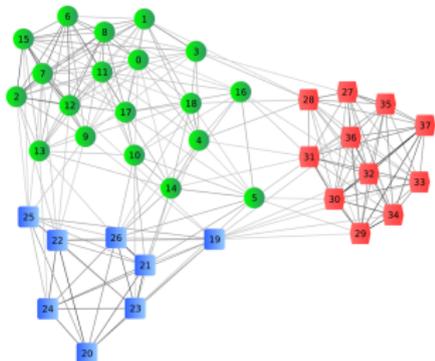
- ▶ Faible densité
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde (loi de puissance)
- ▶ Fort coefficient de clustering : zones denses en arêtes

## Structures des graphes de terrain



- ▶ Faible densité
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde (loi de puissance)
- ▶ Fort coefficient de clustering : zones denses en arêtes

## Structures des graphes de terrain



- ▶ Faible densité
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde (loi de puissance)
- ▶ Fort coefficient de clustering : zones denses en arêtes

## Structures des graphes de terrain

### Réseaux Petits Mondes Hiérarchiques Clusterisés

- ▶ Faible densité
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde (loi de puissance)
- ▶ Fort coefficient de clustering : zones denses en arêtes

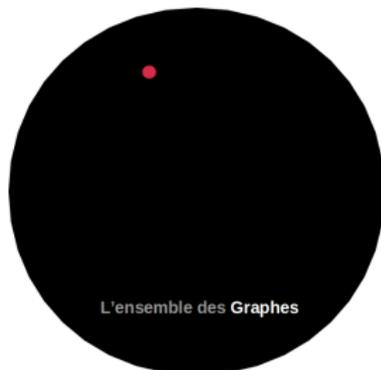
## Structures des graphes de terrain

- Les graphes Lexicaux
- Les graphes d'accointance d'un groupe d'humains
- Le graphe des appels téléphoniques aux USA
- Le graphe du World Wide Web
- Le graphe des collaborations cinématographiques
- Le graphe de *Caenorhabditis elegans*

## Réseaux Petits Mondes Hiérarchiques Clusterisés

- ▶ Faible densité
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde (loi de puissance)
- ▶ Fort coefficient de clustering : zones denses en arêtes

## Structures des graphes de terrain

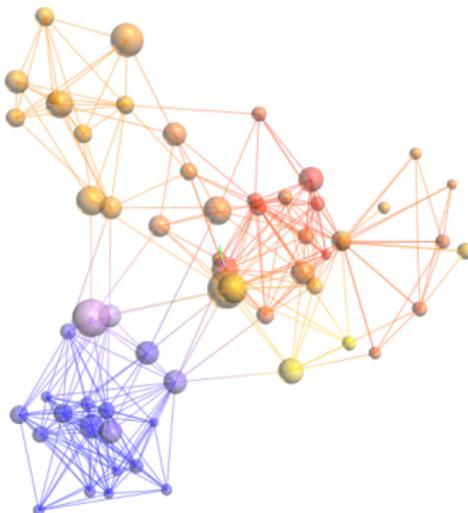


## Réseaux Petits Mondes Hiérarchiques Clusterisés

- ▶ Faible densité
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde (loi de puissance)
- ▶ Fort coefficient de clustering : zones denses en arêtes

Les réseaux Lexicaux extraits des dictionnaires sont des Réseaux Petits Mondes Hiérarchiques Clusterisés

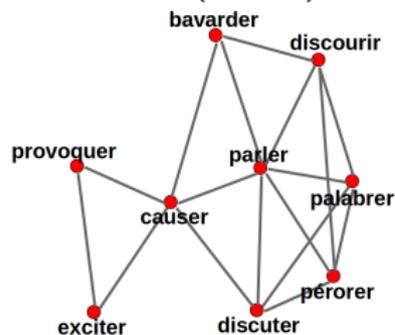
[Gaume, 2004, Steyvers and Tenenbaum, 2005, De Deyne and Storms, 2008, Morais et al., 2013]



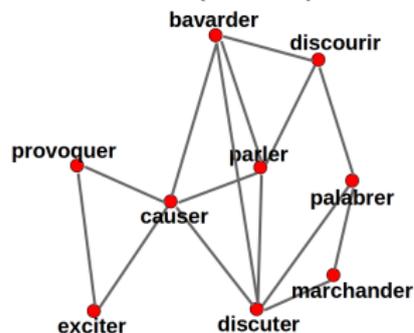
# Comparer $E_1$ et $E_2$ comme des «sacs de liens»

# Comparer $E_1$ et $E_2$ comme des «sacs de liens»

$$G_1 = (V_1, E_1)$$

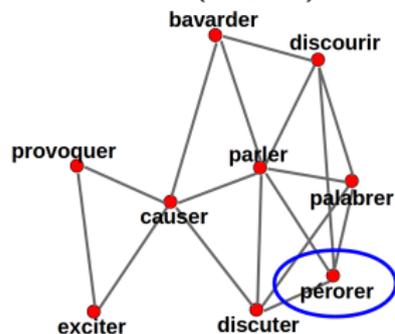


$$G_2 = (V_2, E_2)$$

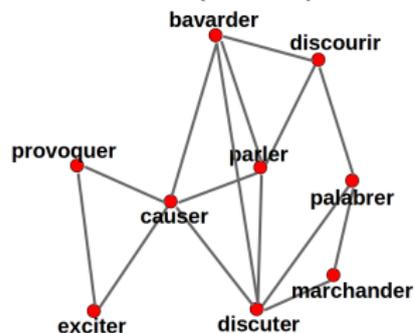


# Comparer $E_1$ et $E_2$ comme des «sacs de liens»

$$G_1 = (V_1, E_1)$$

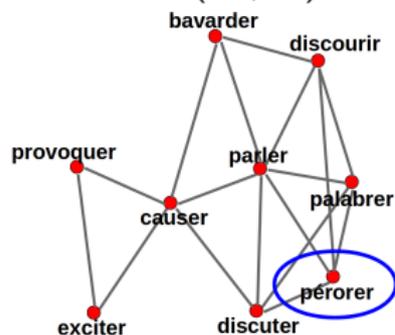


$$G_2 = (V_2, E_2)$$

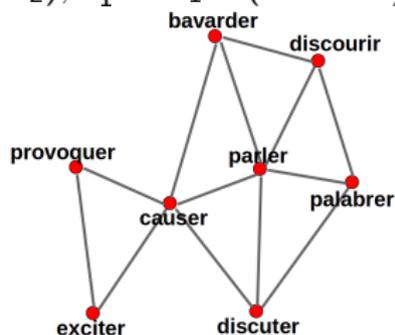


# Comparer $E_1$ et $E_2$ comme des «sacs de liens»

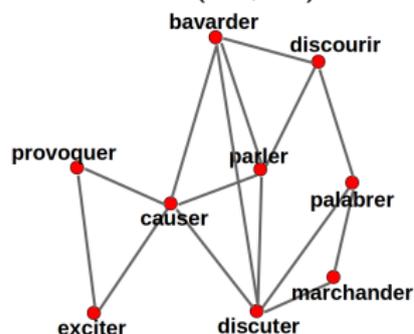
$$G_1 = (V_1, E_1)$$



$$G'_1 = (V' = (V_1 \cap V_2), E'_1 = E_1 \cap (V' \times V'))$$

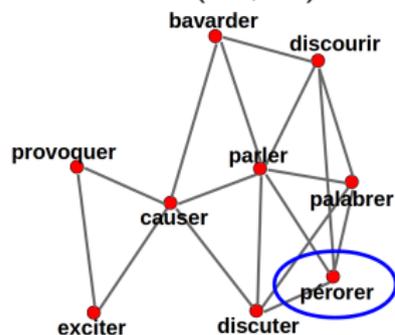


$$G_2 = (V_2, E_2)$$

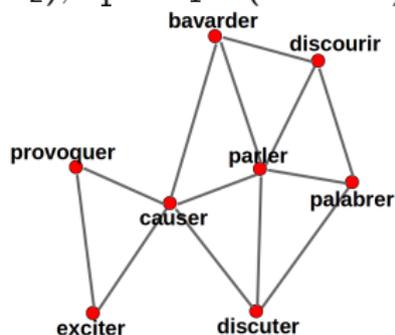


# Comparer $E_1$ et $E_2$ comme des «sacs de liens»

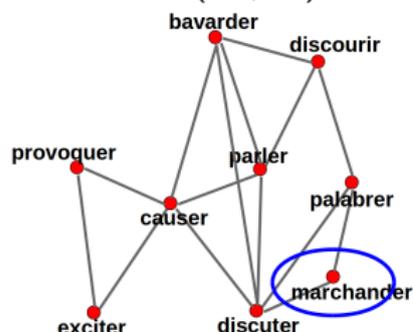
$$G_1 = (V_1, E_1)$$



$$G'_1 = (V' = (V_1 \cap V_2), E'_1 = E_1 \cap (V' \times V'))$$

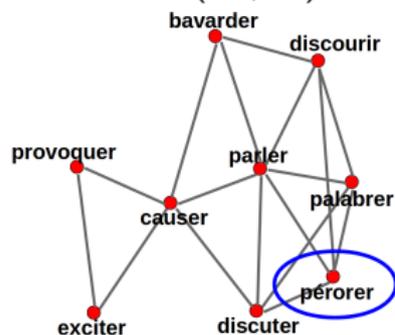


$$G_2 = (V_2, E_2)$$

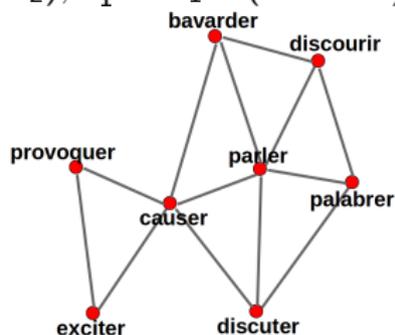


# Comparer $E_1$ et $E_2$ comme des «sacs de liens»

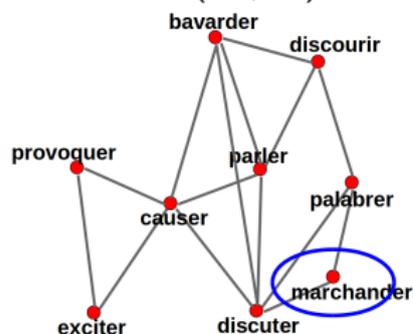
$$G_1 = (V_1, E_1)$$



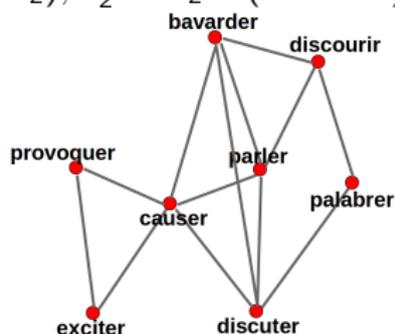
$$G'_1 = (V' = (V_1 \cap V_2), E'_1 = E_1 \cap (V' \times V'))$$



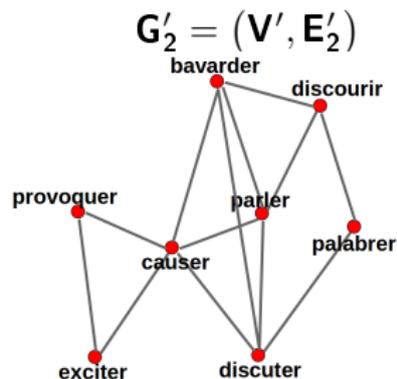
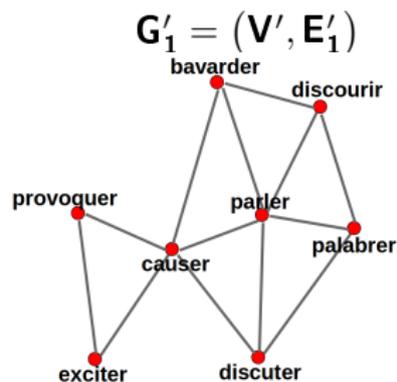
$$G_2 = (V_2, E_2)$$



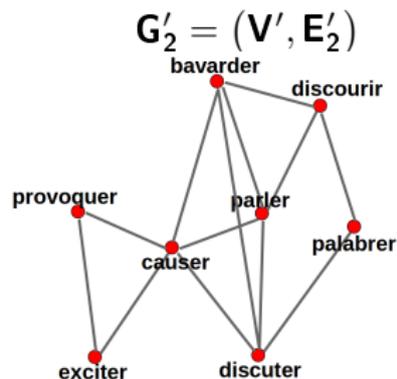
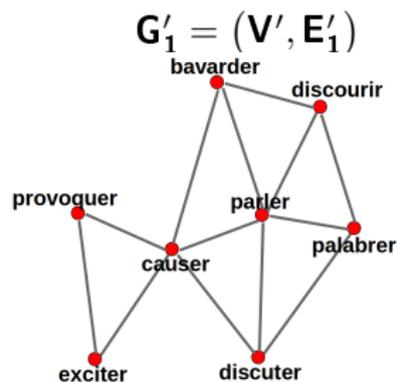
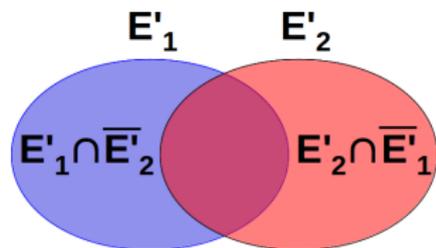
$$G'_2 = (V' = (V_1 \cap V_2), E'_2 = E_2 \cap (V' \times V'))$$



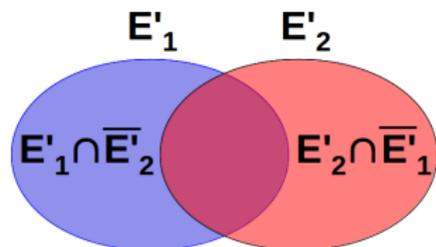
# Comparer $E_1$ et $E_2$ comme des «sacs de liens»



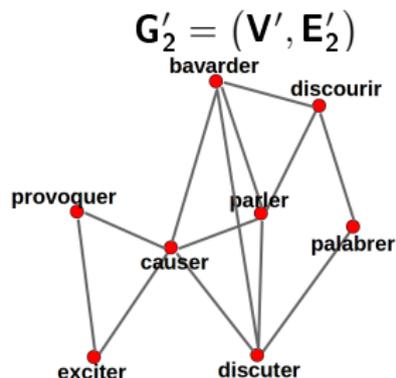
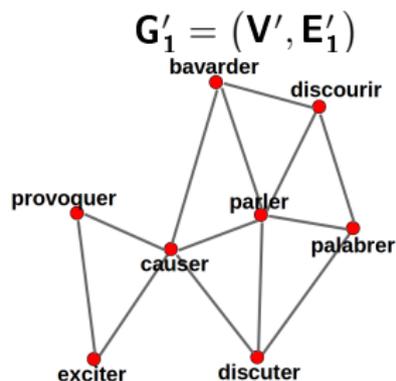
# Comparer $E_1$ et $E_2$ comme des «sacs de liens»



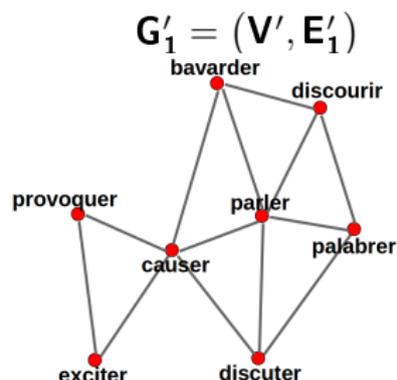
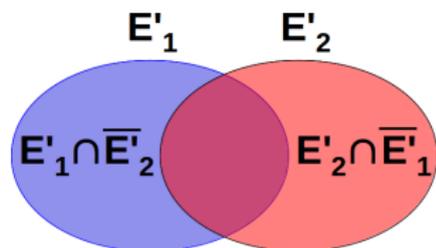
# Comparer $E_1$ et $E_2$ comme des «sacs de liens»



$$GED(G'_1, G'_2) = \frac{|E'_1 \cap \overline{E'_2}| + |E'_2 \cap \overline{E'_1}|}{|E'_1| + |E'_2|} \in [0, 1]$$

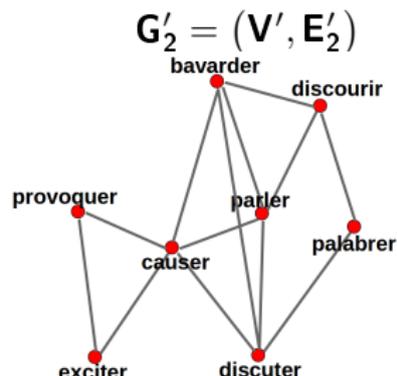


# Comparer $E_1$ et $E_2$ comme des «sacs de liens»

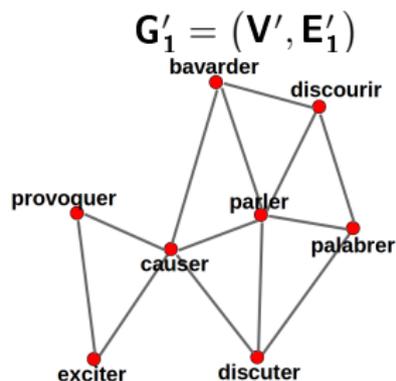
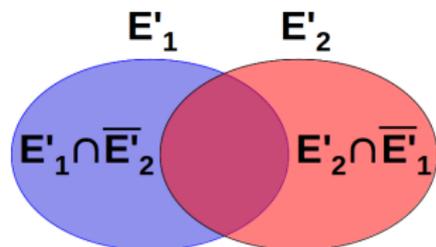


$$GED(G'_1, G'_2) = \frac{|E'_1 \cap \overline{E'_2}| + |E'_2 \cap \overline{E'_1}|}{|E'_1| + |E'_2|} \in [0, 1]$$

$$GED(Lar', Rob') = 0.47$$



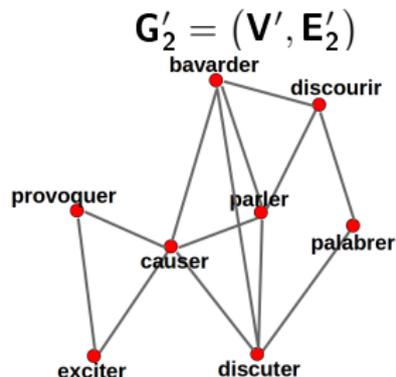
# Comparer $E_1$ et $E_2$ comme des «sacs de liens»



$$GED(G'_1, G'_2) = \frac{|E'_1 \cap \overline{E'_2}| + |E'_2 \cap \overline{E'_1}|}{|E'_1| + |E'_2|} \in [0, 1]$$

$$GED(Lar', Rob') = 0.47$$

Nombreux désaccords entre Larousse et Robert

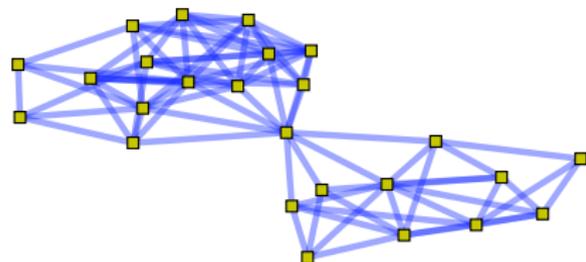


# Hypothèse

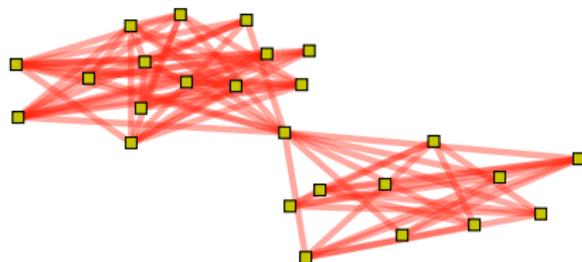
Nombreux désaccords **au niveau des arêtes**

Nombreux désaccords **au niveau des arêtes**

$G_1 = (V, E_1)$



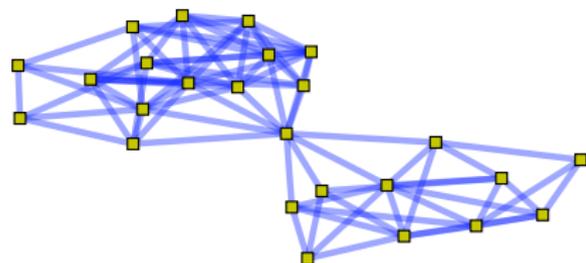
$G_2 = (V, E_2)$



**Aucune arête commune** entre ces deux graphes.

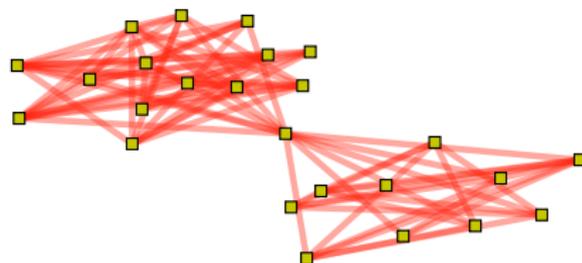
Nombreux désaccords **au niveau des arêtes**

$G_1 = (V, E_1)$



$\text{GED}(G_1, G_2) = 1$

$G_2 = (V, E_2)$



**Aucune arête commune** entre ces deux graphes.

Nombreux désaccords **au niveau des arêtes**

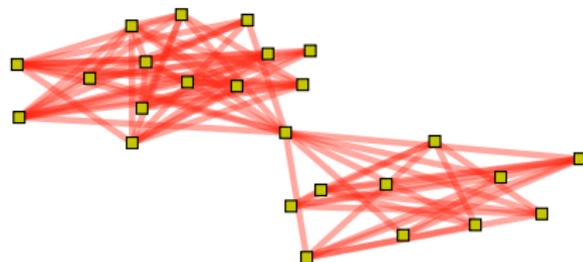
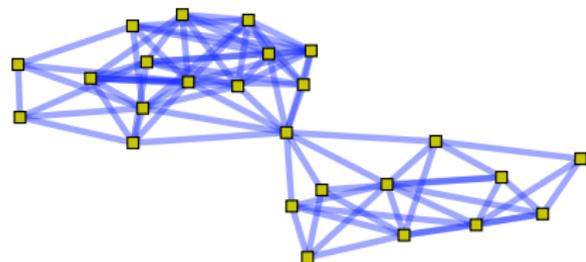
Mais

Bon accord **au niveau des clusters...**

$G_1 = (V, E_1)$

$\text{GED}(G_1, G_2) = 1$

$G_2 = (V, E_2)$



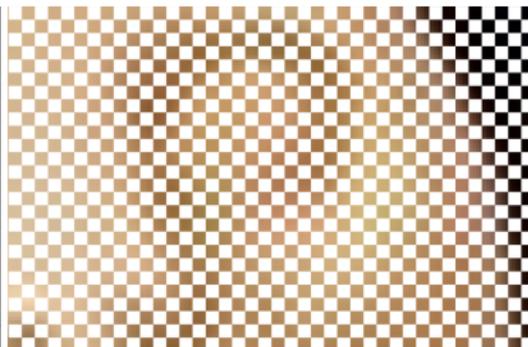
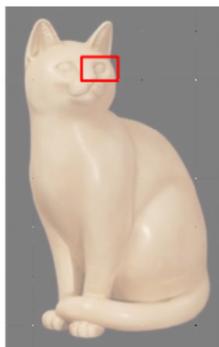
**Aucune arête commune** entre ces deux graphes.

# Métaphore

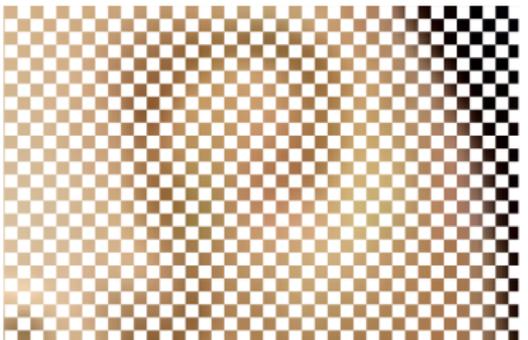
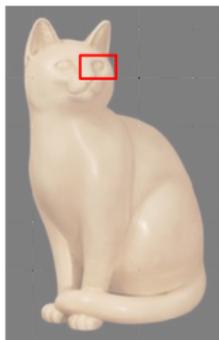
# Métaphore



La photo d'un chat



Chaque pixel **impair** est peint en blanc



Chaque pixel **pair** est peint en blanc

# Métaphore

*Pair =*



# Métaphore

*Pair =*



*Impair =*



# Métaphore

*Pair* =



*Impair* =



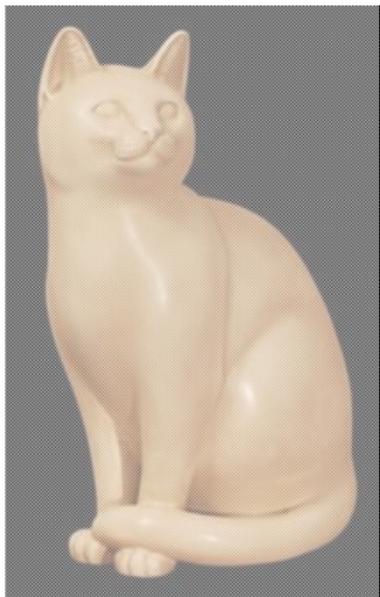
**Aucun pixel commun**  
**Images distantes (au niveau des pixels)**  
**Cependant ces deux images se ressemblent**

# Métaphore

*Pair* =



*Impair* =

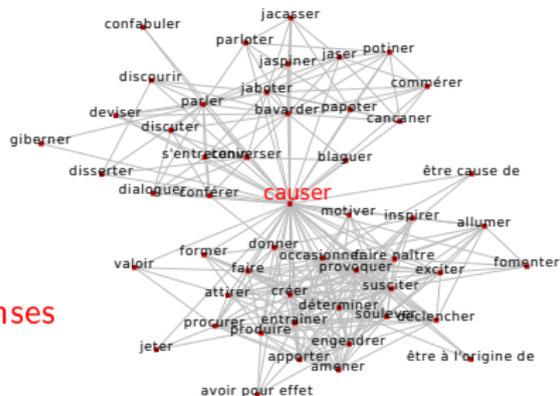


# Marches Aléatoires



## Réseaux Petits Mondes Hiérarchiques

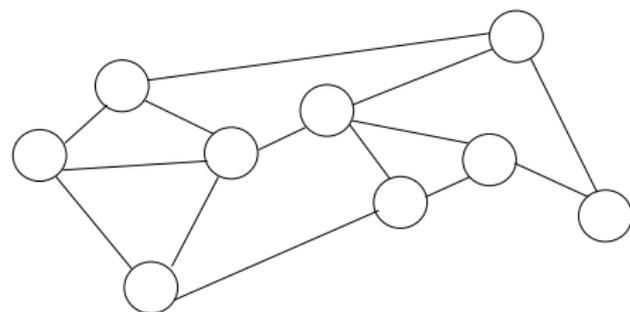
- ▶ Faible densité
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde
- ▶ Fort coefficient de clustering : **zones denses**



## Marches Aléatoires

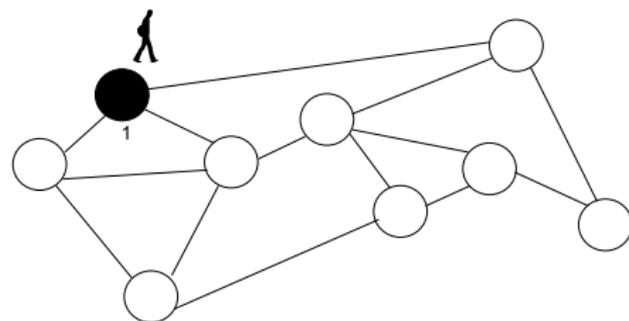
- ▶ **Idée**: si  $(u, v)$  sont **dans un même "cluster"**, alors **beaucoup de chemins** mènent de  $u$  à  $v$  (même si ils ne sont pas adjacents).
- ▶ Un **marcheur aléatoire** a tendance à être capturé par les "clusters"

# Marches Aléatoires



- 1 État initial ( $t=0$ )
- 2  $t = 1$
- 3  $t = 2$
- 4  $t = 3$

# Marches Aléatoires



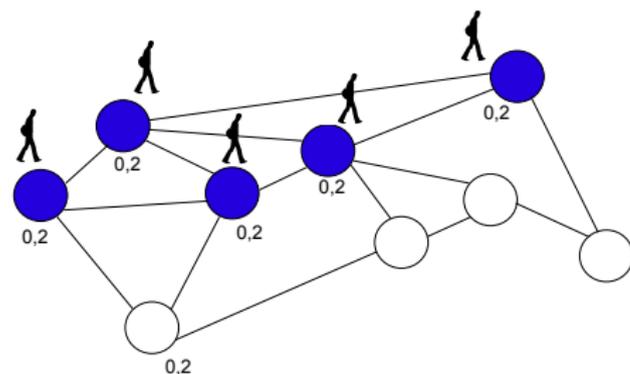
1 État initial ( $t=0$ )

2  $t = 1$

3  $t = 2$

4  $t = 3$

# Marches Aléatoires



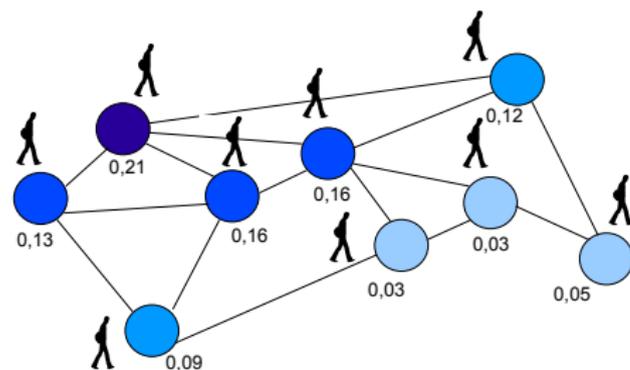
1 État initial ( $t=0$ )

2  $t = 1$

3  $t = 2$

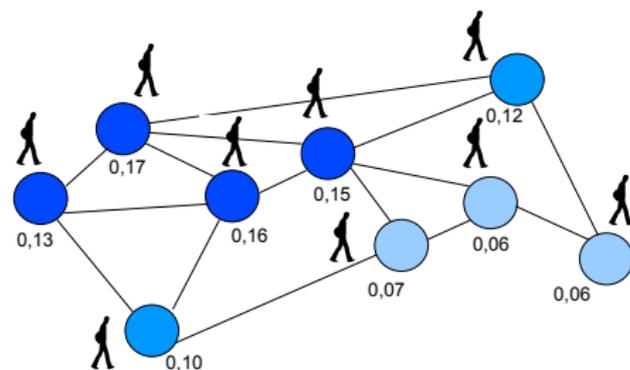
4  $t = 3$

# Marches Aléatoires



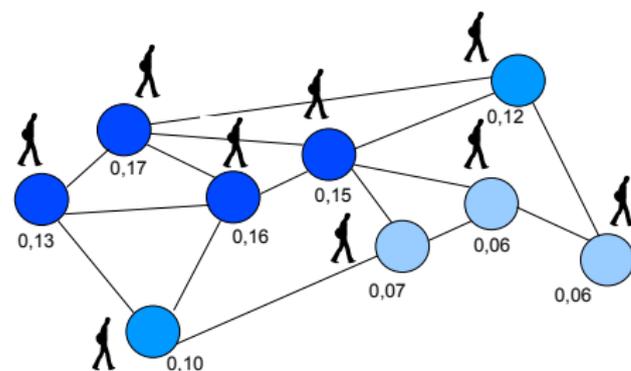
- 1 État initial ( $t=0$ )
- 2  $t = 1$
- 3  $t = 2$
- 4  $t = 3$

# Marches Aléatoires



- 1 État initial ( $t=0$ )
- 2  $t = 1$
- 3  $t = 2$
- 4  $t = 3$

# Marches Aléatoires



- 1 État initial ( $t=0$ )
- 2  $t = 1$
- 3  $t = 2$
- 4  $t = 3$

La Proxémie :  $P_G^t(u, v)$

Dans  $G$ , probabilité d'être sur  $v$  à l'instant  $t$  en partant de  $u$

# Marches Aléatoires

Liste ordonnées des verbes selon leur score de proxémie à partir de *déshabiller*  
( $t = 4$ )

1) dépouiller*	11) dévoiler	21) arracher	31) tailler	41) dire
2) défaire*	12) dégager	22) délacer*	32) développer	42) ouvrir
3) démunir*	13) étaler*	23) voler	33) déchirer	43) déployer
4) déshabiller*	14) ôter*	24) dénouer	34) déchausser	44) dépiauter
5) dévêtir*	15) enlever	25) desserrer	35) vider	45) déposséder
6) révéler	16) défrusquer	26) médire*	36) peler	46) frustrer
7) découvrir*	17) afficher	27) débarrasser	37) dépourvoir	47) indiquer
8) montrer*	18) désaffubler	28) exposer	38) défubler	48) démontrer
9) dénuder*	19) écorcher*	29) couper	39) priver	49) manifester
10) dégarnir*	20) démasquer*	30) exhiber*	40) trahir	50) détacher

**Table:** Les cinquante verbes qui ont la plus forte probabilité d'être atteints en 4 pas depuis *déshabiller* (sur 10000 verbes).

La Proxémie :  $P_G^t(u, v)$

Dans  $G$ , probabilité d'être sur  $v$  à l'instant  $t$  en partant de  $u$

# Marches Aléatoires

Liste ordonnées des verbes selon leur score de proxémie à partir de *déshabiller*  
( $t = 4$ )

1) dépouiller*	11) dévoiler	21) arracher	31) tailler	41) dire
2) défaire*	12) dégager	22) délacer*	32) développer	42) ouvrir
3) démunir*	13) étaler*	23) voler	33) déchirer	43) déployer
4) <b>déshabiller*</b>	14) ôter*	24) dénouer	34) déchausser	44) dépiauter
5) dévêtir*	15) enlever	25) desserrer	35) vider	45) déposséder
6) révéler	16) défrusquer	26) médire*	36) peler	46) frustrer
7) découvrir*	17) afficher	27) débarrasser	37) dépourvoir	47) indiquer
8) montrer*	18) désaffubler	28) exposer	38) défubler	48) démontrer
9) dénuder*	19) écorcher*	29) couper	39) priver	49) manifester
10) dégarnir*	20) démasquer*	30) exhiber*	40) trahir	50) détacher

**Table:** Les cinquante verbes qui ont la plus forte probabilité d'être atteints en 4 pas depuis *déshabiller* (sur 10000 verbes).

La Proxémie :  $P_G^t(u, v)$

Dans  $G$ , probabilité d'être sur  $v$  à l'instant  $t$  en partant de  $u$

# Marches Aléatoires

Liste ordonnées des verbes selon leur score de proxémie à partir de *déshabiller*  
( $t = 4$ )

1) dépouiller*	11) dévoiler	21) arracher	31) tailler	41) dire
2) défaire*	12) dégager	22) délacer*	32) développer	42) ouvrir
3) démunir*	13) étaler*	23) voler	33) déchirer	43) déployer
4) déshabiller*	14) ôter*	24) dénouer	34) déchausser	44) dépiauter
5) dévêtir*	15) enlever	25) desserrer	35) vider	45) déposséder
6) révéler	16) défrusquer	26) médire*	36) peler	46) frustrer
7) découvrir*	17) afficher	27) débarrasser	37) dépourvoir	47) indiquer
8) montrer*	18) désaffubler	28) exposer	38) défubler	48) démontrer
9) dénuder*	19) écorcher*	29) couper	39) priver	49) manifester
10) dégarnir*	20) démasquer*	30) exhiber*	40) trahir	50) détacher

**Table:** Les cinquante verbes qui ont la plus forte probabilité d'être atteints en 4 pas depuis *déshabiller* (sur 10000 verbes).

La Proxémie :  $P_G^t(u, v)$

Dans  $G$ , probabilité d'être sur  $v$  à l'instant  $t$  en partant de  $u$

# Marches Aléatoires

Liste ordonnées des verbes selon leur score de proxémie à partir de *déshabiller*  
( $t = 4$ )

1) dépouiller*	11) dévoiler	21) arracher	31) tailler	41) dire
2) défaire*	12) dégager	22) délacer*	32) développer	42) ouvrir
3) démunir*	13) étaler*	23) voler	33) déchirer	43) déployer
4) déshabiller*	14) ôter*	24) dénouer	34) déchausser	44) dépiauter
5) dévêtir*	15) enlever	25) desserrer	35) vider	45) déposséder
6) révéler	16) défrusquer	26) médire*	36) peler	46) frustrer
7) découvrir*	17) afficher	27) débarrasser	37) dépourvoir	47) indiquer
8) montrer*	18) désaffubler	28) exposer	38) défubler	48) démontrer
9) dénuder*	19) écorcher*	29) couper	39) priver	49) manifester
10) dégarnir*	20) démasquer*	30) exhiber*	40) trahir	50) détacher

**Table:** Les cinquante verbes qui ont la plus forte probabilité d'être atteints en 4 pas depuis *déshabiller* (sur 10000 verbes).

La Proxémie :  $P_G^t(u, v)$

Dans  $G$ , probabilité d'être sur  $v$  à l'instant  $t$  en partant de  $u$



- ▶ **Marches longues:** la probabilité d'atteindre un sommet  $v$  ne dépend que du degrés de  $v$ :

$$\lim_{t \rightarrow \infty} P_G^t(u, v) = \frac{\deg_G(v)}{\sum_{r \in V} \deg_G(r)} = \pi_G(v)$$

- ▶ **Marches longues:** la probabilité d'atteindre un sommet  $v$  ne dépend que du degrés de  $v$ :

$$\lim_{t \rightarrow \infty} P_G^t(u, v) = \frac{\deg_G(v)}{\sum_{r \in V} \deg_G(r)} = \pi_G(v)$$

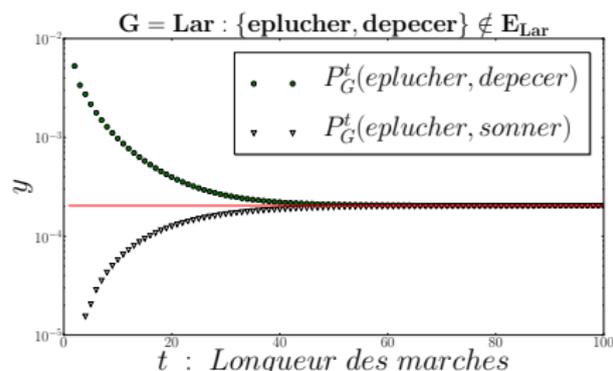
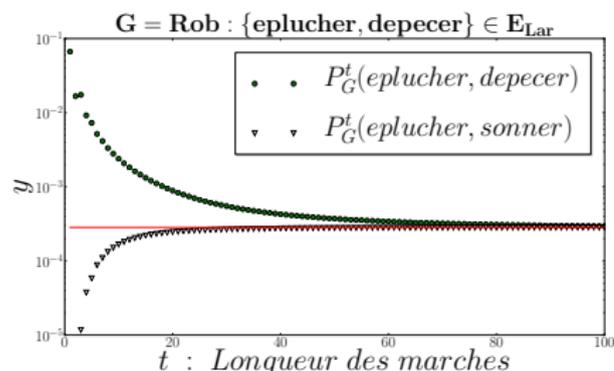
- ▶ **Marches courtes:** forte probabilité de rester dans un cluster:
- $P_G^t(u, v) > \pi_G(v)$  si  $u$  et  $v$  sont dans un même cluster
  - $P_G^t(u, v) < \pi_G(v)$  sinon

- ▶ **Marches longues:** la probabilité d'atteindre un sommet  $v$  ne dépend que du degrés de  $v$ :

$$\lim_{t \rightarrow \infty} P_G^t(u, v) = \frac{\deg_G(v)}{\sum_{r \in V} \deg_G(r)} = \pi_G(v)$$

- ▶ **Marches courtes:** forte probabilité de rester dans un cluster:

- $P_G^t(u, v) > \pi_G(v)$  si  $u$  et  $v$  sont dans un même cluster
- $P_G^t(u, v) < \pi_G(v)$  sinon



- ▶ **Marches longues:** la probabilité d'atteindre un sommet  $v$  ne dépend que du degrés de  $v$ :

$$\lim_{t \rightarrow \infty} P_G^t(u, v) = \frac{\deg_G(v)}{\sum_{r \in V} \deg_G(r)} = \pi_G(v)$$

- ▶ **Marches courtes:** forte probabilité de rester dans un cluster:
  - $P_G^t(u, v) > \pi_G(v)$  si  $u$  et  $v$  sont dans un même cluster
  - $P_G^t(u, v) < \pi_G(v)$  sinon

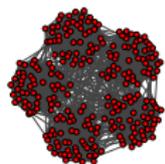
$$t = 5 : \quad \mathbf{CONF}_G(\mathbf{u}, \mathbf{v}) = \frac{P_G^5(\mathbf{u}, \mathbf{v})}{P_G^5(\mathbf{u}, \mathbf{v}) + \pi_G(\mathbf{v})}$$

$CONF_G$  est une mesure de proximité normalisée basée sur les marches aléatoires courtes dans  $G$ :

- ▶ Beaucoup de chemins courts entre  $u$  et  $v$  :  $CONF_G(u, v) > 0,5$ ;
- ▶ Peu de chemins courts entre  $u$  et  $v$  :  $CONF_G(u, v) < 0,5$ ;

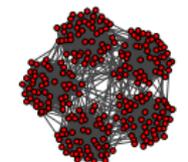
# Confluence: Expérience contrôlée

# Confluence: Expérience contrôlée

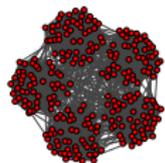


$$G_a = (V, E_a)$$

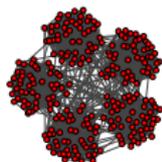
# Confluence: Expérience contrôlée



$$G_b = (V, E_b)$$

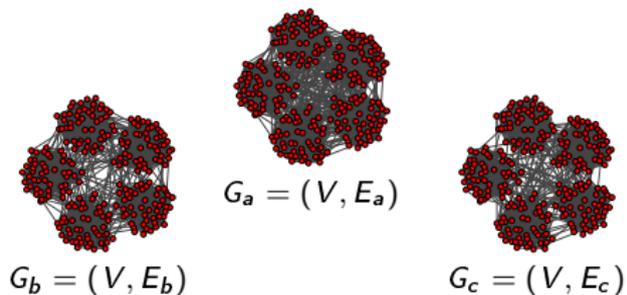


$$G_a = (V, E_a)$$



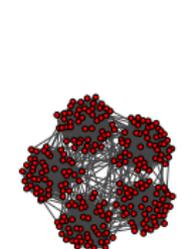
$$G_c = (V, E_c)$$

# Confluence: Expérience contrôlée

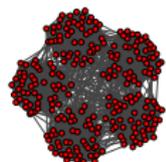


$$E_b \cap E_c = \emptyset \Rightarrow$$
$$GED(G_b, G_c) = 1$$

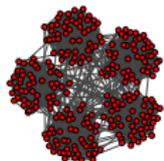
# Confluence: Expérience contrôlée



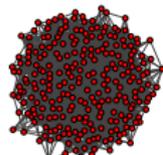
$$G_b = (V, E_b)$$



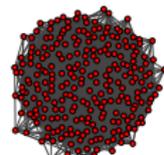
$$G_a = (V, E_a)$$



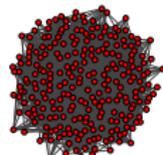
$$G_c = (V, E_c)$$



$$G_b^R = (V, E_b^R)$$



$$G_a^R = (V, E_a^R)$$

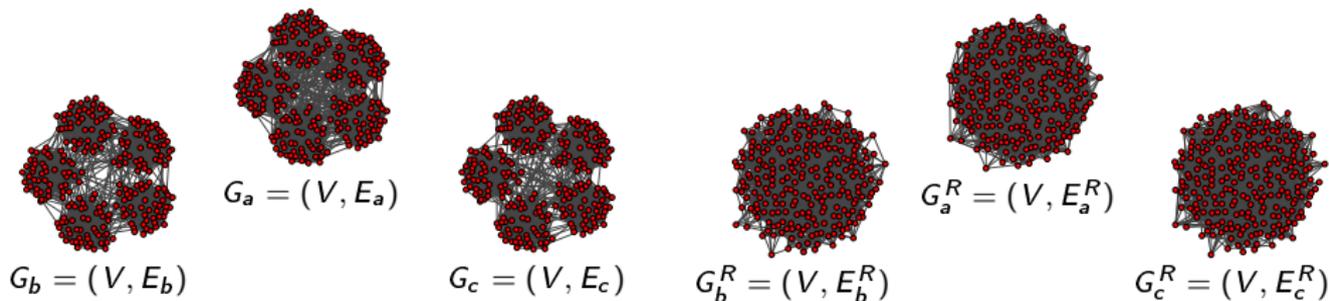


$$G_c^R = (V, E_c^R)$$

$$E_b \cap E_c = \emptyset \Rightarrow \\ GED(G_b, G_c) = 1$$

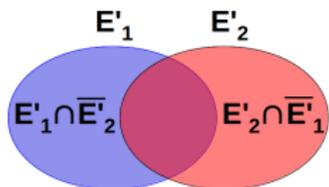
$$E_b^R \cap E_c^R = \emptyset \Rightarrow \\ GED(G_b^R, G_c^R) = 1$$

# Confluence: Expérience contrôlée

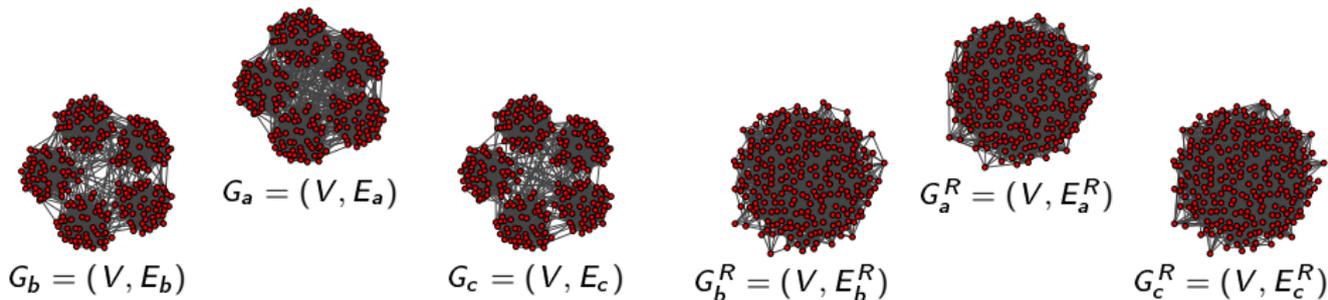


$$E_b \cap E_c = \emptyset \Rightarrow \\ GED(G_b, G_c) = 1$$

$$E_b^R \cap E_c^R = \emptyset \Rightarrow \\ GED(G_b^R, G_c^R) = 1$$

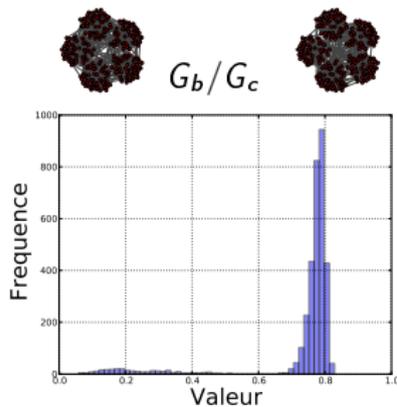
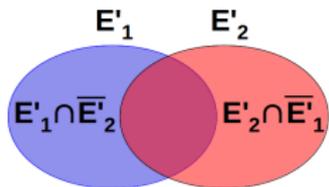


# Confluence: Expérience contrôlée

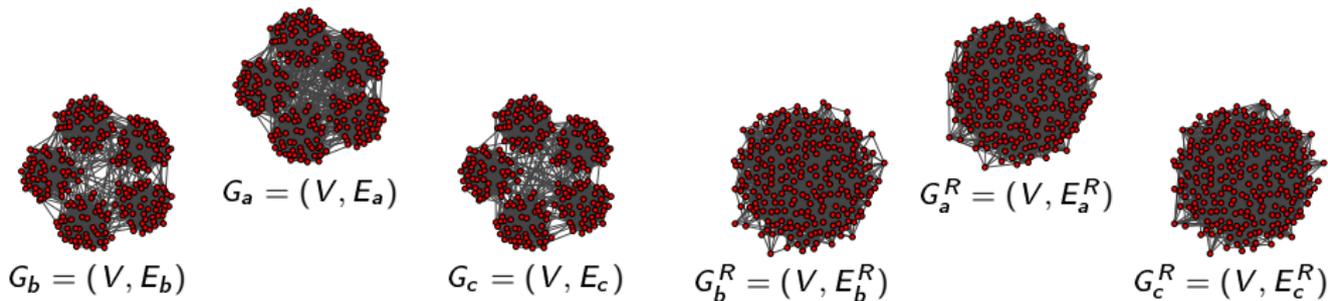


$E_b \cap E_c = \emptyset \Rightarrow$   
 $GED(G_b, G_c) = 1$

$E_b^R \cap E_c^R = \emptyset \Rightarrow$   
 $GED(G_b^R, G_c^R) = 1$

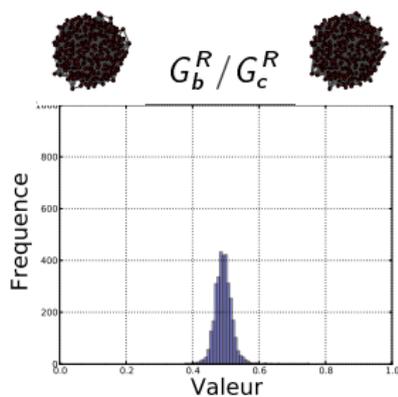
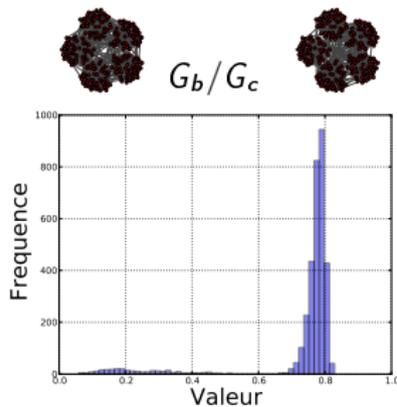
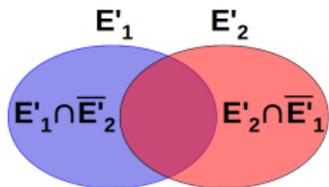


# Confluence: Expérience contrôlée



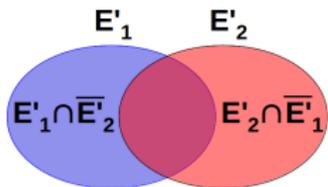
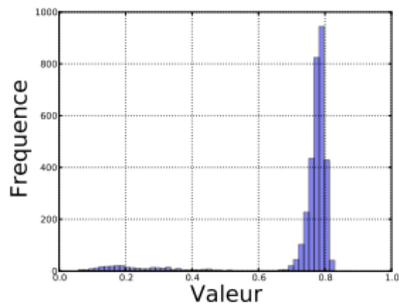
$E_b \cap E_c = \emptyset \Rightarrow$   
 $GED(G_b, G_c) = 1$

$E_b^R \cap E_c^R = \emptyset \Rightarrow$   
 $GED(G_b^R, G_c^R) = 1$

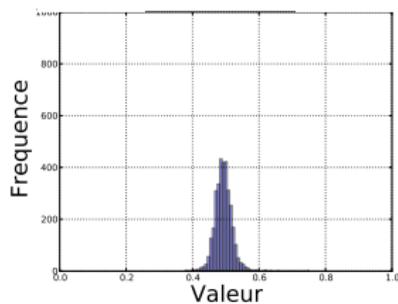


# Confluence: Expérience contrôlée

$$GED(G_b, G_c) = 1$$



$$GED(G_b^R, G_c^R) = 1$$

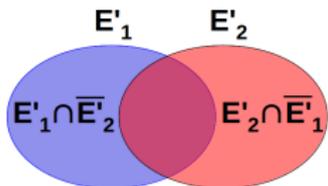
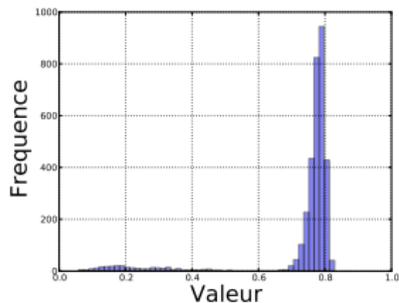


# Confluence: Expérience contrôlée

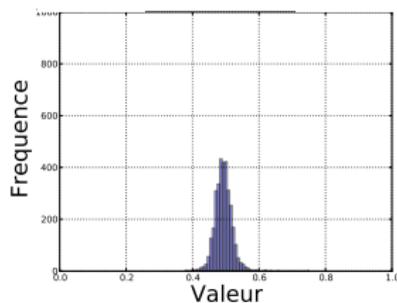
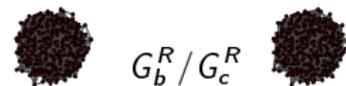
$$\mu(\mathbf{G}_1, \mathbf{G}_2) =$$

$$\frac{1}{|E_1 \cap \bar{E}_2| + |E_2 \cap \bar{E}_1|} \left( \sum_{\{u,v\} \in (E_2 \cap \bar{E}_1)} \text{CONF}_{G_1}(\{u,v\}) + \sum_{\{u,v\} \in (E_1 \cap \bar{E}_2)} \text{CONF}_{G_2}(\{u,v\}) \right)$$

$$\text{GED}(G_b, G_c) = 1$$



$$\text{GED}(G_b^R, G_c^R) = 1$$



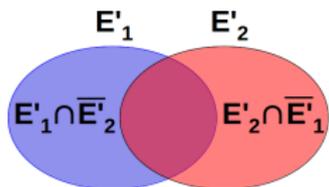
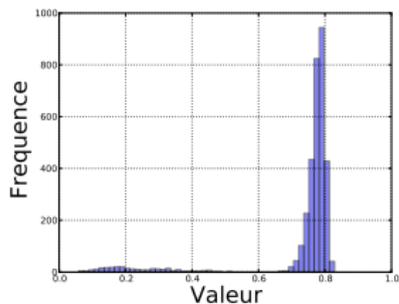
# Confluence: Expérience contrôlée

$$\mu(\mathbf{G}_1, \mathbf{G}_2) =$$

$$\frac{1}{|E_1 \cap \bar{E}_2| + |E_2 \cap \bar{E}_1|} \left( \sum_{\{u,v\} \in (E_2 \cap \bar{E}_1)} \text{CONF}_{G_1}(\{u,v\}) + \sum_{\{u,v\} \in (E_1 \cap \bar{E}_2)} \text{CONF}_{G_2}(\{u,v\}) \right)$$

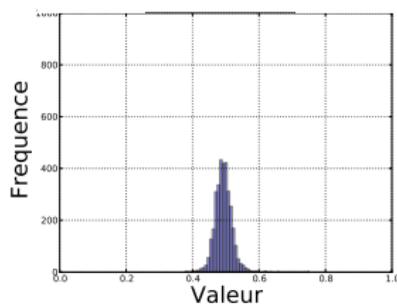
$$\mu(G_b, G_c) = 0.74$$

$$\text{GED}(G_b, G_c) = 1$$



$$\mu(G_b^R, G_c^R) = 0.49$$

$$\text{GED}(G_b^R, G_c^R) = 1$$



# Applications sur les Réseaux lexicaux

$$\mu(\mathbf{G}_1, \mathbf{G}_2) = \frac{1}{|E_1 \cap \bar{E}_2| + |E_2 \cap \bar{E}_1|} \left( \sum_{\{u,v\} \in (E_2 \cap \bar{E}_1)} \text{CONF}_{\mathbf{G}_1}(\{u,v\}) + \sum_{\{u,v\} \in (E_1 \cap \bar{E}_2)} \text{CONF}_{\mathbf{G}_2}(\{u,v\}) \right)$$

Robert / Larousse

# Applications sur les Réseaux lexicaux

$$\mu(\mathbf{G}_1, \mathbf{G}_2) =$$

$$\frac{1}{|E_1 \cap \overline{E_2}| + |E_2 \cap \overline{E_1}|} \left( \sum_{\{u,v\} \in (E_2 \cap \overline{E_1})} \text{CONF}_{G_1}(\{u,v\}) + \sum_{\{u,v\} \in (E_1 \cap \overline{E_2})} \text{CONF}_{G_2}(\{u,v\}) \right)$$

Robert / Larousse

$$\text{GED}(G'_{Rob}, G'_{Lar}) = 0.47$$



$$\text{GED}(G'^R_{Rob}, G'^R_{Lar}) = 0.47$$



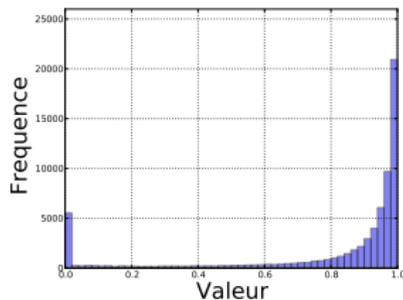
# Applications sur les Réseaux lexicaux

$$\mu(\mathbf{G}_1, \mathbf{G}_2) =$$

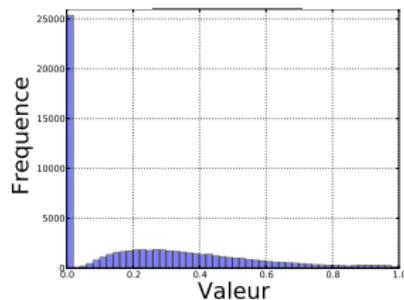
$$\frac{1}{|E_1 \cap \bar{E}_2| + |E_2 \cap \bar{E}_1|} \left( \sum_{\{u,v\} \in (E_2 \cap \bar{E}_1)} \text{CONF}_{G_1}(\{u,v\}) + \sum_{\{u,v\} \in (E_1 \cap \bar{E}_2)} \text{CONF}_{G_2}(\{u,v\}) \right)$$

Robert / Larousse

$$\text{GED}(G'_{Rob}, G'_{Lar}) = 0.47$$



$$\text{GED}(G'^R_{Rob}, G'^R_{Lar}) = 0.47$$



# Applications sur les Réseaux lexicaux

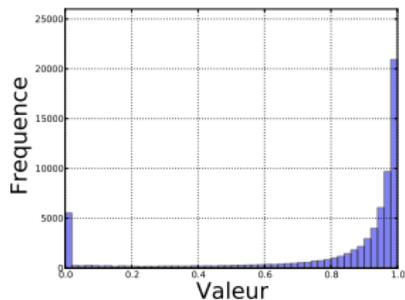
$$\mu(\mathbf{G}_1, \mathbf{G}_2) =$$

$$\frac{1}{|E_1 \cap \bar{E}_2| + |E_2 \cap \bar{E}_1|} \left( \sum_{\{u,v\} \in (E_2 \cap \bar{E}_1)} \text{CONF}_{G_1}(\{u,v\}) + \sum_{\{u,v\} \in (E_1 \cap \bar{E}_2)} \text{CONF}_{G_2}(\{u,v\}) \right)$$

Robert / Larousse

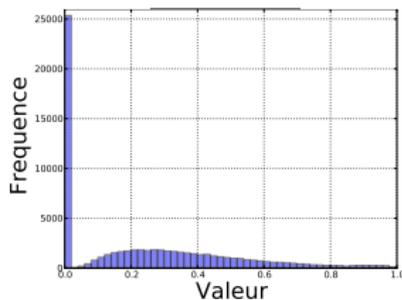
$$\mu(G'_{\text{Rob}}, G'_{\text{Lar}}) = 0.80$$

$$\text{GED}(G'_{\text{Rob}}, G'_{\text{Lar}}) = 0.47$$



$$\mu(G'^R_{\text{Rob}}, G'^R_{\text{Lar}}) = 0.24$$

$$\text{GED}(G'^R_{\text{Rob}}, G'^R_{\text{Lar}}) = 0.47$$



# Applications sur les Réseaux lexicaux

# Applications sur les Réseaux lexicaux

Table: Pédigrés de quelques graphes lexicaux

Graphes lexicaux		n	m	$\langle k \rangle$	C	$L_{lcc}$	$\lambda (r^2)$
Lar	Adjectifs	5510	21147	7,68	0,21	4,92	-2,06 (0,88)
	Noms	12159	31601	5,20	0,20	6,10	-2,39 (0,88)
	Verbes	5377	22042	8,20	0,17	4,61	-1,94 (0,88)
Rob	Adjectifs	7693	20011	5,20	0,14	5,26	-2,05 (0,94)
	Noms	24570	55418	4,51	0,11	6,08	-2,34 (0,94)
	Verbes	7357	26567	7,22	0,12	4,59	-2,01 (0,93)
Jdm	Adjectifs	9859	30087	6,10	0,16	5,44	-2,24 (0,90)
	Noms	29213	56381	3,86	0,14	6,48	-2,66 (0,93)
	Verbes	7658	22260	5,81	0,14	5,06	-2,08 (0,89)
Wik	Adjectifs	6960	6594	1,89	0,15	8,48	-2,46 (0,95)
	Noms	43206	37661	1,74	0,13	10,56	-2,51 (0,89)
	Verbes	7203	7497	2,08	0,25	9,22	-2,28 (0,92)

# Applications sur les Réseaux lexicaux

Table: Mesurer la similarité structurelle entre réseaux lexicaux

GED ( $\mu$ ) ( $\mu^R$ ) sur les paires de graphes			
$G_1/G_2$	$Rob_A$	$Jdm_A$	$Wik_A$
$Lar_A$	<b>0,45</b> ( <b>0,76</b> ) (0,34)	<b>0,47</b> ( <b>0,71</b> ) (0,38)	<b>0,75</b> ( <b>0,41</b> ) (0,06)
$Rob_A$		<b>0,51</b> ( <b>0,70</b> ) (0,29)	<b>0,71</b> ( <b>0,42</b> ) (0,05)
$Jdm_A$			<b>0,54</b> ( <b>0,43</b> ) (0,03)
$G_1/G_2$	$Rob_N$	$Jdm_N$	$Wik_N$
$Lar_N$	<b>0,48</b> ( <b>0,70</b> ) (0,20)	<b>0,48</b> ( <b>0,70</b> ) (0,19)	<b>0,72</b> ( <b>0,31</b> ) (0,03)
$Rob_N$		<b>0,47</b> ( <b>0,70</b> ) (0,13)	<b>0,71</b> ( <b>0,29</b> ) (0,02)
$Jdm_N$			<b>0,46</b> ( <b>0,32</b> ) (0,01)
$G_1/G_2$	$Rob_V$	$Jdm_V$	$Wik_V$
$Lar_V$	<b>0,48</b> ( <b>0,73</b> ) (0,40)	<b>0,46</b> ( <b>0,70</b> ) (0,39)	<b>0,78</b> ( <b>0,25</b> ) (0,05)
$Rob_V$		<b>0,47</b> ( <b>0,70</b> ) (0,37)	<b>0,78</b> ( <b>0,25</b> ) (0,06)
$Jdm_V$			<b>0,55</b> ( <b>0,31</b> ) (0,04)





**Hypothèse** : La synonymie reflète la structure sémantique du lexique commune aux membres d'une même communauté linguistique

**Constat** : Il existe de nombreux désaccords entre membres d'une même communauté linguistique:

$$\text{GED}(G_1, G_2) \approx 0.50$$

- ▶ **Beaucoup** de paires de mots sont synonymes pour l'individu  $I_1$  et pas pour l'individu  $I_2$
- ▶ **Beaucoup** de paires de mots sont synonymes pour l'individu  $I_2$  et pas pour l'individu  $I_1$

**Hypothèse** : La synonymie reflète la structure sémantique du lexique commune aux membres d'une même communauté linguistique

## Incompatible ?

**Constat** : Il existe de nombreux désaccords entre membres d'une même communauté linguistique:

$$\text{GED}(G_1, G_2) \approx 0.50$$

- ▶ **Beaucoup** de paires de mots sont synonymes pour l'individu  $I_1$  et pas pour l'individu  $I_2$
- ▶ **Beaucoup** de paires de mots sont synonymes pour l'individu  $I_2$  et pas pour l'individu  $I_1$

**Hypothèse** : La synonymie reflète la structure sémantique du lexique commune aux membres d'une même communauté linguistique

**Incompatible ?**

**Constat** : Il existe de nombreux désaccords entre membres d'une même communauté linguistique:

**$GED(G_1, G_2) \approx 0.50$**

- ▶ **Beaucoup** de paires de mots sont synonymes pour l'individu  $I_1$  et pas pour l'individu  $I_2$
- ▶ **Beaucoup** de paires de mots sont synonymes pour l'individu  $I_2$  et pas pour l'individu  $I_1$

**Non**

**Parceque** : Ces désaccords sont des désaccords faibles :

**$\mu(G_1, G_2) \approx 0.70$  (accord au niveau structurel)**

**Hypothèse** : La synonymie reflète la structure sémantique du lexique commune aux membres d'une même communauté linguistique

L'analyse d'une relation lexicale doit être faite au niveau de la structure globale dessinée par la relation. C'est ce que la mesure  $\mu$  peut faire: avec un niveau de représentation adéquat, elle réconcilie les jugements portés par deux juges différents sur une même relation lexicale.

1 Structure des réseaux lexicaux

2 Un score d'habilité lexicale: *Skillex* ←←

3 Perspectives

# Le protocole Approx [Duvignau et Gaume, 2004]

				
<b>1 démonter des légos (à la main)</b>	<b>2 couper un pain (avec un couteau)</b>	<b>3 rompre du pain (à la main)</b>	<b>4 déchirer un journal (à la main)</b>	<b>5 scier une planche (avec une scie)</b>
				
<b>6 hacher du persil (avec un couteau)</b>	<b>7 briser un verre (avec un marteau)</b>	<b>8 déshabiller une poupée (à la main)</b>	<b>9 exploser un ballon (à la main)</b>	<b>10 écorcer une buche (à la main)</b>
				
<b>11 écraser une tomate (à la main)</b>	<b>12 peler une banane (à la main)</b>	<b>13 éplucher une carotte (with a peeler)</b>	<b>14 éplucher une orange (à la main)</b>	<b>15 froisser une feuille de papier (à la main)</b>

# Le protocole Approx [Duvignau et Gaume, 2004]



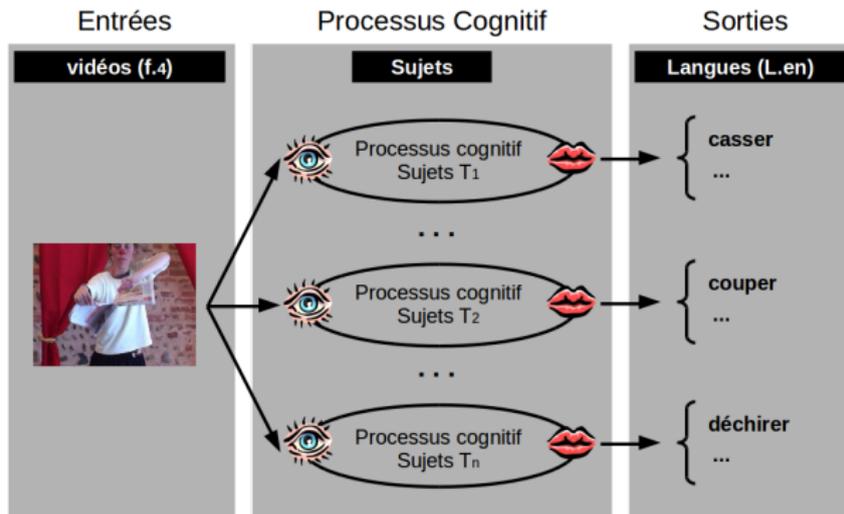
"Qu'est-ce qu'elle a fait la dame ?"

"Ce qu'elle a fait la dame, dis le moi d'une autre manière, avec d'autres mots"



## Exemple

# Le protocole Approx [Duvignau et Gaume, 2004]



	2-5 ans	18-40 ans
% d'approximations	37%	6%
% de verbes spécifiques	40%	84%

Table: Résultats (à la main)

Les productions des enfants typiques :

- ▶ Des verbes plus génériques que les adultes typiques.  
Hypothèses  $H_1$  : Le **degré** : la généricité du verbe  $u$  est corrélée à  $deg(u)$ .
- ▶ Des verbes plus approximatifs que les adultes typiques.  
Hypothèses  $H_2$  : La **confluence**: la proximité sémantique entre  $u$  et  $v$  est corrélée à  $CONF_G(u, v)$ .

	2-5 ans	18-40 ans
% d'approximations	37%	6%
% de verbes spécifiques	40%	84%

Table: Résultats (à la main)

Les productions des enfants typiques :

- ▶ Des verbes plus génériques que les adultes typiques.

Hypothèses  $H_1$  : **Le degrés** : la généricité du verbe  $u$  est corrélée à  $deg(u)$ .

- ▶ Des verbes plus approximatifs que les adultes typiques.

Hypothèses  $H_2$  : **La confluence**: la proximité sémantique entre  $u$  et  $v$  est corrélée à  $CONF_G(u, v)$ .

	2-5 ans	18-40 ans
% d'approximations	37%	6%
% de verbes spécifiques	40%	84%

Table: Résultats (à la main)

Les productions des enfants typiques :

- ▶ Des verbes plus génériques que les adultes typiques.

Hypothèses  $H_1$  : **Le degrés** : la généricité du verbe  $u$  est corrélée à  $deg(u)$ .

- ▶ Des verbes plus approximatifs que les adultes typiques.

Hypothèses  $H_2$  : **La confluence**: la proximité sémantique entre  $u$  et  $v$  est corrélée à  $CONF_G(u, v)$ .

	2-5 ans	18-40 ans
% d'approximations	37%	6%
% de verbes spécifiques	40%	84%

Table: Résultats (à la main)

Les productions des enfants typiques :

- ▶ Des verbes plus génériques que les adultes typiques.

Hypothèses  $H_1$  : **Le degrés** : la généricité du verbe  $u$  est corrélée à  $deg(u)$ .

- ▶ Des verbes plus approximatifs que les adultes typiques.

Hypothèses  $H_2$  : **La confluence**: la proximité sémantique entre  $u$  et  $v$  est corrélée à  $CONF_G(u, v)$ .

	2-5 ans	18-40 ans
% d'approximations	37%	6%
% de verbes spécifiques	40%	84%

Table: Résultats (à la main)

Les productions des enfants typiques :

- ▶ Des verbes plus génériques que les adultes typiques.

Hypothèses  $H_1$  : **Le degrés** : la généricité du verbe  $u$  est corrélée à  $deg(u)$ .

- ▶ Des verbes plus approximatifs que les adultes typiques.

Hypothèses  $H_2$  : **La confluence**: la proximité sémantique entre  $u$  et  $v$  est corrélée à  $CONF_G(u, v)$ .

# Zone lexico-semantic d'une action

A chaque film  $F_i$  est associé une zone lexico-sémantique dans le graphe de synonymie (départ de la balade aléatoire)  $\Delta(F_i)$ :

- ▶  $(\Delta(F_i))_v$  la probabilité que le verbe  $v$  soit utilisé par un **échantillon de population** pour décrire l'action du film  $F_i$

	Français	Mandarin
<i>24-59 mois</i>	74	29
<i>18-40 ans</i>	76	60

Table: Populations étudiées (Duvignau & Cheung)

- ▶ **échantillon de population** = 25 adultes et 25 enfants (aucun avantage à une quelconque population  $\implies$  pas de biais méthodologique)

# Zone lexico-semanticque d'une action

A chaque film  $F_i$  est associé une zone lexico-sémantique dans le graphe de synonymie (départ de la balade aléatoire)  $\Delta(F_i)$ :

- ▶  $(\Delta(F_i))_v$  la probabilité que le verbe  $v$  soit utilisé par un **échantillon de population** pour décrire l'action du film  $F_i$

	Français	Mandarin
<i>24-59 mois</i>	74	29
<i>18-40 ans</i>	76	60

Table: Populations étudiées (Duvignau & Cheung)

- ▶ **échantillon de population** = 25 adultes et 25 enfants (aucun avantage à une quelconque population  $\implies$  pas de biais méthodologique)

# Zone lexico-semanticque d'une action

A chaque film  $F_i$  est associé une zone lexico-sémantique dans le graphe de synonymie (départ de la balade aléatoire)  $\Delta(F_i)$ :

- ▶  $(\Delta(F_i))_v$  la probabilité que le verbe  $v$  soit utilisé par un **échantillon de population** pour décrire l'action du film  $F_i$

	Français	Mandarin
<i>24-59 mois</i>	74	29
<i>18-40 ans</i>	76	60

Table: Populations étudiées (Duvignau & Cheung)

- ▶ **échantillon de population** = 25 adultes et 25 enfants (aucun avantage à une quelconque population  $\implies$  pas de biais méthodologique)

Deux mesures sont associées à chaque participant  $p$  qui produit le verbe  $v$  pour décrire l'action du film  $F_i$  de zone lexico-sémantique  $\Delta(F_i)$ :

- ▶ Un score de Degrés :  $deg_G(v)$
- ▶ Un score de Confluence :  $CONF_G(\Delta(F_i), v)$

Deux mesures sont associées à chaque participant  $p$  qui produit le verbe  $v$  pour décrire l'action du film  $F_i$  de zone lexico-sémantique  $\Delta(F_i)$ :

- ▶ Un score de Degrés :  $deg_G(v)$
- ▶ Un score de Confluence :  $CONF_G(\Delta(F_i), v)$

Deux mesures sont associées à chaque participant  $p$  qui produit le verbe  $v$  pour décrire l'action du film  $F_i$  de zone lexico-sémantique  $\Delta(F_i)$ :

- ▶ Un score de Degrés :  $deg_G(v)$
- ▶ Un score de Confluence :  $CONF_G(\Delta(F_i), v)$

Deux mesures sont associées à chaque participant  $p$  qui produit le verbe  $v$  pour décrire l'action du film  $F_i$  de zone lexico-sémantique  $\Delta(F_i)$ :

- ▶ Un score de Degrés :  $deg_G(v)$
- ▶ Un score de Confluence :  $CONF_G(\Delta(F_i), v)$

	Français	Mandarin
<i>24-59 mois</i>	74	29
<i>18-40 ans</i>	76	60

**Table:** Populations étudiées (Duvignau & Cheung)

# Hypothèse $H_1$ (Moyenne des degrés)

# Hypothèse $H_1$ (Moyenne des degrés)

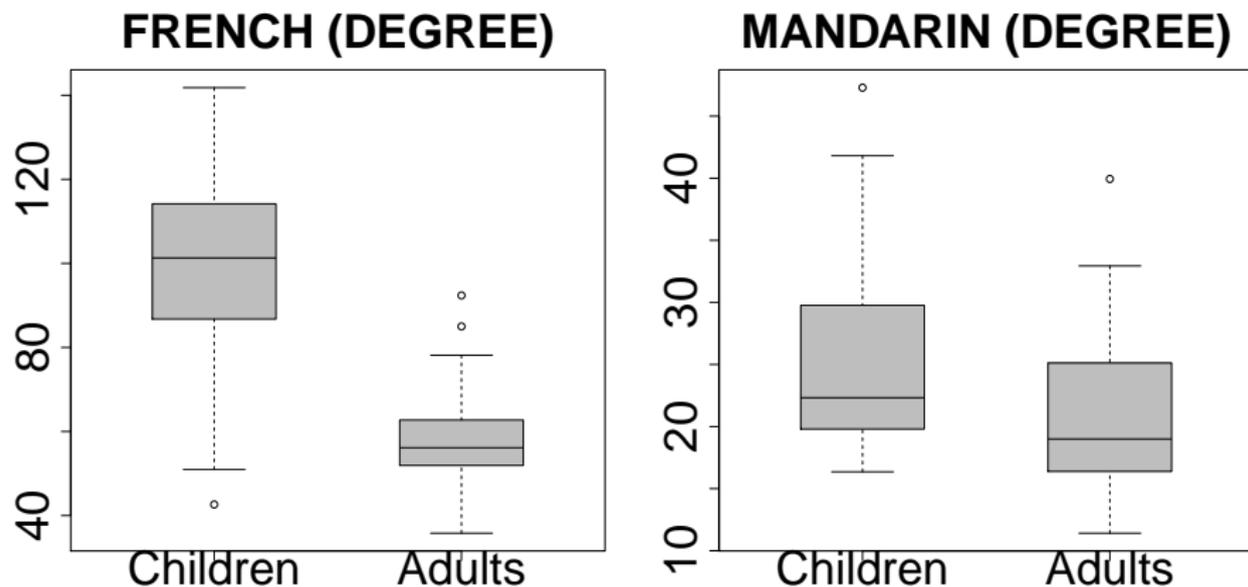


Figure: Scores de degrés pour le Français et pour le Mandarin

# Hypothèse $H_1$ (Moyenne des degrés)

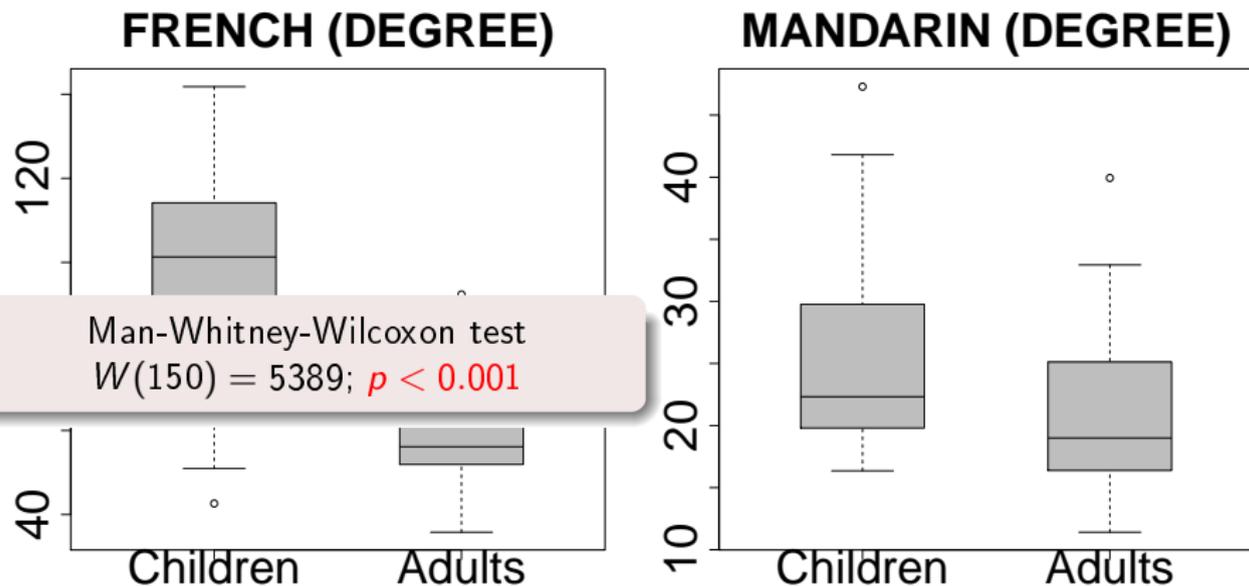


Figure: Scores de degrés pour le Français et pour le Mandarin

# Hypothèse $H_1$ (Moyenne des degrés)

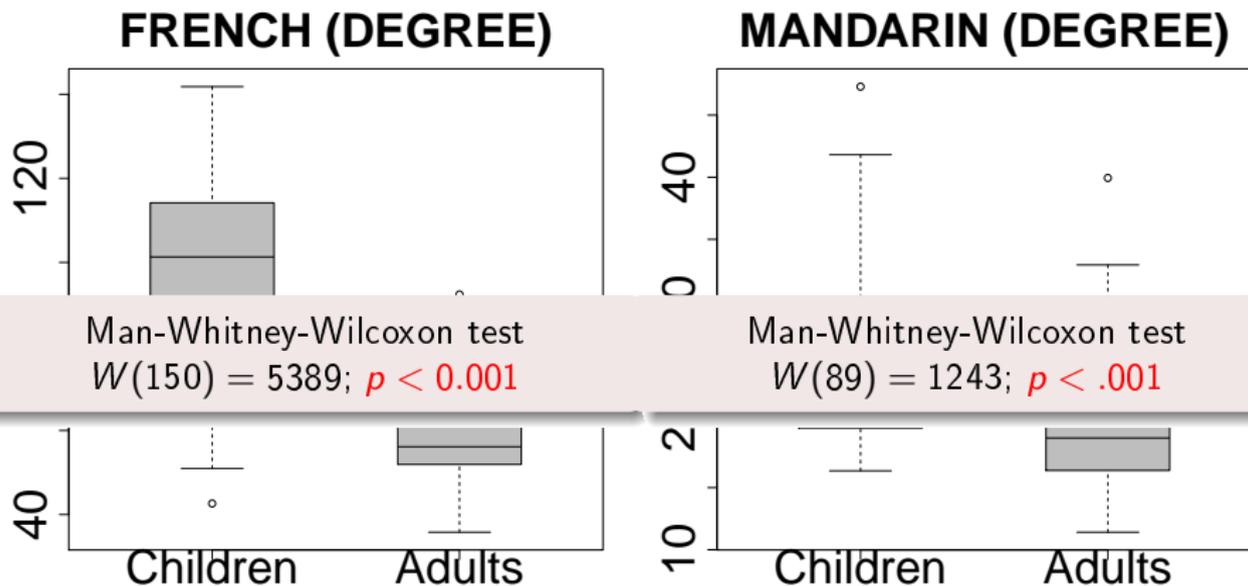


Figure: Scores de degrés pour le Français et pour le Mandarin

# Hypothèse $H_1$ (Moyenne des degrés)

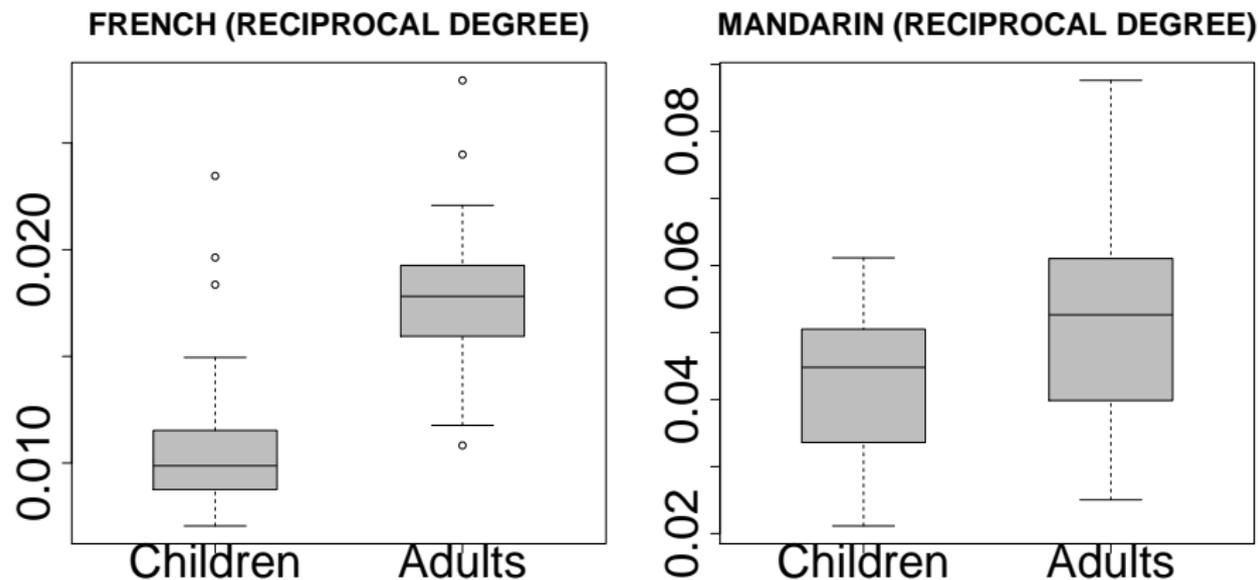
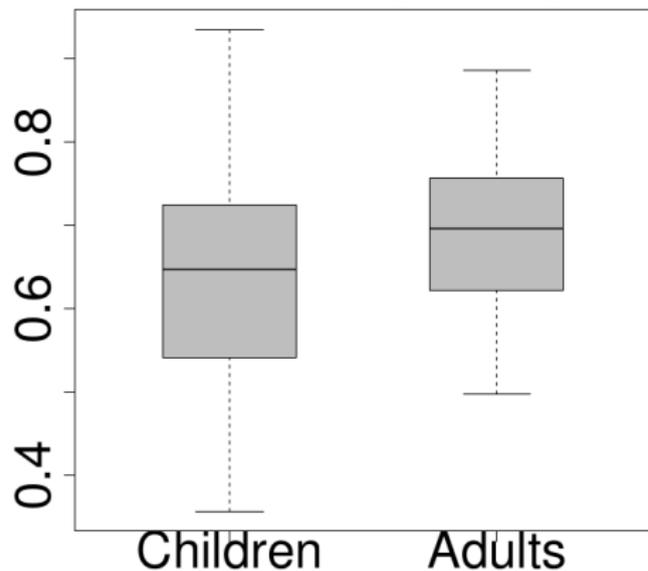


Figure: (Inverse des scores de degrés pour le Français et pour le Mandarin)

# Hypothèse $H_2$ (Moyenne des Confluences)

# Hypothèse $H_2$ (Moyenne des Confluences)

## FRENCH (CONF)



## MANDARIN (CONF)

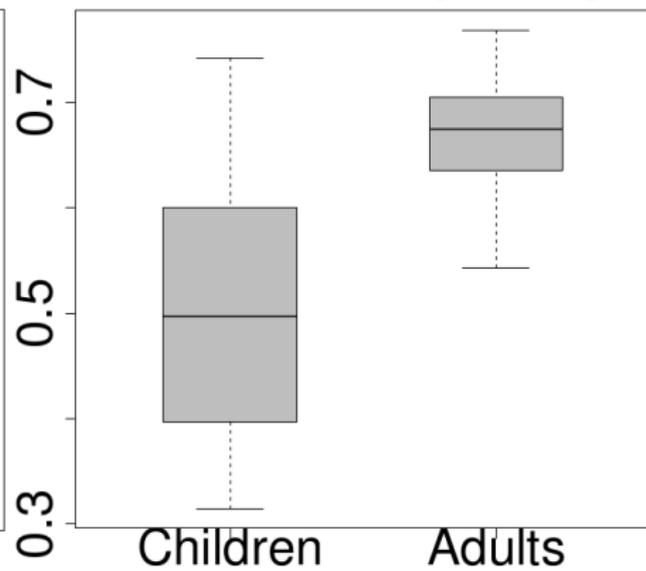
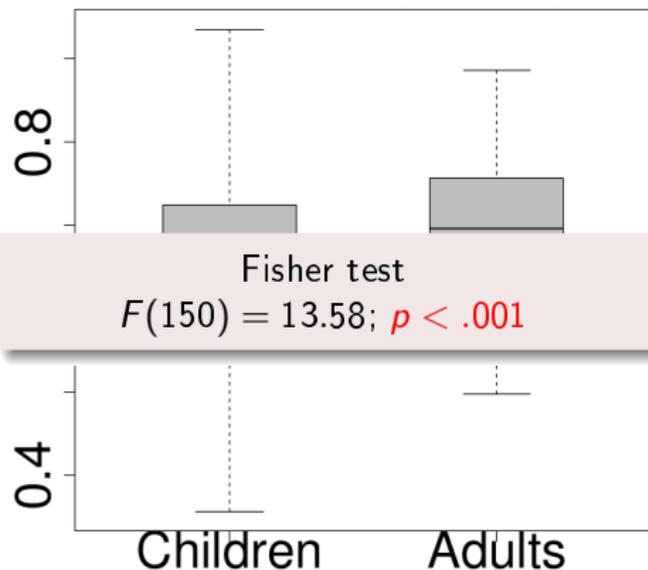


Figure: Scores de Confluence pour le Français et pour le Mandarin

# Hypothèse $H_2$ (Moyenne des Confluences)

## FRENCH (CONF)



Fisher test  
 $F(150) = 13.58; p < .001$

## MANDARIN (CONF)

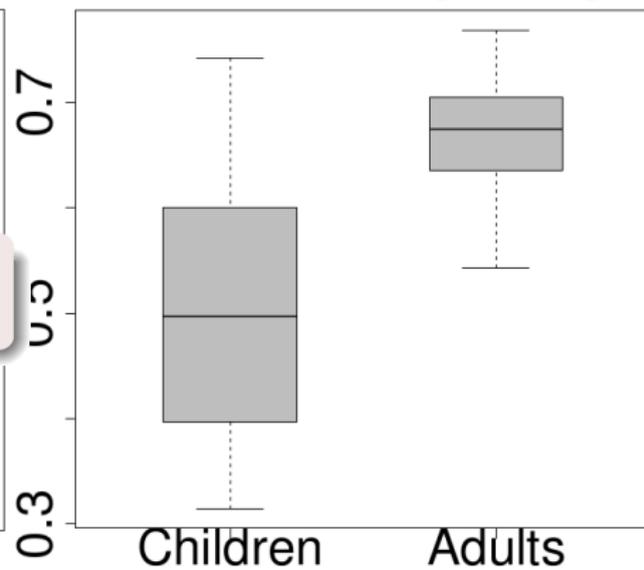
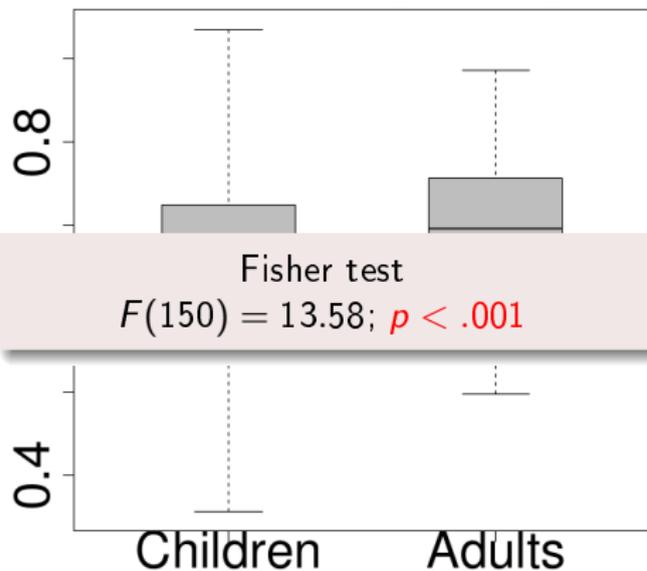


Figure: Scores de Confluence pour le Français et pour le Mandarin

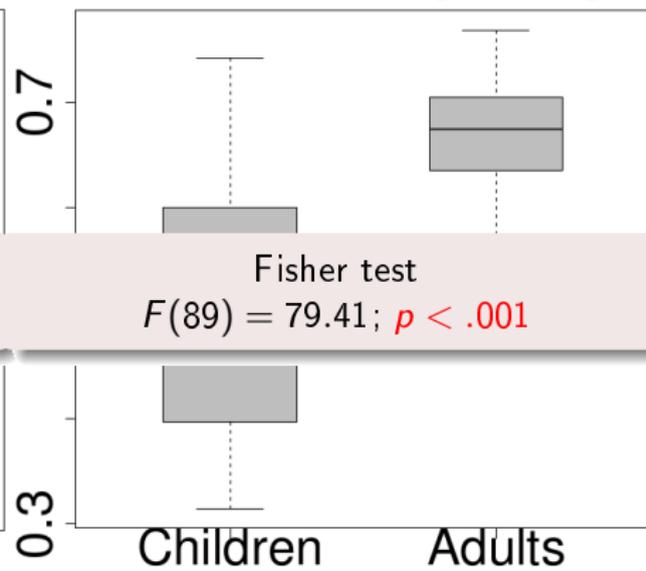
# Hypothèse $H_2$ (Moyenne des Confluences)

## FRENCH (CONF)



Fisher test  
 $F(150) = 13.58; p < .001$

## MANDARIN (CONF)



Fisher test  
 $F(89) = 79.41; p < .001$

Figure: Scores de Confluence pour le Français et pour le Mandarin

Le score **Skillex** est associé à chaque participant  $p$  qui produit le verbe  $v$  pour décrire l'action du film  $F_i$  de zone lexico-sémantique  $\Delta(F_i)$ :

$$\blacktriangleright \text{Skillex}_G(\Delta(F_i), v) = \frac{\text{CONF}_G(\Delta(F_i), v)}{\text{deg}_G(v)}$$

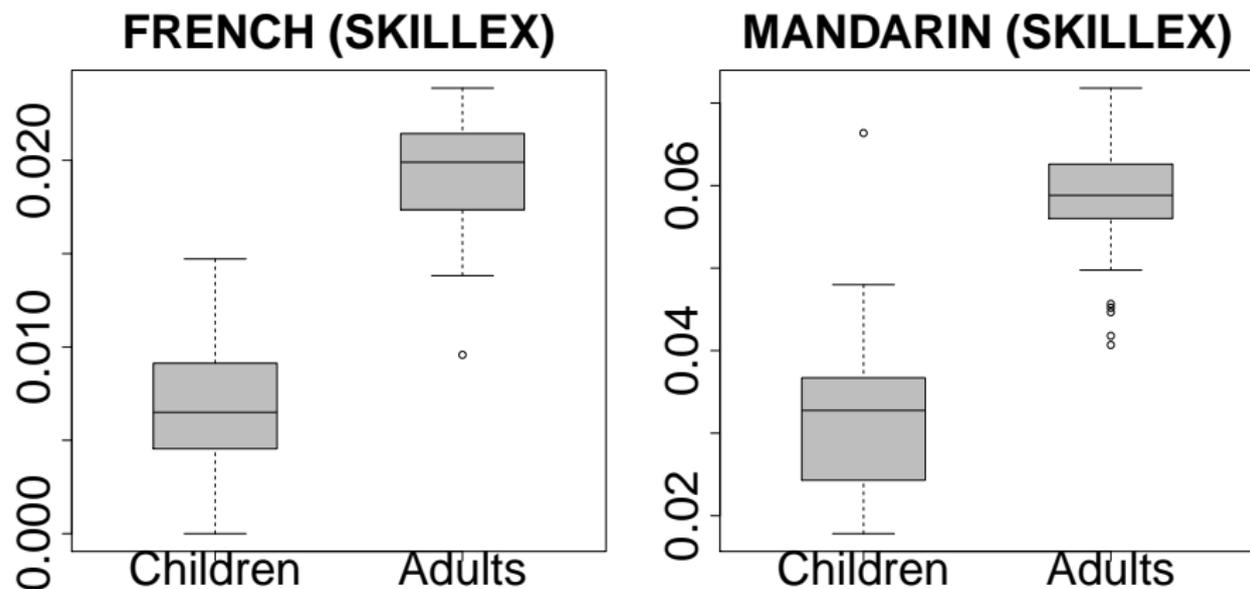
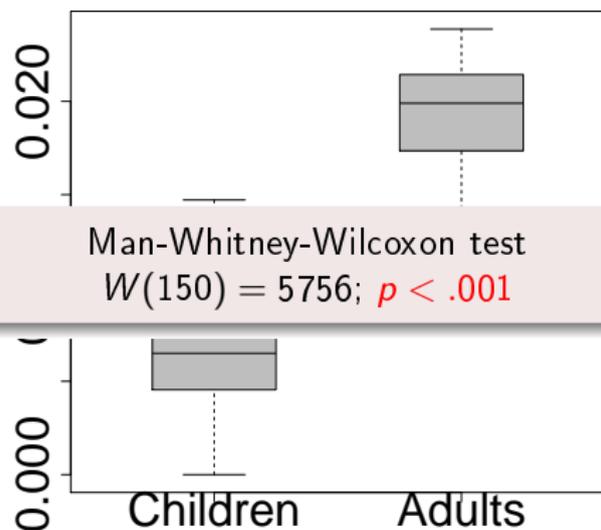


Figure: Scores *Skillex* pour le Français et pour le Mandarin

## FRENCH (SKILLEX)



Man-Whitney-Wilcoxon test  
 $W(150) = 5756$ ;  $p < .001$

## MANDARIN (SKILLEX)

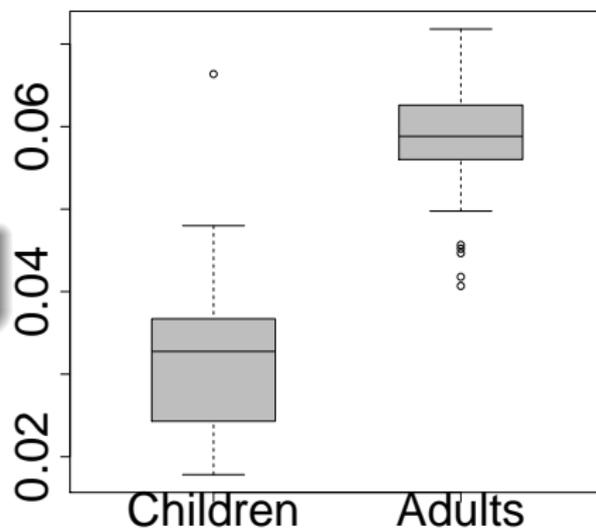


Figure: Scores Skillex pour le Français et pour le Mandarin

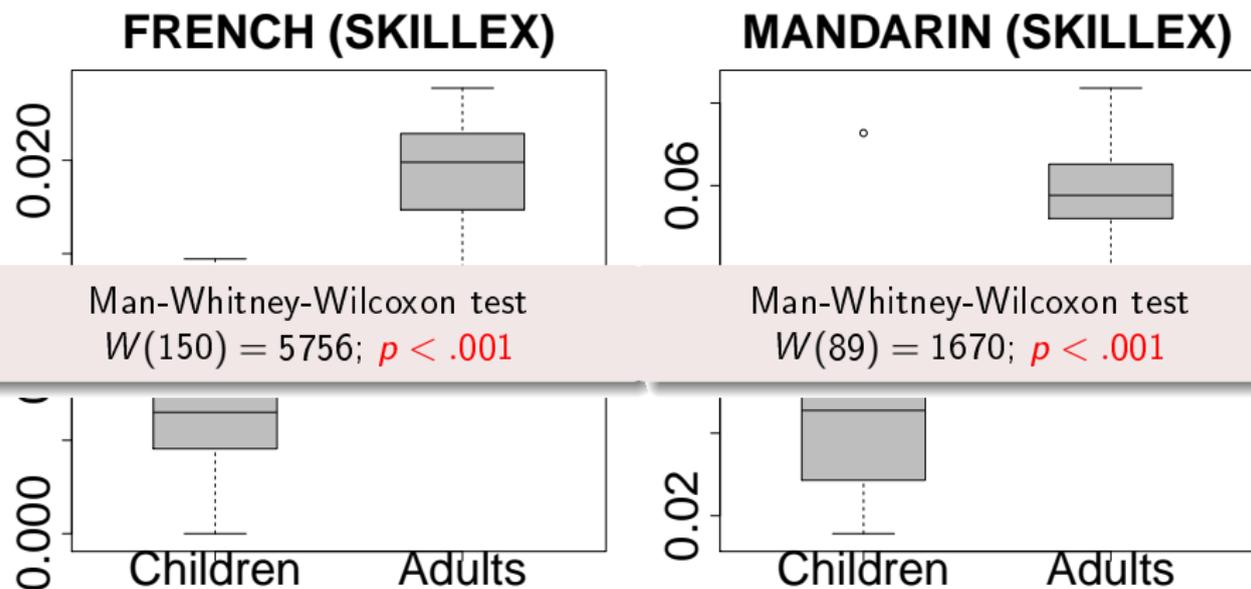


Figure: Scores *Skillex* pour le Français et pour le Mandarin

Est-ce que le score *Skillex* permet une classification automatique enfants VS adultes des participants au protocole *Approx* ?

# Classification automatique: jeunes enfants VS Adultes

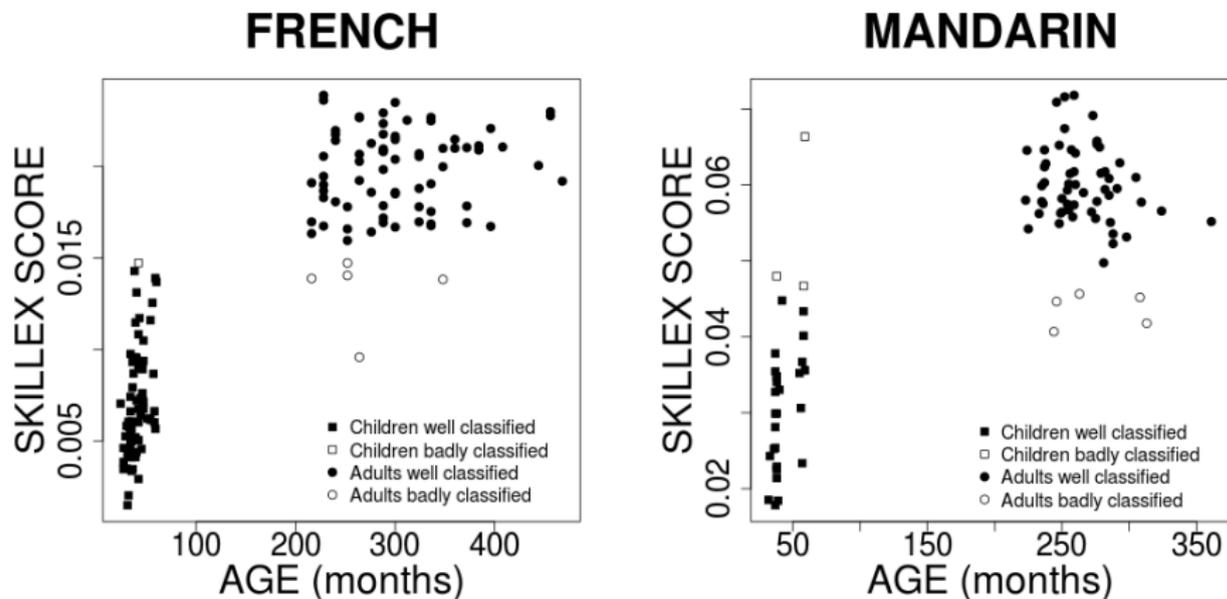


Figure: Scores *Skillex* selon l'âge des participants au protocole *Approx*

# Classification automatique: jeunes enfants VS Adultes

## FRENCH



Score	Precision	$\kappa$
<i>Degree</i>	.91	.81
<i>Conf</i>	.61	.21
<i>Skillex</i>	<b>.96</b>	<b>.92</b>

Table: 2-means clustering: French

AGE (months)

## MANDARIN

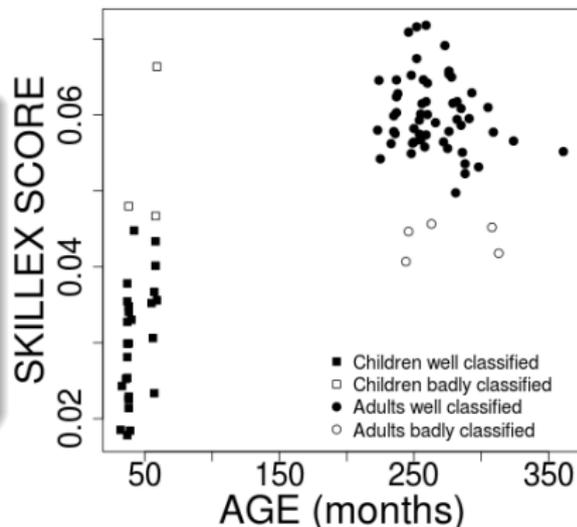


Figure: Scores *Skillex* selon l'âge des participants au protocole *Approx*

# Classification automatique: jeunes enfants VS Adultes

## FRENCH



Score	Precision	$\kappa$
<i>Degree</i>	.91	.81
<i>Conf</i>	.61	.21
<i>Skillex</i>	<b>.96</b>	<b>.92</b>

## MANDARIN



Score	Precision	$\kappa$
<i>Degree</i>	.64	.17
<i>Conf</i>	.84	.62
<i>Skillex</i>	<b>.91</b>	<b>.80</b>

Table: 2-means clustering: French

Table: 2-means clustering: Mandarin

100 200 300 400  
AGE (months)

50 150 250 350  
AGE (months)

Figure: Scores *Skillex* selon l'âge des participants au protocole *Approx*

Les films sont sans biais culturels [Cheung et al., 2010]:

Analyses statistiques d'un protocole culturel [Cheung et al., 2010]:

- ▶ 20 participants Français (Duvignau);
- ▶ 20 participants Taiwanais (Cheung);

Les scores de degrés, ni de confluence ne sont sensibles à la variabilité des réseaux lexicaux pour une langue donnée [Gaillard et al., 2011, Gaume et al., 2014]

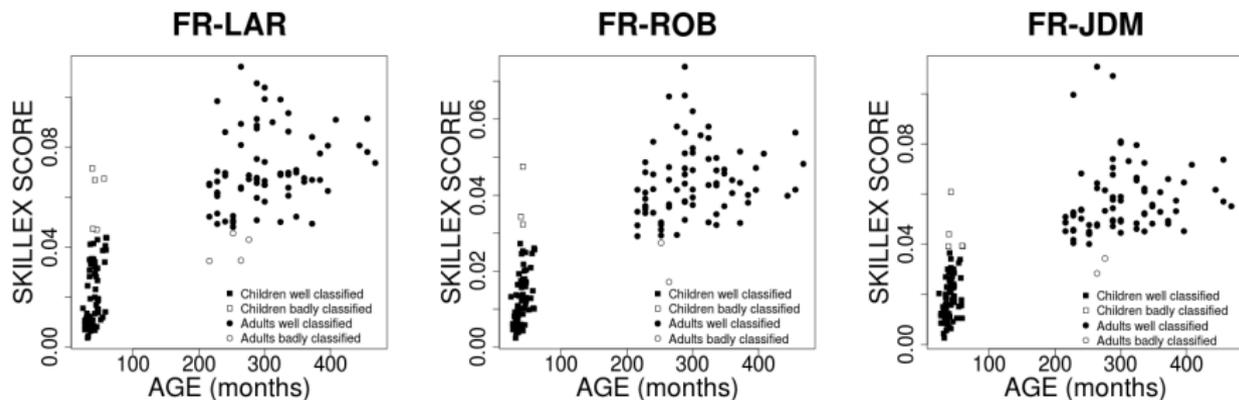
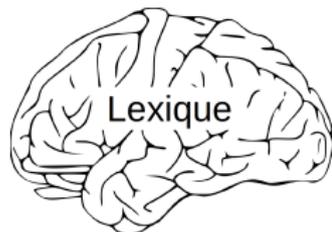
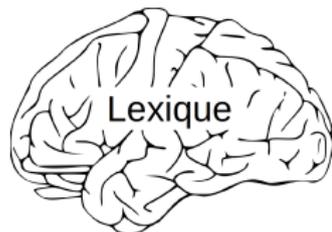


Figure: Scores *Skillex* selon l'âge des participants au protocole *Approx*

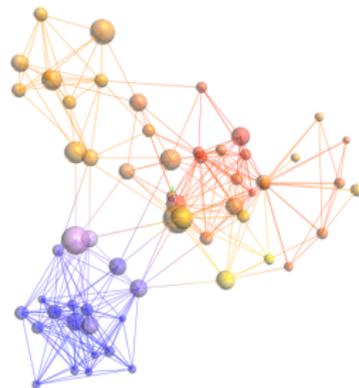


Artefact Cognitif





Artefact Cognitif



- 1 Structure des réseaux lexicaux
- 2 Un score d'habilité lexicale: *Skillex*
- 3 Perspectives



# Travaux futurs : **typologie**

## Autres langues et autres catégories d'actions

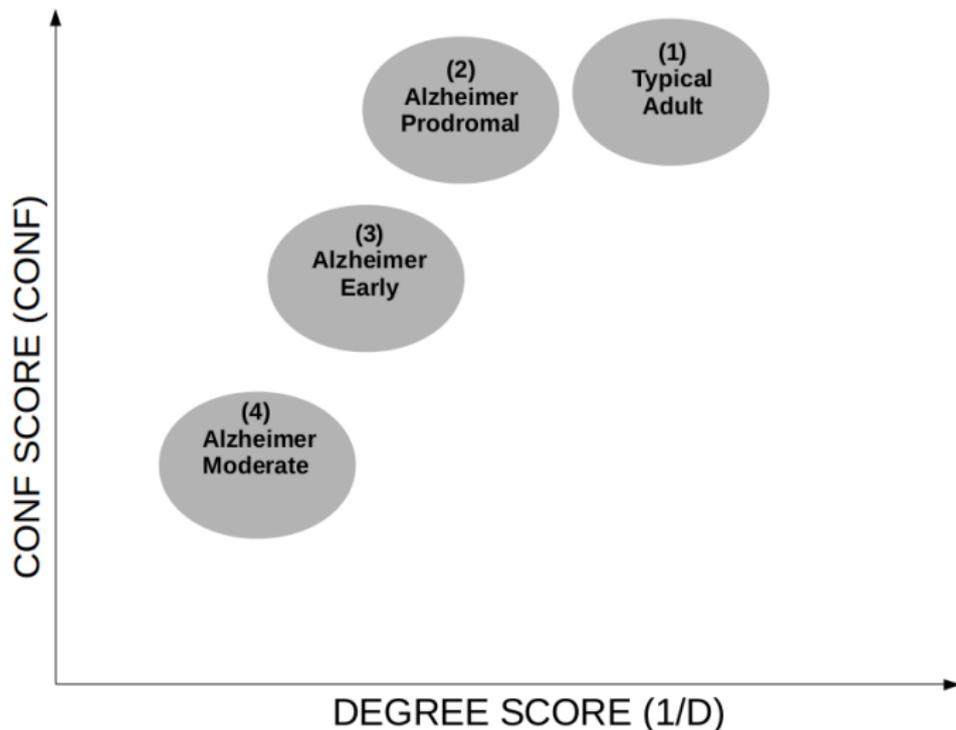
- ▶ Etendre les analyses à d'autres langues  
⇒ typologie des dynamiques d'acquisition lexicale selon les langues
- ▶ Etendre le protocole à d'autres catégories d'actions  
(i.e. verbes de mouvement)

# Travaux futurs : **clinique**

Aide au diagnostic de troubles provoquant des troubles langagiers

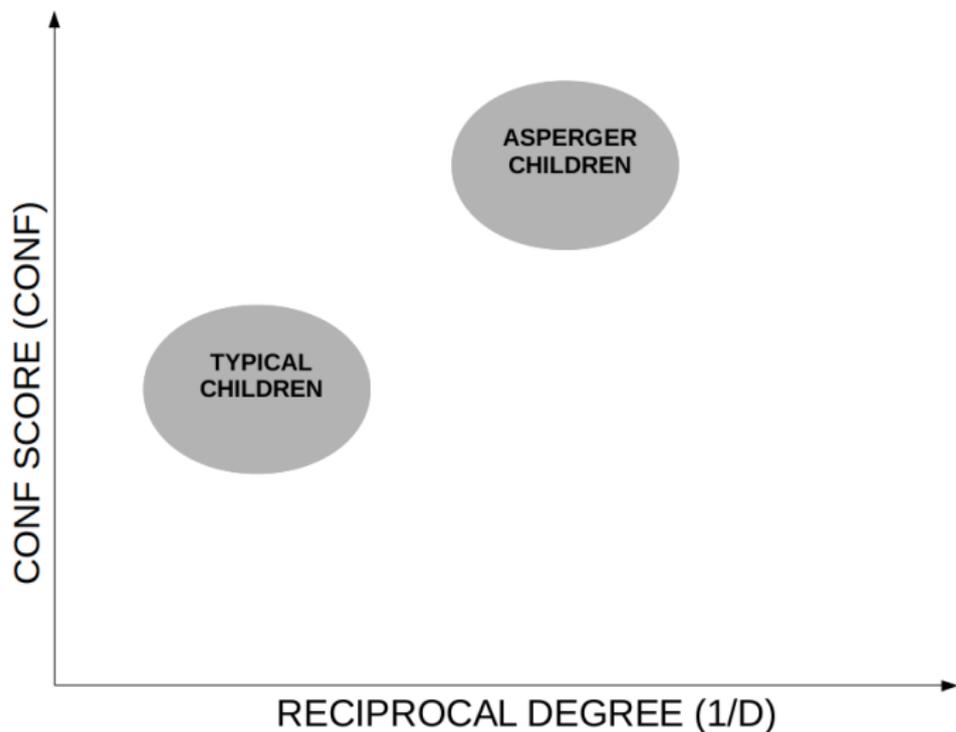
# Travaux futurs : clinique

Aide au diagnostic de troubles provoquant des troubles langagiers



# Travaux futurs : clinique

Aide au diagnostic de troubles provoquant des troubles langagiers



Merci  
谢谢  
Thank you



Cheung, H., Desalle, Y., Duvignau, K., Gaume, B., Chang, C.-H., and Magistry, P. (2010).  
The use of a cultural protocol for quantifying cultural variations in verbs semantic between  
chinese and french.

In Proceedings of 24th Pacific Asia Conference on Language, Information and  
Computation: Workshop on Model and Measurement of Meaning (M3), Sendai, Japan.



De Deyne, S. and Storms, G. (2008).

Word associations: Network and semantic properties.

Behavior Research Methods, 40(1):213–231.



Duvignau, K. (2008).

Acquisition du lexique verbal par proximité sémantique: premiers pas avec les troubles  
spécifiques du langage oral et ouverture sur la dyslexie.

ANAE: Revue de Neuropsychologie du Développement et des Apprentissages,  
20(96–97):65–70.



Gaillard, B., Gaume, B., and Navarro, E. (2011).

Invariant and variability of synonymy networks: Self mediated agreement by confluence.

In Proceedings of the TextGraph-6 Workshop, pages 15–23. Association for Computational  
Linguistics.



Gaume, B. (2004).

Balades aléatoires dans les petits mondes lexicaux.

Information Engineering Sciences, 4(2):39–96.

