



**HAL**  
open science

## (In)Cohérence de la conversation

Maxime Amblard

► **To cite this version:**

Maxime Amblard. (In)Cohérence de la conversation. TAL et Linguistique de corpus, Oct 2021, Nanterre, France. hal-03619163

**HAL Id: hal-03619163**

**<https://inria.hal.science/hal-03619163>**

Submitted on 24 Mar 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# (In)Cohérence de la conversation

## TAL et Linguistique de corpus

---

Maxime Amblard

19 octobre 2021



Introduction

ODiM/MePheSTO

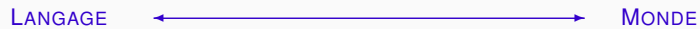
Le TAL au service du diagnostic médical

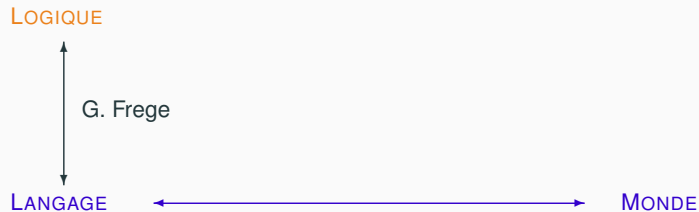
Perspectives

## Introduction

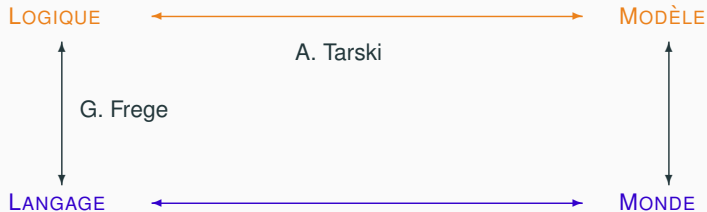
---

1. Rendre compte des phénomènes de la langue naturelle par des cadres formels
2. Troubles de la pensée et du langage



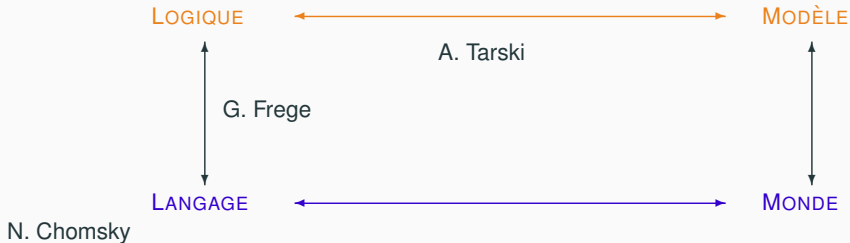


Principe de compositionnalité



Principe de compositionnalité  
Satisfiabilité

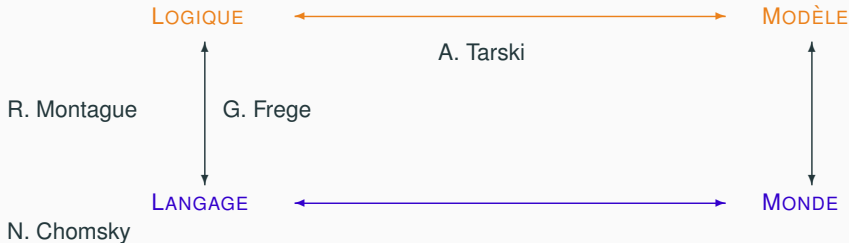




Principe de compositionnalité

Satisfiabilité

Linguistique computationnelle



Principe de compositionnalité

Satisfiabilité

Linguistique computationnelle

Sémantique computationnelle

- Semantique

- Semantique
- Compositionnalité (Frege)

- Semantique
- Compositionnalité (Frege)
- Approches logique (Montague)

- Semantique
- Compositionnalité (Frege)
- Approches logique (Montague)
  - (1) John loves Mary

- Semantique
- Compositionnalité (Frege)
- Approches logique (Montague)

(1) John loves Mary

**love(John, Mary)**

- Semantique
- Compositionnalité (Frege)
- Approches logique (Montague)

(1) John loves Mary

**love(John, Mary)**

- Comment utiliser ces représentations ?
- Utilité de ces représentations ?



- Semantique
- Compositionnalité (Frege)
- Approches logique (Montague)

(1) John loves Mary

**love(John, Mary)**

- Comment utiliser ces représentations ?
  - Utilité de ces représentations ?
- (2) If a farmer owns a donkey, he beats it

- Semantique
- Compositionnalité (Frege)
- Approches logique (Montague)

(1) John loves Mary

**love(John, Mary)**

- Comment utiliser ces représentations ?
- Utilité de ces représentations ?

(2) If a farmer owns a donkey, he beats it

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

- Semantique
- Compositionnalité (Frege)
- Approches logique (Montague)

(1) John loves Mary

**love(John, Mary)**

- Comment utiliser ces représentations ?
- Utilité de ces représentations ?

(2) If a farmer owns a donkey, he beats it

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

- Réalité cognitive, réalité conceptuelle ? ...

## Peut-on comprendre la folie ?

Article "The Conversation France" 31.05.2017



<http://theconversation.com/peut-on-comprendre-la-folie-78510>

**ODiM/MePheSTO**

---

Identification et Informatisation de signes prodromiques et diagnostiques  
dans les maladies mentales

### Identification et Informatisation de signes prodromiques et diagnostiques dans les maladies mentales

Responsables du projet :

- Maxime Amblard MCF HDR informatique, Loria UMR 7503,
  - équipe projet Sémagramme Inria Nancy Grand-Est
  - associé équipe Discours ATILF
- Michel Musiol, Pr psychopathologie de la cognition,
  - équipe Discours ATILF UMR 7118
  - associé équipe projet Sémagramme INRIA Grand-Est

### Permanents

- Philippe de Groot (leader) - DR Inria
- Bruno Guillaume - CR Inria
- Sylvain Pogodalla - CR Inria



### Permanents

- Philippe de Groote (leader) - DR Inria
- Bruno Guillaume - CR Inria
- Sylvain Pogodalla- CR Inria

### Non-permanents

- Samuel Buchel (doctorant, ATILF - INRIA)
- Vincent-Thomas Barrouillet (doctorant, ATILF – INRIA)
- Amandine Lecomte (doctorante, ATILF - INRIA)
- Pierre Lefebvre (Ingénieur Informaticien)

## Partenaires scientifiques

- DFKI (German Research Center for Artificial Intelligence, Saarbrücken)
- équipe-projet STARS INRIA Nice Sophia-Antipolis
- équipe CoBTeK (Cognition Behaviour Technology, EA 7276, Université de Nice)

## Partenaires scientifiques

- DFKI (German Research Center for Artificial Intelligence, Saarbrücken)
- équipe-projet STARS INRIA Nice Sophia-Antipolis
- équipe CoBTeK (Cognition Behaviour Technology, EA 7276, Université de Nice)

## Hospitaliers

- Le Vinatier, CHU Lyon
- Centre Hospitalier Tizi Ouzou (Algérie)
- Hôpital Montperrin (Aix-en-Provence)
- Centre Psychothérapique de Nancy
- Centre Hospitalier Universitaire de Nice, Hôpital Pasteur
- University Clinic of Saarland
- Dept. Of Psychiatry, University of Oldenburg

- psychiatrie
- psychopathologie
- linguistique
- psycholinguistique
- sémantique formelle
- sciences du numérique

## Historique du projet (moyens)

- Groupe de Recherche sur les Communications, axe **conversations pathologiques** (Alain Trognon, Michel Musiol, etc..)

[1990-2010]

## Historique du projet (moyens)

- Groupe de Recherche sur les Communications, axe **conversations pathologiques** (Alain Trognon, Michel Musiol, etc..) [1990-2010]
- SLAM - **Schizophrénie et Langage : Analyse et Modélisation** avec Manuel Rebuschi - MCF HDR Philosophie Univ. Lorraine, UMR 7117 [2011-2018]

## Historique du projet (moyens)

- Groupe de Recherche sur les Communications, axe **conversations pathologiques** (Alain Trognon, Michel Musiol, etc..) [1990-2010]
- SLAM - **Schizophrénie et Langage : Analyse et Modélisation** avec Manuel Rebuschi - MCF HDR Philosophie Univ. Lorraine, UMR 7117 [2011-2018]
- ODIM - **Outils informatisés d'aide au Diagnostic des Maladies mentales** Action Exploratoire Inria [2019-2021]

## Historique du projet (moyens)

- Groupe de Recherche sur les Communications, axe **conversations pathologiques** (Alain Trognon, Michel Musiol, etc..) [1990-2010]
- SLAM - **Schizophrénie et Langage : Analyse et Modélisation** avec Manuel Rebuschi - MCF HDR Philosophie Univ. Lorraine, UMR 7117 [2011-2018]
- ODIM - **Outils informatisés d'aide au Diagnostic des Maladies mentales** Action Exploratoire Inria [2019-2021]
- MePheSTO - **Digital Phenotyping 4 Psychiatric Disorders from Social Interaction** Projet Inria-DFKI - Analyse multimodale de l'interaction patient-clinicien. [2020-2022]



### Groupe des psychoses (Schizophrénie)

- La schizophrénie affecte env 1% de la population mondiale (600 000 personnes en France)
- La guérison est exceptionnelle (voire inexistante)
- La maladie est considérée pour la personne et pour ses proches comme un handicap sévère
- Le coût pour la société est très élevé
- Le 1er épisode psychotique franc survient après l'âge de 18 ans

### Groupe des psychoses (Schizophrénie)

- La schizophrénie affecte env 1% de la population mondiale (600 000 personnes en France)
- La guérison est exceptionnelle (voire inexistante)
- La maladie est considérée pour la personne et pour ses proches comme un handicap sévère
- Le coût pour la société est très élevé
- Le 1<sup>er</sup> épisode psychotique franc survient après l'âge de 18 ans
- Il n'existe toujours pas de signes pathognomoniques
- Le diagnostic psychiatrique est établi de manière a-théorique (ex DSM 5 outil universel) les listes de critères (i.e. symptômes) sont générales et de fait imprécises, exemples : 'discours désorganisé' , 'idées délirantes'

### Objectifs principaux - Phase 1

- **Compléter** le paradigme classique des Troubles du Langage et de la Pensée (TLP)
- **Par** un modèle sémantico-formel des Troubles du Discours (TDD)

G42 (../..) je m'aurais mariée en grand blanc... en grande cérémonie... je  
devais faire un mariage de 3000 invités... ça vous dit rien ?

A43 : 3000 ?

G43 : 3000 invités... alors vous voyez que j'en ai

A44 : mon dieu en effet... et un mariage avec qui ?

G44 : celui qui voudra de moi... si c'est L. c'est L.... si c'est pas L.... cela sera  
un autre... un autre américain... ah je m'en fais plus maintenant

- Les schizophrènes sont logiquement cohérents

- Les schizophrènes sont logiquement cohérents
- Les ruptures sont au niveau du processus de construction de la représentation conversationnelle sur la dimension pragmatique

- Les schizophrènes sont logiquement cohérents
- Les ruptures sont au niveau du processus de construction de la représentation conversationnelle sur la dimension pragmatique
  
- La sous-spécification (ambiguïté) est centrale dans la rupture

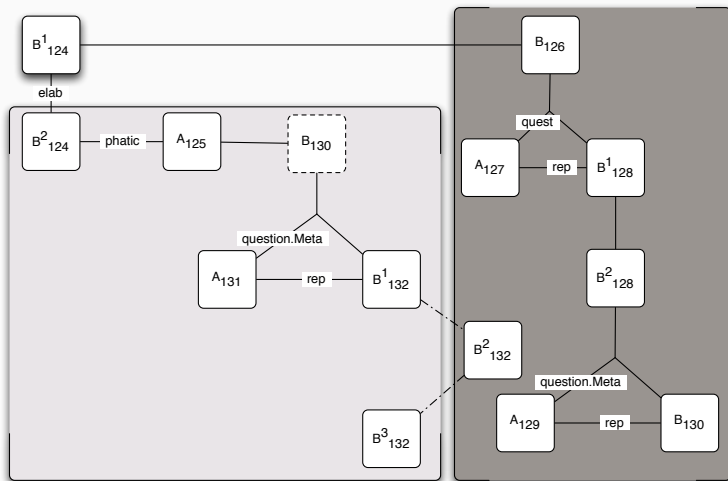
- Les schizophrènes sont logiquement cohérents
- Les ruptures sont au niveau du processus de construction de la représentation conversationnelle sur la dimension pragmatique
- La sous-spécification (ambiguïté) est centrale dans la rupture  
Un choix n'est jamais définitif !



- Les schizophrènes sont logiquement cohérents
- Les ruptures sont au niveau du processus de construction de la représentation conversationnelle sur la dimension pragmatique
  
- La sous-spécification (ambiguïté) est centrale dans la rupture  
Un choix n'est jamais définitif !  
phonologique, morphologique, lexical, référent de discours ...

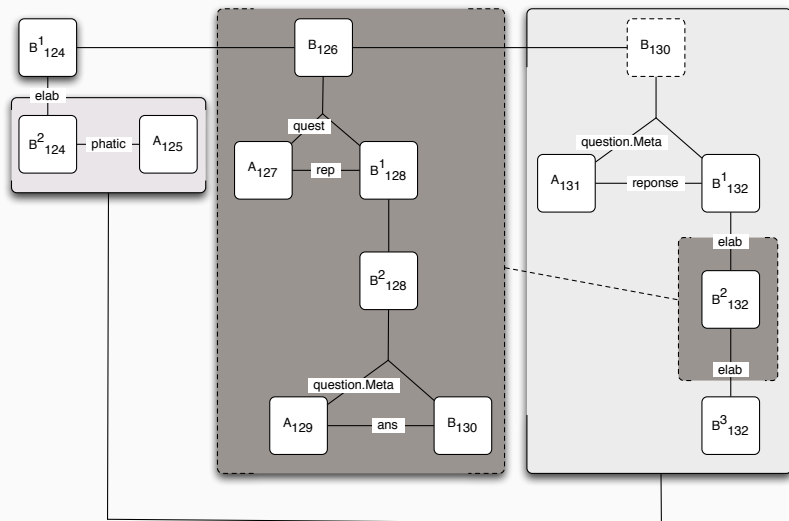
# Analyse inspirée de la SDRT

Point de vue du patient



# Analyse inspirée de la SDRT

## Point de vue du psychologue



### Objectifs principaux - Phase 2

- Élaborer des outils informatisés d'aide au diagnostic de schizophrénie
- Adapter et utiliser ces outils au problème du dépistage des personnes à risque
- Identifier des prodromes (signes annonciateurs de l'entrée dans la maladie / dépistage précoce = prise en charge + efficace)

Déclenchement de la maladie / épidémiologie

## Déclenchement de la maladie / épidémiologie

- suite à désocialisation, prise de stupéfiants, épisodes de délinquance, etc

## Déclenchement de la maladie / épidémiologie

- suite à désocialisation, prise de stupéfiants, épisodes de délinquance, etc
- les symptômes (actuellement décrits sur le plan clinique 'inhibitions, hallucinations..)

## Déclenchement de la maladie / épidémiologie

- suite à désocialisation, prise de stupéfiants, épisodes de délinquance, etc
- les symptômes (actuellement décrits sur le plan clinique 'inhibitions, hallucinations..)
- sont supposés apparaître 1 à 2 ans avant le 1er épisode psychotique franc et des prodromes (jusqu'à 5 ans auparavant)



## Déclenchement de la maladie / épidémiologie

- suite à désocialisation, prise de stupéfiants, épisodes de délinquance, etc
- les symptômes (actuellement décrits sur le plan clinique 'inhibitions, hallucinations..)
- sont supposés apparaître 1 à 2 ans avant le 1er épisode psychotique franc et des prodromes (jusqu'à 5 ans auparavant)
- **prédictibilité des outils cliniques actuels : 50 %**

Mettre au jour des **signes langagiers et des structures de discours** décrivant le syndrome schizophrénique et l'entrée dans la maladie

Mettre au jour des **signes langagiers et des structures de discours** décrivant le syndrome schizophrénique et l'entrée dans la maladie

**le plus spécifiquement possible**

(approche pathognomonique)

Mettre au jour des **signes langagiers et des structures de discours** décrivant le syndrome schizophrénique et l'entrée dans la maladie

**le plus spécifiquement possible**

(approche pathognomonique)

**le plus précisément possible**

(indices modélisés/signes informatisables)

## Analyse et Modélisation des données

- analyse lexicale, morphosyntaxique, pragmatique, Analyse outillée. . .  
Traitement Automatique Langues
- Analyse hiérarchique ou sémantico-formelle du discours.. (contraintes intra ou inter-interventions, SDRT. . . , rupture de la frontière droite)

## Analyse et Modélisation des données

- Articulation d'indices
  - neuropsychologiques (mémoire de travail, flexibilité cognitive, planification de l'action..)
  - neurocognitifs : oculomoteurs (zones de fixation, fréquence saccadique. . . ), expressions faciales.
  - linguistiques (prosodie. . . )
  - psycholinguistiques (inférences. . . )

## **Le TAL au service du diagnostic médical**

---

Si ce n'est pas un bon psy, c'est possiblement une bonne aide.

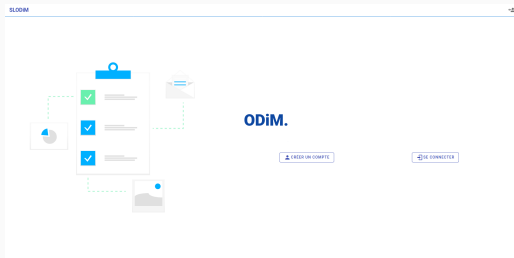


Si ce n'est pas un bon psy, c'est possiblement une bonne aide.

Pas pour toutes les situations, toutes les maladies, mais il est possible de se concentrer sur des pathologies particulières

## Analyse automatique

- Nombre de mots, longueur, etc.
- Analyse lexicale
- Catégorisation des mots (catégorie grammaticale, déictique)
- Relations syntaxiques (UD, distribution)
- Sémantique (embeddings, backchannel, open repair class, etc.)



**SLODIM** Tableau de bord > Dossier patient

**RAPPORT GLOBAL**      **ENTRETIENS**      **TESTS NEUROCOGNITIFS**

**Résumé**

**Jean**  
 ADRESSE  
 jean@mail.fr

**Historique**

**Analyse brute**

Participant	Nombre de tours de parole	Nombre de mot	Nombre de moyen mots/TSP	Longueur moyenne des mots	Ratio stopwords
Y	388 vs 388	1040 vs 1040	4.47 vs 4.46	4.67 vs 4.70	4.48 vs 4.48
O	31 vs 18	31 vs 18	4.26 vs 4.4	4.26 vs 4.4	0 vs 0



## L'action exploratoire

### Description

ODiM est un projet interdisciplinaire, à l'interface de : la psychiatrie-psychopathologie, la linguistique, la sémantique formelle et les sciences du numérique. Il tend à remplacer le paradigme des Troubles du Langage et de la Pensée (TLP) tel qu'on l'utilise dans le secteur de la Santé mentale par un modèle sémantico-formel des Troubles du Discours (TDD). Il s'agit de traduire ces troubles en signes diagnostiques ainsi que de dépistage des personnes vulnérables dites "à risques".



### Téléchargement du fichier

**i** Formats acceptés : [voir la liste](#)

 Fichier

SUIVANT

**2** Choix des analyses

**3** Format de sortie

## Études contrastives et longitudinales

- Analyse des productions de différents groupes
- Analyse de l'évolution dans le temps

## Construction de modèles de classification

- Par apprentissage automatique
- À partir des spécificités du dialogue

## Construction d'autres corpus

- Réseaux sociaux
- Expériences fermées

## Phase de tests sur des données réelles

- **Caractérisation** des profils
- Tests auprès de **cliniciens**
- Approfondissement des **analyses sémantiques/dialogiques**
  - En diversifiant les critères
  - En intégrant les embeddings et la classification
  - En faisant émerger des structures de dialogue



## Digital Phenotyping 4 Psychiatric Disorders from Social Interaction

Projet Inria-DFKI

Intelligence Artificielle pour l'identification et la classification de **phénotypes numériques objectifs**, et donc mesurables, des troubles psychiatriques.

- Analyses **multimodales** : son, discours, vidéo
- **Multilinguisme** - multiculturalisme - spécificité d'une langue
- Ouverture à d'**autres pathologies** proches (bipolarité, dépression, etc.)

## Objectifs

- Identification de symptômes pathologiques
- Accompagnement et suivi des patients
- Définition de processus de remédiation

## Utilisation

- Approches symboliques (sémantique formelle, logique, représentation)
- Approches numériques (modèles du dialogue)

- **Volume très réduit** de données pour la complexité de la tâche

- **Volume très réduit** de données pour la complexité de la tâche
- **Impossibilité de l'anonymisation** - partage des données

- **Volume très réduit** de données pour la complexité de la tâche
- **Impossibilité de l'anonymisation** - partage des données  
Qui complexifie la reproductibilité des résultats

- **Volume très réduit** de données pour la complexité de la tâche
- **Impossibilité de l'anonymisation** - partage des données  
Qui complexifie la reproductibilité des résultats
- **Manque de données** - accéder aux patients et à leur production

- **Volume très réduit** de données pour la complexité de la tâche
- **Impossibilité de l'anonymisation** - partage des données  
Qui complexifie la reproductibilité des résultats
- **Manque de données** - accéder aux patients et à leur production
- **Biais et interprétation** des résultats

- **Volume très réduit** de données pour la complexité de la tâche
- **Impossibilité de l'anonymisation** - partage des données  
Qui complexifie la reproductibilité des résultats
- **Manque de données** - accéder aux patients et à leur production
- **Biais et interprétation** des résultats
- **Annotation** d'un grand volume de données



## De la réalité des patients

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

- Le problème du diagnostic ne réside pas seulement dans la **puissance de calcul trop faible** mais dans la découverte de **signes prodromiques**
- Quand bien même, difficulté de mener une **thérapie adaptée à une personne** (sensibilité, empathie, conscience,...)

- Le problème du diagnostic ne réside pas seulement dans la **puissance de calcul trop faible** mais dans la découverte de **signes prodromiques**
- Quand bien même, difficulté de mener une **thérapie adaptée à une personne** (sensibilité, empathie, conscience,...)
  
- **La langue est un phénomène complexe**  
Ici on entre dans la complexité de son fonctionnement
  
- Il s'agit donc d'**épauler les spécialistes** en leur proposant des indices à considérer pour leur diagnostic.  
**En aucun cas de se substituer à eux**

## Perspectives

---

- Augmenter les phénomènes analysés avec SLAMtk  
En particulier sur la syntaxe et la statistique lexicale
- Essayer des approches réseaux de neurones sur le corpus SLAM  
Nécessite plus de ressources en français
- Étudier plus profondément les annotation humaines sur le corpus
- Augmenter le corpus en volume et en nombre de pathologies étudiées  
Collecte de données à l'hôpital Montperrin d'Aix-En-Provence
- Définir un processus de remédiation
- Affiner l'analyse des dysfonctionnements, s'ouvrir vers une interprétation cognitive et donner un contexte plus complexe pour l'interprétation

- Définir des grammaires sémantiques robustes
- Définir une représentation formelle des dialogues

Travaux en cours sur la relation question/réponse

**Merci !**