

# **Où va la recherche en éducation ? Analyse factorielle de résumés de communications**

**Ph. Dessus • DEACT, UPMF & IUFM Grenoble**

*Séminaire R2I • 6 juillet 98*

## **I. Travail sur le corpus des Biennales**

- **Problématique**
- **Études sur l'état de la recherche en éducation**
- **Hypothèses & méthode**
- **Résultats & discussion**

## **II. Présentation de LSA**

- **Fonctionnement, un exemple**
- **Études évaluatives**
- **Projets de travail**

# I.1. Problématique

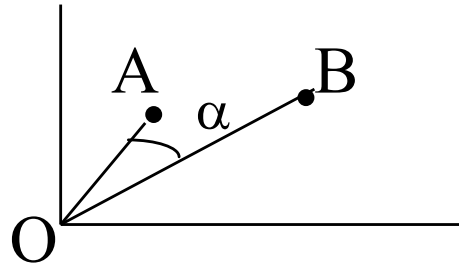
- **Peu d'études sur les perspectives de la recherche en éducation basées sur un matériau représentatif.**
- **Le corpus des résumés de communications (1500 résumés, 400 000 mots) aux différentes Biennales de l'éducation et de la formation (92-98) paraît l'être :**
  - **représente de nombreux courants de recherche ;**
  - **communications validées par un comité de lecture ;**
  - **longueur des résumés (260 mots en moyenne).**
- **L'évolution des préoccupations des chercheurs pourra être analysée à partir du lexique utilisé.**

## I.2. Études sur l'état de la recherche en éducation

- **Mialaret 97 : 400 questionnaires à des chercheurs en éducation francophones sur objet, méthode et public.**
  - Classement, sans synthèse. Validation par directeur ? par publications ? Formulation lapidaire des thèmes.
- **Beillerot & Demori 98 : Analyse lexicométrique des titres des 400 thèses en sci. éduc. (90-94).**
  - Occurrences max. : *scolaire, social, pédagogie, éducation* dans primaire et professionnel.
  - Titres courts, donc généraux, formes lexicales ne tiennent pas compte des différences sing./plur., champ sci. éduc. trop restreint.

## I.3. Méthode d'analyse statistique

- **LSA (*latent semantic analysis*)**, met en œuvre un type d'analyse factorielle sur de larges corpus de textes. Il permet de rendre compte du sens des mots d'un texte à partir de leur utilisation.
  - la proximité sémantique de deux termes : ils seront d'autant plus proches qu'ils auront été souvent relevés ensemble • dans des textes différents • dans le même texte • ou comme voisins des mêmes termes.
  - la centralité sémantique d'un terme : son degré de typicalité au sein d'un ensemble de textes.



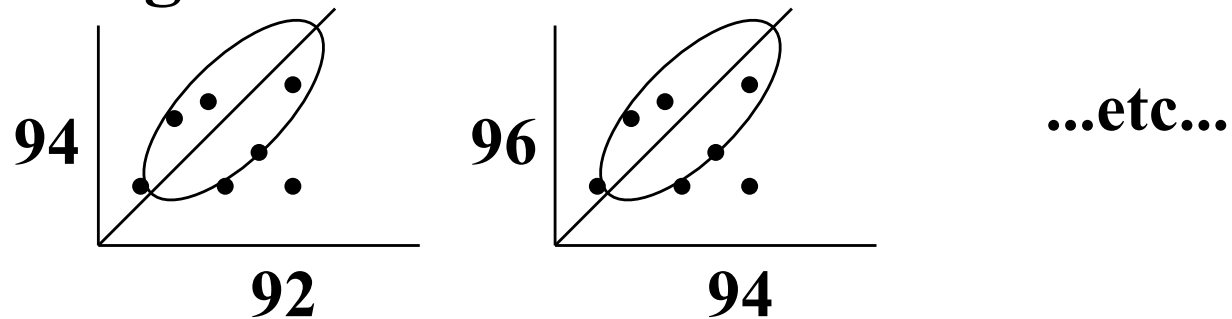
Proximité de A et B :  $\cos \alpha$ .  
Centralité de A :  $\|OA\|$

## I.4.a Hypothèses

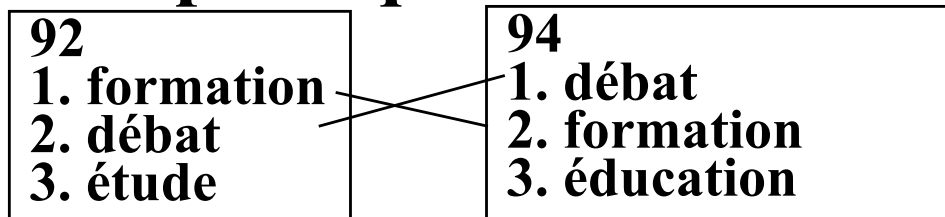
- ① **Pas d'évolution globale dans le rang de centralité des termes les plus centraux (typiques) interbiennales. L'utilisation des termes les plus centraux ne change pas interbiennale.**
- ② **L'évolution spécifique de centralité de certains termes précédents est fluctuante, ceux « sujets à débat ».**
- ③ **Un réseau sémantique représentant l'évolution des plus fortes proximités des termes « sujets à débat » mettra au jour des thèmes de recherche indépendants et datés.**

## I.4.b. Hypothèses, vue graphique

① Liaison globale entre les centralités des termes.

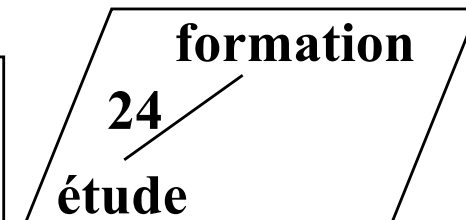


② Évolution spécifique de centralité des termes.



③ Réseau de proximité des termes.

92 <i>formation</i> proche de <i>étude, niveau</i>	94 <i>formation</i> proche de <i>étude, formateur</i>
--	---



## **I.5.a. Résultats : corrélations de centralités globales interbiennale**

- **Très forte corrélation interbiennale de la centralité des 100 termes les plus centraux. En supprimant les 15 termes les plus centraux, on remarque une absence de liaison entre les éditions 94 et 96 : « rupture » dans la centralité.**
- **Corrélations (\*  $p < .01$ ) :**
  - 92-94,  $r = .79^*$**
  - 94-96,  $r = .11$**
  - 96-98,  $r = .43^*$**

## I.5.b. Résultats : évolution spécifique de la centralité de certains termes

Termes	92-94	94-96	96-98
Savoir	↗	↘	↗
Compétences	↘	↗	↘
Formateurs	↘	↗	↘
Recherches	↘	↗	↘
Temps	↘	↗	↗
Entretiens	↗		↗
Sens		↘	↗
Représentations		↘	↗
Sociale		↘	↗

**Très  
fluctuants**

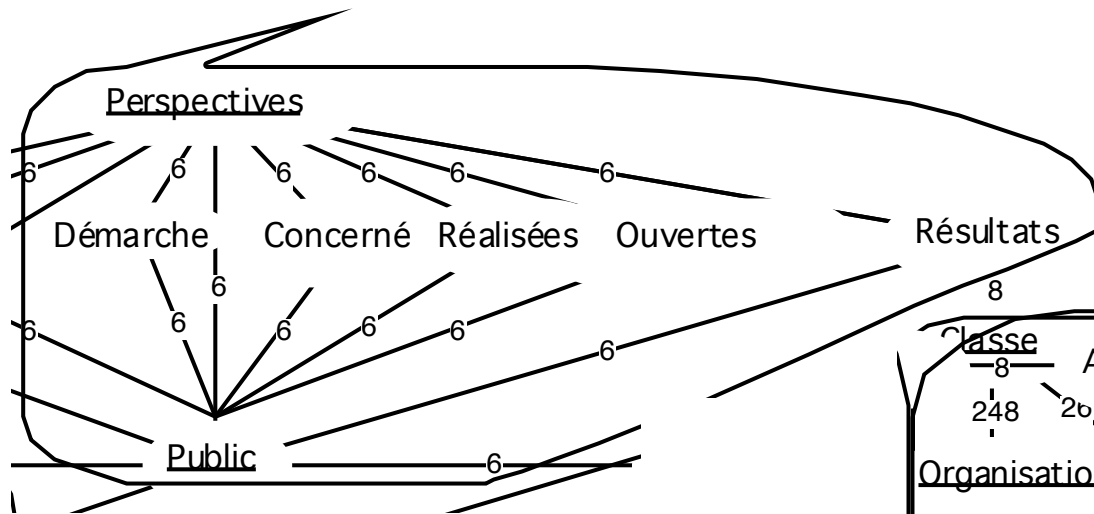
**Fluctuants  
en hausse  
92-98**



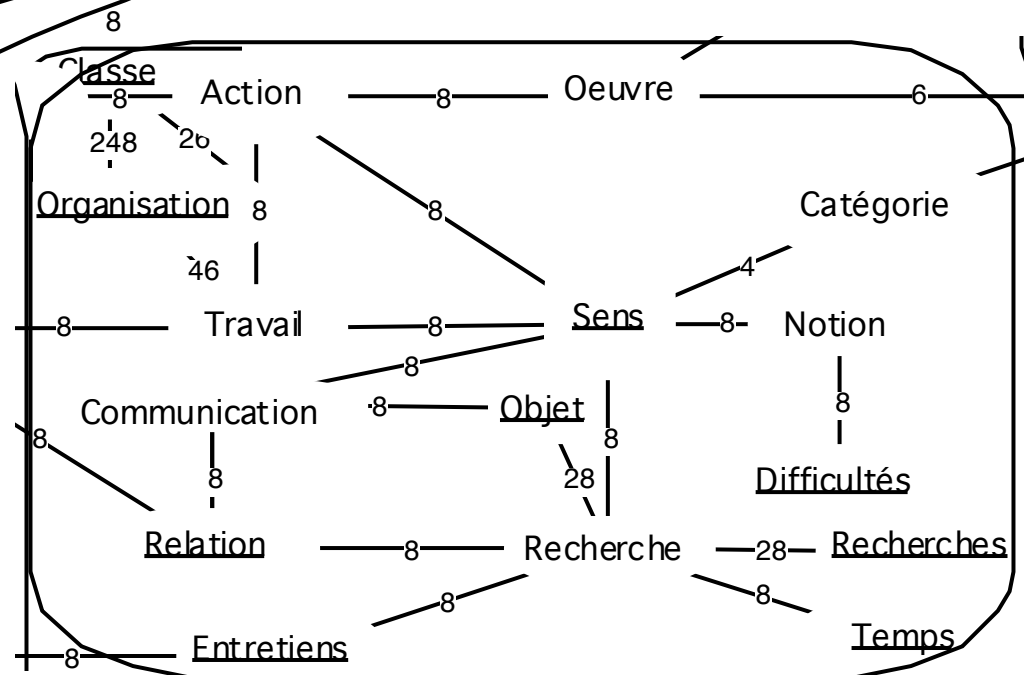
## **I.5.c. Résultats : commentaires sur l'évolution de centralité**

- **Cause de la rupture précédente : il y a un nombre plus grand de termes variant de centralité entre 1994 et 1996.**
- **Très faible nombre de termes très fluctuants, tous en lien avec la recherche en éducation.**
- **Prise en compte croissante de la dimension temporelle et du sens.**
- **Hypothèse 2 non vérifiée : les termes à propos de la recherche varient également de centralité (recherches, entretiens...).**

# I.5.d. Résultats : carte de concepts (extraits)



**Recherche 96**



**Recherche 98**

## **I.5.e. Résultats : commentaires sur la carte de concepts**

- **Sept zones (*voir poly.*) indépendantes sont mises au jour. Hypothèse validée.**
- **Réseau large sur la didactique et daté 92-98, importance de ces apports.**
- **Réseau sur le système éducatif restreint, mais composé de nombreux termes sujets à débat, ce qui montre un débat continu sur ces termes.**
- **Réseau sur la pédagogie très restreint.**

## I.6. Discussion

- **Méthode systématique, répliquable sur d'autres contenus (colloques, contenus éditoriaux).**
- **Caractéristiques des résumés de communications**
  - rupture 94-96 dans la centralité des termes ;
  - des termes exclusivement du domaine variant régulièrement de centralité ;
  - deux pôles fortement liés sémantiquement et datés : la méthodologie de la recherche 96 et 98.

## II.1.1. Présentation du fonctionnement de LSA : un exemple, les textes de départ

**9 phrases tirées de titres de mémos sur l'interaction homme-ordinateur (c), sur les graphes (m).**

**c1 : *Human machine interface for Lab 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, unordered 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.1.2. Présentation de LSA : un exemple, générer une matrice

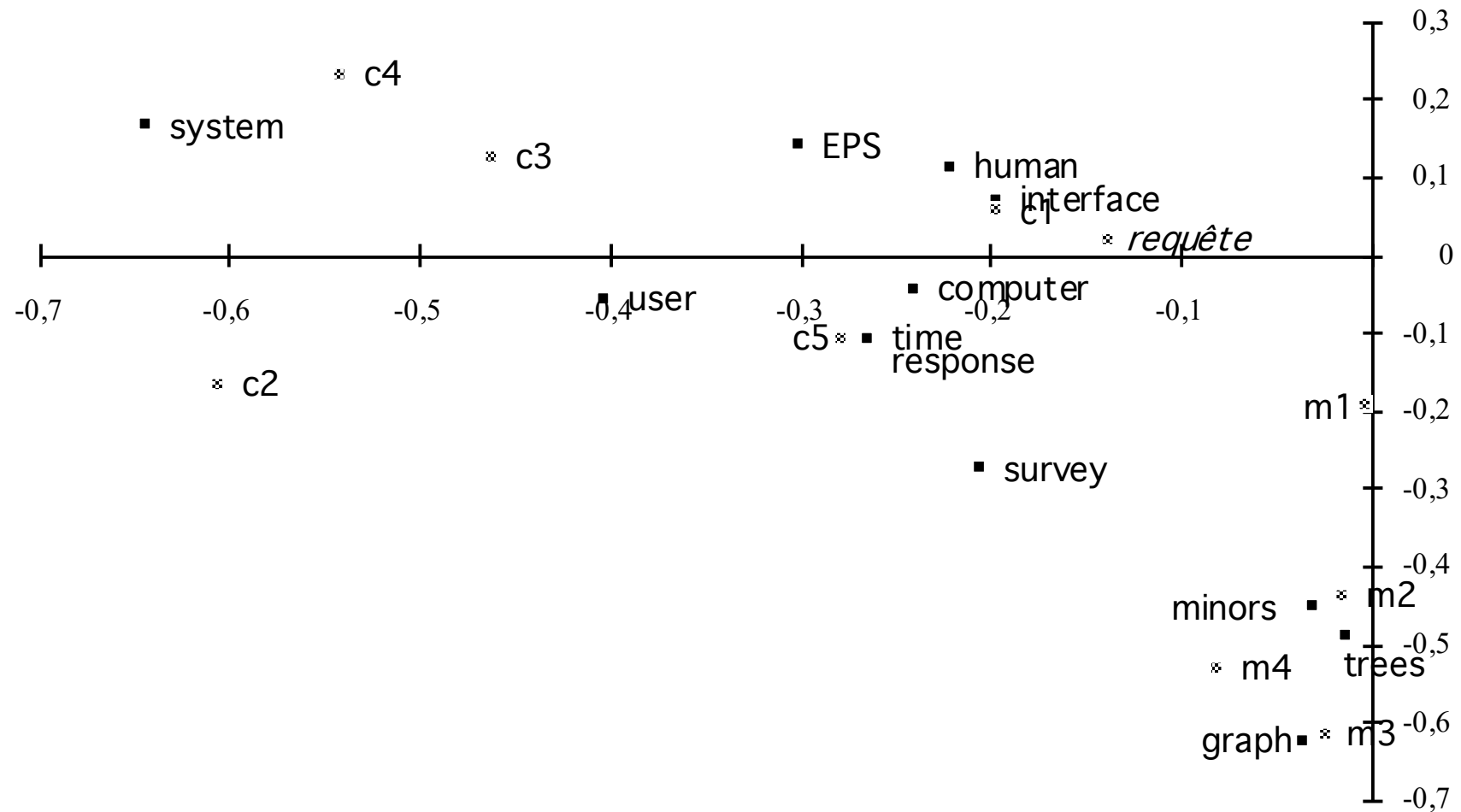
- LSA génère une matrice configurée avec les occurrences du terme  $i$  dans le contexte  $j$ .

<i>i/j</i>	<b>c1</b>	<b>c2</b>	<b>c3</b>	<b>c4</b>	<b>c5</b>	<b>m1</b>	...
<b>human</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	...
<b>interface</b>	<b>1</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	...
<b>computer</b>	<b>1</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	...
<b>user</b>	<b>0</b>	<b>1</b>	<b>1</b>	<b>0</b>	<b>1</b>	<b>0</b>	...
<b>system</b>	<b>0</b>	<b>1</b>	<b>1</b>	<b>2</b>	<b>0</b>	<b>0</b>	...
<b>response</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>0</b>	...
...							

## **II.1.3. Présentation de LSA : traitement de la matrice**

- **Chaque terme et document est représenté par un vecteur dans un espace à 9 dimensions. LSA réalise une analyse factorielle de cette matrice.**
- **LSA opère ensuite une approximation de la matrice, en ne conservant que les 100 dimensions les plus significatives. C'est ce filtrage qui induit des relations sémantiques.**
- **À partir de cette matrice réduite, on calcule la centralité (importance) de termes et la proximité entre 2 termes.**

## II.1.4. Présentation de LSA : une réduction à deux dimensions





## **II.2.1. Quelques études sur LSA : rechercher des informations**

- **LSA conçu pour rechercher des informations dans de larges bases de données documentaires.**
- **Une recherche informatisée par mots-clés ne récupère que 20 % de l'information pertinente ; de plus, 50 % de l'information récupérée ne l'est pas.**
- **Dumais 91 teste LSA avec de larges bases de données (4 000 documents, 25 000 termes). Les documents récupérés avec LSA sont 20 % plus pertinents que ceux avec une requête par mots-clés.**

## **II.2.2. Quelques études sur LSA : acquérir des connaissances**

- **Landauer & Dumais 97 « entraînent » LSA avec une encyclopédie électronique (30 000 articles).**
- **Il testent la capacité de LSA d'apparier un mot avec un synonyme (test du TOEFL).**
- **Le score obtenu par LSA est semblable au niveau d'entrée de candidats non anglophones à une université américaine.**

## **II.2.3. Quelques études sur LSA : évaluer la qualité d'un écrit**

- **Landauer *et al.* 97 ont demandé à des étudiants d'écrire un court texte sur le cœur humain, texte noté indépendamment par deux experts.**
- **Ils ont « entraîné » LSA avec une trentaine d'articles sur le sujet, puis calculé la proximité de chaque texte d'étudiant avec l'ensemble des articles.**
- **La corrélation entre la note des experts et celle de LSA est significative.**

## **II.3. Projets de travail en cours**

- **Analyse des questions posées par les jurys du concours PE lors de l'entretien professionnel 97.**
  - termes les plus centraux : école, apprentissage, élève(s), place, classe, apprentissages, enfant, rôle.
  - proximités maximales : maître-autorité, évaluation-légitime, lecture-méthode, lecture-plaisir, enfants-conseils, enfants-service.
- **Acquisition de connaissances sur la dyslexie à partir d'un texte : LSA vs étudiants (avec Bianco & Lemaire).**
  - Lecture d'un texte, puis listage de mots voisins de mots-cibles. Comparaison étudiants vs LSA. Selon que LSA connaît le français et un cours supplémentaire

## **II.4. Projets de travail**

- **Analyse des communications du colloque IUFM.**
- **Détermination de parties principales et secondaires d'un texte en vue de son hypertextualisation (avec Bollon, Dubois, Lemaire).**
- **Proximités sémantiques de termes liés par hypertexte : LSA vs liens d'une encyclopédie sur l'astronomie.**