



# Motivation

- Étude linguistique de la pathologie mentale [[Chaika, 1974](#)] et [[Fromkin, 1975](#)]
- *Discontinuités pragmatiques* dans l'accomplissement de l'interaction verbale, [[Musiol and Trognon, 1996](#)]
- Usage pathologique de la planification du discours chez les schizophrènes paranoïdes, [[Verhaegen, 2007](#)]

**Mots clés :** troubles psychiatriques, corpus, psycholinguistique, formalisme, épistémologie

# Plan

- 1 Contexte : SLAM
- 2 Traitements automatiques
- 3 Représentation formelle
- 4 Conclusion

# Plan

- 1 Contexte : SLAM
  - Le projet SLAM
  - Corpus
- 2 Traitements automatiques
- 3 Représentation formelle
- 4 Conclusion

# SLAM - Schizophrénie et Langage :

## Analyse et modélisation

Le projet vise à systématiser l'étude des conversations pathologiques dans le cadre d'une approche interdisciplinaire

- Constitution d'une ressource linguistique sur la pathologie mentale
- Études épistémologique et philosophique (norme, folie, rationalité)
- Identifier ces usages par l'utilisation de :
  - modèles formels (type SDRT)
  - outils et méthodes du TAL

# SLAM - Schizophrénie et Langage :

## Analyse et modélisation

Le projet vise à systématiser l'étude des conversations pathologiques dans le cadre d'une approche interdisciplinaire

- Constitution d'une ressource linguistique sur la pathologie mentale
- Études épistémologique et philosophique (norme, folie, rationalité)
- Identifier ces usages par l'utilisation de :
  - modèles formels (type SDRT)
  - outils et méthodes du TAL
- [Rebuschi et al., 2014] : des corrélations explicites
- [Amblard and Fort, 2014] : usage pathologique des disfluences

# Structure de SLAM

- Module Corpus [coord. Michel Musiol]
- Module Formalisation [coord. Maxime Amblard]
- Module Epistémologie [coord. Manuel Rebuschi]

# Structure de SLAM

- Module Corpus [coord. Michel Musiol]
  - structurer la collecte de données conversationnelles,
  - transcription et étiquetage
  - première analyse pragmatique
- Module Formalisation [coord. Maxime Amblard]
- Module Epistémologie [coord. Manuel Rebuschi]

# Structure de SLAM

- Module Corpus [coord. Michel Musiol]
  
- Module Formalisation [coord. Maxime Amblard]
  - interroger la réalité cognitive des modèles sémantico-pragmatiques
  - identifier automatiquement des usages non usuels de la langue
- Module Epistémologie [coord. Manuel Rebuschi]

# Structure de SLAM

- Module Corpus [coord. Michel Musiol]
- Module Formalisation [coord. Maxime Amblard]
- Module Epistémologie [coord. Manuel Rebuschi]
  - interroger les concepts normatifs de rationalité et de logicité
  - étudier le fonctionnement de l'interprétation en interaction linguistique, et le statut des normes implicites

# Objectifs sociétaux

- Définir des indices pouvant participer à la pose d'un diagnostic
  - Définir des processus de remédiation
  - Mesurer l'avancée de la remédiation
- 
- Produire un corpus (autant que possible accessible)
  - Outiller la collecte du corpus

# De la réalité des patients

- Analyse formelle de le langage = définir une norme
- Dévier = dysfonctionnement
  
- Or, tout locuteur est confronté quotidiennement à des troubles du langage provenant de personnes saines.
  
- Le diagnostic ne peut souffrir d'approximations

# Rejouer les ambiguïtés linguistiques

- G82** l'an dernier euh (→) j'savais pas comment faire **j'étais perdue** et pourtant j'avais pris mes médicaments j'suis dans un état vous voyez même ma bouche elle est sèche j'suis dans un triste état
- V83** Vous êtes quand même bien (↑)
- G84** J'pense que ma tête est bien mais on croirait à moitié (↓) la moitié qui va et la moitié qui va pas j'ai l'impression de ça vous voyez (↑)
- V85** D'accord
- G86** Ou alors c'est la conscience peut être la conscience est ce que c'est ça (↑)
- V87** Vous savez **ça arrive à tout le monde d'avoir des moments biens et des moments où on est perdu**
- G88** **Oui j'ai peur de perdre tout le monde**
- V89** Mais ils vont plutôt bien vos enfants (↑)
- G90** Ils ont l'air ils ont l'air mais ils ont des allergies ils ont (→) mon petit fils il s'est cassé le bras à l'école tout ça

# Entretien semi-dirigé schizophrène/psychologue

- Contenu de l'entretien (transcription)
- Capacités neuro-cognitives :
  - Wechsler Adult Intelligence Scale-III  
(mesure du quotient intellectuel, ou QI)
  - California Verbal Learning Test  
(capacité cognitive et de stratégie)
  - Trail Making Test  
(dépréciation de la flexibilité cognitive et de l'inhibition).
- Comportement oculomoteur (double système d'*eye-tracker*)
- Activité de l'encéphale (EEG)

# Mesures physiologiques

Les mesures physiologiques sont reprises par d'autres projets, ou d'autres aspects du projets.

- Double système d'Eye-trackers : identifier un comportement pathologique dans la tenue du regard
- EEG : identifier des différentiels de potentiel en fonction de la production langagière.

Les deux types de mesures seront corrélées avec les indices mis en avant par le volet langage.

# Un corpus relativement important

|               | corpus Ville1 |        |       | corpus Ville2 |        |       | total |
|---------------|---------------|--------|-------|---------------|--------|-------|-------|
|               | hommes        | femmes | total | hommes        | femmes | total |       |
| schizophrènes | 15            | 3      | 18    | 20            | 10     | 30    | 48    |
| témoins       | 15            | 8      | 23    | 4             | 4      | 8     | 31    |
| total         | 30            | 11     | 41    | 24            | 14     | 38    | 79    |

## Un corpus relativement important

|               | corpus Ville1 |        |       | corpus Ville2 |        |       | total |
|---------------|---------------|--------|-------|---------------|--------|-------|-------|
|               | hommes        | femmes | total | hommes        | femmes | total |       |
| schizophrènes | 15            | 3      | 18    | 20            | 10     | 30    | 48    |
| témoins       | 15            | 8      | 23    | 4             | 4      | 8     | 31    |
| total         | 30            | 11     | 41    | 24            | 14     | 38    | 79    |

- L'identification des pathologies est extraite des dossiers médicaux et suit le **DSM-IV**
- La catégorie principale est les **schizophrènes paranoïdes**, sans exclusion des autres formes de schizophrénie
- Corpus Ville1 en grande partie constitué en centre de remédiation
- Le corpus Ville2 comporte toute sorte de patients

# Taille du corpus

31 575 tours de parole / 375 000 mots

|              | corpus Ville1      |                   | corpus Ville2      |                 |
|--------------|--------------------|-------------------|--------------------|-----------------|
|              | nb tours de parole | nb de mots        | nb tours de parole | nb de mots      |
| <i>S</i>     | 3 863              | 46 859            | 4 062              | 66 725          |
| <i>T</i>     | 7 282 } 11 145     | 72 903 } 119 762  | 371 } 4 433        | 12 356 } 79 081 |
| <i>P + S</i> | 3 819              | 30 293            | 4 098              | 33 686          |
| <i>P + T</i> | 7 698 } 11 517     | 108 278 } 138 571 | 382 } 4 480        | 4 156 } 37 842  |
| <i>total</i> | 22 662             | 258 333           | 8 913              | 116 923         |

# Tradition propriétaire des corpus

Difficulté de mise en perspective de ce type de travaux :

- Impossibilité d'accéder à des corpus d'autres projets / plus anciens

# Tradition propriétaire des corpus

Difficulté de mise en perspective de ce type de travaux :

- Impossibilité d'accéder à des corpus d'autres projets / plus anciens

⇒ Nécessité de transférer les bonnes pratiques de la linguistique en matière de gestion des corpus vers les autres disciplines.

# Un corpus difficile à constituer

Démarches administratives lourdes :

- CPP de la région de l'institution médicale
  - Description finalisée du protocole
  - Intrusion du protocole
  - Plusieurs mois d'instruction
  - Contraction d'une assurance
- CNIL
- Les données ne doivent pas être utilisées pour/contre le patient.

# Un corpus difficile à constituer

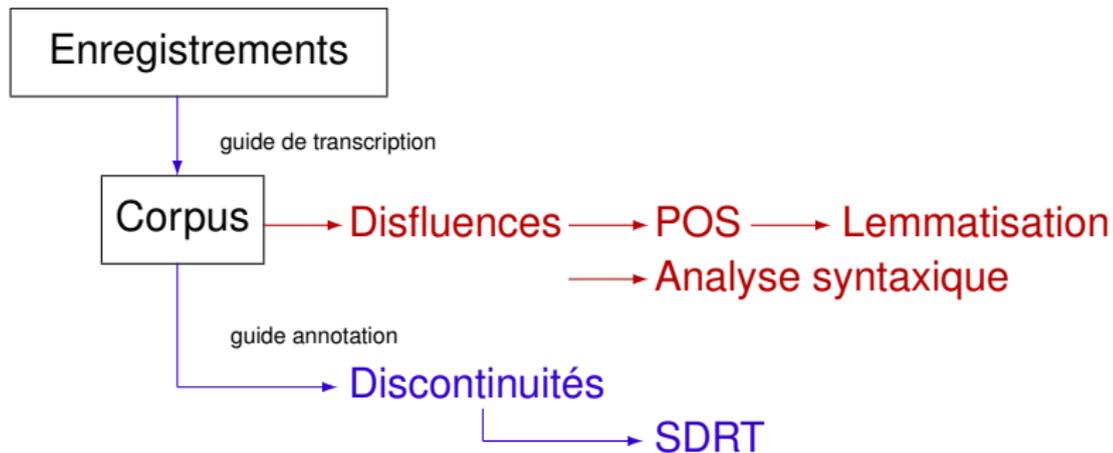
## Participation des patients :

- Identifier les patients capables de participer
- Identifier ceux qui acceptent de participer
- Gérer les inquiétudes des patients
  - identification des patients et obtention de leur accord (crainte de divulgation d'éléments biographiques) (perte ~ 45 %)
  - abandons (perte ~ 10 %)

# Un corpus difficile à constituer

- Transcription
  - Automatique : tests de plusieurs systèmes entraînés pour le français oral non concluants
  - Manuelle : deux annotateurs (dont le/la psychologue)
- Guide de transcription

# Annotation de la ressource



# Normalisation des corpus

- Normalisation : trentaine d'expressions régulières
  - sous-corpus Ville2 : MS Word → Distagger
  - sous-corpus Ville1 : CLAN → Distagger
- Identification des tours de parole par numéro unique
- Statut du sujet (patient vs témoin) : dans la structure du corpus
- Psychologue pris en compte pour la dynamique de l'interaction

# Plan

- 1 Contexte : SLAM
- 2 Traitements automatiques**
  - Disfluences
  - POS/Lemmes
- 3 Représentation formelle
- 4 Conclusion

# Traitements automatiques ?

- Corpus de grande taille
- Limiter les interventions humaines pour l'identification automatique :
  - des disfluences dans les transcriptions avec `Distagger` [Constant and Dister, 2010]
  - des POS et des lemmes avec `MElt` [Denis et Sagot, 2012]

# Distagger

f-score : 95,5 %, précision : 95,3 %, rappel : 95,8 %

[Constant and Dister, 2010]

## 1 'euh'

(1) *moi ça m'est presque plus euh difficile et euh anti-naturel de parler*

## 2 Répétitions

(2) *j' arrive à être à être concentrée quand il faut faire quelque chose*

## 3 Autocorrections

(3) *enfin je sais pas trop le les termes*

## 4 Amorces (interruption de morphèmes)

(4) *pis progressivement vous av- pouvez travailler sur votre concentration*

# Distagger

- Sept étiquettes

{*IGN+EUH*}, {*IGN+REP*}, {*IGN+CORR*}, {*IGN+FRAG*},  
{*IGN+short\_pause*}, {*IGN+slot*} et {*IGN+speaker*}

# Distagger

- Sept étiquettes

{*IGN+EUH*}, {*IGN+REP*}, {*IGN+CORR*}, {*IGN+FRAG*},  
{*IGN+short\_pause*}, {*IGN+slot*} et {*IGN+speaker*}

⇒ repérer les tours de parole et les interlocuteurs

# Distagger

- Sept étiquettes

{*IGN+EUH*}, {*IGN+REP*}, {*IGN+CORR*}, {*IGN+FRAG*},  
{*IGN+short\_pause*}, {*IGN+slot*} et {*IGN+speaker*}

⇒ repérer les tours de parole et les interlocuteurs

⇒ volumes très faibles (respectivement 1 et 5 étiquettes)

# Distagger

- Sept étiquettes

{*IGN+EUH*}, {*IGN+REP*}, {*IGN+CORR*}, {*IGN+FRAG*},  
{*IGN+short\_pause*}, {*IGN+slot*} et {*IGN+speaker*}

⇒ repérer les tours de parole et les interlocuteurs

⇒ volumes très faibles (respectivement 1 et 5 étiquettes)

- Scripts Python

- pré-traitement
- application de l'outil
- post-traitement sur les corpus

# Analyse des étiquettes

- Certaines annotations sont écartées : '*mmh mmh*', '*oui oui*', '*non non*', '*vous vous*'
- ~ 1 500 lignes de code
- Automatisation de la production
  - de graphiques :
    - par entretien : disfluences par tour de parole (catégories et total)
    - par sous-corpus : une surface
  - de données :
    - moyennes de disfluences par tour de parole et par nombre de mots
    - indice de significativité (sous-corpus)

# Répartition en absolu

Nombre max de disfluences dans un tour de parole

S : [6, 4, 13, 9, 6, 3, 4, 4, 3, 7, 5, 9, 13, 7, 6, 5, 7, 9, 6]

T : [6, 5, 4, 25, 6, 7, 4, 7, 7, 4, 7, 7, 4, 13, 12, 5, 15, 13, 8, 12, 5, 5]

# Répartition en absolu

Nombre max de disfluences dans un tour de parole

S : [6, 4, 13, 9, 6, 3, 4, 4, 3, 7, 5, 9, 13, 7, 6, 5, 7, 9, 6]

T : [6, 5, 4, 25, 6, 7, 4, 7, 7, 4, 7, 7, 4, 13, 12, 5, 15, 13, 8, 12, 5, 5]

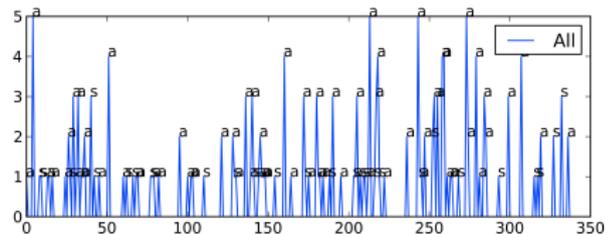
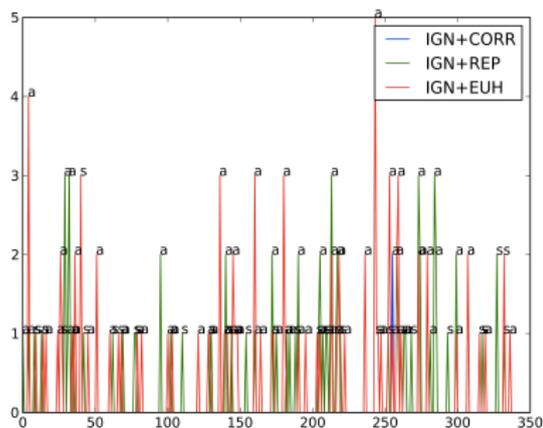
Les témoins ont des tours de parole contenant plus de disfluences :

$$\max(S) = 13 \text{ vs } \max(T) = 25$$

# Visualisation des résultats

par entretien

## Répartition des disfluences dans un entretien



## Résultats (ratio de disfluences)

|                    | S      | T      | S+T    | P+S    | P+T    | P      |
|--------------------|--------|--------|--------|--------|--------|--------|
| Corpus Ville2      |        |        |        |        |        |        |
| par tour de parole | 0,5417 | 0,5589 | 0,545  | 0,1400 | 0,1513 | 0,1424 |
| par / nb mots      | 0,032  | 0,0168 | 0,0288 | 0,0144 | 0,0138 | 0,0142 |
| Corpus Ville1      |        |        |        |        |        |        |
| par tour de parole | 0,7117 | 0,484  | 0,5842 | 0,3338 | 0,7369 | 0,5599 |
| par / nb mots      | 0,0595 | 0,0468 | 0,0524 | 0,0421 | 0,0496 | 0,0463 |

- Par nombre de tours de parole : variabilité importante
- Par nombre de mots :
  - témoins et psychologue : valeurs du même ordre
  - schizophrènes : valeurs bien supérieures

## Significativité

[de Mareüil et al., 2013]

$$s = \frac{(p_1 - p_2)}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

où :

$$p = (n_1 p_1 + n_2 p_2) / (n_1 + n_2)$$

Valeurs significatives : &gt; 1,96

|          | corpus Ville1  | corpus Ville2 |
|----------|----------------|---------------|
| S et Psy | 10,6806923083  | 19,4197596818 |
| T et Psy | 0,422898291704 | 3,23530253756 |
| S et T   | 10,2827554261  | 16,0376100956 |

# Biais

- Différences entre les sous-corpus (transcription différente)  
⇒ pas d'évaluation qualitative globale
  - `Distagger` sur le corpus de parole spontanée TCOF-POS  
[Benzitoun et al., 2012] : 4,3 %
- Différences d'âge et de QI

|                 | âge          | QI           | études   |
|-----------------|--------------|--------------|----------|
| Schizophrènes   | 29 ans       | 95           | 12,4 ans |
| Témoins         | 23 ans       | 103          | 13 ans   |
| Test de Student | $p = 0,0058$ | $p = 0,0203$ |          |

- Patients sous traitement

# POS / Lemmes

Complexité de la production

MEIt entraîné sur le corpus oral TCOF-POS

# POS / Lemmes

## Distribution des étiquettes POS

| nb moyen |   | étiq. | tours | ratio  | diff étiq. |
|----------|---|-------|-------|--------|------------|
| Ville1   | T | 3471  | 316   | 10,98  | 40         |
|          | S | 2757  | 210   | 13,12  | 38         |
|          | P | 3647  | 280   | 13,025 | 39         |
| Ville2   | T | 1462  | 46    | 31,78  | 35         |
|          | S | 2269  | 131   | 17,32  | 37         |
|          | P | 1065  | 113   | 9,42   | 35         |

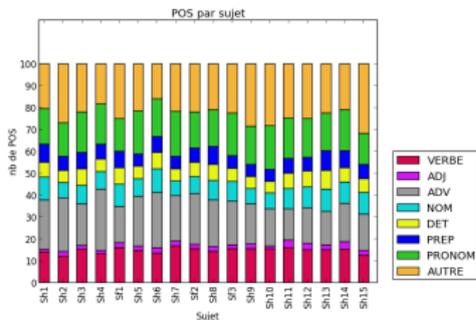
# POS/Lemme

Répartition des catégories morpho-syntaxiques en grandes catégories

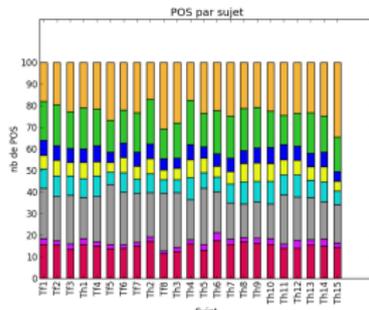
|        |    | VER | ADJ | ADV | NOM | DET | PRP | PRO | AUTRE |
|--------|----|-----|-----|-----|-----|-----|-----|-----|-------|
| Ville1 | T  | 524 | 83  | 711 | 297 | 234 | 218 | 617 | 785   |
|        | S  | 416 | 61  | 537 | 238 | 183 | 190 | 497 | 632   |
|        | Ps | 575 | 91  | 795 | 318 | 247 | 247 | 634 | 738   |
| Ville2 | T  | 247 | 45  | 173 | 199 | 146 | 141 | 265 | 243   |
|        | S  | 378 | 58  | 243 | 266 | 201 | 182 | 440 | 498   |
|        | Ps | 192 | 27  | 135 | 117 | 86  | 80  | 231 | 194   |

# POS/Lemme

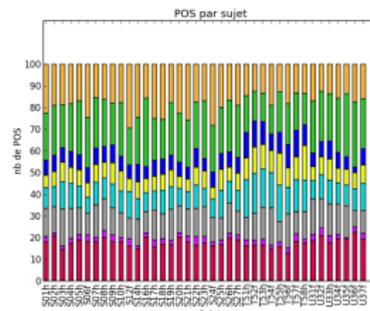
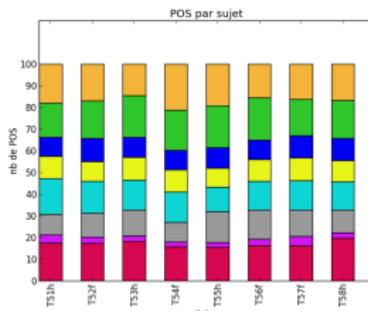
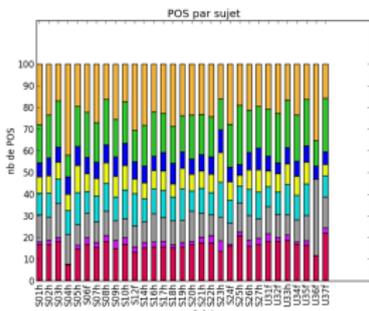
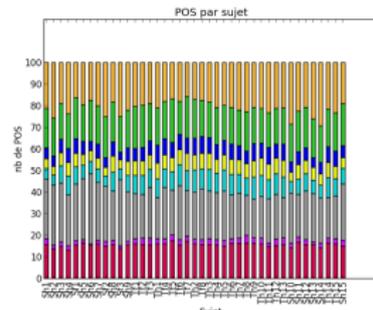
S



T



P



# POS/Lemme

$$\text{Richesse lexicale} = \frac{\text{nombre de lemmes}}{\text{nombre total de formes}}$$

$$\text{Diversité lexicale} = \frac{\text{nombre de lemmes}}{\text{nombre total de formes différentes (types)}}$$

# POS/Lemme

|   | Ville1 |      | Ville2 |      |
|---|--------|------|--------|------|
|   | RL     | DL   | RL     | DL   |
| T | 0,045  | 0.68 | 0,118  | 0.73 |
| S | 0,05   | 0.69 | 0,06   | 0.70 |
| P | 0,022  | 0.64 | 0,06   | 0.68 |

|   | Ville2 |      |      |      |                 |      |                 |      |
|---|--------|------|------|------|-----------------|------|-----------------|------|
|   | H      |      | F    |      | avec traitement |      | sans traitement |      |
|   | RL     | DL   | RL   | DL   | RL              | DL   | RL              | DL   |
| T | 0,15   | 0.76 | 0,14 | 0.74 |                 |      |                 |      |
| S | 0,07   | 0.72 | 0,08 | 0.71 | 0,067           | 0.71 | 0,10            | 0.72 |

# POS/Lemme

|   | Ville1 |      | Ville2 |      |
|---|--------|------|--------|------|
|   | RL     | DL   | RL     | DL   |
| T | 0,045  | 0.68 | 0,118  | 0.73 |
| S | 0,05   | 0.69 | 0,06   | 0.70 |
| P | 0,022  | 0.64 | 0,06   | 0.68 |

|   | Ville2 |      |      |      |                 |      |                 |      |
|---|--------|------|------|------|-----------------|------|-----------------|------|
|   | H      |      | F    |      | avec traitement |      | sans traitement |      |
|   | RL     | DL   | RL   | DL   | RL              | DL   | RL              | DL   |
| T | 0,15   | 0.76 | 0,14 | 0.74 |                 |      |                 |      |
| S | 0,07   | 0.72 | 0,08 | 0.71 | 0,067           | 0.71 | 0,10            | 0.72 |

# POS/Lemme

|   | Ville1 |      | Ville2 |      |
|---|--------|------|--------|------|
|   | RL     | DL   | RL     | DL   |
| T | 0,045  | 0.68 | 0,118  | 0.73 |
| S | 0,05   | 0.69 | 0,06   | 0.70 |
| P | 0,022  | 0.64 | 0,06   | 0.68 |

|   | Ville2 |      |      |      |                 |      |                 |      |
|---|--------|------|------|------|-----------------|------|-----------------|------|
|   | H      |      | F    |      | avec traitement |      | sans traitement |      |
|   | RL     | DL   | RL   | DL   | RL              | DL   | RL              | DL   |
| T | 0,15   | 0.76 | 0,14 | 0.74 |                 |      |                 |      |
| S | 0,07   | 0.72 | 0,08 | 0.71 | 0,067           | 0.71 | 0,10            | 0.72 |

# POS/Lemme

|   | Ville1 |      | Ville2 |      |
|---|--------|------|--------|------|
|   | RL     | DL   | RL     | DL   |
| T | 0,045  | 0.68 | 0,118  | 0.73 |
| S | 0,05   | 0.69 | 0,06   | 0.70 |
| P | 0,022  | 0.64 | 0,06   | 0.68 |

|   | Ville2 |      |      |      |                 |      |                 |      |
|---|--------|------|------|------|-----------------|------|-----------------|------|
|   | H      |      | F    |      | avec traitement |      | sans traitement |      |
|   | RL     | DL   | RL   | DL   | RL              | DL   | RL              | DL   |
| T | 0,15   | 0.76 | 0,14 | 0.74 |                 |      |                 |      |
| S | 0,07   | 0.72 | 0,08 | 0.71 | 0,067           | 0.71 | 0,10            | 0.72 |

# POS/Lemme

|   | Ville1 |      | Ville2 |      |
|---|--------|------|--------|------|
|   | RL     | DL   | RL     | DL   |
| T | 0,045  | 0.68 | 0,118  | 0.73 |
| S | 0,05   | 0.69 | 0,06   | 0.70 |
| P | 0,022  | 0.64 | 0,06   | 0.68 |

|   | Ville2 |      |      |      |                 |      |                 |      |
|---|--------|------|------|------|-----------------|------|-----------------|------|
|   | H      |      | F    |      | avec traitement |      | sans traitement |      |
|   | RL     | DL   | RL   | DL   | RL              | DL   | RL              | DL   |
| T | 0,15   | 0.76 | 0,14 | 0.74 |                 |      |                 |      |
| S | 0,07   | 0.72 | 0,08 | 0.71 | 0,067           | 0.71 | 0,10            | 0.72 |

# POS/Lemme

- RL et DL équivalents entre Témoins et Schizophrènes
- RL plus basse chez les Schizophrènes dans Ville 2
- Mais les sans traitements sont équivalents contrairement à ceux avec traitement (DL toujours équivalent)
- Le groupe des sans-traitement étant composé majoritairement de femme
- Mais le comportement est similaire entre les genres

# POS/Lemme

## Significativité

|        | Ville1 |       | Ville2 |       |
|--------|--------|-------|--------|-------|
|        | RL     | DL    | RL     | DL    |
| T et P | 21.38  | 18.24 | 21.14  | 11.17 |
| S et P | 24.27  | 20.08 | 0.14   | 7.85  |
| T et S | 4.72   | 4.28  | 22.70  | 6.95  |

|        | Ville2 |      |       |      |                 |      |                 |      |
|--------|--------|------|-------|------|-----------------|------|-----------------|------|
|        | H      |      | F     |      | avec traitement |      | sans traitement |      |
|        | RL     | DL   | RL    | DL   | RL              | DL   | RL              | DL   |
| T et P | 21.70  | 7.67 | 8.44  | 7.08 | 17.57           | 8.01 | 4.76            | 5.92 |
| S et P | 1.42   | 6.97 | 4.79  | 2.53 | 0.61            | 5.52 | 0.46            | 3.90 |
| T et S | 22.19  | 4.12 | 13.79 | 5.51 | 18.56           | 4.77 | 4.47            | 2.29 |

# Plan

- 1 Contexte : SLAM
- 2 Traitements automatiques
- 3 Représentation formelle**
- 4 Conclusion

# Objectif, position et contexte

Comprendre la folie ?

- Paraît impossible dans les cas les plus radicaux :  
déli de réalité, pensée apparemment contradictoire...
- Approche externe (explication à la troisième personne).

# Objectif, position et contexte

## Comprendre la folie ?

- Paraît impossible dans les cas les plus radicaux :  
déli de réalité, pensée apparemment contradictoire...
- Approche externe (explication à la troisième personne).

## Position défendue :

- Sans nier la pertinence des approches à la troisième personne
- Refus du *réductionnisme* (neurobiologique, psychanalytique...)
- Rationalité du sujet (à la première personne)

# Converser avec des schizophrènes

- Point commun avec Sass :
  - le contenu n'est pas le problème de la pensée schizophrène
  - ⇒ postulat : version forte du principe charité (à la Quine)

# Converser avec des schizophrènes

- Point commun avec Sass :
  - le contenu n'est pas le problème de la pensée schizophrène
  - ⇒ postulat : version forte du principe charité (à la Quine)
  
- Stratégie suivie (Musiol & Rebuschi 2007, Amblard et al. 2010) :
  - pas de nouveaux types d'états mentaux
  - déviance dans les règles d'usage du langage

# Converser avec des schizophrènes

Deux interlocuteurs et deux points de vue (spontanés) sur la conversation pathologique (où apparaissent des ruptures)

# Converser avec des schizophrènes

Deux interlocuteurs et deux points de vue (spontanés) sur la conversation pathologique (où apparaissent des ruptures)

| <b>Sujet normal</b><br>(à la 3e personne)                     | <b>Schizophrène</b><br>(à la 1e personne)   |
|---|---|
| correction pragmatique<br>↓                                   | déviante pragmatique<br>↑                   |
| déviante sémantique   | correction sémantique                       |
| contenu contradictoire :<br><i>apparence de contradiction</i> | contenu cohérent :<br><i>tout va bien !</i> |

# Converser avec des schizophrènes

Deux interlocuteurs et deux points de vue (spontanés) sur la conversation pathologique (où apparaissent des ruptures)

| <b>Sujet normal</b><br>(à la 3e personne)                     | <b>Schizophrène</b><br>(à la 1e personne)   |
|---|---|
| correction pragmatique<br>↓                                   | déviante pragmatique<br>↑                   |
| déviante sémantique   | correction sémantique                       |
| contenu contradictoire :<br><i>apparence de contradiction</i> | contenu cohérent :<br><i>tout va bien !</i> |

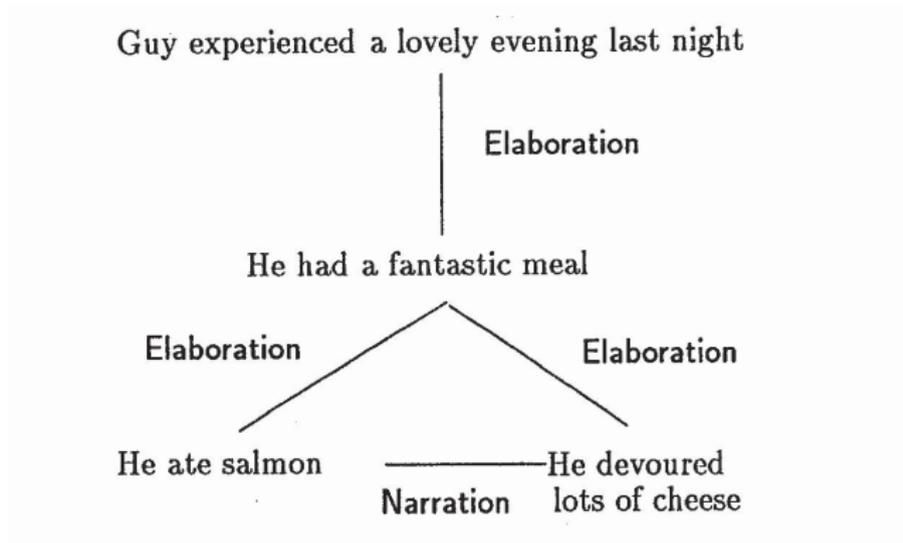
⇒ La représentation conversationnelle *ne peut pas* se borner au seul contenu (sémantique).

# Présentation du formalisme

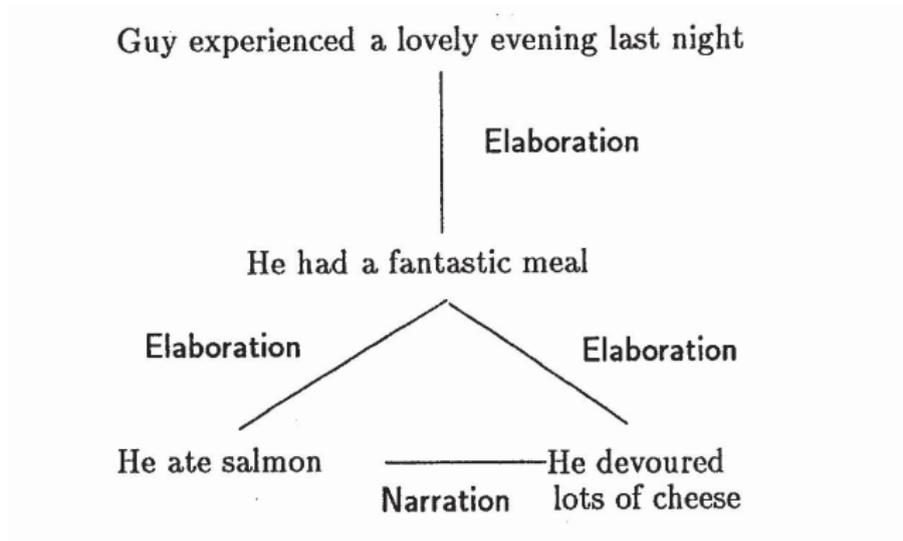
Formalisme utilisé :

- DRT (Discourse Representation Theory) : formalisme sémantique développé à partir des années 1980 [Kamp & Reyle 1993]
- S-DRT (Segmented DRT) : extension de la DRT qui incorpore les relations rhétoriques entre actes, développée à partir des années 1990 [Asher & Lascarides 2003]

## Structure rhétorique en S-DRT : plusieurs types de relations (subordination ou coordination)



## Structure rhétorique en S-DRT : plusieurs types de relations (subordination ou coordination)



Contraintes sur les sites de rattachement (règle de la frontière droite).  
Ex : “He found **it** really marvelous”

# Spécificité des conversations schizophrènes

Conjecture 1 :

Les schizophrènes sont logiquement consistants.

# Spécificité des conversations schizophrènes

## Conjecture 1 :

Les schizophrènes sont logiquement consistants.

Par conséquent les ruptures interviennent au niveau du processus de construction de la représentation conversationnelle sur la dimension pragmatique (sur les relations rhétoriques des S-DRS).

# Spécificité des conversations schizophrènes

## Conjecture 1 :

Les schizophrènes sont logiquement consistants.

Par conséquent les ruptures interviennent au niveau du processus de construction de la représentation conversationnelle sur la dimension pragmatique (sur les relations rhétoriques des S-DRS).

## Conjecture 2 :

La sous-spécification (ambiguïté) est centrale dans la rupture

# Spécificité des conversations schizophrènes

## Conjecture 1 :

Les schizophrènes sont logiquement consistants.

Par conséquent les ruptures interviennent au niveau du processus de construction de la représentation conversationnelle sur la dimension pragmatique (sur les relations rhétoriques des S-DRS).

## Conjecture 2 :

La sous-spécification (ambiguïté) est centrale dans la rupture

*Slogan* : “Un choix n’est jamais définitif !”

# Exemples

- B124** Oh ouais (↑) et pis compliqué (↓) et c'est vraiment très très compliqué (→) **la politique** c'est quelque chose quand on s'en occupe **faut être gagnant** parce qu'autrement quand on est perdant c'est fini quoi (↓)
- A125** Oui
- B126** J. C. D. **est mort**, L. **est mort**, P. **est mort** euh (...)
- A127** **Ils sont morts parce qu'ils ont perdu à votre avis** (↑)
- B128** Non ils gagnaient mais si ils sont morts, **c'est la maladie** quoi c'est c'est (→)
- A129** Ouais c'est parce qu'ils étaient malades, c'est pas parce qu'ils faisaient de la politique (↑)
- B130** Si enfin (→)
- A131** Si vous pensez que c'est parce qu'ils faisaient de la politique (↑)
- B132** Oui tiens oui il y a aussi **C. qui a accompli un meurtre là** (→) il était présent lui aussi qui est à B. mais enfin (→) c'est encore à cause de la politique ça

# Analyse de l'exemple 1

Le sujet schizophrène bascule à deux reprises d'un thème à l'autre :

- La mort politique (ou symbolique)
- La mort physique (ou littérale)

# Analyse de l'exemple 1

Le sujet schizophrène bascule à deux reprises d'un thème à l'autre :

- La mort politique (ou symbolique)
- La mort physique (ou littérale)

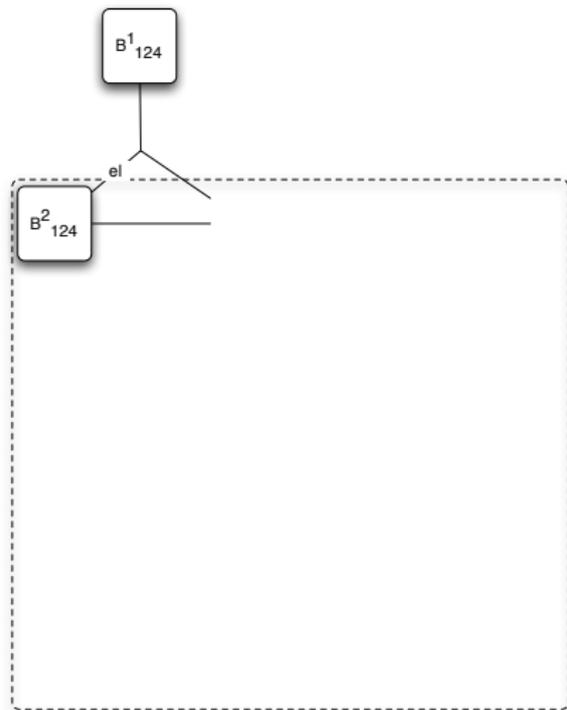
Les deux thèmes sont directement liés, mais ils expriment des réalités différentes.

Cet extrait est le lieu d'une **rupture de la frontière droite**.

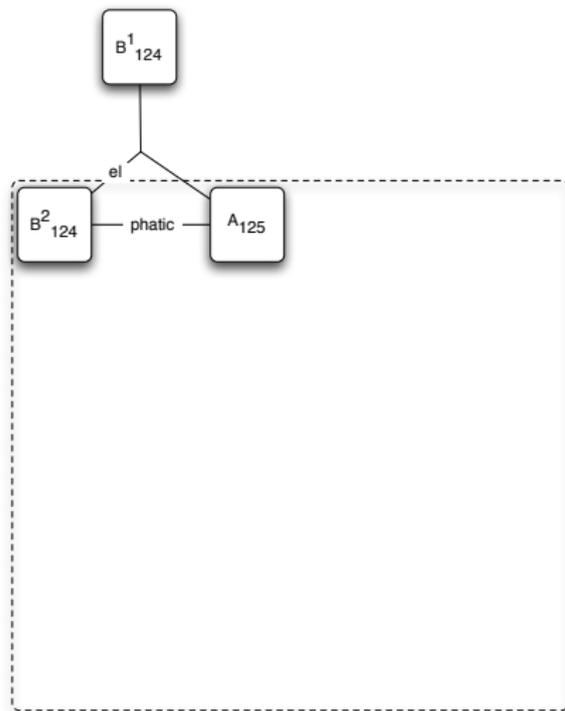
(B124) Oh ouais ( $\uparrow$ ) et pis compliqué ( $\downarrow$ ) et c'est vraiment très très compliqué  
( $\rightarrow$ )

B<sup>1</sup><sub>124</sub>

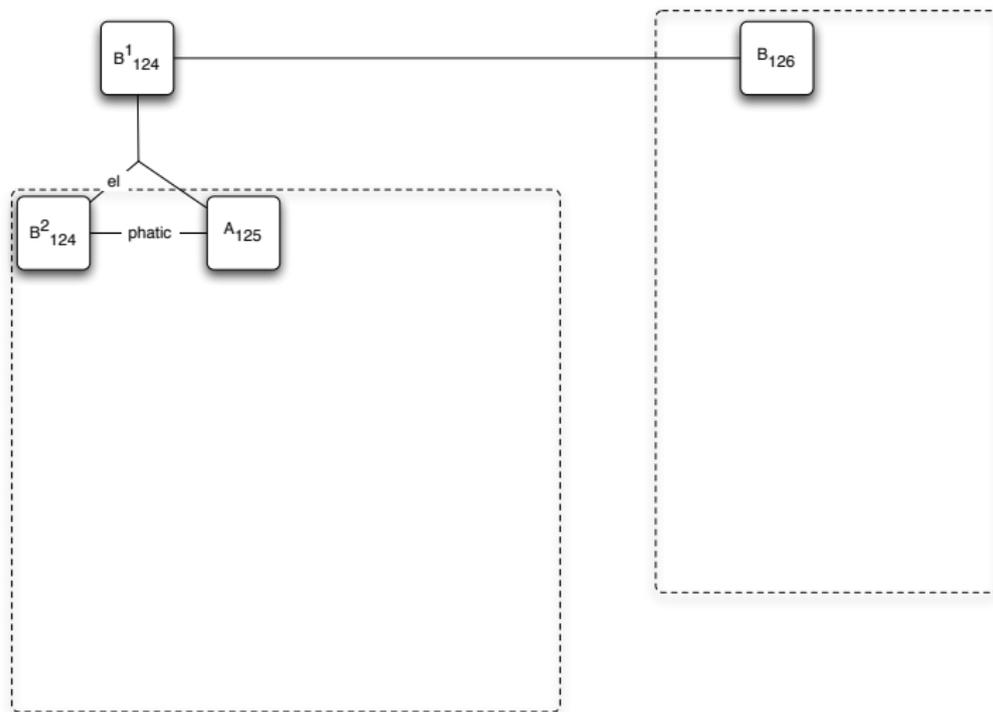
la politique c'est quelque chose quand on s'en occupe faut être gagnant parce qu'autrement quand on est perdant c'est fini quoi (↓)



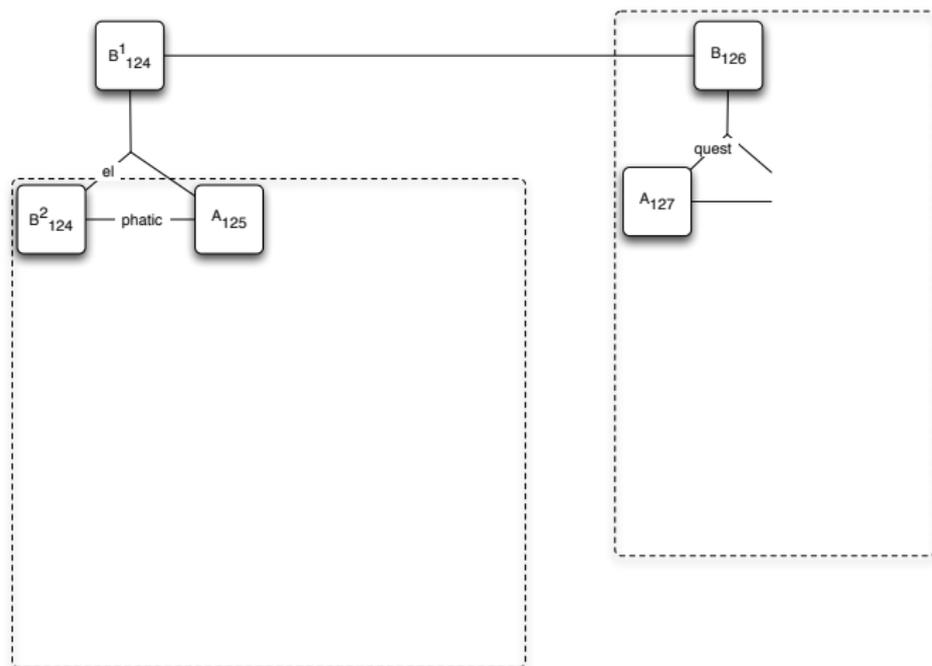
(A125) Oui



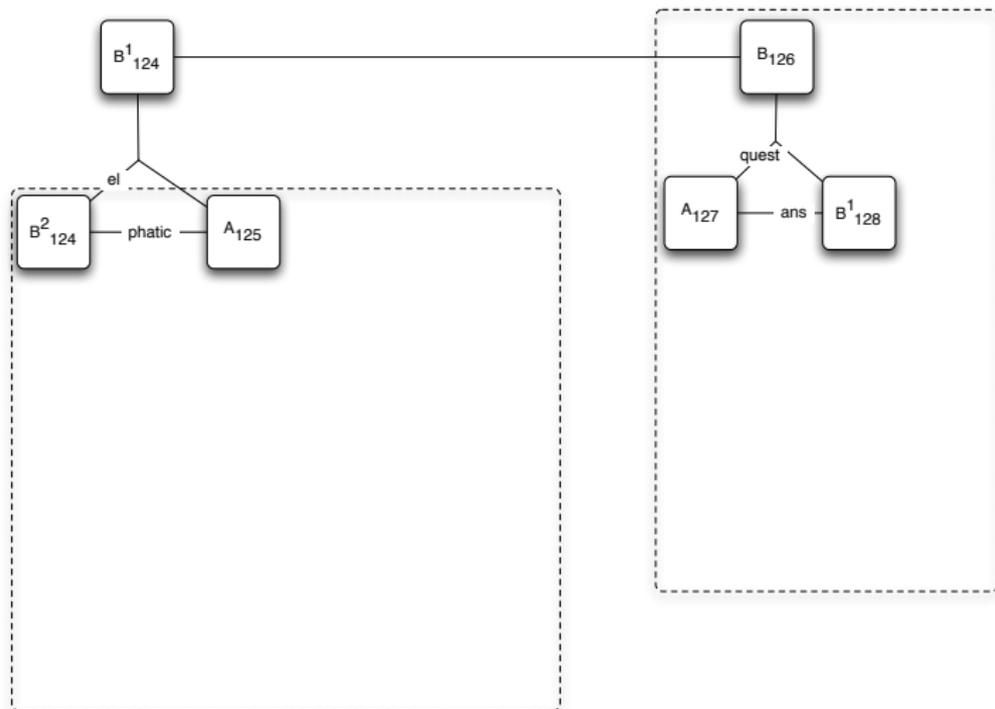
(B126) J. C. D. est mort, L. est mort, P. est mort euh (...)



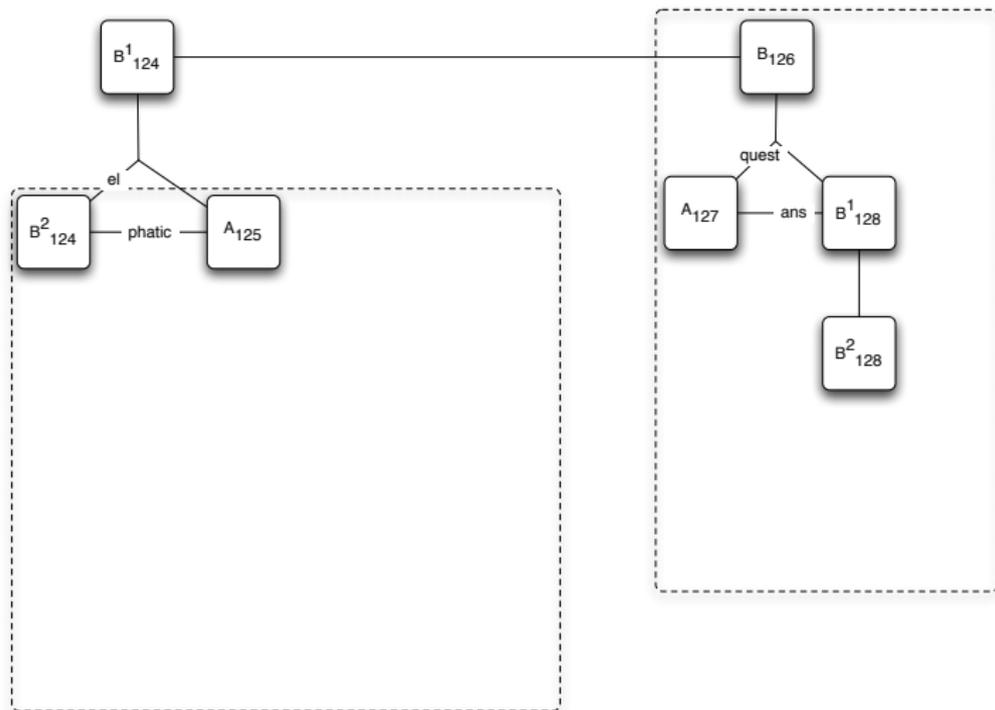
(A127) Ils sont morts parce qu'ils ont perdu à votre avis (↑)



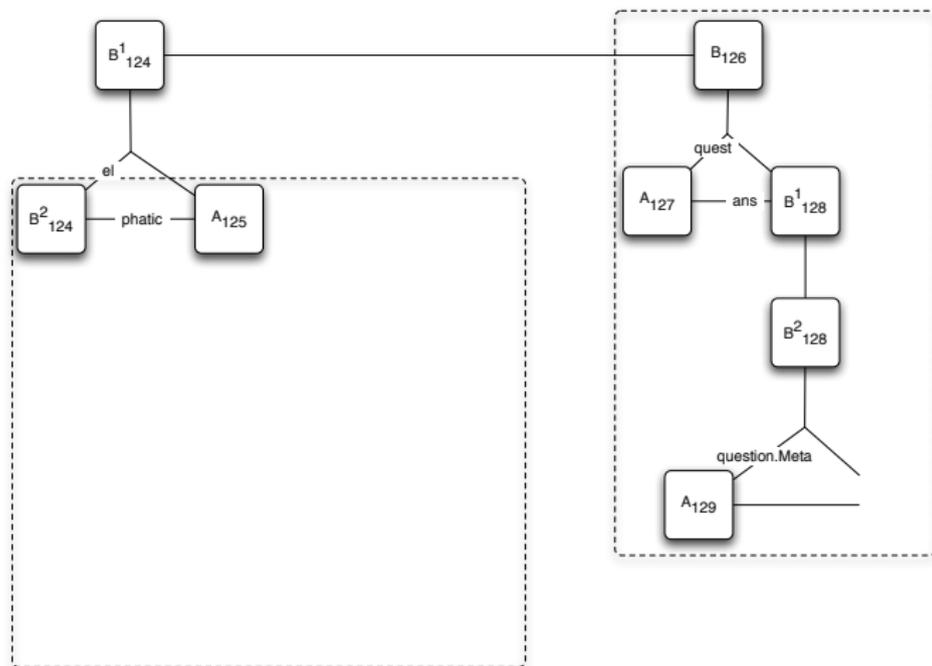
(B128) Non ils gagnaient mais si ils sont morts, c'est la maladie quoi c'est c'est (→)



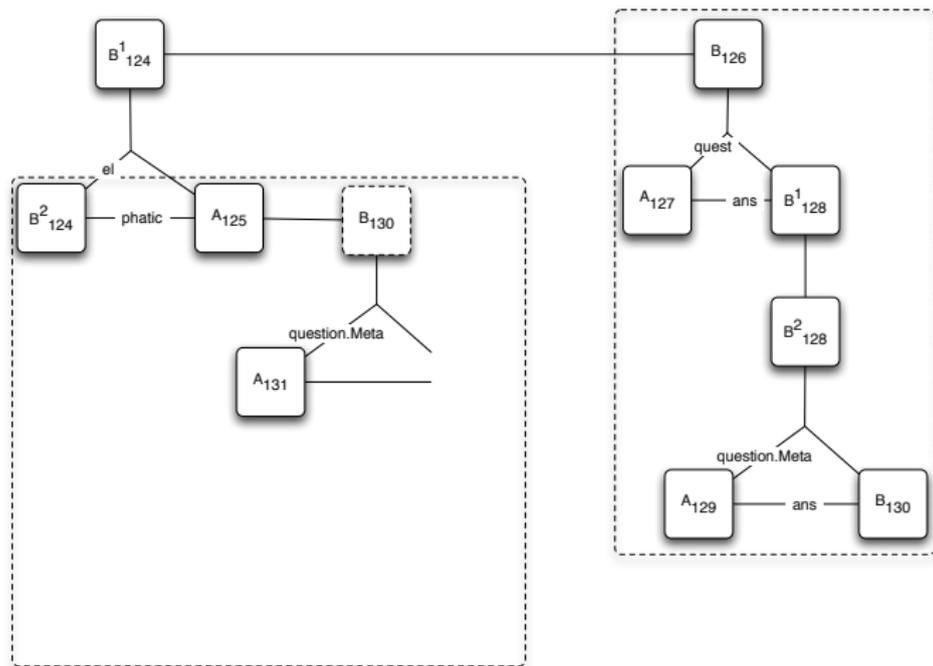
(B128) Non ils gagnaient mais si ils sont morts, c'est la maladie quoi c'est c'est (→)



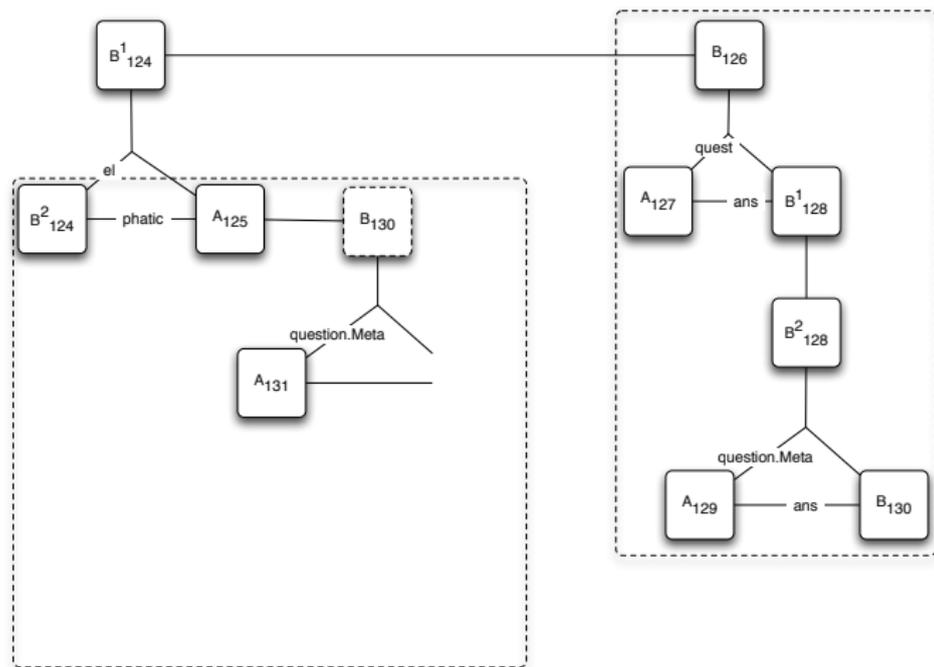
(A129) Ouais c'est parce qu'ils étaient malades, c'est pas parce qu'ils faisaient de la politique (↑)



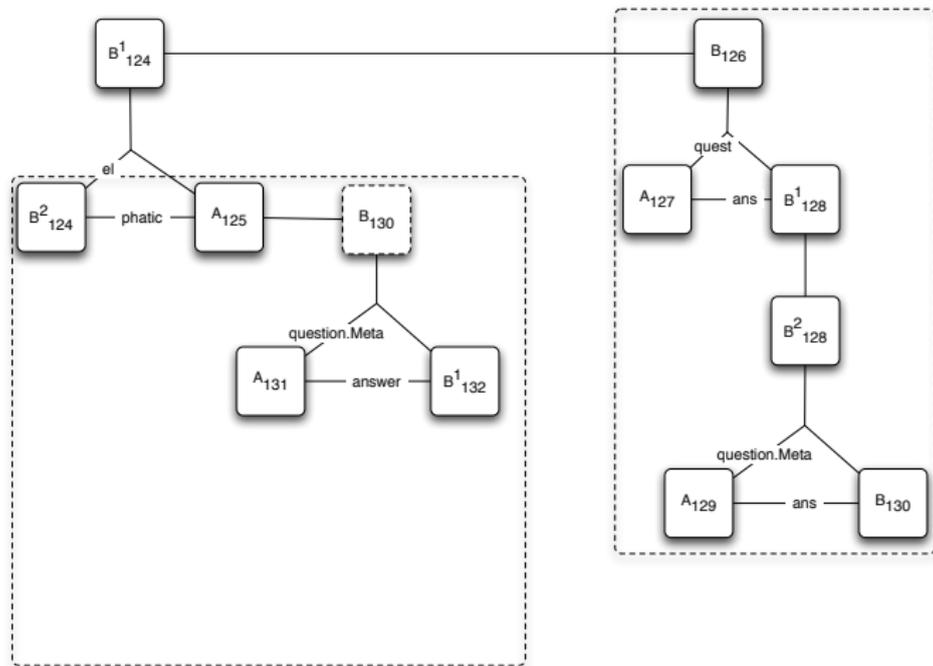
(B130) Si enfin (→)



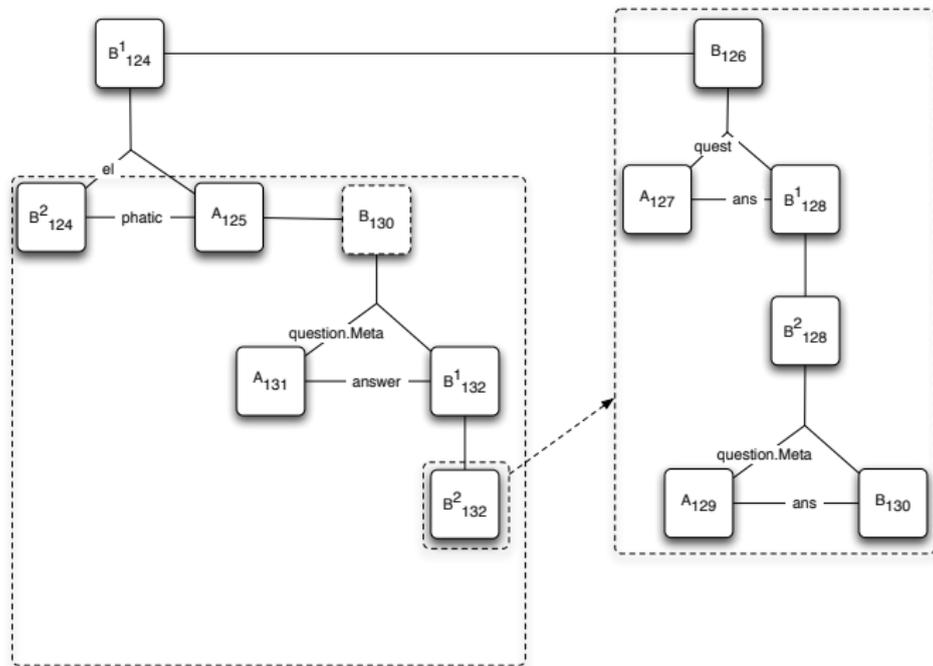
(A131) Si vous pensez que c'est parce qu'ils faisaient de la politique (↑)



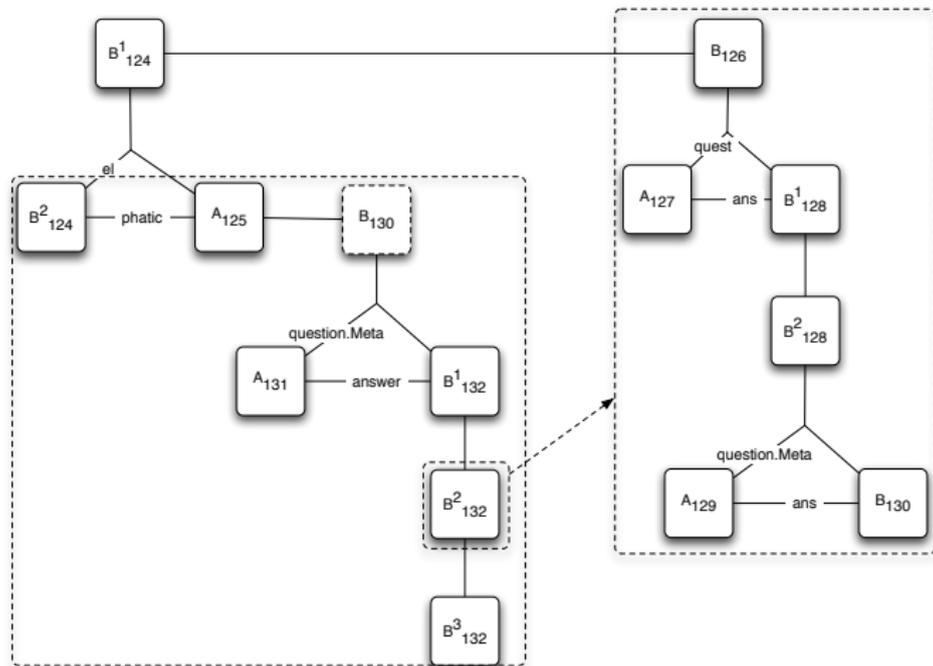
(B132) **Oui tiens oui il y a aussi C. qui a accompli un meurtre là** ( $\rightarrow$ ) il était présent lui aussi qui est à B. mais enfin ( $\rightarrow$ ) c'est encore à cause de la politique ça



(B132) Oui tiens oui il y a aussi C. qui a accompli un meurtre là ( $\rightarrow$ ) **il était présent lui aussi qui est à B. mais enfin** ( $\rightarrow$ ) c'est encore à cause de la politique ça



(B132) Oui tiens oui il y a aussi C. qui a accompli un meurtre là ( $\rightarrow$ ) il était présent lui aussi qui est à B. mais enfin ( $\rightarrow$ ) **c'est encore à cause de la politique ça**



# Inconsistance de la clôture

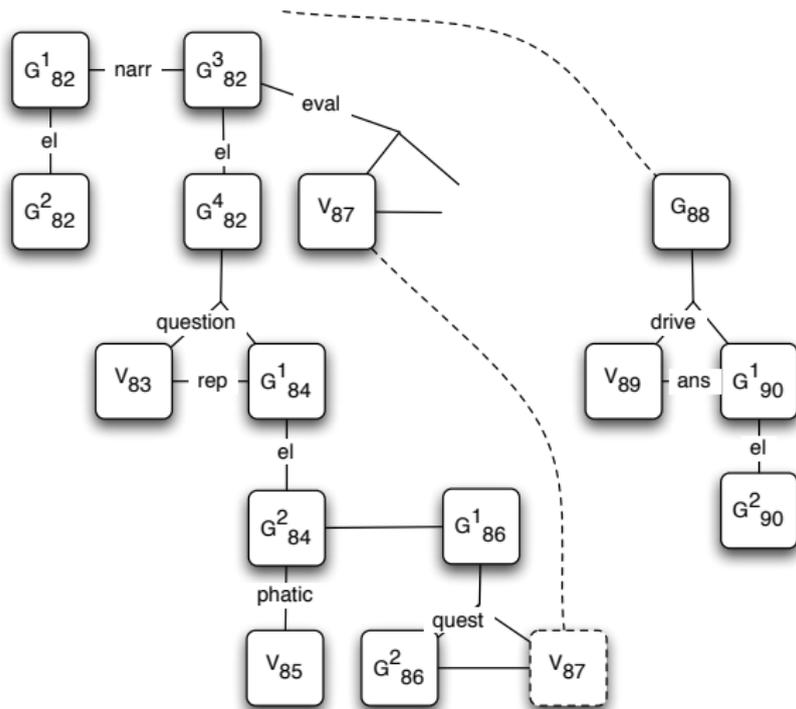
Les deux interlocuteurs tentent une remontée dans la structure :

- Le psychologue utilise la règle de remontée
- Le schizophrène **enfreint la règle de remontée** : il n'effectue pas de clôture basse acceptable.

# Remontée à travers la structure

- G82** l'an dernier euh (→) j'savais pas comment faire **j'étais perdue** et pourtant j'avais pris mes médicaments j'suis dans un état vous voyez même ma bouche elle est sèche j'suis dans un triste état
- V83** Vous êtes quand même bien (↑)
- G84** J'pense que ma tête est bien mais on croirait à moitié (↓) la moitié qui va et la moitié qui va pas j'ai l'impression de ça vous voyez (↑)
- V85** D'accord
- G86** Ou alors c'est la conscience peut être la conscience est ce que c'est ça (↑)
- V87** Vous savez **ça arrive à tout le monde d'avoir des moments biens et des moments où on est perdu**
- G88** **Oui j'ai peur de perdre tout le monde**
- V89** Mais ils vont plutôt bien vos enfants (↑)
- G90** Ils ont l'air ils ont l'air mais ils ont des allergies ils ont (→) mon petit fils il s'est cassé le bras à l'école tout ça

## Remontée à travers la structure



# Plan

- 1 Contexte : SLAM
- 2 Traitements automatiques
- 3 Représentation formelle
- 4 Conclusion
  - Le problème de l'anonymisation
  - Conclusion

# Anonymisation traditionnelle

- Identification et substitution des entités nommées
  - outil performant mais non opérationnel  
[Grouin and Zweigenbaum, 2013]
  - scripts Python
- Intervention humaine importante
  - 10 catégories
  - identification par les majuscules
  - vérification manuelle

spk1 Donc vous habitez sur Ville1 ↑

spk2 Ville2

spk1 Ville2 D'accord alors moi qui suis pas du tout d'ici, c'est pas très loin

- Capacité d'ajouter des "bip" sonores sur la bande son

spk1 Et donc euh j' avais j' ai pendant trois trois quatre ans **j' avais commencé des études** j' ai fait un peu différentes choses parce que

...

spk1 Euh **dans une école d' ingénieur à Ville1 dans dans le nord** euhh et donc euhh euhh ouais donc j' ai je c' est là où j' ai commencé à être malade en fait juste [...]

spk1 donc du coup ben là c' est je j' ai j' ai repéré deux trois le le c' était quand même assez stressant euh **la la prépa**

spk2 Mmh mmh

spk1 donc euh donc du coup ouais euh et bon pour euh en ce qui concerne les études donc du coup après j' ai j' ai arrêté le le le l' école d' ingénieur enfin la prépa **je suis revenue à Ville2**

spk2 Mmh mmh

spk1 **j' ai fait euh une une une fac de de maths** je suis allé en fac de maths

spk1 à à avoir des délires de persécution tout ça j' ai commencé à à penser à la schizophrénie Euhh mais bon en même temps **juste avant de au lycée je je faisais quand même une grosse dépression**

spk1 Donc euh et donc euh du coup euhh ouais donc euh **à Ville2** pareil y avait encore la la dépression qui s' installait euh j' étais dans **un appart en fait j' étais place Lieux3**

...

spk1 et c'était très très gênant

spk2 Ben **c'est le centre de Ville2**

# Impossibilité de l'anonymisation

- Tâche avec faible contexte : *randomiser* les tours de paroles
- Impossibilité d'anonymiser l'histoire et la géographie
  - limiter le nombre d'intervenants sur la ressource : analyses sémantico-pragmatiques, bandes sons, etc.

# Conclusion

- Production d'une ressource normalisée, riche en méta-données malgré les **questions éthiques**
- Usage pathologique des disfluences chez les sujets schizophrènes par des outils issus du TAL.
- Formalisation des conversations pathologiques avec la SDRT
  
- Respecter l'anonymat des personnes sans renoncer aux enjeux scientifiques
- Identification des conséquences des conclusions
- Corréler ces indices avec d'autres mesures (résultats aux tests psycho-cognitifs, mesures oculométriques et EEG)



Amblard, M. and Fort, K. (2014).

Étude quantitative des disfluences dans le discours de schizophrènes : automatiser pour limiter les biais.

In *TALN - Traitement Automatique des Langues Naturelles*, pages 292–303, Marseille, France.



Benzitoun, C., Fort, K., and Sagot, B. (2012).

TCOF-POS : un corpus libre de français parlé annoté en morphosyntaxe.

In *Traitement Automatique des Langues Naturelles (TALN)*, pages 99–112, Grenoble, France.



Chaika, E. (1974).

A linguist looks at “schizophrenic” language.

*Brain and Language*, 1(3) :257–276.



Constant, M. and Dister, A. (2010).

Automatic detection of disfluencies in speech transcriptions.

In Pettorino, M., Giannini, A., Chiari, I., and Dovetto, F., editors, *Spoken Communication*, volume 1, pages 259–272. Cambridge Scholars Publishing.



de Mareüil, P. B., Adda, G., Adda-Decker, M., Barras, C., Habert, B., and Paroubek, P. (2013).

Une étude quantitative des marqueurs discursifs, disfluences et chevauchements de parole dans des interviews politiques.

*TIPA. Travaux interdisciplinaires sur la parole et le langage [En ligne]*, 29.

mis en ligne le 19 décembre 2013, consulté le 14 février 2014.



Fromkin, V. A. (1975).

A linguist looks at “a linguist looks at ‘schizophrenic language’”.

*Brain and Language*, 2(0) :498 – 503.



Grouin, C. and Zweigenbaum, P. (2013).

Automatic de-identification of french clinical records : Comparison of rule-based and machine-learning approaches.

In *Stud Health Technol Inform, Proc of MEDINFO*, volume 192, pages 476–80, Copenhagen, Denmark.



Musiol, M. and Trognon, A. (1996).

L'accomplissement interactionnel du trouble schizophrénique.  
*Raisons Pratiques* 7, pages 179–209.



Rebuschi, M., Amblard, M., and Musiol, M. (2014).

Using SDRT to analyze pathological conversations. Logicity, rationality and pragmatic deviances.

In *Interdisciplinary Works in Logic, Epistemology, Psychology and Linguistics : Dialogue, Rationality, and Formalism*, Logic, Argumentation & Reasoning, pages 343–368. Springer.



Verhaegen, F. (2007).

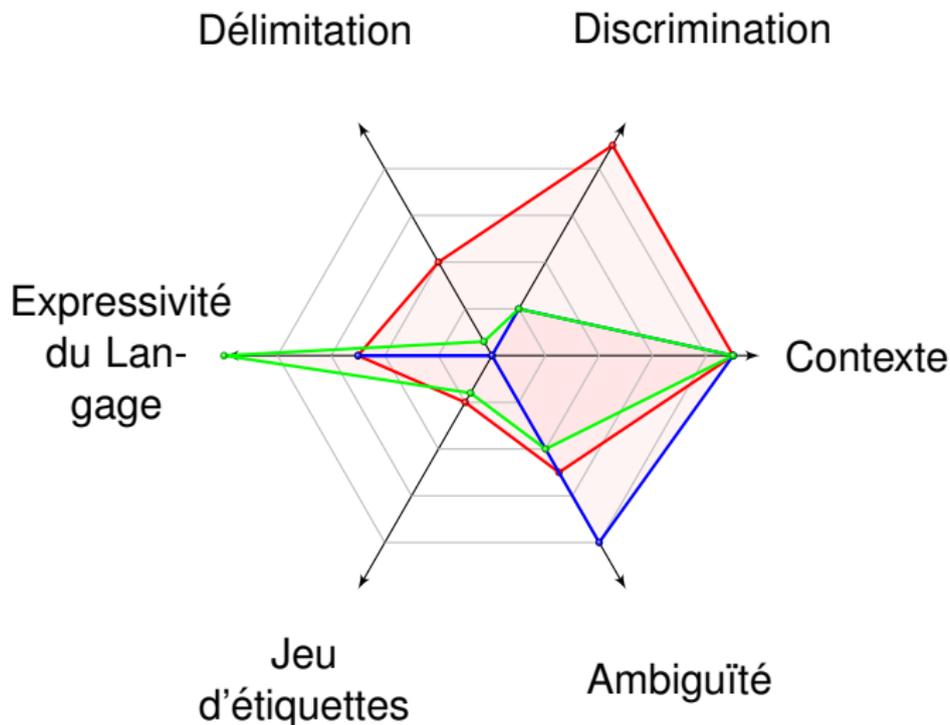
*Psychopathologie cognitive des processus intentionnels schizophréniques dans l'interaction verbale.*

PhD thesis, Université Nancy 2, France.

# Modéliser une campagne d'annotation

- Éviter les biais
- Définir les mesures des accords inter-annotateurs
- Identifier les outils techniques (et donc la documentation et l'accès à la documentation)
- Prévoir les coûts
- Définir la structure des campagnes

# Complexité de l'annotation



Discontinuités

Syntaxe

S-DRT