

# VERS UNE ANALYSE AUTOMATIQUE DES DISCUSSIONS LORS D'ATELIERS-PHILO



PHILIPPE DESSUS, LAB. SCIENCES DE L'ÉDUCATION  
PRÉSENTATION AU LIDILEM - 17 JUIN 2015

UNIVERSITÉ  
GRENOBLE  
ALPES

espe

École supérieure  
du professorat  
et de l'éducation  
Académie de Grenoble



# PROPOS

- ❖ **Question générale** : En quoi les approches interdisciplinaires de sciences cognitives (au sens large) peuvent aider à mieux comprendre et analyser ce qui se dit dans un atelier-philo
- ❖ **Point d'entrée** : Analyse automatique du contenu sémantique des discussions, de la complexité des textes soumis aux élèves
- ❖ **But** : Réfléchir à des questions de recherche sur ces sujets
- ❖ Références disponibles à <http://bit.ly/atelier-philo>

# PLAN

- ❖ Sciences cognitives appliquées à l'éducation et LSA (*Latent Semantic Analysis*)
- ❖ Qu'est-ce qui se passe dans les ateliers-philo et comment l'analyser ?
- ❖ Présentation de *ReaderBench*, un “couteau suisse” pour analyser différents paramètres liés à la lecture/ compréhension de textes
- ❖ Point sur le travail en cours et à faire

# SCIENCES COGNITIVES ET LATENT SEMANTIC ANALYSIS



COMMENT UNE TECHNIQUE DE RECHERCHE D'INFORMATION À LA GOOGLE  
EST DEVENUE UN OUTIL DE MODÉLISATION COGNITIVE

# LE DÉBUT DE L'HISTOIRE

Copyright 1997 by the American Psychological Association, Inc.  
0033-295X/97/\$3.00

Psychological Review  
1997, Vol. 104, No. 2, 211-240

## A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge

Susan T. Dumais  
Bellcore

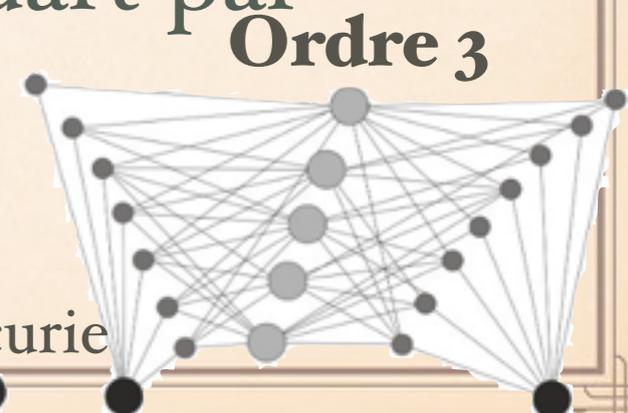
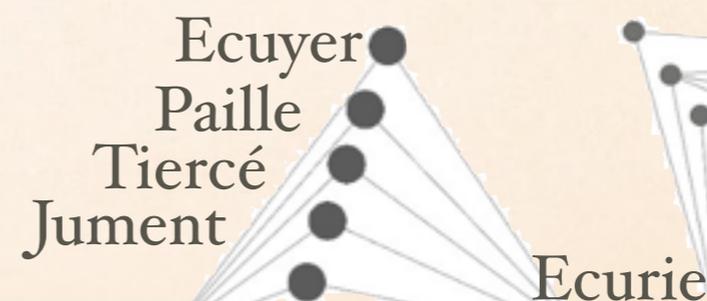
Thomas K Landauer  
University of Colorado at Boulder

How do people know as much as they do with as little information as they get? The problem takes many forms; learning vocabulary from text is an especially dramatic and convenient case for research. A new general theory of acquired similarity and knowledge representation, latent semantic analysis (LSA), is presented and used to successfully simulate such learning and several other psycholinguistic phenomena. By inducing global knowledge indirectly from local co-occurrence data in a large body of representative text, LSA acquired knowledge about the full vocabulary of English at a comparable rate to schoolchildren. LSA uses no prior linguistic or perceptual similarity knowledge; it is based solely on a general mathematical learning method that achieves powerful inductive effects by extracting the right number of dimensions (e.g., 300) to represent objects and contexts. Relations to other theories, phenomena, and problems are sketched.

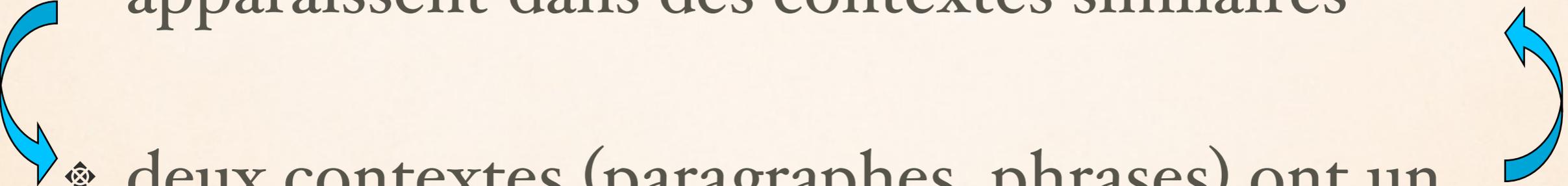
◆ Comment peut-on apprendre tant en étant exposé à si peu ?

# LA PAUVRETÉ DU STIMULUS

- ❖ En lisant des textes, on apprend des choses sur les mots du texte, mais aussi sur des mots qui n'y figurent pas
- ❖ Des mots jamais co-occurents dans des documents peuvent être jugés sémantiquement proches : co-occurrences de second...  $n$  ordre
- ❖ Un élève de fin de primaire apprend par la lecture env. 10 nouveaux mots par jour, dont seulement le quart par lecture directe



# COMMENT FONCTIONNE L'ANALYSE SÉMANTIQUE LATENTE ? (1/2)

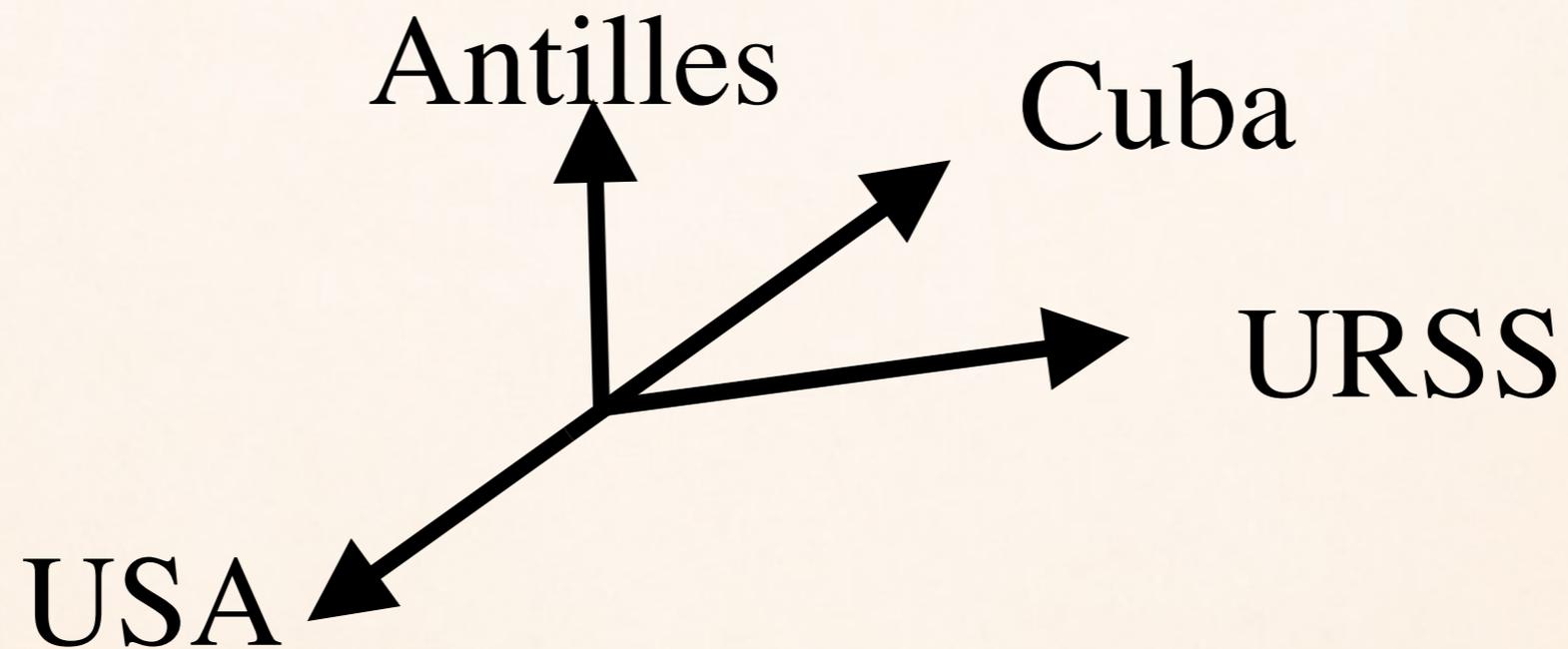
- ❖ LSA méthode d'analyse statistique de grands corpus textuels (type d'analyse factorielle). Part du principe que
    - ❖ deux mots ont un sens similaire s'ils apparaissent dans des contextes similaires
    - ❖ deux contextes (paragrapes, phrases) ont un sens similaire (contiennent des informations similaires) s'ils contiennent des mots de sens similaire
- 

# COMMENT FONCTIONNE LSA (2/2) ?

- ❖ Approche «paquets de mots», pas de prise en compte de la ponctuation, de la syntaxe, du style, etc.
- ❖ Pas de prise en compte de «mots outils» (déterminants, prépositions)
- ❖ Importance de la composition du corpus choisi pour «entraîner» LSA : tout mot absent du corpus ne peut être analysé

## PRISE EN COMPTE DE CERTAINES RELATIONS SÉMANTIQUES

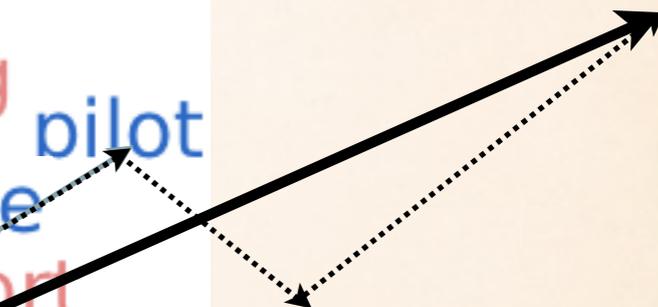
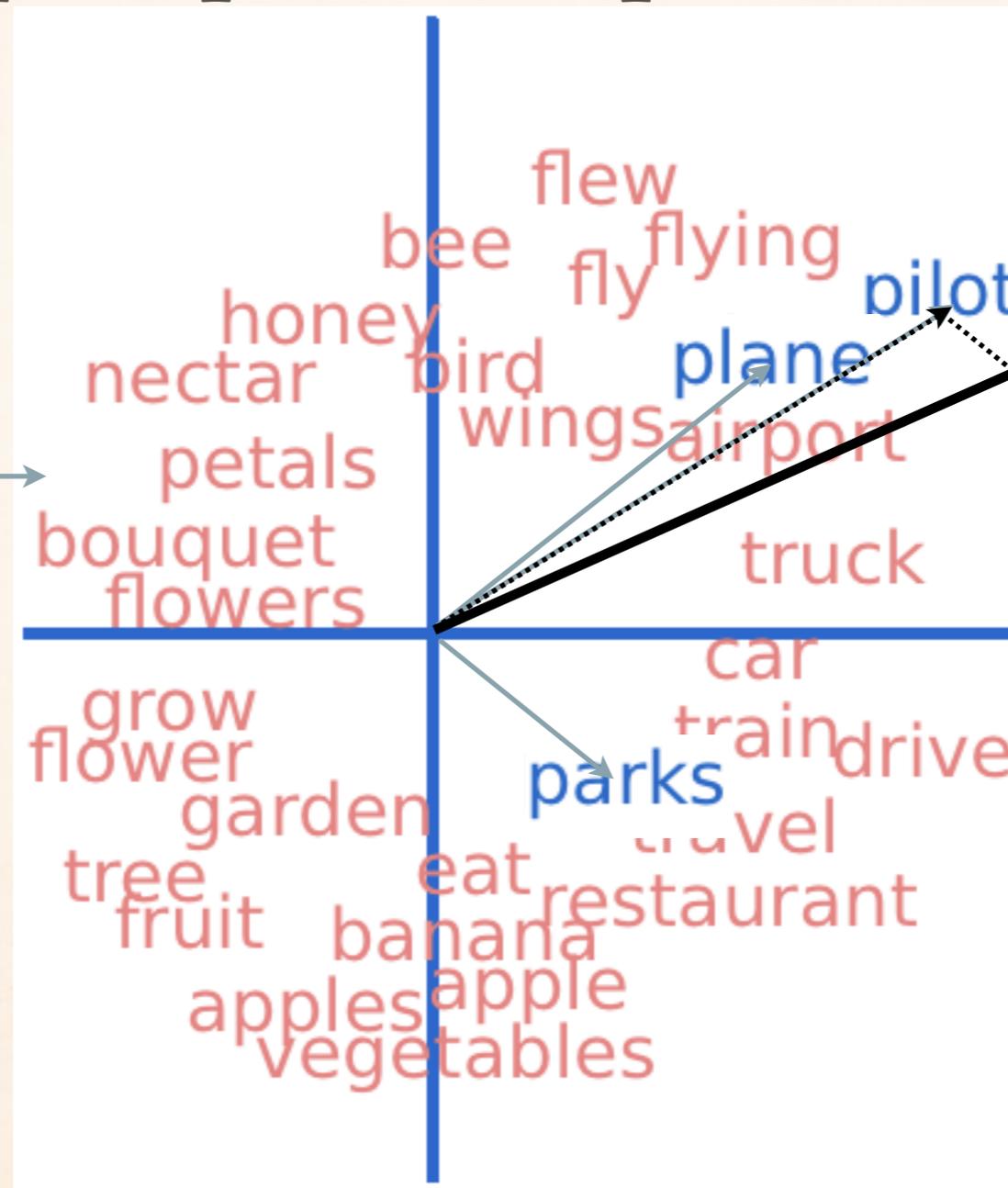
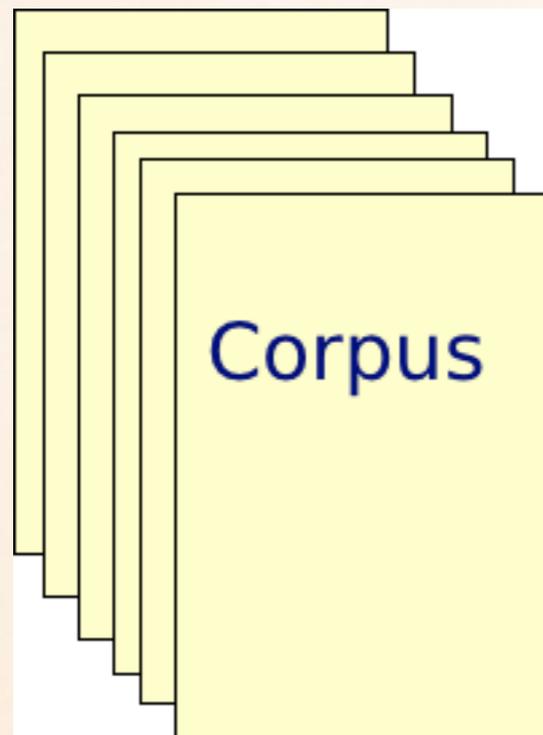
- \* Règle certaines contraintes liées au sens des mots établies par [Tversky 77] dans la comparaison de pays (représentation fictive)



# REPRÉSENTATION VECTORIELLE DES PHRASES, PARAGRAPHES

📖 [Lemaire & Denhière 05]

## “The pilot parks the plane”



UTILISATIONS DE LSA  
À DES FINS ÉDUCATIVES  
 [Dessus 09; Lemaire & Dessus 03]

- ❖ Très nombreuses utilisations de LSA à des fins éducatives pour
  - ❖ Sélectionner ou produire des textes (p. ex., résumés)
  - ❖ Evaluer des textes produits (p. ex., par des élèves)
  - ❖ Evaluer ou détecter les capacités d'autorégulation, les intentions (p. ex., d'opérateurs)

# VUE DE NOS RECHERCHES

**1997-99**

Premiers essais sur LSA :  
prédiction de liens hypertextes ; de réponses à QCM (B. Lemaire)

[Dessus 00]

**2000-2**

Systemes d'éval. automatique de résumés de cours :  
*Apex 1.0 & Apex 2.0*  
(A. Vernier, B. Lemaire)

[Lemaire & Dessus 01]

**2005-9**

Systeme d'entraî-  
nement au résumé :  
*Résum'Web*  
(S. Mandin, B. Lemaire, G. Denhière)

[Lemaire et al. 05 ; Mandin 10]

**2010-pr.**

Systeme d'aide à la lecture :  
*ReaderBench*  
(M. Dascalu, S. Trausan-Matu)

[Dascalu et al. 13]

# SCIENCES COGNITIVES APPLIQUÉES À L'ÉDUCATION



[Dessus & Gentaz 06 ; Gentaz & Dessus 04]

- ❖ **Psychologie, sciences du langage** : Disposer de modèles cognitifs de l'activité à étudier. Comprendre un texte [McNamara 04] ou le résumer [Kintsch & van Dijk 75]
- ❖ **Informatique** : Réaliser une implémentation informatique qui simule et puisse aider l'activité des élèves, de l'enseignant
- ❖ **Psycho, Education** : Valider (en amont ou aval) par comparaison avec comportements humains (enseignants, élèves, experts)
- ❖ **Education** : Utiliser et valider l'implémentation dans des contextes de laboratoire ou scolaires, à des fins d'étude de l'enseignement/  
apprentissage

# LES ATELIERS-PHILO



QUELLES ACTIVITÉS COGNITIVES (INDIVIDUELLES ET COLLECTIVES)  
SONT À L'OEUVRE DANS LES ATELIERS-PHILO ?

# DÉFINITION D'UN ATELIER-PHILO

- ❖ « L'atelier philo est un espace de **parole collective** où peuvent s'exprimer les questions universelles des enfants et s'élaborer des **réflexions critiques** donnant du sens aux **contenus** scolaires. Rapport au monde, **multiplicité des points de vue**, réflexion sur les préjugés et l'opinion, l'atelier philo a pour but de favoriser l'émergence de sujets autonomes et responsables. »

source <http://www.citoyendedemain.net/pratiques/latelier-philo>

# LES TYPES DE RÉFLEXIONS DANS LES ATELIERS-PHILO

- ❖ Réflexion...
- ❖ à partir de *contenus scolaires* (nécessité de disposer de corpus de manuels scolaires)
- ❖ en *parole collective* (déterminer qui parle, pour dire quoi, par rapport à soi, par rapport aux autres ?)
- ❖ multiplicité des *points de vue* (essayer de déterminer des “voix” ?)

# LES ATELIERS-PHILO : QUELQUES QUESTIONS

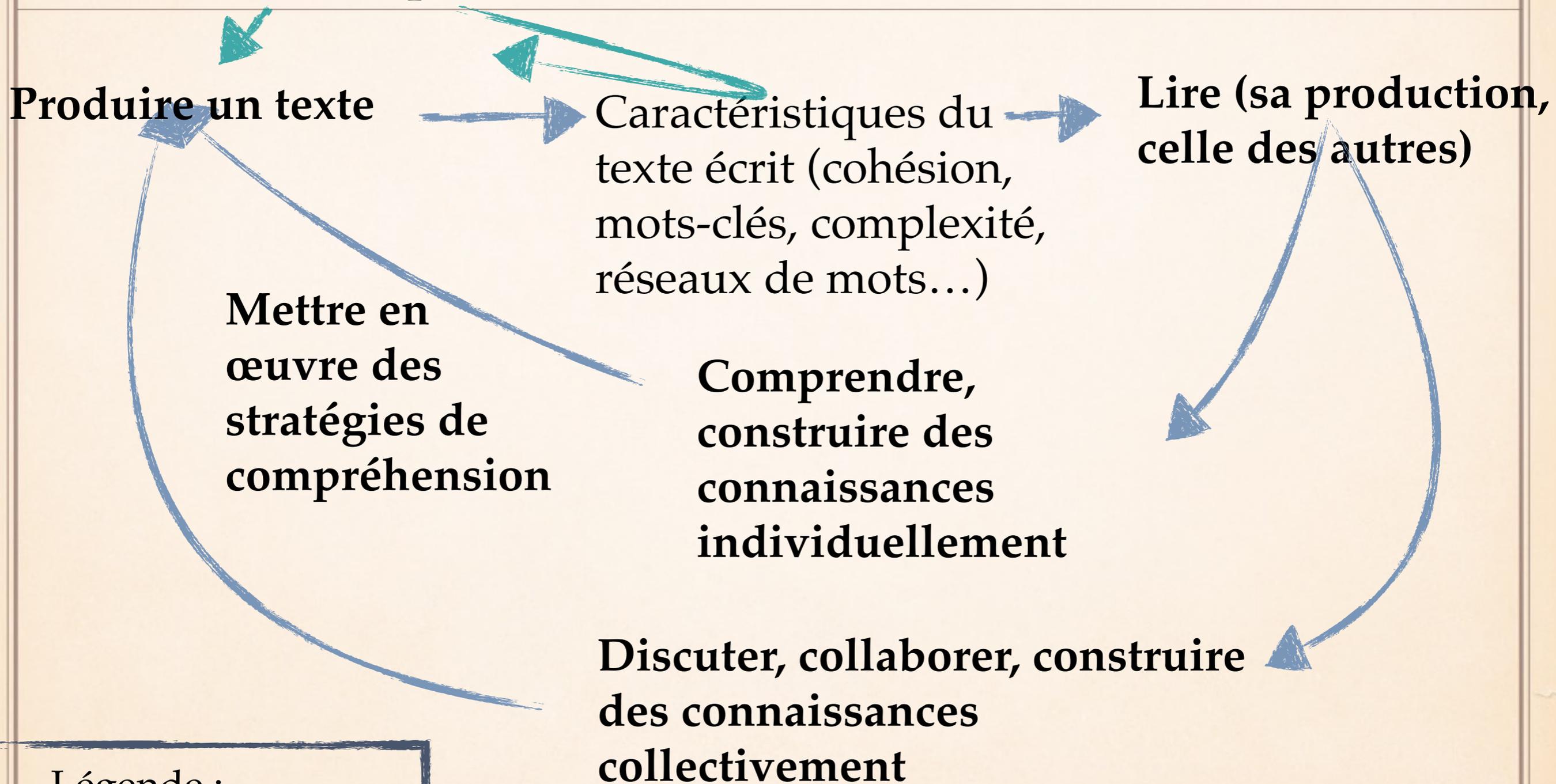
 {Auriac 07}

- ❖ L'enseignant et les élèves peuvent-ils rapidement (et de manière fiable) analyser
  - ❖ le contenu du texte travaillé (de quoi parle-t-il) ?
  - ❖ le niveau des textes proposés (correspond-il à celui des élèves) ?
  - ❖ les stratégies de compréhension mises en œuvre par les élèves quand ils en prennent connaissance ?
  - ❖ le contenu des points de vue exprimés dans la discussion ?
  - ❖ la qualité comparée des contributions de la discussion ?

argumenter raisonner

# LIENS ENTRE LES ACTIVITÉS

**Lire un texte philo**



Légende :  
**Actions/Paramètres**

# QUESTIONS DE RECHERCHE

- ❖ Quelles activités cognitives mises en œuvre dans des activités de lecture/écriture pour comprendre et/ou collaborer ?
  - ❖ centration sur la compréhension et la discussion
- ❖ Comment les simuler, les analyser ?
- ❖ Très coûteux de procéder manuellement. Intérêt de développer des systèmes automatisés pouvant aider
- ❖ Ne pas avoir seulement un système qui *fonctionne*, mais qui donne des résultats proches de ceux d'humains



# READERBENCH



UN ATELIER MULTILINGUE POUR ÉVALUER DIVERSES INFORMATIONS  
À PROPOS DE LECTURE ET COMPRÉHENSION (EN INPUT, EN OUTPUT),  
INDIVIDUELLEMENT ET COLLABORATIVEMENT

# LA COHÉSION COMME MESURE DE NOMBREUX PHÉNOMÈNES

- ❖ La cohésion (recouvrement sémantique de 2 propositions) peut être utilisée pour rendre compte des écarts entre mots, propositions, paragraphes, etc.
- ❖ Elle rentre donc très souvent dans le calcul de nombreux paramètres de *ReaderBench* : cohérence inter-phrases, inter-paragraphes, mais aussi de complexité
- ❖ De nombreux travaux montrent que des techniques d'analyse sémantique comme LSA capturent de manière assez satisfaisante la cohésion (Bellissens *et al.* 10)
- ❖ Validation : 10 histoires dont liens interparag. évalués par étudiants ; lien jugement humain vs. LSA/LDA/*WordNet* calculé :  $r = .54$

# HISTOIRE DE READERBENCH

- ❖ **2008-11** Début d'une collaboration avec l'univ. de Bucarest (WP commun du FP7 LTfLL) : *Pensum*/France (avec S. Mandin puis M. Loiseau) et *PolyCAFe*/Roumanie (équipe de S. Trausan-Matu)
- ❖ **2011-13** Thèse en co-tutelle (UPMF-Univ. Bucarest) de Mihai Dascalu (co-dir avec S. Trausan-Matu), partiellement intégrée au projet DEVCOMP (LSE, porté par M. Bianco, LSE). *ReaderBench*. But : analyse des stratégies en lecture, analyse de la complexité et raffinement de *PolyCAFe*
- ❖ **2013-pr.** Ajout/raffinement de modules, avec la participation d'autres étudiants en informatique roumains : Luisa Stavarache, Bogdan Oprescu, Ionut Cristian Paraschiv (dir. S. Trausan-Matu & M. Dascalu)
- ❖ **2015.** Séjour Fulbright de M. Dascalu à l'univ. d'Arizona, collaboration avec D. McNamara

## 4 GRANDS TYPES DE FONCTIONNALITÉS DE READERBENCH

1. Identification de mots-clés, réseau de mots, mots inférés
2. Analyse de la complexité textuelle
3. Identification de stratégies de compréhension en lecture
4. Analyse de la participation/collaboration dans les forums

# 1. IDENTIFICATION DE MOTS-CLÉS

 [Dascalu 13]

Cognitis enim pilatorum caesorumque funeribus nemo deinde ad has stationes appulit navem, sed ut Scironis praerupta letalia declinantes litoribus Cypriis contigui navigabant, quae Isauriae scopulis sunt controversa.

Oportunum est, ut arbitror, explanare nunc causam, quae ad exitium praecipitem Aginatium inpulit iam inde a priscis maioribus nobilem, ut locuta est pertinacior fama. nec enim super hoc ulla documentorum rata est fides.

Paphius quin et Cornelius senatores, ambo venenorum artibus pravis se polluisse confessi, eodem pronuntiante Maximino sunt interfecti. pari sorte etiam procurator monetae extinctus est. *Sericum enim et Asbolium supra dictos, quoniam cum hortaretur passim nominare, quos vellent, adiecta religione firmarat, nullum igni vel ferro se puniri iussurum, plumbi validis ictibus interemit. et post hoc flammis Campensem aruspicem dedit, in negotio eius nullo sacramento constrictus.*

Et olim licet otiosae sint tribus pacataeque centuriae et nulla suffragiorum certamina set Pompiliani redierit securitas temporis, per omnes tamen quotquot sunt partes terrarum, ut domina suscipitur et regina et ubique patrum reverenda cum auctoritate canities populi que Romani nomen circumspectum et verecundum.

1. Classe gramm. du mot
2. Score d'un mot =  $tf^*idf$  (fréq. locale pondérée par fréq. inv. globale) + similarité mot avec **texte lu** + similarité moyenne mot avec chaque **phrase** + similarité moyenne mot avec chaque **paragraphe**
3. Présenter les mots par score décroissant et par classe (option)

# 1. COPIE D'ÉCRAN ANALYSE DU TEXTE

**ReaderBench - Document Visualization**

**Title:** Subject departments as professional communities?  
**Source:** LSE   **URI:**   **Sentiment polarity:**

**Contents**

a growing body of literature suggests that when schools become professional communities there are expected benefits in terms of teacher learning, school improvement and student achievement. in this article the concept of professional communities is examined for certain subject departments in dutch secondary schools. the authors report on research into the extent to which mathematics departments operate as professional communities. at the same time, it was investigated whether the level at which departments operate as professional communities is related to student achievement. the results indicate that departments are cohesive bodies regulating teacher behaviour in several respects. however, as professional communities they do not focus on improving the quality of their teachers and instruction. some characteristics of professional communities prove to be beneficial for student achievement, while others are not. the authors offer recommendations on how departments can develop into more professionally organised communities. [31.949]

**Topics**

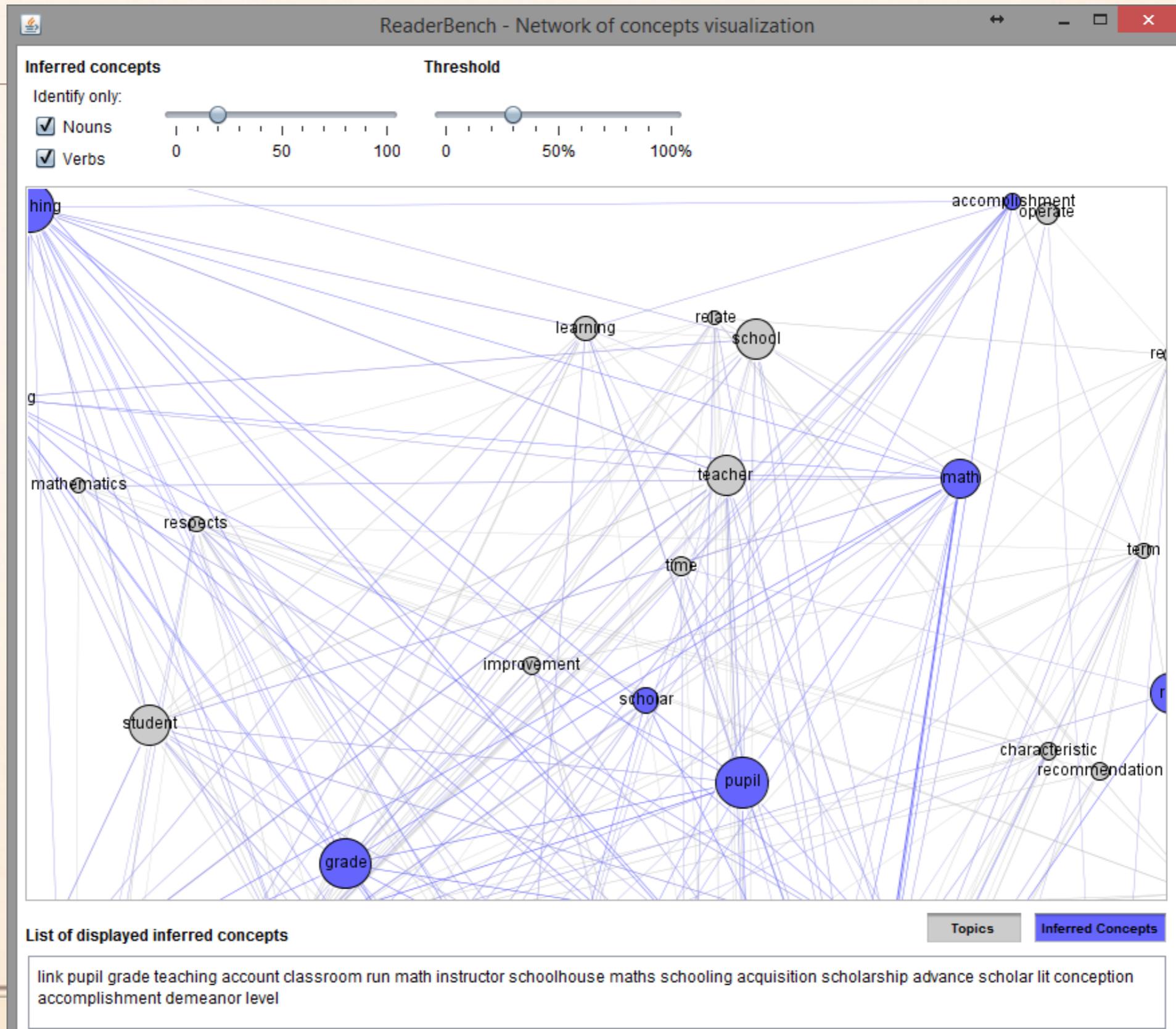
Filter only:  Nouns  Verbs

0 25 50

Topics	Relevance
community	7.05
student	3.9
school	3.86
teacher	3.84
department	3.52
achievement	2.74
body	2.35
learning	1.73
author	1.69
operate	1.43
literature	1.4
time	1.23
recommendation	1.11
improvement	1.08
instruction	1.05
characteristic	1.03
extent	0.96
respects	0.95

Advanced View   Visualize Multi-Layered Cohesion Graph   Select Voices   Display Voice Inter-animation   Generate network of concepts

# 1. COPIE D'ÉCRAN 2 - RÉSEAU DE MOTS



## 2. COMPLEXITÉ TEXTUELLE

 [Dascalu *et al.* 12]

- ❖ Partiellement en français pour l'instant (école élémentaire, nécessité de disposer d'une grande base de textes pré-calibrés pour entraîner *ReaderBench*)
- ❖ But : Essayer de quantifier, sur un nombre de facteurs important, des indices de complexité de textes donnés à lire. Pouvoir aisément ajouter de nouveaux facteurs ; multilingue (fr, anglais, italien, bientôt espagnol)
- ❖ Nécessité de combiner plusieurs niveaux d'évaluation (lexical, morphosyntaxique, sémantique), car chaque niveau en lui-même n'est pas suffisamment relié à la complexité [Bailin & Grafstein 01]

## 2. QUELQUES NIVEAUX DE COMPLEXITÉ

 [Dascalu *et al.* 12; Dascalu *et al.* 15]

### ❖ **De surface**

- ❖ Mesure de lisibilité
- ❖ Entropie (diversité)
- ❖ Complexité des mots (longueur moyenne des mots, etc.)

### ❖ **Morphologie et syntaxe**

- ❖ Mesure intégrée de complexité lexicale et syntaxique (Schultze 2010)
- ❖ Profondeur et taille de l'arbre d'analyse (*parsing tree*)

### ❖ **Sémantique**

- ❖ Cohésion (proximité sémantique interphrases) avec LSA ou LDA
  - ❖ Chaînes lexicales (mots sémantiquement reliés), montrent l'importance d'un concept
- ❖ La technique de classification *Support Vector Machine* permet de classifier des documents après «entraînement» avec un grand nombre de doc. pré-classifiés

## 2. LIEN ENTRE COMPLEXITÉ, STRATÉGIES ET COMPRÉHENSION

 [Dascalu *et al.* 15]

- ❖ Transcription et analyse de 149 résumés de compréhension de 2 textes de Cycle 3. Mesure de la fluence orale des élèves et de leur compréhension effective du texte lu (questions littérales, de vocabulaire, d'inférence, sur le thème de l'histoire)
- ❖ Analyse automatique de la complexité de ces résumés, et des stratégies par *ReaderBench*
- ❖ SVM pour créer 3 classes de niveau de compréhension, pour essayer de prédire la compréhension finale du texte lu

## 2. COMPLEXITÉ, STRATÉGIES ET COMPRÉHENSION : RÉSULTATS

 [Dascalu *et al.* 15a]

	Facteurs / Classes de compréhension	C1	C2	C3	Moy.
1	Toutes les stratégies	.60	.45	.61	.55
2	Toutes les stratégies & Cohésion avec le document lu	.57	.45	.55	.52
3	Tous les indices de complexité du résumé	.17	.14	.94	.42
4	Facteurs 2 & 3 (stratégies / cohésion / complexité)	.52	.49	.53	.51
5.	2 & fluence orale	.67	.53	.78	.66
6	5 & complexité morphologique et sémantique	.74	.55	.78	.68

### 3. AUTO-EXPLICATION DE STRATÉGIES

 [Dascalu *et al.* 13; Bianco *et al.* 13]

- ❖ Le soleil brillait de toutes ses forces et les dieux qui vivaient sur la Terre trouvaient qu'il faisait trop chaud. Les dieux racontèrent à Indra qu'ils avaient vu le grand serpent Ati étirer sa tête jusque dans le ciel pour avaler les nuages qui passaient. Ati avait si soif qu'il buvait toute l'eau des nuages : voilà pourquoi il ne tombait plus une goutte de pluie.



"Qu'as-tu compris de ce passage ?"

- ❖ Indra ne trouva pas cela drôle du tout. À cause de ce glouton, la Terre entière mourait de soif !

Etc...

# 3. EXEMPLES DE STRATÉGIES

 [Nardy *et al.* sous presse]

## Paraphrases

[P1, CE2] *Ben en fait il y avait des dieux et ben ils trouvaient qu'il faisait trop chaud*

## Inférences textuelles

[P1, CM2] *et y a un serpent qui a Ati qui qui mange enfin qui boit l'eau des nuages et c'est pour ça qu'il fait aussi chaud*

## Inférences de connaissances

[P2, CM1] *il veut délivrer un peu l'eau donc du coup il veut combattre Ati le serpent*

[P1, CM2] *Alors c'est dans le ciel y a des Dieux et Ati c'est un grand serpent qui va prendre l'eau dans les nuages*

## Auto-évaluation (contrôle)

[P3, CM2] *Bon là déjà on sait que Indra c'est un garçon pas une fille*

[P5, CE2] *Et j'pense que après le serpent il était mort...c'est tout.*

### 3. HEURISTIQUES D'IDENTIFICATION DES STRATÉGIES

- ❖ Heuristiques d'identification des stratégies
  - ❖ **Causalité** : mise en évidence de mots-clés («parce que», «pour», «donc», «à cause de», etc.)
  - ❖ **Contrôle** : idem («je me souviens», «je crois», «ils racontent», etc.)
  - ❖ **Paraphrase de mots** : Similarité sémantique + similarité des lemmes via *WordNet*
  - ❖ **Inférence de connaissances** : Mot non paraphrasé mais : proche d'un mot du texte, ces deux étant proches du texte lu.
  - ❖ **Reprise (*bridging*)** : Segment de phrase du texte évoqué de manière non littérale (établir une relation avec un passage précédent, qui a été repris)

# 3. COPIE D'ÉCRAN - ANALYSE DE STRATÉGIES

Document title: Matilda [config/LSA/lemonde\_fr, config/LDA/lemonde\_fr]

[View document](#)

Verbalization: [REDACTED] Alice.xml)

## Contents

Text	Causality	Control	Paraphr...	Knowle...	Bridging	Cohesion
la mère[8] devint toute blanche . elle dit[5] à son mari il y a quelqu' un dans la maison[2] . ils arrêtèrent[9] tous de manger[10] . ils étaient tous sur le qui - vive . la voix[7] reprit[11] salut[6] , salut[6] , salut[6] . le frère[12] se mit à crier ça recommence[13] ! matilda se leva et alla éteindre la télévision[3] .						0.315
Je ai compris[4] que c' est une famille[2] la famille[2] dans laquelle il ? suis qui dinent[1] devant la télé[3] . et qui . tout de un coup il z entendent[4] une voix[7] qui leur dit[5] salut[6] . et du coup ils ont peur donc parce que la mère[8] de matilda ? donc c' est que je pense que ils ont peur . alors ils arrêtent[9] de manger[10] . puis le frère[12] commence à comprendre quelque cho quelque chose en disant ça recommence[13]	5	1	13	0	1	
la mère , paniquée , dit à son mari : henri , des voleurs[15] , ils sont dans le salon , tu devrais[14] y aller . le père , raide sur sa chaise ne bougea pas . il n' avait pas envie de jouer au héros . sa femme lui dit : alors , tu te décides ? ils doivent[14] être en train de faucher l' argenterie[16] !						0.294
alors je pense que c' est une famille[*] peut - être assez riche parce que il y a de l' argenterie[16] . et qui pensent que ceux qui doit[14] être riche ou que y a beaucoup de voleurs[15] dans notre dans leur maison donc	2	1	3	1	1	
monsieur verdebois s' essuya nerveusement les lèvres avec sa serviette et proposa d' aller[17] voir[18] tous ensemble . la mère attrapa un tisonnier au coin de la cheminée . le père[19] s' arma d' une canne de golf posée dans un coin . le frère attrapa un tabouret . matilda prit[9] le couteau avec lequel elle mangeait . puis ils se dirigèrent tous les quatre vers la porte du salon en marchant sur la pointe des pieds .						0.399
à ce moment - là , ils entendirent à nouveau la voix . matilda fit alors irruption dans la pièce en brandissant son couteau et cria haut[20] les mains[21] , vous êtes pris[9] ! les autres la suivirent en agitant leurs armes .						0.189
donc la c' est on sait déjà comment s' appelle la famille . et puis ils racontent que là vu que le père[19] veut pas y aller[17] tout seul . il est accompagné de toute sa famille pour aller[17] voir s' y a un voleur . et y a la le la parole[*] ça le bruit aussi ? qui recommence . et du coup elle , la petite fille[*] qui s' appelle matilda commence à avoir peur . donc elle lui dit haut[20] les mains[21] vous êtes pris[9]	4	2	5	2	1	

### 3. ETUDE SUR 19 ÉLÈVES DE CYCLE 3

 [Bianco *et al.* 13]

- ❖ Les verbalisations ont été analysées «à la main» par des juges humains, puis, indépendamment, par *ReaderBench*
  
- ❖ Corrélation par stratégie :
  - ❖ Contrôle : .83
  - ❖ Causalité : .77
  - ❖ Paraphrase : .73
  - ❖ Inférence de connaissance : .60
  
- ❖ Une stratégie plus complexe à déterminer automatiquement, le *bridging* (lien entre plusieurs propositions) n'est pas détectée

## 4. PARTICIPATION ET COLLABORATION DANS LES FORUMS DE DISCUSSION

 [Dascalu et al. 15b, Dascalu et al., soumis]

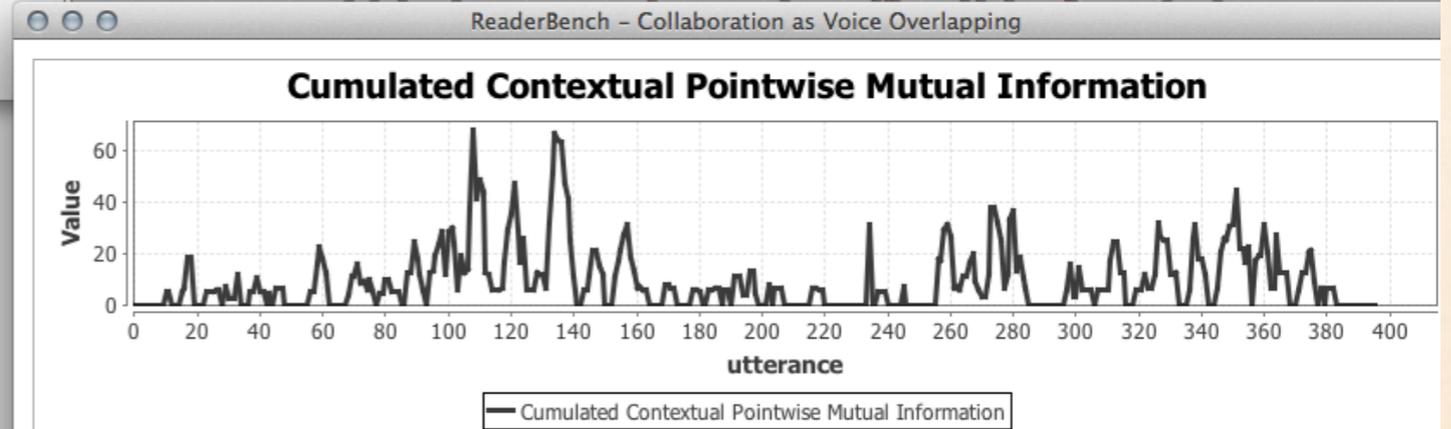
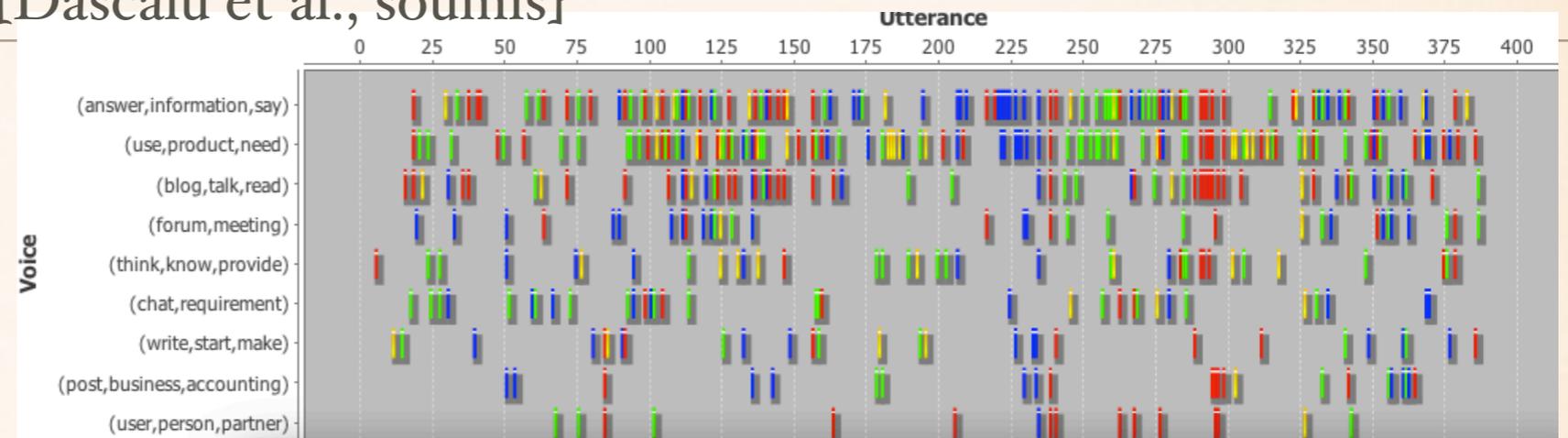
- ❖ Des participants à un débat (*via* forum, chat, ou *de visu*...)
- ❖ Analyse de la multivocalité (de multiples voix, points de vue sont exprimés) et de la polyphonie (chaque voix interagit avec les autres, interanimation) [Bakhtine 81]...
- ❖ ... par la cohésion des tours de parole entre eux (liens lexicaux, syntaxiques, sémantiques)
- ❖ Identification de la *participation* (nombre de tours de parole d'un participant) et de la collaboration (tours de parole en réponse à d'autres)

## 4. PARTICIPATION ET COLLABORATION DANS LES FORUMS DE DISCUSSION

- ❖ Définir le score d'un tour de parole en fonction de 3 paramètres liés aux mots le composant : leur fréquence dans le texte, leur similarité sémantique par rapport au tour de parole, à la discussion, la « couverture sémantique » représentée par la chaîne sémantique du mot tout au long de la discussion (extension sémantique de la chaîne lexicale de [Galley & McKeown 03])
- ❖ *Pointwise Mutual Information* (mesure du recouvrement de 2 variables) pour déterminer les pics de synergie entre voix (interanimation)
- ❖ Analyse de la construction de connaissances
  - ❖ individuelle : cohésion des voix d'*un même* participant
  - ❖ collaborative : cohésion des voix *interparticipants*

# 4. PARTICIPATION ET COLLABORATION DANS LES FORUMS DE DISCUSSION/CHATS

 [Dascalu et al., soumis]



## Automatically identified intense collaboration zones

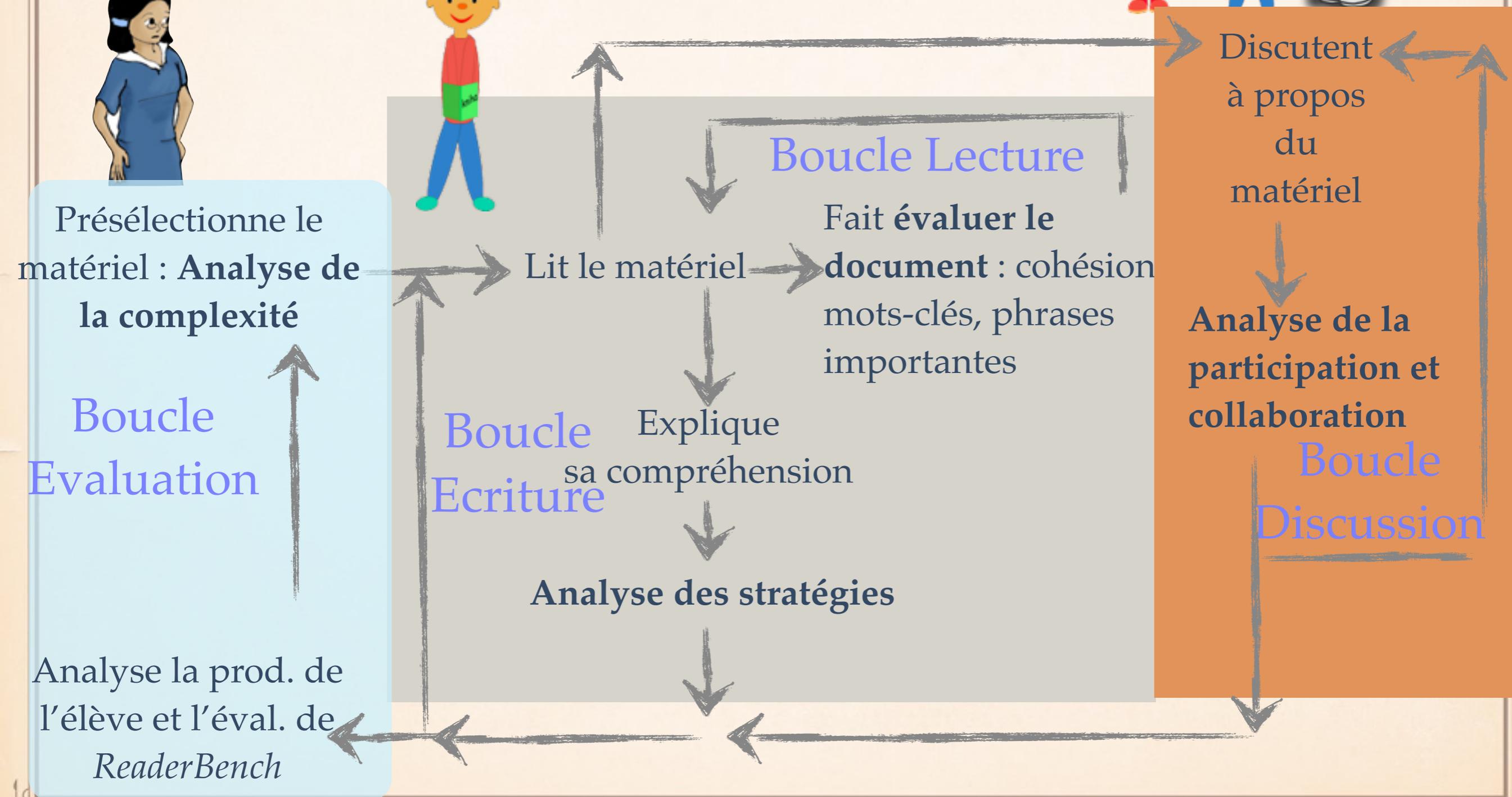
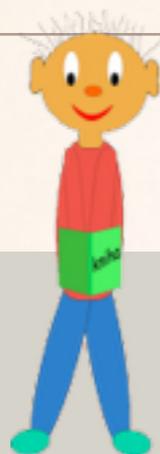
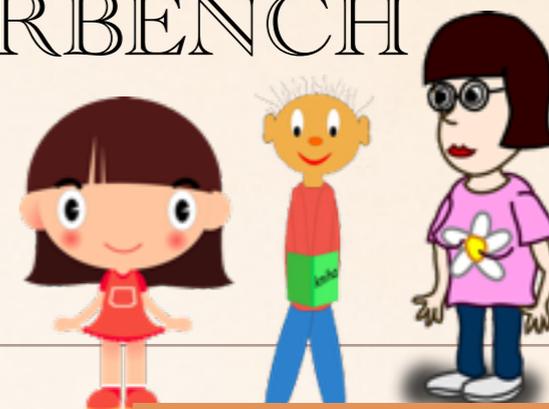
- [256; 283] - 28 utterances with 17.623 average social KB (493.437 overall social KB)
- [58; 159] - 102 utterances with 16.597 average social KB (1692.883 overall social KB)
- [311; 375] - 65 utterances with 14.152 average social KB (919.878 overall social KB)
- [191; 197] - 7 utterances with 9.712 average social KB (67.982 overall social KB)

TRAVAIL EN COURS  
ET À FAIRE



# UN FLUX DE TRAVAIL AVEC READERBENCH

 [D'après Dascalu 13]



# RECHERCHES EN COURS/À VENIR

## Réseaux

Analyse sémantique des résumés d'articles en éducation pour déterminer leur proximité et des tendances scientométriques. Collaboration avec ENS-Lyon (EducMap)

## Complexité

Lien entre la complexité des explications et d'autres variables socio-linguistiques (liaison facultative, réalisation du / dans il(s)). [Nardy *et al.* 15]

## Discussion

Analyse des contributions des élèves en séance de philosophie (cycle 3). Coll. avec J.-P. Simon, LIDILEM

## Réseaux

Analyse de la cohésion de verbalisation comme indice de l'expertise

## Complexité

Analyse des tours de parole des enseignants en situation d'enseignement. Complexité comme indice de leur charge cognitive

## Validation *in situ*

Scénarisation et test de *ReaderBench* en situation scolaire

## Stratégies

Analyser les auto-explications des élèves suite à la lecture de cours ou de MOOC (Ecrire-Pour-Apprendre)

## Complexité

Ajout d'un module d'analyse de sentiments, de calcul de complexité évoluées (*word maturity*)

En cours

En projet

Plus tard

DÉMO DE  
READERBENCH

# MERCI DE VOTRE ATTENTION ! DES QUESTIONS ?

- ❖ Références à <http://bit.ly/atelier-philo>
- ❖ Ces travaux ont été réalisés avec la participation de : Maryse Bianco, Mihai Dascalu, Danielle McNamara, Aurélie Nardy, Bodgdan Oprescu, I.C. Paraschiv, Luisa Stavarache, & Stefan Trausan-Matu
- ❖ et avec des financements des projets EC-FP7 LTfLL & ERRIC, et de l'ANR DEVCOMP
- ❖ Diapos disponibles à <http://webcom.upmf-grenoble.fr/sciedu/pdessus/>

# ANNEXES

# DIAGNOSTIC DE L'AUTO-EXPLICATION DE STRATÉGIES DE LECTURE (MCNAMARA 2004)

- ❖ Une phrase verbalisée... (Phc = Phrase cible du document lu)
  - ❖ ... est la **paraphrase** de Phc si plusieurs mots de la phrase sont identiques (ou synonymes)
  - ❖ ... utilise une **connaissance inférée**
  - ❖ ... utilise un **bridging avec** Phc : si  $\text{prox}(\text{Ph}, \text{Phc}) > \text{seuil}_b$  ET  $\text{prox}(\text{Ph}, \text{Ph}_k) > \text{seuil}_b$
  - ❖ utilise un **contrôle** : si des *patterns* comme «j'ai compris que», «je pense que...» apparaissent dans la phrase.

## II. LSA ET L'INDUCTION (LANDAUER & DUMAIS 97)

Titres d'articles de 2 ensembles disjoints c : informatique ; m : maths

c1: *Human machine interface for ABC computer applications*

c2: *A survey of user opinion of computer system response time*

c3: *The EPS user interface management system*

c4: *System and human system engineering testing of EPS*

c5: *Relation of user perceived response time to error measurement*

m1: *The generation of random, binary, ordered trees*

m2: *The intersection graph of paths in trees*

m3: *Graph minors IV: Widths of trees and well-quasi-ordering*

m4: *Graph minors: A survey*

# II. LE FONCTIONNEMENT DE LSA

(LANDAUER & DUMAIS 97)

	c1	c2	c3	c4	c5	m1	m2	m3	m4
human	1	0	0	1	0	0	0	0	0
interface	1	0	1	0	0	0	0	0	0
computer	1	1	0	0	0	0	0	0	0
user	0	1	1	0	1	0	0	0	0
system	0	1	1	2	0	0	0	0	0
response	0	1	0	0	1	0	0	0	0
time	0	1	0	0	1	0	0	0	0
EPS	0	0	1	1	0	0	0	0	0
survey	0	1	0	0	0	0	0	0	1
trees	0	0	0	0	0	1	1	1	0
graph	0	0	0	0	0	0	1	1	1
minors	0	0	0	0	0	0	0	1	1

$\cos(\text{human.user}) = -.38$  &  $\cos(\text{human.minors}) = -.29$

# APRÈS DÉCOMPOSITION

	c1	c2	c3	c4	c5	m1	m2	m3	m4
<b>human</b>	0.16	0.40	0.38	0.47	0.18	-0.05	-0.12	-0.16	-0.09
<b>interface</b>	0.14	0.37	0.33	0.40	0.16	-0.03	-0.07	-0.10	-0.04
<b>computer</b>	0.15	0.51	0.36	0.41	0.24	0.02	0.06	0.09	0.12
<b>user</b>	0.26	0.84	0.61	0.70	0.39	0.03	0.08	0.12	0.19
<b>system</b>	0.45	1.23	1.05	1.27	0.56	-0.07	-0.15	-0.21	-0.05
<b>response</b>	0.16	0.58	0.38	0.42	0.28	0.06	0.13	0.19	0.22
<b>time</b>	0.16	0.58	0.38	0.42	0.28	0.06	0.13	0.19	0.22
<b>EPS</b>	0.22	0.55	0.51	0.63	0.24	-0.07	-0.14	-0.20	-0.11
<b>survey</b>	0.10	0.53	0.23	0.21	0.27	0.14	0.31	0.44	0.42
<b>trees</b>	-0.06	0.23	-0.14	-0.27	0.14	0.24	0.55	0.77	0.66
<b>graph</b>	-0.06	0.34	-0.15	-0.30	0.20	0.31	0.69	0.98	0.85
<b>minors</b>	-0.04	0.25	-0.10	-0.21	0.15	0.22	0.50	0.71	0.62

$\cos(\text{human.user}) = .94$  &  $\cos(\text{human.minors}) = -.83$

