

Sémantique : modélisation formelle et interprétation de données empiriques.

Maxime Amblard

21 octobre 2016
Séminaire Atifl

LORIA - UMR 7503
Université de Lorraine

MSH-Lorraine - USR 3261
Inria Nancy Grand Est

CNRS

Plan

- 1 Sémantique : éléments en faveur de la logique
- 2 Dynamique et discours
- 3 Données empiriques
- 4 Conclusion et Perspectives

Plan

- 1 Sémantique : éléments en faveur de la logique
- 2 Dynamique et discours
- 3 Données empiriques
- 4 Conclusion et Perspectives

Contexte

Modélisation de la langue naturelle :

- phonologie, morphologie, syntaxe, sémantique, pragmatique

Contexte

Modélisation de la langue naturelle :

- phonologie, morphologie, syntaxe, **sémantique**, pragmatique

Contexte

Modélisation de la langue naturelle :

- phonologie, morphologie, syntaxe, **sémantique**, pragmatique

- ca veut dire quoi sémantique ?

Contexte

Modélisation de la langue naturelle :

- phonologie, morphologie, syntaxe, **sémantique**, pragmatique

- ca veut dire quoi sémantique ?

liens entre :

LANGUAGE < - - - - - > WORLD

Des sémantiques

- sémantique lexicale (relations synonymie, antonymie, etc.)
- sémantique compositionnelle
- phénomènes sémantiques

Emmon Bach : Natural Languages Metaphysics

Des sémantiques

- sémantique lexicale (relations synonymie, antonymie, etc.)
explorer les relations structurelles entre les mots
- sémantique compositionnelle
- phénomènes sémantiques
Emmon Bach : Natural Languages Metaphysics

Des sémantiques

- sémantique lexicale (relations synonymie, antonymie, etc.)
explorer les relations structurelles entre les mots
- sémantique compositionnelle
comment construire les représentations ?
- phénomènes sémantiques
Emmon Bach : Natural Languages Metaphysics

Des sémantiques

- sémantique lexicale (relations synonymie, antonymie, etc.)
explorer les relations structurelles entre les mots
- sémantique compositionnelle
comment construire les représentations ?
- phénomènes sémantiques
Emmon Bach : Natural Languages Metaphysics
décrire la langue du point de vue de la réalité du monde

Des sémantiques

- sémantique lexicale (relations synonymie, antonymie, etc.)
explorer les relations structurelles entre les mots
- sémantique compositionnelle
comment construire les représentations ?
- phénomènes sémantiques
Emmon Bach : Natural Languages Metaphysics
décrire la langue du point de vue de la réalité du monde

Leibniz



Le seul moyen de rectifier notre raisonnement est de le rendre **aussi tangible que celui des mathématiciens**, ainsi il devient aisé de trouver nos erreurs, et en cas de désaccord, il suffit de calculer pour voir qui a raison

Frege, Peano, Russell, ...



- inspiration mathématique, grâce à la possibilité d'un calcul précis
- la vue idéale de Leibniz est partiellement réalisée à partir de la fin du 19^{ième} avec les travaux de Frege, Peano, Russell, etc.
- notations formelles + règles de manipulation = logique formelle
- utilisation de la logique du premier ordre (LPO)

LPO adéquate pour la sémantique ?

LPO adéquate pour la sémantique ?

- “non”. Après tout, de nombreuses théories ont été proposées (DRT, RST, DPL, *etc*)

LPO adéquate pour la sémantique ?

- “non”. Après tout, de nombreuses théories ont été proposées (DRT, RST, DPL, *etc*)
- “oui”, c’est un bon point de départ.

LPO adéquate pour la sémantique ?

- “non”. Après tout, de nombreuses théories ont été proposées (DRT, RST, DPL, *etc*)
- “oui”, c’est un bon point de départ.

Pourquoi ?

LPO adéquate pour la sémantique ?

- “non”. Après tout, de nombreuses théories ont été proposées (DRT, RST, DPL, *etc*)
- “oui”, c’est un bon point de départ.

Pourquoi ?

- beaucoup d’autres formalismes sont des **sous-logiques** ou **divergent** sur les notations de la **logique du premier ordre**.

LPO adéquate pour la sémantique ?

- “non”. Après tout, de nombreuses théories ont été proposées (DRT, RST, DPL, *etc*)
- “oui”, c’est un bon point de départ.

Pourquoi ?

- beaucoup d’autres formalismes sont des **sous-logiques** ou **divergent** sur les notations de la **logique du premier ordre**.
- **beaucoup d’outils informatiques** (theorem provers, model builders, model checkers) existent pour travailler avec la LPO

LPO adéquate pour la sémantique ?

- “non”. Après tout, de nombreuses théories ont été proposées (DRT, RST, DPL, *etc*)
- “oui”, c’est un bon point de départ.

Pourquoi ?

- beaucoup d’autres formalismes sont des **sous-logiques** ou **divergent** sur les notations de la **logique du premier ordre**.
- **beaucoup d’outils informatiques** (theorem provers, model builders, model checkers) existent pour travailler avec la LPO
- mais surtout, la LPO permet de **parler de tout**
(temps, modalités, pluriel, événements,...)

Logique du premier ordre

- 1 symboles du vocabulaire (les symboles **non-logique** du langage).
- 2 **variables** x, y, z, w, \dots
- 3 des opérateurs (booléens)
 - \neg (**negation**),
 - \rightarrow (**implication**),
 - \vee (**disjunction**),
 - \wedge (**conjunction**).
- 4 les quantificateurs
 - \forall (**universel**),
 - \exists (**existential**).
- 5 le symbole d'**égalité** $=$
- 6 les **parenthèses** $)$ et $($ et le point $'$

Le tournant de la sémantique

Autour des années 1930, la vision syntaxique est étendue avec le développement de la théorie des modèles et le tournant sémantique



Tarski (logicien polonais) émigré aux états-unis en 1939 : définition célèbre de la **satisfaction** et introduction de la **théorie des modèles**

Une théorie est valide si on peut définir un univers dans laquelle est vraie

⇒ introduction de la **notion de vérité**

Donc transformer une phrase S en langage naturelle en une formule ϕ fournit une interprétation du sens de S .



Objectifs

- représenter le sens des énoncés grâce à des formules logiques (proposition, premier ordre, classiques, intuitionistes, ...)
- développer des algorithmes pour produire des représentations logiques et utiliser ces représentations
- techniques fondamentales pour construire des représentations sémantiques : λ -calcul

Plusieurs difficultés

- Pourquoi utiliser des représentations ? Pourquoi ne pas utiliser directement la langue naturelle ?

Plusieurs difficultés

- Pourquoi utiliser des représentations ? Pourquoi ne pas utiliser directement la langue naturelle ?
- Est-ce qu'elles peuvent être utiles ? pour la pragmatique ?

Plusieurs difficultés

- Pourquoi utiliser des représentations ? Pourquoi ne pas utiliser directement la langue naturelle ?
- Est-ce qu'elles peuvent être utiles ? pour la pragmatique ?
- Est-ce que ces représentations ont une réalité philosophique, cognitive ou conceptuelle ?
- Ou est-ce seulement un moyen de jouer avec des symboles ?

Plusieurs difficultés

- Pourquoi utiliser des représentations ? Pourquoi ne pas utiliser directement la langue naturelle ?
- Est-ce qu'elles peuvent être utiles ? pour la pragmatique ?
- Est-ce que ces représentations ont une réalité philosophique, cognitive ou conceptuelle ?
- Ou est-ce seulement un moyen de jouer avec des symboles ?
- Est-ce si pratique ? Le raisonnement logique est mathématiquement difficile.

Première étape vers la construction sémantique



Principe de **compositionnalité** de Frege :
*Le sens d'un énoncé est une fonction du sens
de ses parties.*

Première étape vers la construction sémantique



Principe de **compositionnalité** de Frege :
*Le sens d'un énoncé est une fonction du sens
de ses parties.*

- items lexicaux = représentation logique
- sémantique synchronisée sur la syntaxe

Première étape vers la construction sémantique



Principe de **compositionnalité** de Frege :
*Le sens d'un énoncé est une fonction du sens
 de ses parties.*

- items lexicaux = représentation logique
- sémantique synchronisée sur la syntaxe



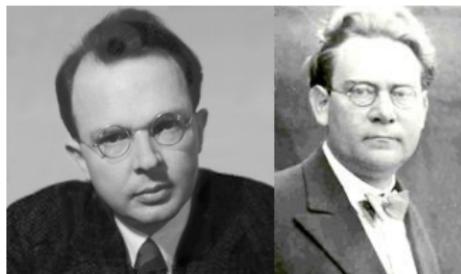
Historiquement récent

Long consensus pour penser que le lien était purement **analogique**
(Frege et Tarski très septiques de l'usage de logique pour les langues)

Historiquement récent

Long consensus pour penser que le lien était purement **analogique**
(Frege et Tarski très septiques de l'usage de logique pour les langues)

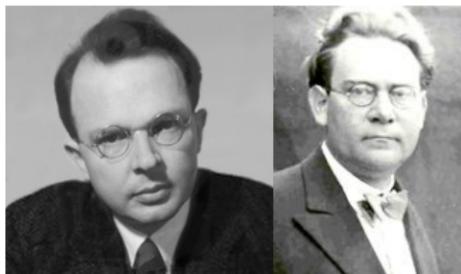
1930-1950 : certains logiciens ont pris ce lien au sérieux : Carnap
(modalités) et Reichenbach (temps)



Historiquement récent

Long consensus pour penser que le lien était purement **analogique** (Frege et Tarski très septiques de l'usage de logique pour les langues)

1930-1950 : certains logiciens ont pris ce lien au sérieux : Carnap (modalités) et Reichenbach (temps)

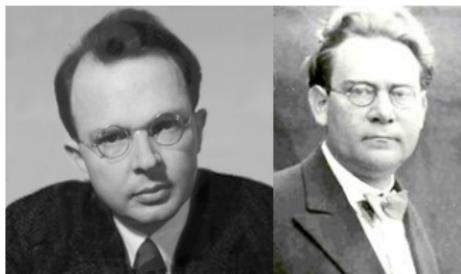


Mais **vision abstraite** et **non calculatoire** (algorithmique) de la logique

Historiquement récent

Long consensus pour penser que le lien était purement **analogique** (Frege et Tarski très septiques de l'usage de logique pour les langues)

1930-1950 : certains logiciens ont pris ce lien au sérieux : Carnap (modalités) et Reichenbach (temps)

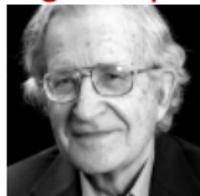


Mais **vision abstraite** et **non calculatoire** (algorithmique) de la logique

50 - 60 : nombreux philosophes se sont positionnés contre une vision partagée de la logique et de la langue naturelle

De Chomski à Montague

Linguistique computationnelle



En 1957, Chomski propose une formalisation de la syntaxe des langues

Sémantique computationnelle



- English as a Formal Language
- The Proper Treatment of Quantification in Ordinary English
- Universal Grammar

Le travail de Montague

- logique intentionnelle
- quantificateurs généralisés (most, few, three, ...)
- premier modèle pour l'ambiguïté de portée des quantificateurs
- définition d'une interface syntaxe-sémantique rigoureuse

⇒ Il remplace l'analogie par l'algorithmique

Représentation sémantique à la van benthem

- 1 définition d'un **fragment de l'anglais** avec les grammaires catégorielles
- 2 spécification du sens des items lexicaux avec des termes du **λ -calcul**
- 3 construction des **représentations sémantiques** avec l'application fonctionnelle et la β -réduction

1 minute pour le λ -calcul

Vision fonctionnelle du calcul :

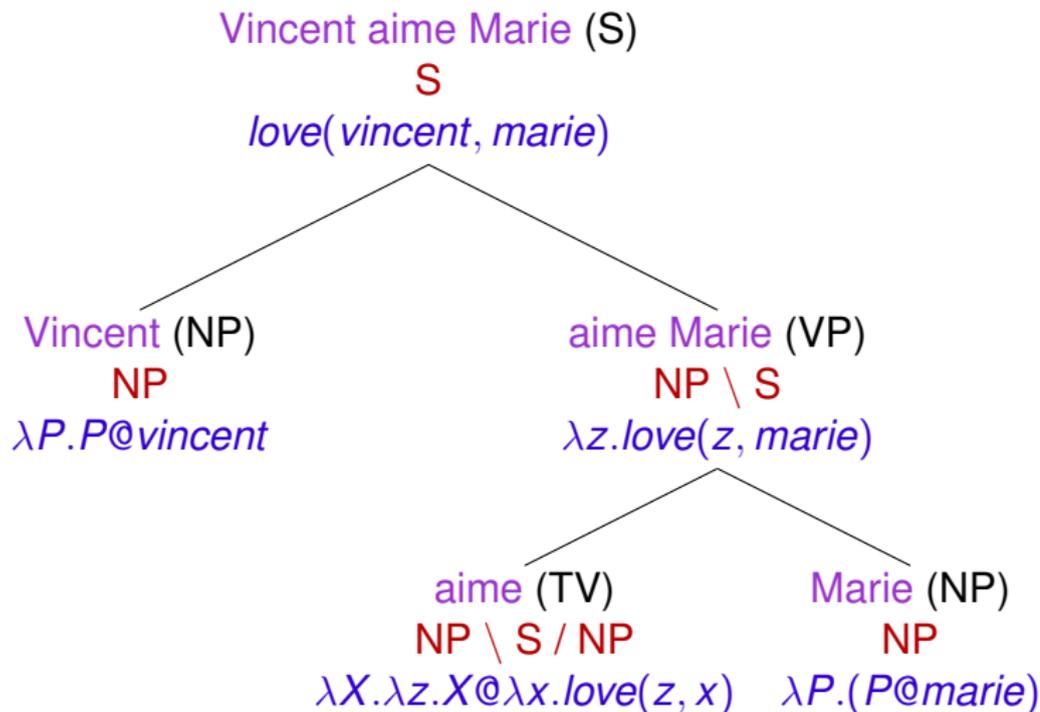
- les variables sont liées par l'opérateur λ

$$\lambda x.man(x)$$

- elles sont des marqueurs dans des formules
- deux termes sont composés par l'application fonctionnelle
- la β -conversion, l' α -conversion et l' η -expansion réalisent le calcul

$$((\lambda x.man(x))@(vincent)) \rightsquigarrow man(Vincent)$$

[Task3] Isomorphisme de Curry-Howard



Mais, il reste beaucoup à faire

- modélisation des noms propres

Vincent vs $\lambda P.P@Vincent$

Mais, il reste beaucoup à faire

- modélisation des noms propres
- modélisation des pronoms

Vincent vs $\lambda P.P@Vincent$

x

Mais, il reste beaucoup à faire

- modélisation des noms propres
- modélisation des pronoms
- ambiguïté de portés des quantificateurs

Vincent vs $\lambda P.P@Vincent$

x

Tout boxer aime une femme

Mais, il reste beaucoup à faire

- modélisation des noms propres

Vincent vs $\lambda P.P@Vincent$

- modélisation des pronoms

x

- ambiguïté de portés des quantificateurs

Tout boxer aime une femme

- ellipse verbale

Marie est allée à la fête et Vincent aussi

Mais, il reste beaucoup à faire

- modélisation des noms propres

Vincent vs $\lambda P.P@Vincent$

- modélisation des pronoms

x

- ambiguïté de portés des quantificateurs

Tout boxer aime une femme

- ellipse verbale

Marie est allée à la fête et Vincent aussi

- gapping

Marie aime Vincent, et Honey-Bunny Pumpkin

Mais, il reste beaucoup à faire

- modélisation des noms propres

Vincent vs $\lambda P.P@Vincent$

- modélisation des pronoms

x

- ambiguïté de portés des quantificateurs

Tout boxer aime une femme

- ellipse verbale

Marie est allée à la fête et Vincent aussi

- gapping

Marie aime Vincent, et Honey-Bunny Pumpkin

- presupposition

Mais, il reste beaucoup à faire

- modélisation des noms propres

Vincent vs $\lambda P.P@Vincent$

- modélisation des pronoms

x

- ambiguïté de portés des quantificateurs

Tout boxer aime une femme

- ellipse verbale

Marie est allée à la fête et Vincent aussi

- gapping

Marie aime Vincent, et Honey-Bunny Pumpkin

- presupposition

- phénomène entre les phrases

Mais, il reste beaucoup à faire

- modélisation des noms propres

Vincent vs $\lambda P.P@Vincent$

- modélisation des pronoms

x

- ambiguïté de portés des quantificateurs

Tout boxer aime une femme

- ellipse verbale

Marie est allée à la fête et Vincent aussi

- gapping

Marie aime Vincent, et Honey-Bunny Pumpkin

- presupposition

- phénomène entre les phrases

- ...

Et de l'autre côté du diagramme



On ajoute des informations :

- time, tense and aspect : logique du temps à la Allen, Reichenbach, van Benthem
- événements : réification à la Davidson
- pluriel
- modalités
- ...

Plan

- 1 Sémantique : éléments en faveur de la logique
- 2 Dynamique et discours**
- 3 Données empiriques
- 4 Conclusion et Perspectives

Limites des approches montageovienne

- donkey sentences

Every farmer who owns a donkey beats it

$$(\exists x \exists y. (\text{farmer } x \wedge \text{donkey } y \wedge \text{own } x \ y)) \rightarrow \text{beat } x \ y$$

- résolution d'anaphores inter-phrastiques

A man walks in the park. He whistle.

$$\exists x. (\text{man } x \wedge \text{walk_in_the_park } x) \wedge (\text{whistle } x)$$

Sémantique dynamique



- **Context Change Potential (CCP)** [Heim1983]
L'interprétation se fait en **fonction du contexte** et
le contexte est **modifié par l'interprétation**

Sémantique dynamique



- **Context Change Potential (CCP)** [Heim1983]
L'interprétation se fait en **fonction du contexte** et le contexte est **modifié par l'interprétation**



- **Discourse Representation Theory (DRT)**
[Kamp1981]/**File Change Semantics (FCS)**
[Heim1982]
niveaux intermédiaires entre représentation et valeurs de vérités

Sémantique dynamique



- **Context Change Potential (CCP)** [Heim1983]
L'interprétation se fait en **fonction du contexte** et le contexte est **modifié par l'interprétation**



- **Discourse Representation Theory (DRT)**
[Kamp1981]/**File Change Semantics (FCS)**
[Heim1982]
niveaux intermédiaires entre représentation et valeurs de vérités



- **Dynamic Predicate Logic (DPL)**
[Groenendijk1991]

Type Theoretic Dynamic Logic



- **Type Theoretic Dynamic Logic (TTDL)** [deGroote2006] : purement montagovien pour le discours avec dynamique par les continuations dans le λ -calcul

Type Theoretic Dynamic Logic



- **Type Theoretic Dynamic Logic (TTDL)** [deGroote2006] : purement montagovien pour le discours avec dynamique par les continuations dans le λ -calcul

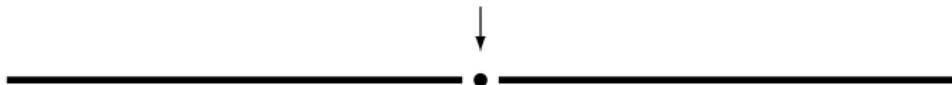
- types primitifs
 - ι : individu / entité
 - o : proposition / valeur de vérité
 - γ : contexte gauche

Type Theoretic Dynamic Logic



- **Type Theoretic Dynamic Logic (TTDL)** [deGroote2006] : purement montagovien pour le discours avec dynamique par les continuations dans le λ -calcul

- types primitifs
 - ι : individu / entité
 - o : proposition / valeur de vérité
 - γ : contexte gauche



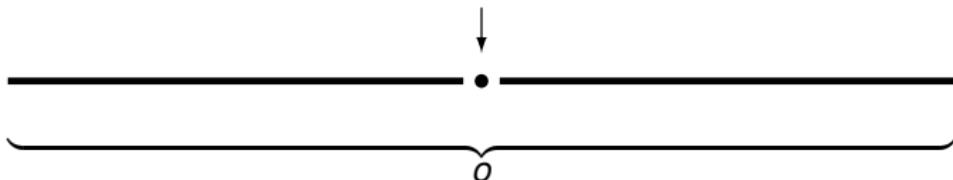
Type Theoretic Dynamic Logic



- **Type Theoretic Dynamic Logic (TTDL)** [deGroote2006] : purement montagovien pour le discours avec dynamicit  par les continuations dans le λ -calcul

- types primitifs

- ι : individu / entit 
- o : proposition / valeur de v rit 
- γ : contexte gauche



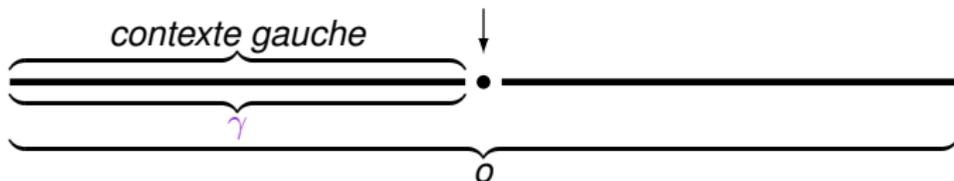
Type Theoretic Dynamic Logic



- **Type Theoretic Dynamic Logic (TTDL)** [deGroote2006] : purement montagovien pour le discours avec dynamicit  par les continuations dans le λ -calcul

- types primitifs

- ι : individu / entit 
- o : proposition / valeur de v rit 
- γ : contexte gauche



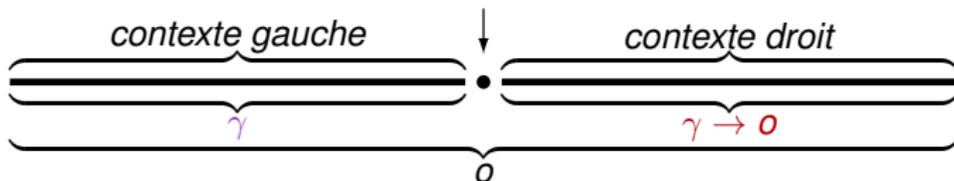
Type Theoretic Dynamic Logic



- **Type Theoretic Dynamic Logic (TTDL)** [deGroote2006] : purement montagovien pour le discours avec dynamicit  par les continuations dans le λ -calcul

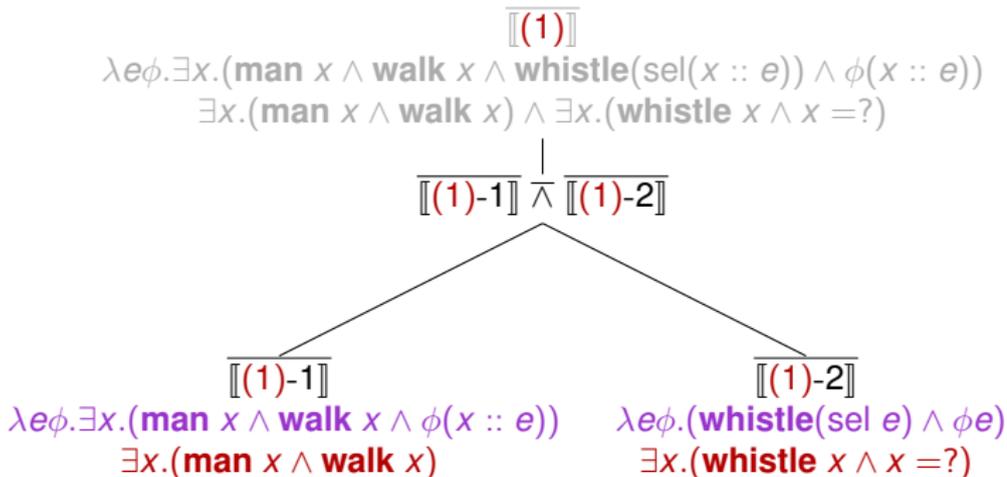
- types primitifs

- ι : individu / entit 
- o : proposition / valeur de v rit 
- γ : contexte gauche



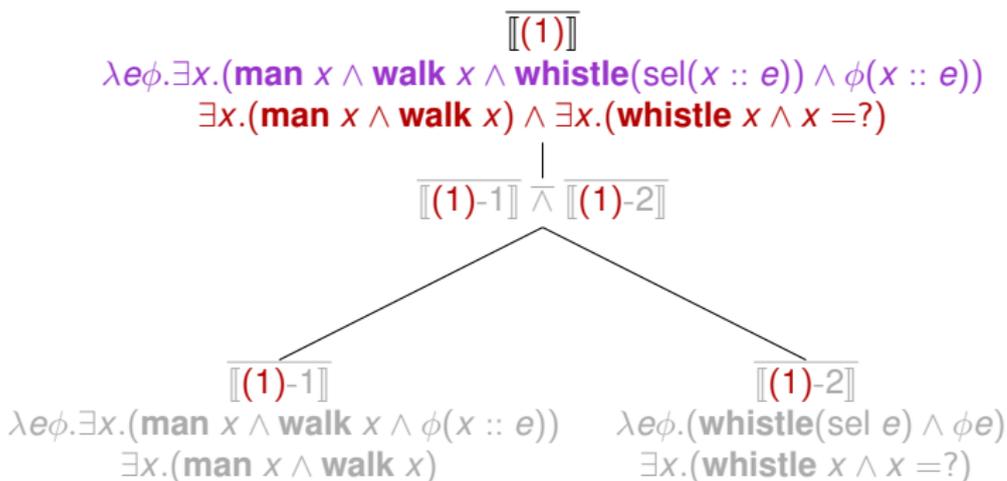
Anaphore discursive en TTDL *cont.*

(1) A man_j walks in the park. He_j whistles.



Anaphore discursive en TTDL *cont.*

(1) A man_i walks in the park. He_i whistles.



Plan

- 1 Sémantique : éléments en faveur de la logique
- 2 Dynamique et discours
- 3 Données empiriques**
- 4 Conclusion et Perspectives

SLAM - Schizophrénie et Language :

Analyse et modélisation

Motivation

- étude linguistique de la pathologie mentale (Chaika 1974) et (Fromkin 1975)
- *discontinuités pragmatiques* dans l'accomplissement de l'interaction verbale, (Musiol et Trognon 1996)
- usage pathologique de la planification du discours chez les schizophrènes paranoïdes, (Verhaegen 2007)

SLAM - Schizophrénie et Language :

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 Language :

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, Amblard et Musiol 2014) : des corrélations explicites
- (Amblard et Fort 2014) : usage pathologique des disfluences

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)

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

31 575 tours de parole / 375 000 mots

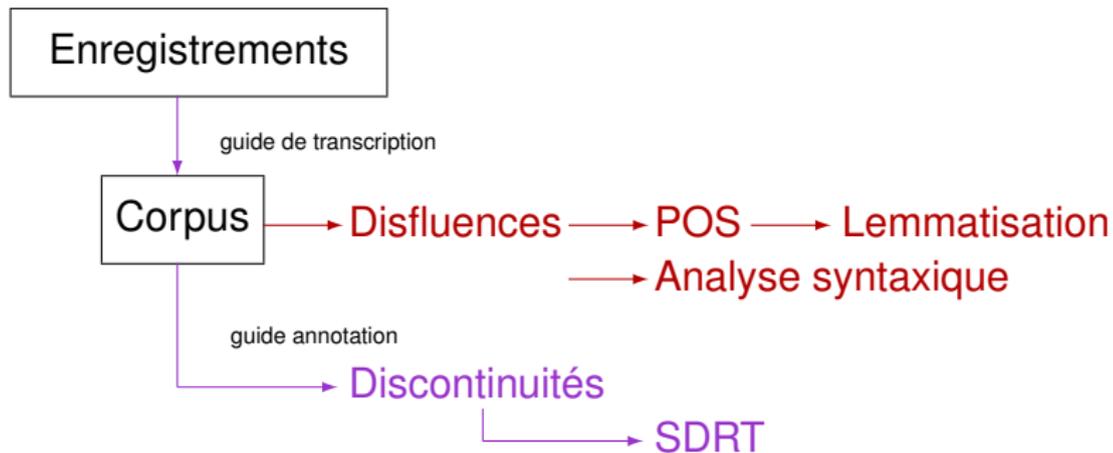
	corpus Ville1				corpus Ville2			
	nb tours de parole		nb de mots		nb tours de parole		nb de mots	
<i>S</i>	3 863	} 11 145	46 859	} 119 762	4 062	} 4 433	66 725	} 79 081
<i>T</i>	7 282		72 903		371		12 356	
<i>P + S</i>	3 819	} 11 517	30 293	} 138 571	4 098	} 4 480	33 686	} 37 842
<i>P + T</i>	7 698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

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.
- participation des patients (pertes importantes de participation)

Contexte du projet SLAM



Traitements automatiques

- corpus de grande taille
 - limiter les interventions humaines pour identifier automatiquement :
- disfluences dans les transcriptions avec `Distagger` (Constant et Dister 2010)
- POS et lemmes avec `MELt` [Denis et Sagot, 2012]

Pourquoi les disfluences ?

- pour étudier leurs usages (conventionnel vs pathologique)
- pour reconstruire des tours de paroles syntaxiquement plus cohérents

Distagger

f-score : 95,5 %, précision : 95,3 %, rappel : 95,8 % (Constant et Dister 2010)

❶ 'euh'

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

❷ Répétitions

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

❸ Autocorrections

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

❹ Amorces (interruption de morphèmes)

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

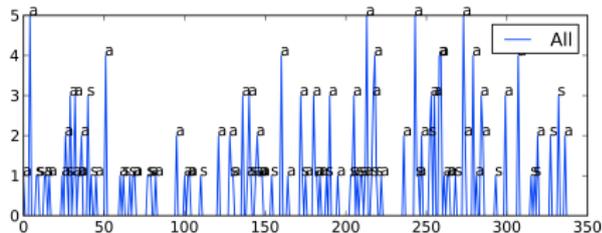
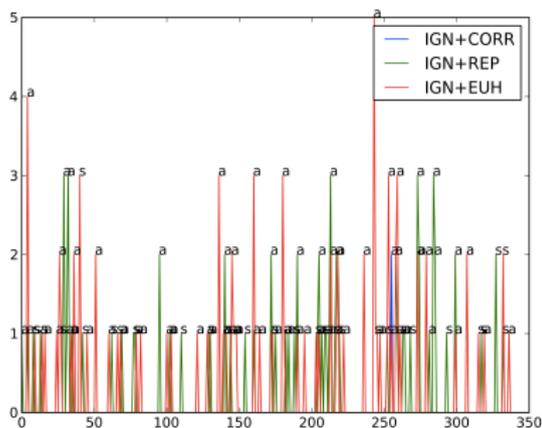
Distagger

- étiquettes $\{IGN+EUH\}, \{IGN+REP\}, \{IGN+CORR\}$
- scripts Python (pré- et post-traitements)
 - ~ 1 500 lignes de code
- certaines annotations sont écartées :
 - 'mmh mmh', 'oui oui', 'non non', 'vous vous'*
- 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)

Visualisation des résultats

par entretien

Répartition des disfluences dans un entretien



Résultats (% de disfluences)

	S	T	S+T	P+S	P+T	P
Corpus Ville2						
par tours 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 tours 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

Résultats (% de disfluences)

	S	T	S+T	P+S	P+T	P
Corpus Ville2						
par tours 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 tours 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

	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

Valeurs significatives : > 1,96

POS/Lemme : pas de comportement significatif

Cat	# cat	par mots	par speech
Schizos	23.667	0.011	0.132
Control	24.217	0.009	0.083
Pst (S+C)	23.975	0.009	0.104

POS/Lemme : pas de comportement significatif

Cat	# cat	par mots	par speech
Schizos	23.667	0.011	0.132
Control	24.217	0.009	0.083
Pst (S+C)	23.975	0.009	0.104

Form	# form	par mots	par speech
Schizos	521.0	0.21	2.672
Control	628.0	0.194	2.053
Pst (S+C)	581.0	0.201	2.324

POS/Lemme : pas de comportement significatif

Cat	# cat	par mots	par speech
Schizos	23.667	0.011	0.132
Control	24.217	0.009	0.083
Pst (S+C)	23.975	0.009	0.104

Form	# form	par mots	par speech
Schizos	521.0	0.21	2.672
Control	628.0	0.194	2.053
Pst (S+C)	581.0	0.201	2.324

Lemmas	# lem	par mots	par speech
Schizos	409.0	0.166	2.103
Control	491.0	0.153	1.611
Pst (S+C)	455.0	0.158	1.827

Biais

- Différences entre les sous-corpus (transcription différente)
⇒ pas d'évaluation qualitative globale
 - `Distagger` sur le corpus de parole spontanée TCOF-POS (Benzitoun, Fort et Sagot 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

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 :
déni 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)

Sujet normal (à la 3e personne)	Schizophrène (à la 1e personne)
correction pragmatique ↓	déviante pragmatique ↑
déviante sémantique	correction sémantique
contenu contradictoire : <i>apparence de contradiction</i>	contenu cohérent : <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)

Sujet normal (à la 3e personne)	Schizophrène (à la 1e personne)
correction pragmatique ↓	déviante pragmatique ↑
déviante sémantique	correction sémantique
contenu contradictoire : <i>apparence de contradiction</i>	contenu cohérent : <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]

SDRT [Asher & Lascarides 2003] + îlots sémantiques

Guy experienced a lovely evening last night

Elaboration

He had a fantastic meal

Elaboration

Elaboration

He ate salmon

Narration

He devoured
lots of cheese

SDRT [Asher & Lascarides 2003] + îlots sémantiques

Guy experienced a lovely evening last night

Elaboration

He had a fantastic meal

Elaboration

Elaboration

He ate salmon

Narration

He devoured
lots of cheese

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.
Principe de Charité de Quine fort.

Spécificité des conversations schizophrènes

Conjecture 1 :

Les schizophrènes sont logiquement consistants.

Principe de Charité de Quine fort.

Les ruptures sont 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.

Principe de Charité de Quine fort.

Les ruptures sont 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.

Principe de Charité de Quine fort.

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

Conjecture 2 :

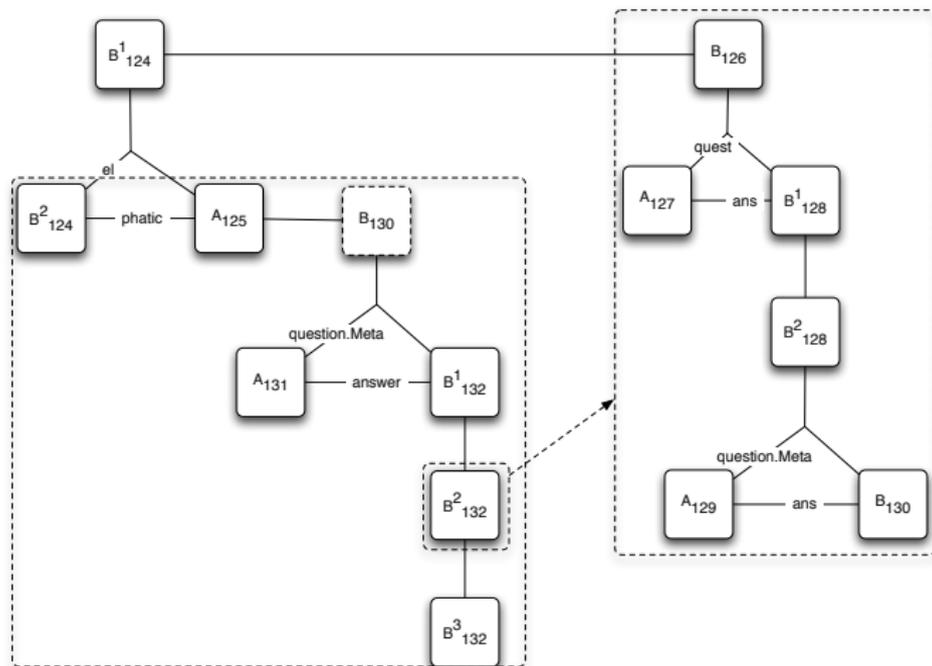
La sous-spécification (ambiguïté) est centrale dans la rupture

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

Phonologique, morphologique, lexical, référent de discours ...

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



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.

Impossibilité de l'anonymisation

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**

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
(comme celles des outils du TAL)

Plan

- 1 Sémantique : éléments en faveur de la logique
- 2 Dynamique et discours
- 3 Données empiriques
- 4 Conclusion et Perspectives**

Conclusion

- retour sur les usages de la logique pour la sémantique
- introduction de la dynamicité par les continuations
- gestion de l'accessibilité des référents de discours (double négation)
- cadre empirique (SLAM)
 - 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

Perspectives

- Évolution SLAM
 - Affiner l'analyse du dysfonctionnement, en ouvrant vers une interprétation cognitive.
 - Augmenter la couverture du corpus en volume et en nombre de pathologies analysées
 - Augmenter les phénomènes analysés dans l'outil
- Réconcilier les deux axes :
 - Construction de grammaire sémantique à large couverture
 - définition d'un formalisme basé sur TTDL inspiré de la SDRT pour rendre compte de l'interaction dialogique

Disjonction en DN-TTDL

?? Either Jones does **not** own a car_{*j*} or he hides it_{*j*}.

$$\begin{aligned} \overline{\overline{[[??-1]]}} &= \overline{\overline{[[\text{own}]]([\text{a}][\text{car}]][\text{Jones}]})} \\ &\rightarrow_{\beta} \langle \lambda e \phi. (\neg(\exists x. (\text{car } x \wedge \text{own jones } x)) \wedge \phi e), \\ &\quad \lambda e \phi. (\exists x. (\text{car } x \wedge \text{own jones } x \wedge \phi(x :: e))) \rangle \end{aligned}$$

Disjonction en DN-TTDL

?? Either Jones does not own a car_{*i*} or he hides it_{*i*}.

$$\begin{aligned} \overline{\overline{[[??-1]]}} &= \overline{\overline{[[own]]([a][car])[Jones]}} \\ &\rightarrow_{\beta} \langle \lambda e \phi. (\neg(\exists x. (\mathbf{car} \ x \wedge \mathbf{own} \ \mathbf{jones} \ x)) \wedge \phi e), \\ &\quad \lambda e \phi. (\exists x. (\mathbf{car} \ x \wedge \mathbf{own} \ \mathbf{jones} \ x \wedge \phi(x :: e))) \rangle \end{aligned}$$

$$\begin{aligned} \overline{\overline{[[??-2]]}} &= \overline{\overline{[[hide]]([it])[Jones]}} \\ &\rightarrow_{\beta} \langle \lambda e \phi. \mathbf{hide} \ \mathbf{jones} \ (\mathbf{sel} \ e) \wedge \phi e, \lambda e \phi. \neg \mathbf{hide} \ \mathbf{jones} \ (\mathbf{sel} \ e) \wedge \phi e \rangle \end{aligned}$$

Disjonction en DN-TTDL

?? Either Jones does not own a car_{*j*} or he hides it_{*j*}.

$$\begin{aligned} \overline{\overline{[??-1]}} &= \overline{\overline{([own]([a][car])[Jones])}} \\ &\rightarrow_{\beta} \langle \lambda e \phi. (\neg(\exists x. (\mathbf{car} \ x \wedge \mathbf{own} \ \mathbf{jones} \ x)) \wedge \phi e), \\ &\quad \lambda e \phi. (\exists x. (\mathbf{car} \ x \wedge \mathbf{own} \ \mathbf{jones} \ x \wedge \phi(x :: e))) \rangle \end{aligned}$$

$$\begin{aligned} \overline{\overline{[??-2]}} &= \overline{\overline{([hide][it][Jones])}} \\ &\rightarrow_{\beta} \langle \lambda e \phi. \mathbf{hide} \ \mathbf{jones} \ (\mathbf{sel} \ e) \wedge \phi e, \lambda e \phi. \neg \mathbf{hide} \ \mathbf{jones} \ (\mathbf{sel} \ e) \wedge \phi e \rangle \end{aligned}$$

$$\begin{aligned} \overline{\overline{[??]}} &= \overline{\overline{[??-1] \vee [??-2]}} \\ &\rightarrow_{\beta} \langle \lambda e \phi. (\forall x. \mathbf{car} \ x \rightarrow ((\mathbf{own} \ \mathbf{jones} \ x) \rightarrow (\mathbf{hide} \ \mathbf{jones} \ (\mathbf{sel}(x :: e))))) \wedge \phi e, \\ &\quad \lambda e \phi. \exists x. \mathbf{car} \ x \wedge \mathbf{own} \ \mathbf{jones} \ x \wedge (\neg \mathbf{hide} \ \mathbf{jones} \ (\mathbf{sel}(x :: e))) \wedge \phi(x :: e) \rangle \end{aligned}$$

POS / Lemmes

Moyenne sur les 41 entretiens de Ville1

	# cat	# form	# lemm	# lemm/#form
Schizos	23.667	521.0	409.0	0.785
Control	24.217	628.0	491.0	0.781
Pst (S+C)	23.975	581.0	455.0	0.783

POS / Lemmes

Moyenne sur les 41 entretiens de Ville1

	# cat	# form	# lemm	# lemm/#form
Schizos	23.667	521.0	409.0	0.785
Control	24.217	628.0	491.0	0.781
Pst (S+C)	23.975	581.0	455.0	0.783

Rien de particulier...

POS/Lemme

Cat	# cat	par mots	par speech
Schizos	23.667	0.011	0.132
Control	24.217	0.009	0.083
Pst (S+C)	23.975	0.009	0.104

POS/Lemme

Cat	# cat	par mots	par speech
Schizos	23.667	0.011	0.132
Control	24.217	0.009	0.083
Pst (S+C)	23.975	0.009	0.104

Form	# form	par mots	par speech
Schizos	521.0	0.21	2.672
Control	628.0	0.194	2.053
Pst (S+C)	581.0	0.201	2.324

POS/Lemme

Cat	# cat	par mots	par speech
Schizos	23.667	0.011	0.132
Control	24.217	0.009	0.083
Pst (S+C)	23.975	0.009	0.104

Form	# form	par mots	par speech
Schizos	521.0	0.21	2.672
Control	628.0	0.194	2.053
Pst (S+C)	581.0	0.201	2.324

Lemmas	# lem	par mots	par speech
Schizos	409.0	0.166	2.103
Control	491.0	0.153	1.611
Pst (S+C)	455.0	0.158	1.827

Significativité

		Schizo / Témoin	Schizo / Psy	Témoin / Psy
Cat	words	0.698	0.373	-0.324
	speech	1.710	0.860	-0.842
Forms	words	1.496	0.812	-0.689
	speech	-	-	-
Lemmas	words	1.32	0.720	-0.612
	speech	-	-	-

CAT

SCHIZO

(u'V', 6879)
 (u'PONCT', 5905)
 (u'ADV', 5557)
 (u'CLS', 4920)
 (u'NC', 4827)
 (u'P', 3789)
 (u'DET', 3365)
 (u'CC', 2230)
 (u'PRO', 2195)
 (u'ADJ', 1919)
 (u'VPP', 1432)
 (u'CS', 1349)
 (u'VINF', 1289)
 (u'CLO', 1266)

PSY

(u'V', 20750)
 (u'PONCT', 17687)
 (u'ADV', 17629)
 (u'NC', 14248)
 (u'CLS', 13347)
 (u'P', 11230)
 (u'DET', 10161)
 (u'PRO', 7328)
 (u'CC', 7117)
 (u'ADJ', 6555)
 (u'VINF', 4642)
 (u'CS', 4078)
 (u'CLO', 3424)
 (u'VPP', 2948)

Témoïn

(u'V', 10723)
 (u'PONCT', 9416)
 (u'ADV', 8976)
 (u'NC', 7743)
 (u'CLS', 7441)
 (u'P', 5634)
 (u'DET', 5436)
 (u'CC', 3897)
 (u'PRO', 3578)
 (u'ADJ', 3377)
 (u'CS', 2229)
 (u'VINF', 1978)
 (u'CLO', 1884)
 (u'VPP', 1806)

FORM

SCHIZO

(u'...', 2758)
 (u'...', 1658)
 (u'eu'h', 1535)
 (u'/', 1413)
 (u'je', 1367)
 (u'est', 1292)
 (u'c'", 1159)
 (u'j'", 993)
 (u'ça', 983)
 (u'de', 937)
 (u'pas', 812)
 (u'ouais', 611)
 (u'et', 606)
 (u'que', 604)

PSY

(u'...', 9049)
 (u'est', 5650)
 (u'c'", 4863)
 (u'eu'h', 4508)
 (u'...', 4377)
 (u'ça', 3554)
 (u'de', 2724)
 (u'/', 2667)
 (u'et', 2540)
 (u'pas', 2467)
 (u'vous', 1938)
 (u'je', 1754)
 (u'en', 1710)
 (u'que', 1663)

Témoin

(u'...', 4310)
 (u'...', 3090)
 (u'est', 2669)
 (u'c'", 2296)
 (u'eu'h', 2095)
 (u'ça', 1668)
 (u'je', 1593)
 (u'/', 1463)
 (u'de', 1301)
 (u'ouais', 1300)
 (u'pas', 1233)
 (u'et', 980)
 (u'que', 940)
 (u'le', 928)

Lemme

SCHIZO

(u'cln', 3356)
 (u'...', 2758)
 (u'être', 2327)
 (u'.'. , 1658)
 (u'ce', 1583)
 (u'avoir', 1497)
 (u'*euh', 1471)
 (u'le', 1470)
 (u'*/', 1211)
 (u'de', 1099)
 (u'cela', 987)
 (u'un', 961)
 (u'pas', 819)
 (u'que', 770)

PSY

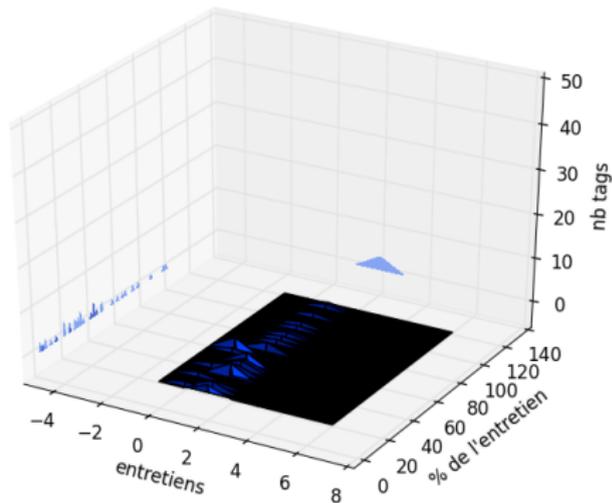
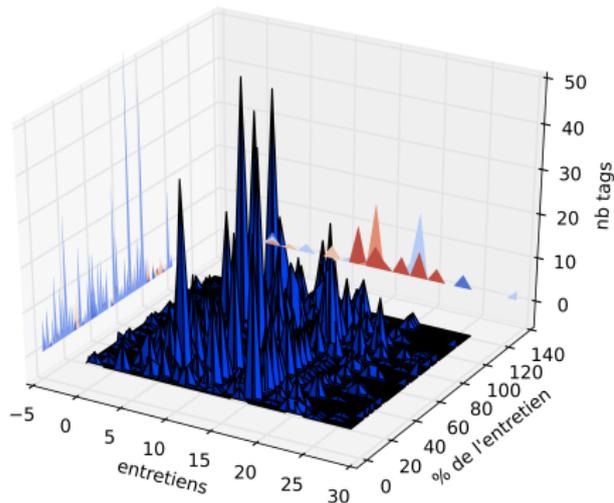
(u'...', 9049)
 (u'être', 7882)
 (u'cln', 6768)
 (u'ce', 6629)
 (u'le', 4943)
 (u'.'. , 4377)
 (u'*euh', 4253)
 (u'avoir', 3601)
 (u'cela', 3554)
 (u'de', 3281)
 (u'un', 2941)
 (u'et', 2913)
 (u'pas', 2474)
 (u'que', 2461)

Témoïn

(u'cln', 4322)
 (u'...', 4310)
 (u'être', 3783)
 (u'.'. , 3090)
 (u'ce', 2909)
 (u'le', 2745)
 (u'avoir', 2112)
 (u'*euh', 1936)
 (u'cela', 1670)
 (u'de', 1531)
 (u'un', 1466)
 (u'*/', 1327)
 (u'que', 1286)
 (u'*ouais', 1277)

Visualisation des résultats

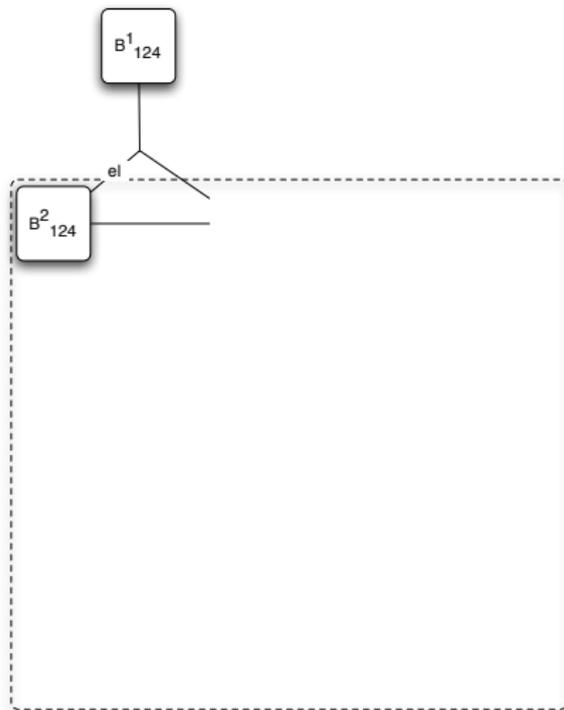
par sous-corpus



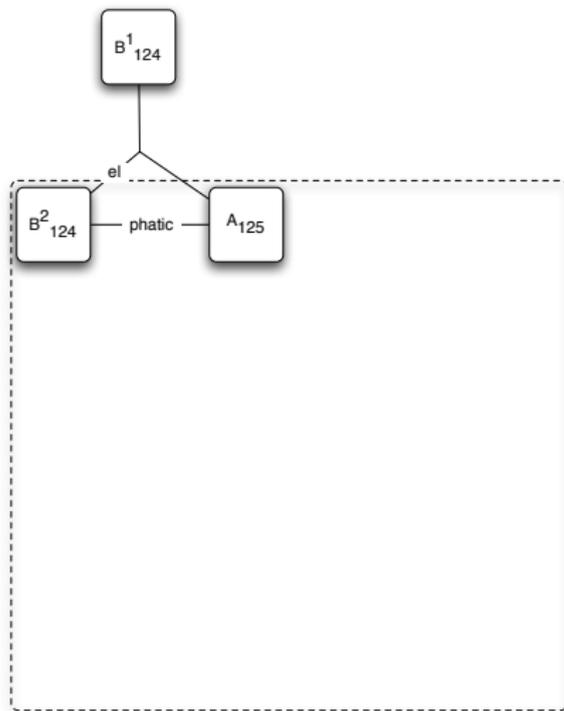
(B124) Oh ouais (\uparrow) et pis compliqué (\downarrow) et c'est vraiment très très compliqué
(\rightarrow)

B¹₁₂₄

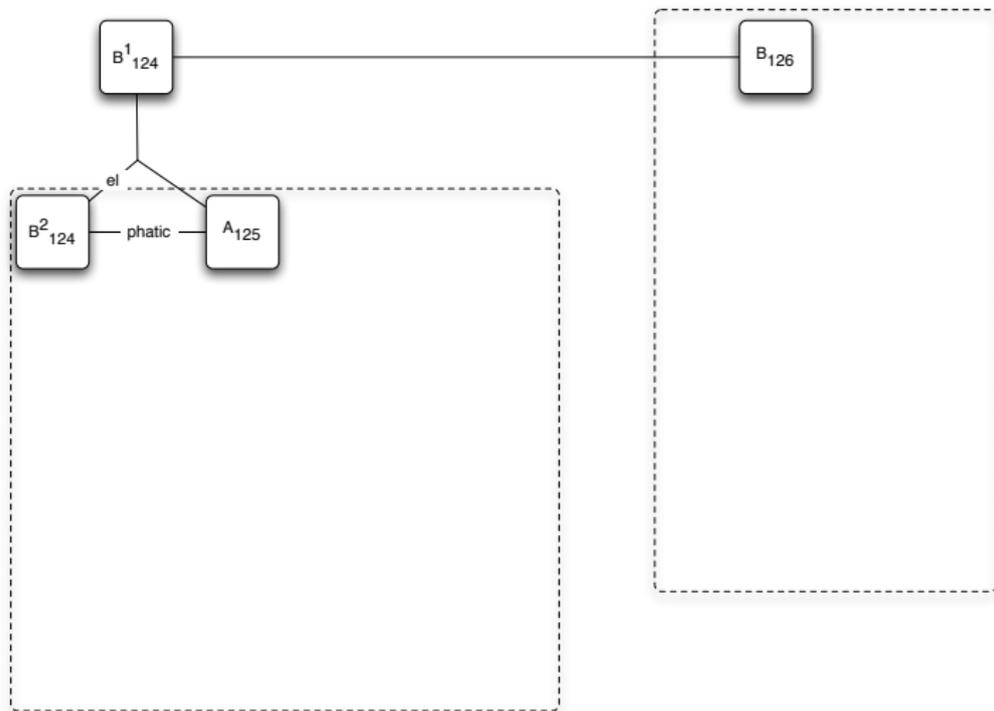
la politique c'est quelque chose quand on s'en occupe faut être gagnant parce qu'autrement quand on est perdant c'est fini quoi (↓)



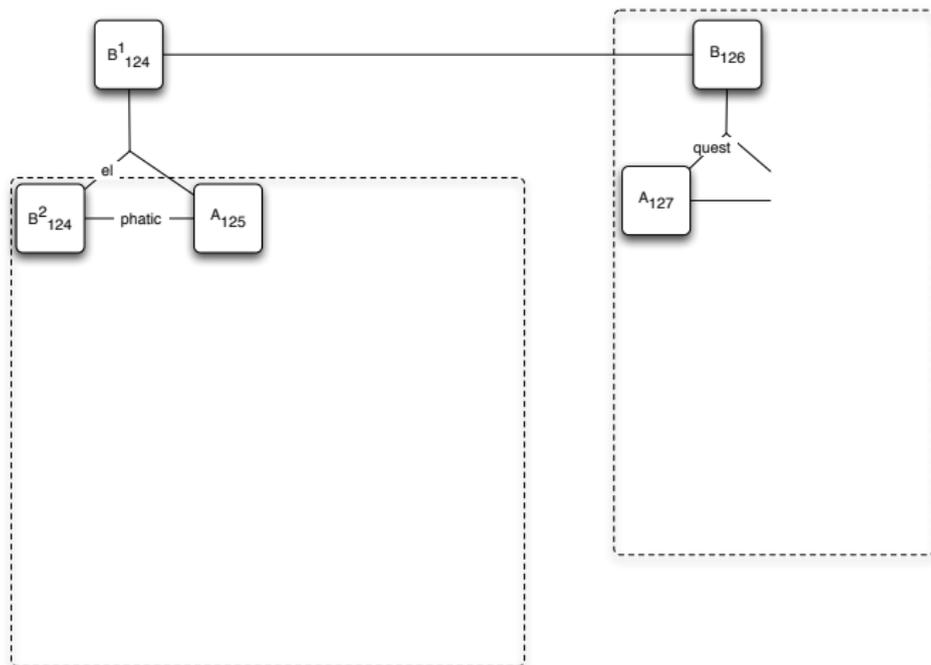
(A125) Oui



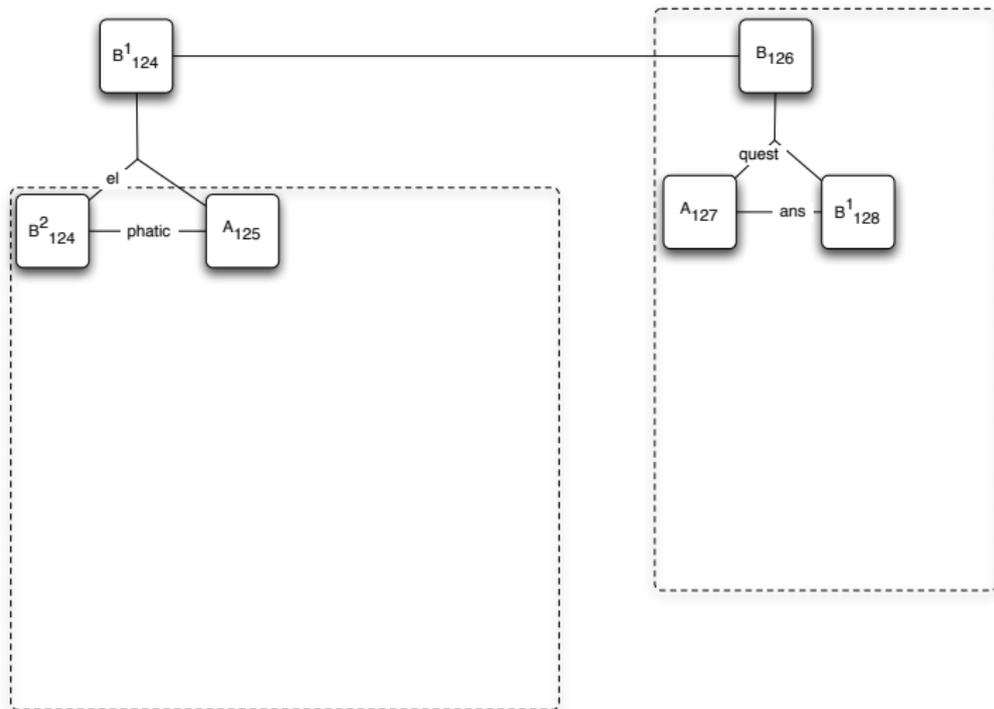
(B126) J. C. D. est mort, L. est mort, P. est mort euh (...)



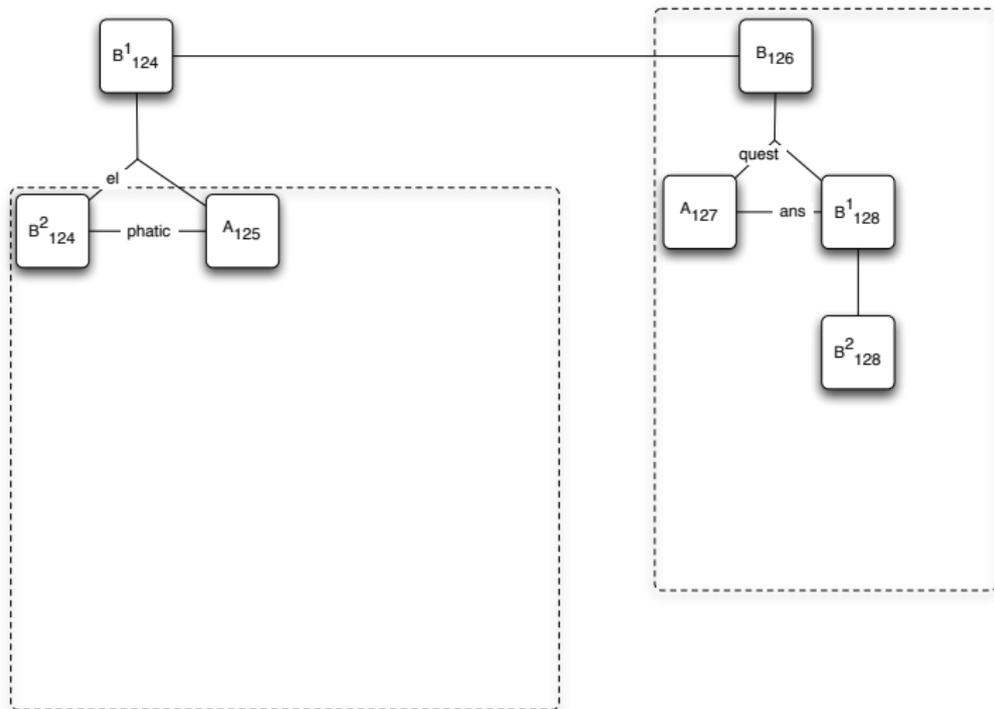
(A127) Ils sont morts parce qu'ils ont perdu à votre avis (↑)



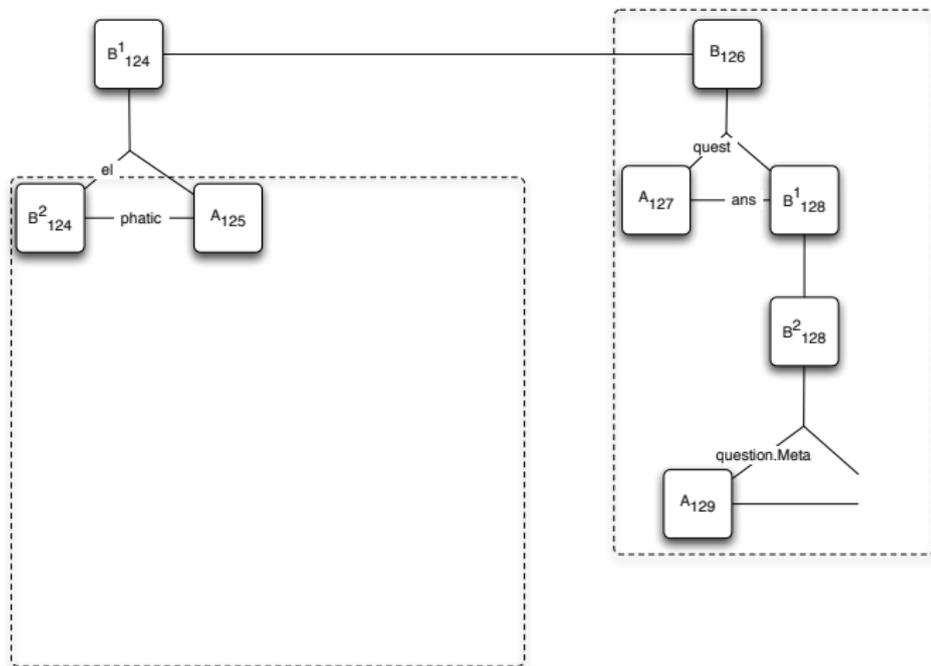
(B128) Non ils gagnaient mais si ils sont morts, c'est la maladie quoi c'est c'est (→)



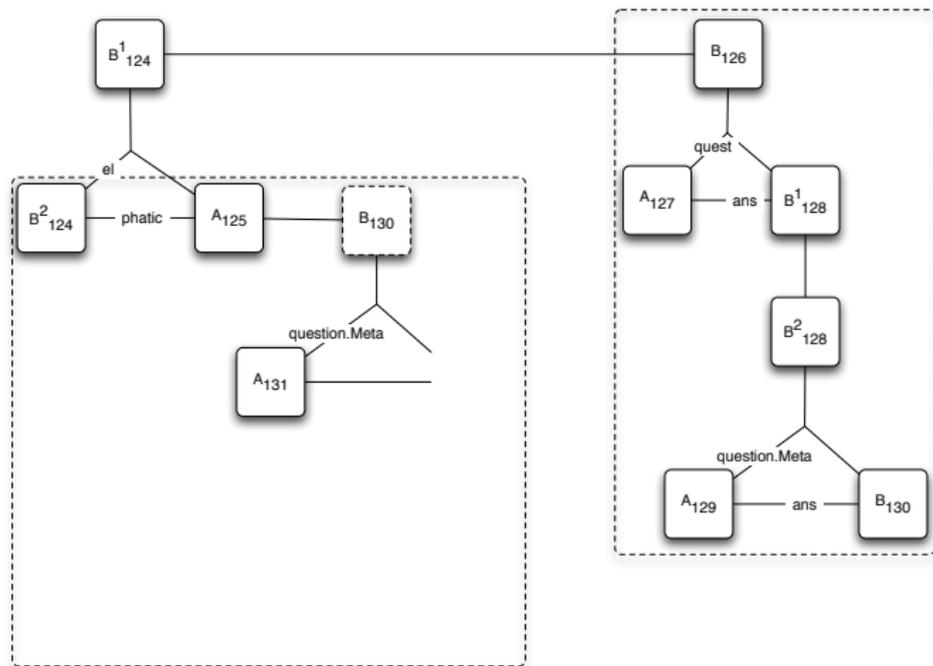
(B128) Non ils gagnaient mais si ils sont morts, c'est la maladie quoi c'est c'est (→)



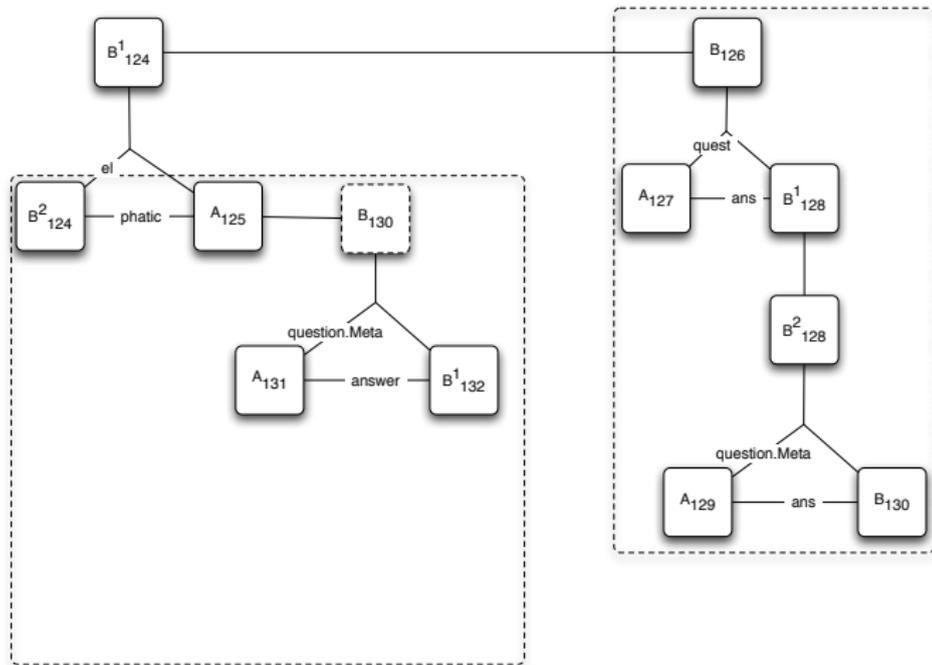
(A129) Ouais c'est parce qu'ils étaient malades, c'est pas parce qu'ils faisaient de la politique (↑)



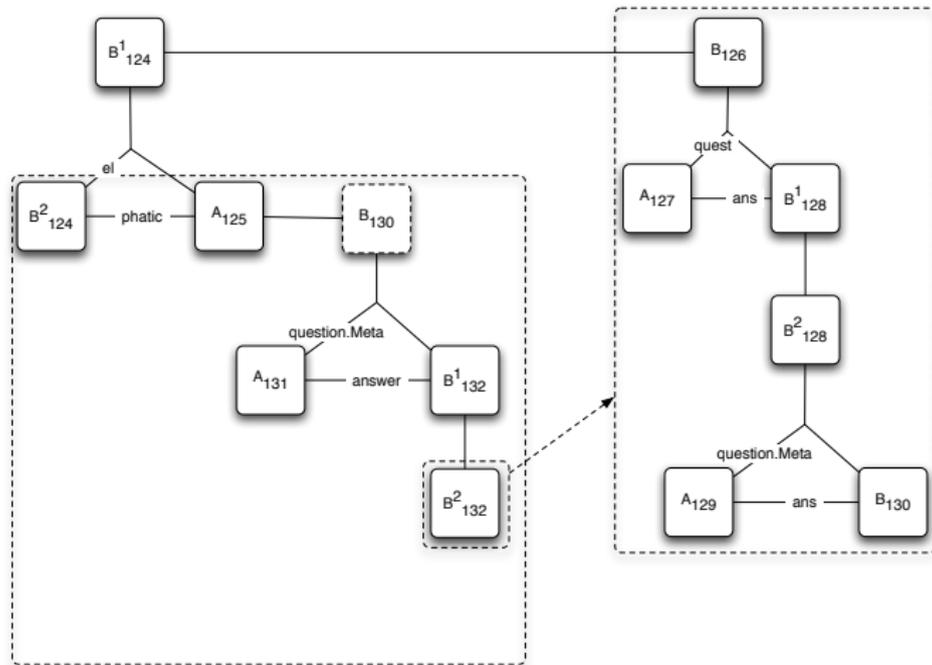
(A131) Si vous pensez que c'est parce qu'ils faisaient de la politique (↑)



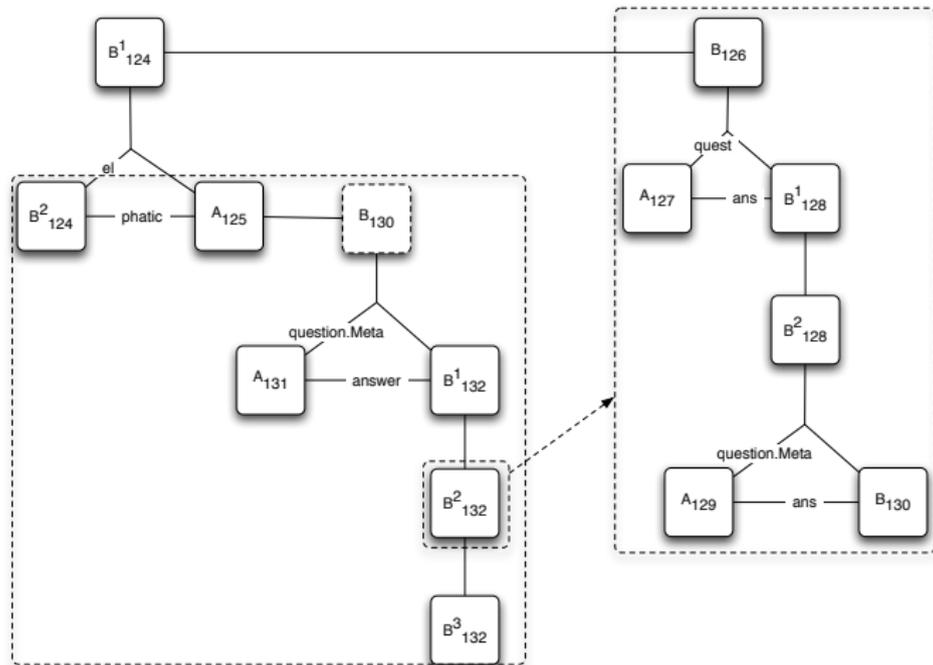
(B132) **Oui tiens oui il y a aussi C. qui a accompli un meurtre là** (\rightarrow) il était présent lui aussi qui est à B. mais enfin (\rightarrow) c'est encore à cause de la politique ça



(B132) Oui tiens oui il y a aussi C. qui a accompli un meurtre là (→) **il était présent lui aussi qui est à B. mais enfin** (→) 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**



- Amblard, Maxime et Karën Fort (2014)**. “Étude quantitative des disfluences dans le discours de schizophrènes : automatiser pour limiter les biais”. In : *TALN - Traitement Automatique des Langues Naturelles*. Marseille, France, p. 292–303. HAL archive ouverte : [hal-01054391](https://hal.archives-ouvertes.fr/hal-01054391) (cf. p. 66, 67).
- Benzitoun, Christophe, Karën Fort et Benoît Sagot (2012)**. “TCOF-POS : un corpus libre de français parlé annoté en morphosyntaxe”. In : *Traitement Automatique des Langues Naturelles (TALN)*. Grenoble, France, p. 99–112. Anthologie ACL : F12-2008 (cf. p. 83).
- Chaika, Elaine (1974)**. “A linguist looks at “schizophrenic” language”. In : *Brain and Language* 1.3, p. 257–276 (cf. p. 65).
- Constant, Matthieu et Anne Dister (2010)**. “Automatic detection of disfluencies in speech transcriptions”. In : *Spoken Communication*. Sous la dir. de M. Pettorino et al. T. 1. Cambridge Scholars Publishing, p. 259–272. HAL archive ouverte : [hal-00636983](https://hal.archives-ouvertes.fr/hal-00636983). URL : <http://hal-upec-upem.archives-ouvertes.fr/hal-00636983> (cf. p. 74, 75).

Fromkin, Victoria A. (1975). “A linguist looks at “a linguist looks at ‘schizophrenic language’””. In : *Brain and Language* 2, p. 498–503.

DOI :

[http://dx.doi.org/10.1016/S0093-934X\(75\)80087-3](http://dx.doi.org/10.1016/S0093-934X(75)80087-3).

URL : <http://www.sciencedirect.com/science/article/pii/S0093934X75800873> (cf. p. 65).

Musiol, Michel et Alain Trognon (1996). “L’accomplissement interactionnel du trouble schizophrénique”. In : *Raisons Pratiques* 7, p. 179–209 (cf. p. 65).

Rebuschi, Manuel, Maxime Amblard et Michel Musiol (2014). “Using SDRT to analyze pathological conversations. Logicality, rationality and pragmatic deviances”. In : *Interdisciplinary Works in Logic, Epistemology, Psychology and Linguistics: Dialogue, Rationality, and Formalism*. Logic, Argumentation & Reasoning. Springer, p. 343–368. HAL archive ouverte : [hal-00910725](https://hal.archives-ouvertes.fr/hal-00910725) (cf. p. 66, 67).

Verhaegen, Frédéric (2007). “Psychopathologie cognitive des processus intentionnels schizophréniques dans l’interaction verbale”. Thèse de doct. Université Nancy 2, France (cf. p. 65).