

# La « mise en variable » des textes : un sujet de controverses

Faire parler les  
questions ouvertes



# Enquête SD 2001

- ★ Champ de l'enquête : personnes de plus de 18 ans fréquentant au moins une fois un service d'hébergement ou de restauration gratuite pendant la période de l'enquête
- ★ Du 15 janvier au 15 février 2001
- ★ 4 084 questionnaires



# Enquête SD 2001

- \* Caractéristiques socio-démographiques
- \* Fréquentations des services
- \* Conditions de vie dans le lieu d'habitation
- \* Parcours résidentiel
- \* Revenus et difficultés financières
- \* Occupation principale (emploi, chômage, formation, expérience)
- \* Relations
- \* Santé
- \* Famille
- \* Recours aux services



# Question ouverte

- ★ Souhaitez-vous ajouter des informations que ce questionnaire n'a pas permis de recueillir ?
- ★ Contexte : question finale
- ★ Intention des concepteurs : redonner la main aux enquêtés
- ★ Conséquence : espace de liberté important du fait d'une question très ouverte (avantage et inconvénient)



# Que mesure-t-on ?

- ★ Question ouverte ou question fermée ?
- ★ Opinion ou niveau d'éducation ?
- ★ Opinion construite ou opinion conjoncturelle ?



# Statistique textuelle

- ★ Une méthode : la lexicométrie
- ★ Pourquoi quantifier ou l'utopie  
« Benzécriste »
- ★ Forme hybride : quantitatif ou qualitatif



# Choix de la méthode

(from *data* to *given*)

- ★ Contraintes de la collecte
  - ★ Objectif de l'analyse : essayer de répondre à une critique
- « *Le questionnaire est trop réducteur. Il est trop vague et n'aborde ni la réalité, ni les vrais problèmes. On sent qu'il est destiné seulement à un ordinateur* »



# « Mise en variable »

- ★ Transformation du corpus textuel en tableau
- ★ Objectif : réduire les dimensions du tableau en limitant la perte d'information



# Deux outils de la lexicométrie

- ★ Normalisation : correction et regroupement de mots
- ★ Lemmatisation : de la forme normalisée au vocable



# Normalisation

- ★ Majuscule ou minuscule
- ★ Noms propres
- ★ Dates
- ★ Acronymes
- ★ Discours direct ou indirect



# Problèmes

- ★ Homographes : être (nom masculin) et être (verbe)
- ★ Polysémie : sortir (importance du contexte)



# Lemmatisation

- ★ Suppression des mots outils
- ★ Des formes normalisées aux vocables
- ★ Choix des formes canoniques :
  - Règle lexicométrique (15 occurrences)
  - Regroupement thématique (interprétation)



# Quelques chiffres

- ★ 2186 réponses
- ★ 47879 occurrences
- ★ 4588 mots différents (dont 52% d'hapax)
- ★ Vocabulaire d'étude : 438 mots mais 60% des occurrences



# Analyse textuelle

« Or c'est la connaissance préalable de l'objet, et celle des limites des techniques utilisées, qui fait la différence entre les bonnes et les mauvaises analyses »

(Lalhou S., 1995, Vers une théorie de l'interprétation en analyse statistique des données textuelles, JADT 1995, pp. 221-228)



# Les techniques statistiques

- ★ Classification : regroupement d'individus à partir de traits communs dans les réponses
- ★ Analyse factorielle : lien entre groupe d'individus et réponses



# L'interprétation des résultats

- ★ Un processus abductif...
- ★ Qui doit prendre en compte le côté « boîte noire » des techniques



# Vocabulaire spécifique

- ★ Réponses caractéristiques d'un groupe (variance inter-classe)
- ★ Réponses modales d'un groupe (variance intra-classe)

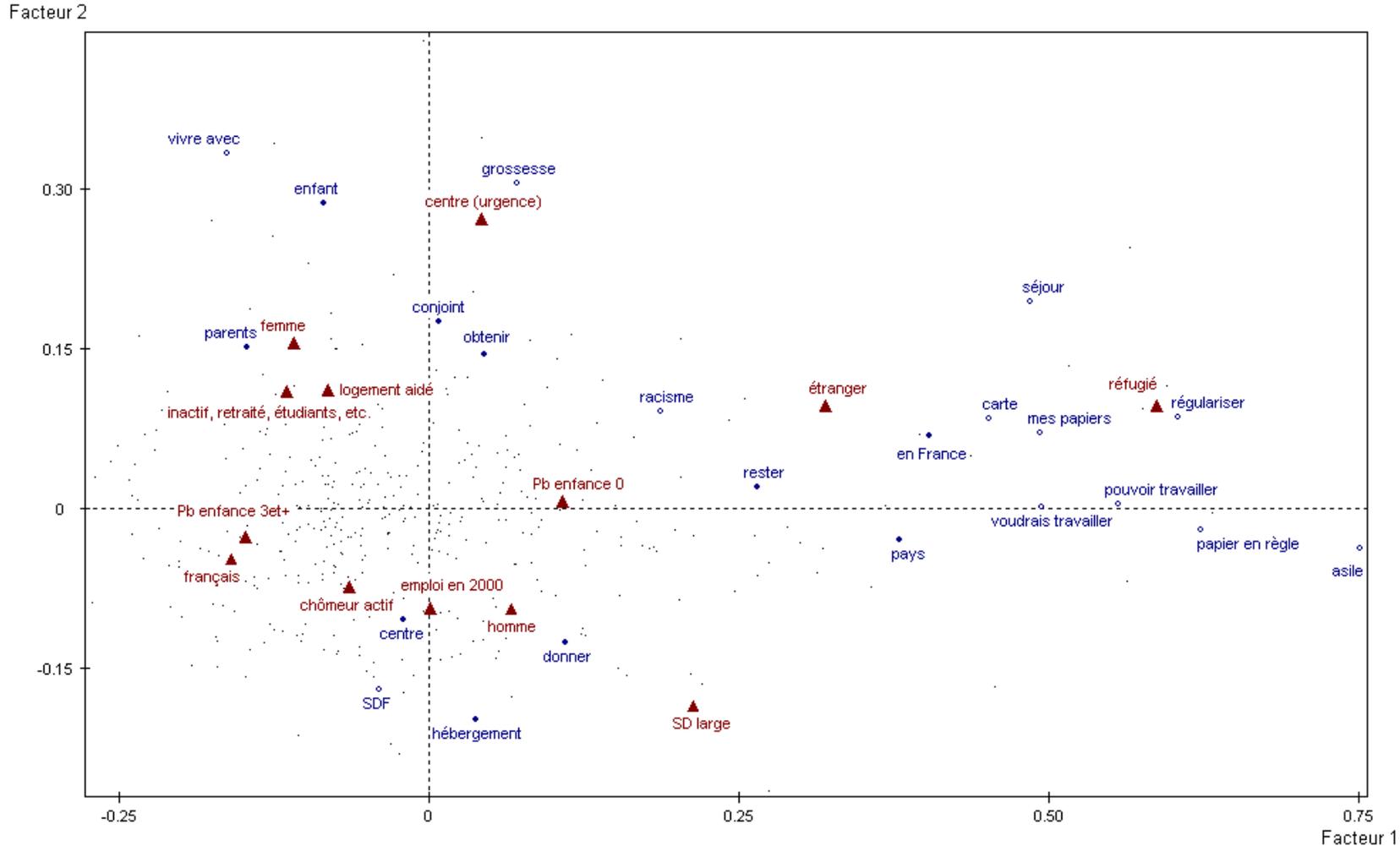


# Quelques résultats

- ★ Enquête : opinion, jugement, etc.
- ★ Logement et travail, les soucis majeurs
- ★ Usager : du quantitatif au qualitatif



# Du discours aux postures



# Conclusions

- ★ Lexicométrie : une goutte d'eau dans l'océan de l'analyse des données textuelles
- ★ La technique : simple outil au service de la recherche
- ★ Théorie de la déviance : une application aux sociologues et statisticiens



# Bibliographie

- \* Jenny J, (1997), Méthodes et pratiques formalisées d'analyse de contenu et de discours dans la recherche sociologique française contemporaine ; états des lieux et classification, Bulletin de Méthodologie Sociologique, n°54, mars 1997, pp. 64-112
- \* Lebart L., Salem A. (1994), « Statistique textuelle », Paris, Dunod, 1994.
- \* Numéro spécial « questions ouvertes », Journal de la Société Française de Statistique, tome 142, vol 4
- \* <http://www.cavi.univ-paris3.fr/lexicometrica/jadt/>

Un site avec beaucoup de textes des journées internationales de l'analyse des données textuelles (JADT)

