

# Vers une mesure de l'acquisition du lexique verbal

Exemple de Catégorisation des Jeunes Enfants-vs-Adultes  
en Français et Mandarin

Y. Desalle

Octogone  
Université de Toulouse, France

UE TAL, 2011

# Plan

- 1 acquisition du lexique verbal
  - L'acquisition du lexique verbal: généralités
  - APPROX: un protocole psycholinguistique.
- 2 Un modèle computationnel du lexique verbal
  - Les dictionnaires: Réseaux Petits Mondes Hiérarchiques (RPMH)
  - 3 mesures disponibles
- 3 Catégorisation automatique
  - Les mesures affectées aux participants
  - Catégorisation et Résultats
  - Flexsem

# Plan

- 1 acquisition du lexique verbal
  - L'acquisition du lexique verbal: généralités
  - APPROX: un protocole psycholinguistique.
- 2 Un modèle computationnel du lexique verbal
  - Les dictionnaires: Réseaux Petits Mondes Hiérarchiques (RPMH)
  - 3 mesures disponibles
- 3 Catégorisation automatique
  - Les mesures affectées aux participants
  - Catégorisation et Résultats
  - Flexsem

# Prédiction de l'âge à partir d'énoncés linguistiques

Quand on entend quelqu'un, comment deviner son âge?

- ▶ critère phonétique/phonologique
- ▶ critère syntaxique
- ▶ critère lexico-sémantique

Comment est utilisé le critère lexico-sémantique?

# Prédiction de l'âge à partir d'énoncés linguistiques

Quand on entend quelqu'un, comment deviner son âge?

- ▶ critère phonétique/phonologique
- ▶ critère syntaxique
- ▶ critère lexico-sémantique

Comment est utilisé le critère lexico-sémantique?

# Prédiction de l'âge à partir d'énoncés linguistiques

Quand on entend quelqu'un, comment deviner son âge?

- ▶ critère phonétique/phonologique
- ▶ critère syntaxique
- ▶ critère lexico-sémantique

Comment est utilisé le critère lexico-sémantique?

# Prédiction de l'âge à partir d'énoncés linguistiques

Quand on entend quelqu'un, comment deviner son âge?

- ▶ critère phonétique/phonologique
- ▶ critère syntaxique
- ▶ critère lexico-sémantique

Comment est utilisé le critère lexico-sémantique?

# Prédiction de l'âge à partir d'énoncés linguistiques

Quand on entend quelqu'un, comment deviner son âge?

- ▶ critère phonétique/phonologique
- ▶ critère syntaxique
- ▶ critère lexico-sémantique

Comment est utilisé le critère lexico-sémantique?



# Acquisition lexicale des noms d'objet: rappels

L'acquisition lexicale des noms par les enfants s'appuie sur le principe d' **extension sémantique**:

- ▶ par similarité de forme à un niveau basique:  
une pomme  $\equiv$  une autre pomme;
- ▶ par extension catégoriel (similarité fonctionnelle):  
une pomme  $\in$  fruit  $\equiv$  une banane  $\in$  fruit;
- ▶ par similarité de forme intra-catégorielle:  
une pomme  $\in$  fruit  $\equiv$  une orange  $\in$  fruit;
- ▶ par similarité de forme inter-catégorielle:  
une pomme  $\in$  fruit  $\equiv$  une balle  $\in$  jeu;

Ces catégories correspondent souvent à des catégories du monde.

Elles sont peu dépendantes de la langue étudiée.

# Acquisition lexicale des noms d'objet: rappels

L'acquisition lexicale des noms par les enfants s'appuie sur le principe d' **extension sémantique**:

- ▶ par similarité de forme à un niveau basique:  
une pomme  $\equiv$  une autre pomme;
- ▶ par extension catégoriel (similarité fonctionnelle):  
une pomme  $\in$  fruit  $\equiv$  une banane  $\in$  fruit;
- ▶ par similarité de forme intra-catégorielle:  
une pomme  $\in$  fruit  $\equiv$  une orange  $\in$  fruit;
- ▶ par similarité de forme inter-catégorielle:  
une pomme  $\in$  fruit  $\equiv$  une balle  $\in$  jeu;

Ces catégories correspondent souvent à des catégories du monde.

Elles sont peu dépendantes de la langue étudiée.

## Acquisition lexicale des noms d'objet: rappels

L'acquisition lexicale des noms par les enfants s'appuie sur le principe d' **extension sémantique**:

- ▶ par similarité de forme à un niveau basique:  
une pomme  $\equiv$  une autre pomme;
- ▶ par extension catégoriel (similarité fonctionnelle):  
une pomme  $\in$  fruit  $\equiv$  une banane  $\in$  fruit;
- ▶ par similarité de forme intra-catégorielle:  
une pomme  $\in$  fruit  $\equiv$  une orange  $\in$  fruit;
- ▶ par similarité de forme inter-catégorielle:  
une pomme  $\in$  fruit  $\equiv$  une balle  $\in$  jeu;

Ces catégories correspondent souvent à des catégories du monde.

Elles sont peu dépendantes de la langue étudiée.

## Acquisition lexicale des noms d'objet: rappels

L'acquisition lexicale des noms par les enfants s'appuie sur le principe d' **extension sémantique**:

- ▶ par similarité de forme à un niveau basique:  
une pomme  $\equiv$  une autre pomme;
- ▶ par extension catégoriel (similarité fonctionnelle):  
une pomme  $\in$  fruit  $\equiv$  une banane  $\in$  fruit;
- ▶ par similarité de forme intra-catégorielle:  
une pomme  $\in$  fruit  $\equiv$  une orange  $\in$  fruit;
- ▶ par similarité de forme inter-catégorielle:  
une pomme  $\in$  fruit  $\equiv$  une balle  $\in$  jeu;

Ces catégories correspondent souvent à des catégories du monde.

Elles sont peu dépendantes de la langue étudiée.

## Acquisition lexicale des noms d'objet: rappels

L'acquisition lexicale des noms par les enfants s'appuie sur le principe d' **extension sémantique**:

- ▶ par similarité de forme à un niveau basique:  
une pomme  $\equiv$  une autre pomme;
- ▶ par extension catégoriel (similarité fonctionnelle):  
une pomme  $\in$  fruit  $\equiv$  une banane  $\in$  fruit;
- ▶ par similarité de forme intra-catégorielle:  
une pomme  $\in$  fruit  $\equiv$  une orange  $\in$  fruit;
- ▶ par similarité de forme inter-catégorielle:  
une pomme  $\in$  fruit  $\equiv$  une balle  $\in$  jeu;

Ces catégories correspondent souvent à des catégories du monde.

Elles sont peu dépendantes de la langue étudiée.

## Acquisition lexicale des noms d'objet: rappels

L'acquisition lexicale des noms par les enfants s'appuie sur le principe d' **extension sémantique**:

- ▶ par similarité de forme à un niveau basique:  
une pomme  $\equiv$  une autre pomme;
- ▶ par extension catégoriel (similarité fonctionnelle):  
une pomme  $\in$  fruit  $\equiv$  une banane  $\in$  fruit;
- ▶ par similarité de forme intra-catégorielle:  
une pomme  $\in$  fruit  $\equiv$  une orange  $\in$  fruit;
- ▶ par similarité de forme inter-catégorielle:  
une pomme  $\in$  fruit  $\equiv$  une balle  $\in$  jeu;

Ces catégories correspondent souvent à des catégories du monde.

Elles sont peu dépendantes de la langue étudiée.

## Acquisition lexicale des noms d'objet: rappels

L'acquisition lexicale des noms par les enfants s'appuie sur le principe d' **extension sémantique**:

- ▶ par similarité de forme à un niveau basique:  
une pomme  $\equiv$  une autre pomme;
- ▶ par extension catégoriel (similarité fonctionnelle):  
une pomme  $\in$  fruit  $\equiv$  une banane  $\in$  fruit;
- ▶ par similarité de forme intra-catégorielle:  
une pomme  $\in$  fruit  $\equiv$  une orange  $\in$  fruit;
- ▶ par similarité de forme inter-catégorielle:  
une pomme  $\in$  fruit  $\equiv$  une balle  $\in$  jeu;

Ces catégories correspondent souvent à des catégories du monde.

Elles sont peu dépendantes de la langue étudiée.

# Acquisition lexicale des verbes d'action: rappels (1)

Une action peut être décrite à l'aide de plusieurs dimensions.  
Par exemple, a minima, lorsqu'un personne déchire une feuille:

- ▶ l'agent qui fait l'action;
- ▶ l'objet (feuille) sur lequel est réalisée l'action;
- ▶ l'instrument à l'aide duquel est réalisée l'action;
- ▶ le résultat de l'action: changement de l'objet;
- ▶ la manière dont est réalisée l'action.



# Acquisition lexicale des verbes d'action: rappels (1)

Une action peut être décrite à l'aide de plusieurs dimensions.  
Par exemple, a minima, lorsqu'un personne déchire une feuille:

- ▶ l'agent qui fait l'action;
- ▶ l'objet (feuille) sur lequel est réalisée l'action;
- ▶ l'instrument à l'aide duquel est réalisée l'action;
- ▶ le résultat de l'action: changement de l'objet;
- ▶ la manière dont est réalisée l'action.

# Acquisition lexicale des verbes d'action: rappels (1)

Une action peut être décrite à l'aide de plusieurs dimensions.  
Par exemple, a minima, lorsqu'un personne déchire une feuille:

- ▶ l'agent qui fait l'action;
- ▶ l'objet (feuille) sur lequel est réalisée l'action;
- ▶ l'instrument à l'aide duquel est réalisée l'action;
- ▶ le résultat de l'action: changement de l'objet;
- ▶ la manière dont est réalisée l'action.

# Acquisition lexicale des verbes d'action: rappels (1)

Une action peut être décrite à l'aide de plusieurs dimensions.  
Par exemple, a minima, lorsqu'un personne déchire une feuille:

- ▶ l'agent qui fait l'action;
- ▶ l'objet (feuille) sur lequel est réalisée l'action;
- ▶ l'instrument à l'aide duquel est réalisée l'action;
- ▶ le résultat de l'action: changement de l'objet;
- ▶ la manière dont est réalisée l'action.

# Acquisition lexicale des verbes d'action: rappels (1)

Une action peut être décrite à l'aide de plusieurs dimensions.  
Par exemple, a minima, lorsqu'un personne déchire une feuille:

- ▶ l'agent qui fait l'action;
- ▶ l'objet (feuille) sur lequel est réalisée l'action;
- ▶ l'instrument à l'aide duquel est réalisée l'action;
- ▶ le résultat de l'action: changement de l'objet;
- ▶ la manière dont est réalisée l'action.

# Acquisition lexicale des verbes d'action: rappels (1)

Une action peut être décrite à l'aide de plusieurs dimensions.  
Par exemple, a minima, lorsqu'un personne déchire une feuille:

- ▶ l'agent qui fait l'action;
- ▶ l'objet (feuille) sur lequel est réalisée l'action;
- ▶ l'instrument à l'aide duquel est réalisée l'action;
- ▶ le résultat de l'action: changement de l'objet;
- ▶ la manière dont est réalisée l'action.

## Acquisition lexicale des verbes d'action: rappels (2)

Une verbe d'action met en jeu une catégorisation sur chacune de ces dimensions. Avec *déchirer*

- ▶ catégorie d'agents;
- ▶ catégorie d'objets;
- ▶ catégorie d'instruments;
- ▶ catégorie de résultats;
- ▶ catégorie de manières.

Ces catégories ne correspondent pas toujours à des catégories du monde  
Elles sont linguistiques.

## Acquisition lexicale des verbes d'action: rappels (2)

Une verbe d'action met en jeu une catégorisation sur chacune de ces dimensions. Avec *déchirer*

- ▶ catégorie d'agents;
- ▶ catégorie d'objets;
- ▶ catégorie d'instruments;
- ▶ catégorie de résultats;
- ▶ catégorie de manières.

Ces catégories ne correspondent pas toujours à des catégories du monde  
Elles sont linguistiques.

## Acquisition lexicale des verbes d'action: rappels (2)

Une verbe d'action met en jeu une catégorisation sur chacune de ces dimensions. Avec *déchirer*

- ▶ catégorie d'agents;
- ▶ catégorie d'objets;
- ▶ catégorie d'instruments;
- ▶ catégorie de résultats;
- ▶ catégorie de manières.

Ces catégories ne correspondent pas toujours à des catégories du monde  
Elles sont linguistiques.



## Acquisition lexicale des verbes d'action: rappels (2)

Une verbe d'action met en jeu une catégorisation sur chacune de ces dimensions. Avec *déchirer*

- ▶ catégorie d'agents;
- ▶ catégorie d'objets;
- ▶ catégorie d'instruments;
- ▶ catégorie de résultats;
- ▶ catégorie de manières.

Ces catégories ne correspondent pas toujours à des catégories du monde  
Elles sont linguistiques.

## Acquisition lexicale des verbes d'action: rappels (2)

Une verbe d'action met en jeu une catégorisation sur chacune de ces dimensions. Avec *déchirer*

- ▶ catégorie d'agents;
- ▶ catégorie d'objets;
- ▶ catégorie d'instruments;
- ▶ catégorie de résultats;
- ▶ catégorie de manières.

Ces catégories ne correspondent pas toujours à des catégories du monde  
Elles sont linguistiques.

## Acquisition lexicale des verbes d'action: rappels (2)

Une verbe d'action met en jeu une catégorisation sur chacune de ces dimensions. Avec *déchirer*

- ▶ catégorie d'agents;
- ▶ catégorie d'objets;
- ▶ catégorie d'instruments;
- ▶ catégorie de résultats;
- ▶ catégorie de manières.

Ces catégories ne correspondent pas toujours à des catégories du monde  
Elles sont linguistiques.

## Acquisition lexicale des verbes d'action: rappels (2)

Une verbe d'action met en jeu une catégorisation sur chacune de ces dimensions. Avec *déchirer*

- ▶ catégorie d'agents;
- ▶ catégorie d'objets;
- ▶ catégorie d'instruments;
- ▶ catégorie de résultats;
- ▶ catégorie de manières.

Ces catégories ne correspondent pas toujours à des catégories du monde  
Elles sont linguistiques.

# Acquisition lexicale des verbes d'action: rappels (3)

Lorsque l'enfant acquiert le lexique des verbes d'action, il utilise également le principe d'extension sémantique mais:

La sur-extension verbale est multidimensionnelle.

# Acquisition lexicale des verbes d'action: rappels (3)

Lorsque l'enfant acquiert le lexique des verbes d'action, il utilise également le principe d'extension sémantique mais:

La sur-extension verbale est multidimensionnelle.

## Acquisition lexicale des verbes d'action: rappels (4)

En cas du manque du verbe conventionnel spécifique (*briser* lorsque l'on brise un verre), l'enfant:

- ▶ utilise un verbe générique: catégories larges sur chaque dimensions  
*casser*
- ▶ **ou** utilise une approximation (sur-extension): erreur de catégorisation sur une ou plusieurs dimensions  
*déchiqueter*

## Acquisition lexicale des verbes d'action: rappels (4)

En cas du manque du verbe conventionnel spécifique (*briser* lorsque l'on brise un verre), l'enfant:

- ▶ utilise un verbe générique: catégories larges sur chaque dimensions  
*casser*
- ▶ **ou** utilise une approximation (sur-extension): erreur de catégorisation sur une ou plusieurs dimensions  
*déchiqueter*



## Acquisition lexicale des verbes d'action: rappels (4)

En cas du manque du verbe conventionnel spécifique (*briser* lorsque l'on brise un verre), l'enfant:

- ▶ utilise un verbe générique: catégories larges sur chaque dimensions  
*casser*
- ▶ **ou** utilise une approximation (sur-extension): erreur de catégorisation sur une ou plusieurs dimensions  
*déchiqueter*

# Acquisition lexicale des verbes d'action: rappels (5)

Les verbes d'action sont acquis plus tard que les noms d'objet:

- ▶ les actions sont plus difficiles à catégoriser que les entités
- ▶ le problème de l'attention conjointe
- ▶ le polymorphisme des verbes

Etudier l'emploi des verbes devrait être utile pour faire de la  
prédiction sur l'âge.

Jeunes Enfants VS Adultes

# Acquisition lexicale des verbes d'action: rappels (5)

Les verbes d'action sont acquis plus tard que les noms d'objet:

- ▶ les actions sont plus difficiles à catégoriser que les entités
- ▶ le problème de l'attention conjointe
- ▶ le polymorphisme des verbes

Etudier l'emploi des verbes devrait être utile pour faire de la  
prédiction sur l'âge.

Jeunes Enfants VS Adultes

# Acquisition lexicale des verbes d'action: rappels (5)

Les verbes d'action sont acquis plus tard que les noms d'objet:

- ▶ les actions sont plus difficiles à catégoriser que les entités
- ▶ le problème de l'attention conjointe
- ▶ le polymorphisme des verbes

Etudier l'emploi des verbes devrait être utile pour faire de la  
prédiction sur l'âge.

Jeunes Enfants VS Adultes

# Acquisition lexicale des verbes d'action: rappels (5)

Les verbes d'action sont acquis plus tard que les noms d'objet:

- ▶ les actions sont plus difficiles à catégoriser que les entités
- ▶ le problème de l'attention conjointe
- ▶ le polymorphisme des verbes

Etudier l'emploi des verbes devrait être utile pour faire de la  
prédiction sur l'âge.  
Jeunes Enfants VS Adultes

# Acquisition lexicale des verbes d'action: rappels (5)

Les verbes d'action sont acquis plus tard que les noms d'objet:

- ▶ les actions sont plus difficiles à catégoriser que les entités
- ▶ le problème de l'attention conjointe
- ▶ le polymorphisme des verbes

Etudier l'emploi des verbes devrait être utile pour faire de la  
prédiction sur l'âge.

Jeunes Enfants VS Adultes

# Plan

- 1 acquisition du lexique verbal
  - L'acquisition du lexique verbal: généralités
  - **APPROX: un protocole psycholinguistique.**
- 2 Un modèle computationnel du lexique verbal
  - Les dictionnaires: Réseaux Petits Mondes Hiérarchiques (RPMH)
  - 3 mesures disponibles
- 3 Catégorisation automatique
  - Les mesures affectées aux participants
  - Catégorisation et Résultats
  - Flexsem

# APPROX: qu'est-ce que c'est?

[Duvignau et al., 2005]

- ▶ 17 films d'action de transformation d'objet;
- ▶ Une question: "qu'est-ce qu'elle a fait, la dame?";
- ▶ L'analyse des réponses;
- ▶ Populations: enfants et adultes en Français et Mandarin (projet M3)



# APPROX: qu'est-ce que c'est?

[Duvignau et al., 2005]

- ▶ 17 films d'action de transformation d'objet;
- ▶ Une question: "qu'est-ce qu'elle a fait, la dame?";
- ▶ L'analyse des réponses;
- ▶ Populations: enfants et adultes en Français et Mandarin (projet M3)

# APPROX: qu'est-ce que c'est?

[Duvignau et al., 2005]

- ▶ 17 films d'action de transformation d'objet;
- ▶ Une question: "qu'est-ce qu'elle a fait, la dame?";
- ▶ L'analyse des réponses;
- ▶ Populations: enfants et adultes en Français et Mandarin (projet M3)

# APPROX: qu'est-ce que c'est?

[Duvignau et al., 2005]

- ▶ 17 films d'action de transformation d'objet;
- ▶ Une question: "qu'est-ce qu'elle a fait, la dame?";
- ▶ L'analyse des réponses;
- ▶ Populations: enfants et adultes en Français et Mandarin (projet M3)

# APPROX: qu'est-ce que c'est?

[Duvignau et al., 2005]

- ▶ 17 films d'action de transformation d'objet;
- ▶ Une question: "qu'est-ce qu'elle a fait, la dame?";
- ▶ L'analyse des réponses;
- ▶ Populations: enfants et adultes en Français et Mandarin (projet M3)

# APPROX: l'analyse des réponses.

Les verbes qui décrivent les actions sont analysés à partir de 2 critères:

- ▶ Critère Conventionnel/Approximation
- ▶ Critère Générique/Spécifique

# APPROX: l'analyse des réponses.

Les verbes qui décrivent les actions sont analysés à partir de 2 critères:

- ▶ Critère Conventionnel/Approximation
- ▶ Critère Générique/Spécifique

# APPROX: l'analyse des réponses.

Les verbes qui décrivent les actions sont analysés à partir de 2 critères:

- ▶ Critère Conventionnel/Approximation
- ▶ Critère Générique/Spécifique

# Premiers résultats en Français

Table: Premiers résultats en Français

	2-5 ans	18-40 ans
% d'approximations	37%	6%
% de verbes verbes génériques	60%	16%

Les enfants sains de 2-5 ans produisent significativement:

- ▶ plus d'approximations verbales que les adultes sains;
- ▶ plus de verbes génériques que les adultes sains;



Une modèle computationnel pertinent de l'organisation du  
lexique verbale  
doit être en adéquation avec ces résultats!

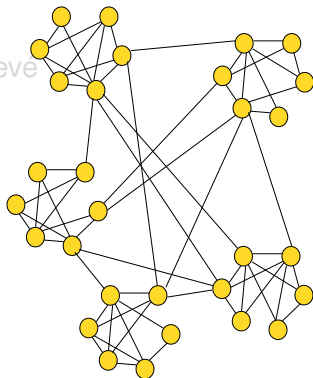
# Plan

- 1 acquisition du lexique verbal
  - L'acquisition du lexique verbal: généralités
  - APPROX: un protocole psycholinguistique.
- 2 Un modèle computationnel du lexique verbal
  - Les dictionnaires: Réseaux Petits Mondes Hiérarchiques (RPMH)
  - 3 mesures disponibles
- 3 Catégorisation automatique
  - Les mesures affectées aux participants
  - Catégorisation et Résultats
  - Flexsem

# Réseaux Petits Mondes

Coefficient de clustering élevé  
Les amis de mes amis  
sont sûrement mes amis.

Court-circuits!  
(Short cuts)

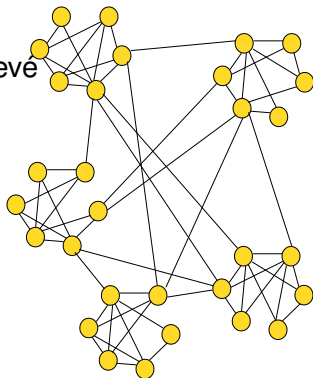


# Réseaux Petits Mondes

**Coefficient de clustering élevé**

Les amis de mes amis  
sont sûrement mes amis.

Court-circuits!  
(Short cuts)

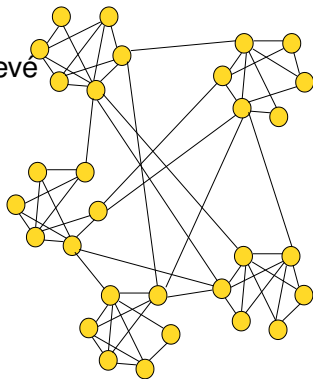


# Réseaux Petits Mondes

Coefficient de clustering élevé

Les amis de mes amis  
sont sûrement mes amis.

Court-circuits!  
(Short cuts)

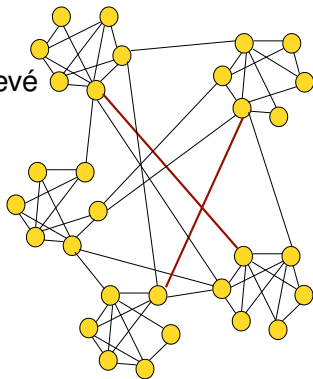


# Réseaux Petits Mondes

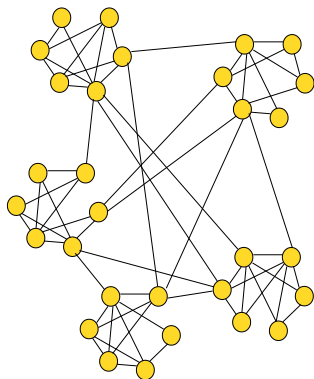
Coefficient de clustering élevé

Les amis de mes amis  
sont sûrement mes amis.

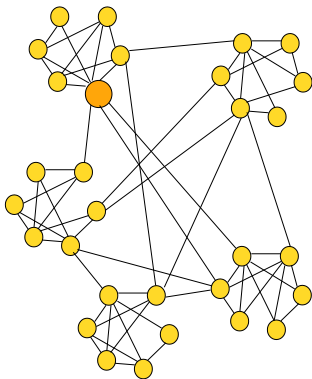
Court-circuits!  
(Short cuts)



# Les Petits Mondes sont Hiérarchiques



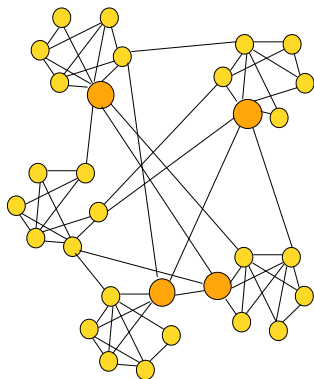
# Les Petits Mondes sont Hiérarchiques





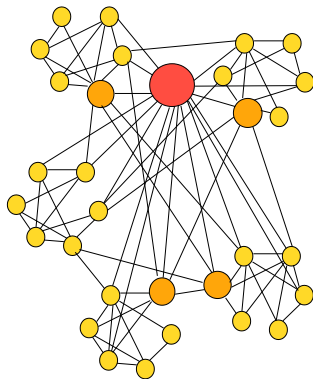
# Les Petits Mondes sont Hiérarchiques

Les RPM sont  
Hiérarchiques.



# Les Petits Mondes sont Hiérarchiques

Les RPM sont  
Hiérarchiques.



# Graphes de Dictionnaires: RPMH

## COMMENT CONSTRUIRE UN GRAPHE LEXICAL?

Les graphes de synonymes sont  
des Réseaux Petits Mondes  
Hiérarchiques.

# Graphes de Dictionnaires: RPMH

## COMMENT CONSTRUIRE UN GRAPHE LEXICAL?

Word A



Les graphes de synonymes sont  
des Réseaux Petits Mondes  
Hiérarchiques.

# Graphes de Dictionnaires: RPMH

COMMENT CONSTRUIRE UN GRAPHE LEXICAL?

Word A



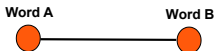
Word B



Les graphes de synonymes sont  
des Réseaux Petits Mondes  
Hiérarchiques.

# Graphes de Dictionnaires: RPMH

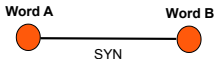
COMMENT CONSTRUIRE UN GRAPHE LEXICAL?



Les graphes de synonymes sont  
des Réseaux Petits Mondes  
Hiérarchiques.

# Graphes de Dictionnaires: RPMH

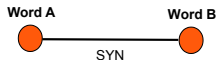
COMMENT CONSTRUIRE UN GRAPHE LEXICAL?



Les graphes de synonymes sont  
des Réseaux Petits Mondes  
Hiérarchiques.

# Graphes de Dictionnaires: RPMH

COMMENT CONSTRUIRE UN GRAPHE LEXICAL?



Les graphes de synonymes sont  
des Réseaux Petits Mondes  
Hiérarchiques.





## Les Dictionnaires Utilisés pour le Projet M3

- ▶ Pour le Français: **DicoSyn.Verb**  
Extrait à partir de la concaténation des synonymes verbaux de 7 dictionnaires de synonymes français.
- ▶ Pour le Mandarin: **CilinCWN.Verb**  
Extrait à partir de la fusion du WordNet chinois et du thesaurus chinois TongYiCi CiLin.

# Plan

- 1 acquisition du lexique verbal
  - L'acquisition du lexique verbal: généralités
  - APPROX: un protocole psycholinguistique.
- 2 Un modèle computationnel du lexique verbal
  - Les dictionnaires: Réseaux Petits Mondes Hiérarchiques (RPMH)
  - 3 mesures disponibles
- 3 Catégorisation automatique
  - Les mesures affectées aux participants
  - Catégorisation et Résultats
  - Flexsem

## Trois mesures

- ▶ L'incidence des sommets;
- ▶ L'indice de proximité sémantique ou proxémie: *Prox*;
- ▶ L'indice de flexibilité sémantique: *ReFlex*;

## Trois mesures

- ▶ L'incidence des sommets;
- ▶ L'indice de proximité sémantique ou proxémie: *Prox*;
- ▶ L'indice de flexibilité sémantique: *ReFlex*;

## Trois mesures

- ▶ L'incidence des sommets;
- ▶ L'indice de proximité sémantique ou proxémie: *Prox*;
- ▶ L'indice de flexibilité sémantique: *ReFlex*;

# L'incidence: hypothèse

Hypothèse: L'incidence correspond au critère  
Générique/Spécifique.

# La proxémie

La probabilité pour le marcheur  
(Mr.Meaning),  
se promenant sur le graphe de  
verbes, d'atteindre un sommet  
dans le graphe.



# La proxémie

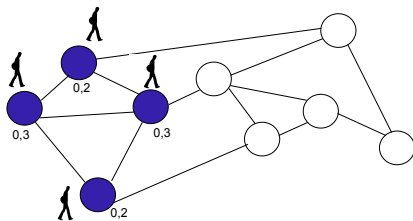
Etat initial  $\Delta_0$  ( $t = 0$ )

$t = 1$

$t = 2$

$t = 3$

$t = 4$  pour des graphes  
lexicaux





# La proxémie

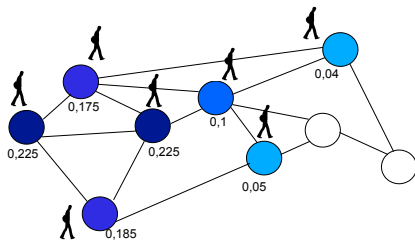
Etat initial  $\Delta_0$  ( $t = 0$ )

$t = 1$

$t = 2$

$t = 3$

$t = 4$  pour des graphes  
lexicaux



# La proxémie

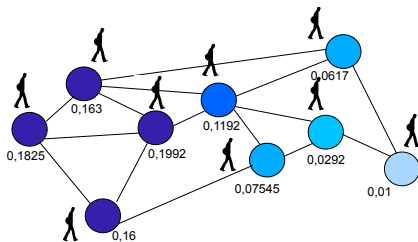
Etat initial  $\Delta_0$  ( $t = 0$ )

$t = 1$

$t = 2$

$t = 3$

$t = 4$  pour des graphes  
lexicaux



# La proxémie

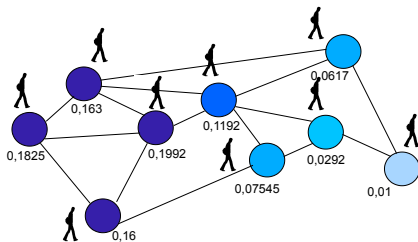
Etat initial  $\Delta_0$  ( $t = 0$ )

$t = 1$

$t = 2$

$t = 3$

$t = 4$  pour des graphes  
lexicaux



# La proxémie

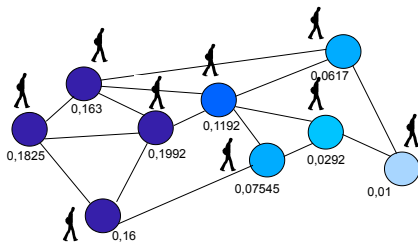
Etat initial  $\Delta_0$  ( $t = 0$ )

$t = 1$

$t = 2$

$t = 3$

$t = 4$  pour des graphes  
lexicaux



# La proxémie: hypothèse

Hypothèse: La proxémie, *Prox*, permet de saisir la distinction verbes conventionnels/verbes approximatifs.

# ReFLex

L'indice de flexibilité sémantique, *ReFlex*, entre l'état initial (verbe virtuel: zone sémantique)  $\Delta_0$  et le verbe  $v$ :

$$ReFlex(G, \Delta_0, v, t) = \frac{Prox(G, \Delta_0, v, t)}{deg(v)}$$

où  $deg(v)$  est l'incidence de  $v$  dans le graphe  $G$

# ReFlex: hypothèse

Hypothèse:

L'indice *ReFlex* permet de saisir la distinction verbes  
conventionnel/verbes approximatifs  
et la distinction entre verbes générique/spécifique.

# Plan

- 1 acquisition du lexique verbal
  - L'acquisition du lexique verbal: généralités
  - APPROX: un protocole psycholinguistique.
- 2 Un modèle computationnel du lexique verbal
  - Les dictionnaires: Réseaux Petits Mondes Hiérarchiques (RPMH)
  - 3 mesures disponibles
- 3 **Catégorisation automatique**
  - **Les mesures affectées aux participants**
  - Catégorisation et Résultats
  - Flexsem



# La distribution initiale

L'objectif est de catégoriser les enfants de *24-59* mois versus des adultes de *18-40* ans en utilisant l'incidence, l'indice de proxémie et l'indice de flexibilité sémantique des verbes produits.

- ▶ L'incidence ne dépend que du graphe;
- ▶ Les indices *Prox* et *ReFlex* dependent du graphe et de la distribution initiale

# La distribution initiale

L'objectif est de catégoriser les enfants de *24-59* mois versus des adultes de *18-40* ans en utilisant l'incidence, l'indice de proxémie et l'indice de flexibilité sémantique des verbes produits.

- ▶ L'incidence ne dépend que du graphe;
- ▶ Les indices *Prox* et *ReFlex* dependent du graphe et de la distribution initiale

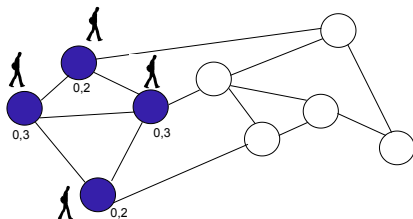
# La distribution initiale

L'objectif est de catégoriser les enfants de *24-59* mois versus des adultes de *18-40* ans en utilisant l'incidence, l'indice de proxémie et l'indice de flexibilité sémantique des verbes produits.

- ▶ L'incidence ne dépend que du graphe;
- ▶ Les indices *Prox* et *ReFlex* dependent du graphe et de la distribution initiale

# L'indice de proximité sémantique

Etat initial ( $t = 0$ )



# La distribution initiale

Pour chaque film, la distribution initiale est calculée:

- ▶ A partir d'une liste  $L_v$  de verbes produits par:
  - 25 adultes aléatoirement sélectionnés;
  - 25 enfants aléatoirement sélectionnés.
- ▶ La distribution initiale  $\Delta_0$  est telle que:

$$[\Delta_0]_{v_i} = \frac{freq(v_i)}{\sum_{v \in L_v} freq(s)}, v_i \in L_v \quad (1)$$

# La distribution initiale

Pour chaque film, la distribution initiale est calculée:

- ▶ A partir d'une liste  $L_v$  de verbes produits par:
  - 25 adultes aléatoirement sélectionnés;
  - 25 enfants aléatoirement sélectionnés.
- ▶ La distribution initiale  $\Delta_0$  est telle que:

$$[\Delta_0]_{v_i} = \frac{\text{freq}(v_i)}{\sum_{v \in L_v} \text{freq}(s)}, v_i \in L_v \quad (1)$$

# La distribution initiale

Pour chaque film, la distribution initiale est calculée:

- ▶ A partir d'une liste  $L_v$  de verbes produits par:
  - 25 adultes aléatoirement sélectionnés;
  - 25 enfants aléatoirement sélectionnés.
- ▶ La distribution initiale  $\Delta_0$  est telle que:

$$[\Delta_0]_{v_i} = \frac{freq(v_i)}{\sum_{v \in L_v} freq(s)}, v_i \in L_v \quad (1)$$

# Une mesure pour Catégoriser les Jeunes Enfants des Adultes





# Une mesure pour Catégoriser les Jeunes Enfants des Adultes



# Une mesure pour Catégoriser les Jeunes Enfants des Adultes

SCORE



SCORE



# Une mesure pour Catégoriser les Jeunes Enfants des Adultes

SCORE



$\neq$

SCORE



# Le score d'incidence: *Degree*<sub>score</sub>

*Degree*<sub>score</sub>



*Degree*<sub>score</sub>



*Degree*<sub>score</sub>  
Incidence moyenne des verbes

## Le score d'incidence: $Degree_{score}$

$Degree_{score}$



$Degree_{score}$



$Degree_{score}$

Incidence moyenne des verbes

## Le score de proxémie: $ProX_{score}$

$ProX_{score}$



$ProX_{score}$



$ProX_{score}$

Moyenne, sur l'ensemble des films, des proxémies des verbes  
à partir de la distribution initiale donnée pour chaque film.

## Le score de proxémie: $ProX_{score}$

$ProX_{score}$



$ProX_{score}$



$ProX_{score}$

Moyenne, sur l'ensemble des films, des proxémies des verbes  
à partir de la distribution initiale donnée pour chaque film.

# Le score de flexibilité *ReFlex*<sub>score</sub>

*Flex*<sub>score</sub>



*Flex*<sub>score</sub>



*Flex*<sub>score</sub>

Moyenne, sur l'ensemble des films, des indices *ReFlex* des verbes à partir de la distribution initiale donnée pour chaque film.



# Le score de flexibilité *ReFlex*<sub>score</sub>

*Flex*<sub>score</sub>



*Flex*<sub>score</sub>



*Flex*<sub>score</sub>

Moyenne, sur l'ensemble des films, des indices *ReFlex* des verbes  
à partir de la distribution initiale donnée pour chaque film.

# Plan

- 1 acquisition du lexique verbal
  - L'acquisition du lexique verbal: généralités
  - APPROX: un protocole psycholinguistique.
- 2 Un modèle computationnel du lexique verbal
  - Les dictionnaires: Réseaux Petits Mondes Hiérarchiques (RPMH)
  - 3 mesures disponibles
- 3 **Catégorisation automatique**
  - Les mesures affectées aux participants
  - **Catégorisation et Résultats**
  - Flexsem

# Catégorisation

	Français	Mandarin
<i>24-59 mois</i>	74	28
<i>18-40 ans</i>	76	60

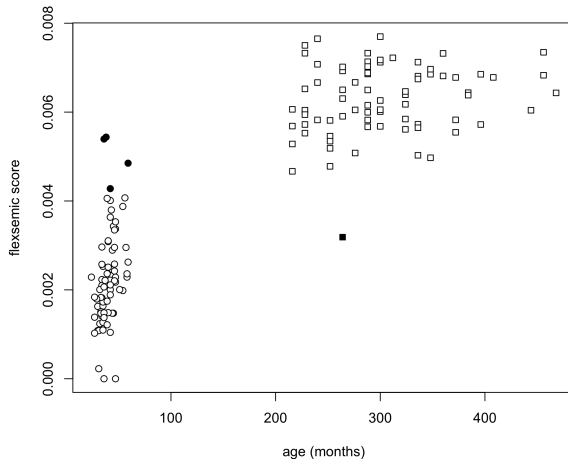
- ▶ Procédure: Clustering 2-means à partir de chacune des mesures;
- ▶ Evaluation: Comparaison avec les populations d'origine.

## Results

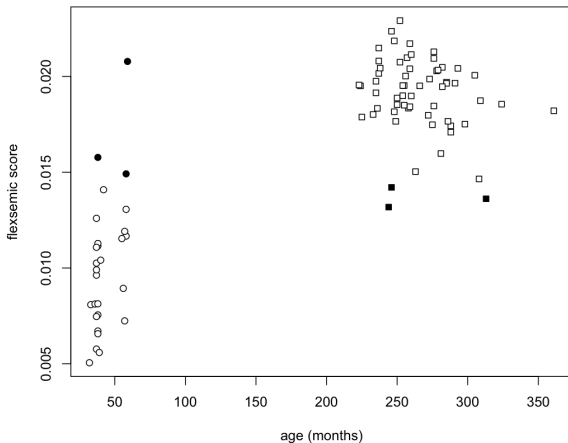
SCORE	LANGUAGE	F-SCORE	$\kappa$
<i>Degree</i> <sub>score</sub>	Mandarin	0,65	0,18
<i>Prox</i> <sub>score</sub>	Mandarin	0,85	0,64
<i>Flex</i> <sub>score</sub>	Mandarin	<b>0,93</b>	<b>0,84</b>
<i>Degree</i> <sub>score</sub>	Français	0,87	0,75
<i>Prox</i> <sub>score</sub>	Français	0,59	0,18
<i>Flex</i> <sub>score</sub>	Français	<b>0,96</b>	<b>0,92</b>

Table: Résultats du clustering 2-means en Français et en Mandarin

# Résultats en Français



# Résultats en Chinois



# Premiers résultats en Français

Table: Premiers résultats en Français

	2-5 ans	18-40 ans
% d'approximations	37%	6%
% de verbes verbes génériques	60%	16%

Les enfants sains de 2-5 ans produisent significativement:

- ▶ plus d'approximations verbales que les adultes sains;
- ▶ plus de verbes génériques que les adultes sains.

# Plan

- 1 acquisition du lexique verbal
  - L'acquisition du lexique verbal: généralités
  - APPROX: un protocole psycholinguistique.
- 2 Un modèle computationnel du lexique verbal
  - Les dictionnaires: Réseaux Petits Mondes Hiérarchiques (RPMH)
  - 3 mesures disponibles
- 3 Catégorisation automatique
  - Les mesures affectées aux participants
  - Catégorisation et Résultats
  - Flexsem






# Flexsem

Flexsem est une plateforme (<http://erss.irit.fr/flexsem>) contenant les données Approx:

- ▶ plusieurs langues: Français, Mandarin, Polonais, Coréen, Portugais . . . ;
- ▶ plusieurs âges;
- ▶ plusieurs populations pathologiques: Alzheimer, Apserger, Aphasiques, Schizophrènes . . . .

# Résumé

- ▶ A partir d'observations psycholinguistiques et de la structure RPMH des graphes lexicaux, nous avons construit une nouvelle mesure de l'acquisition du lexique verbale: la mesure de *flexibilité sémantique*
- ▶ Cette mesure est pertinent pour catégoriser automatiquement les jeunes enfants et les adultes en Français et en Mandarin.
- ▶ Perspectives
  - Cette mesure de *flexibilité* pourrait être utilisé pour catégoriser d'autres populations (e.g. Apserger-vs-jeunes enfants sains)
  - Cette mesure de *flexibilité* pourrait être réutilisée dans des applications prédictives de profilage.

-  Desalle, Y., Gaume, B., and Duvignau, K. (2009).  
SLAM : Solutions lexicales automatique pour métaphores.  
[Traitements Automatiques des Langues](#), 50(1):145–175.
-  Duvignau, K., Gaume, B., and Kern, S. (2005).  
Semantic approximations intraconcept vs. interconcepts in early verbal  
lexicon: flexibility against error.  
[In Proceedings of ELA 2005, Emergence of language abilities: ontogeny  
and phylogeny.](#)
-  Gaume, B., Duvignau, K., Gasquet, O., and Gineste, M.-D. (2002).  
Forms of meaning, meaning of forms.  
[Journal of Experimental and Theoretical Artificial Intelligence](#), 14:61–74.



Gaume, B., Duvignau, K., Prévot, L., and Desalle, Y. (2008).

Toward a cognitive organization for electronic dictionaries, the case for semantic proxemy.

In [Coling 2008: Proceedings of the Workshop on Cognitive Aspects of the Lexicon \(COGALEX 2008\)](#), pages 86–93, Manchester.



Gaume, B., Mathieu, F., and Navarro, E. (2010).

Building Real-World Complex Networks by Wandering on Random Graphs.

[Information - Interaction - Intelligence](#), 10(1).



Sajous, F., Navarro, E., Gaume, B., Prévot, L., and Chudy, Y. (2010).

Semi-automatic Endogenous Enrichment of Collaboratively Constructed Lexical Resources: Piggybacking onto Wiktionary.

In Loftsson, H., Rögvaldsson, E., and Helgadóttir, S., editors, [Proceedings of the 7th International Conference on NLP \(IceTAL 2010\)](#), volume 6233 of [LNAI](#), pages 332–344, Reykjavik, Iceland. Springer-Verlag.