

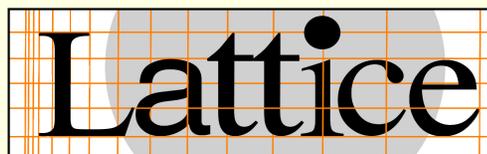
# **L'anaphore à antécédent flou :**

## **une caractérisation et ses conséquences**

### **sur l'annotation des relations anaphoriques**

Frédéric Landragin

Journée ATALA du 16 juin 2007



Paris

# Contenu de la présentation

1. Objectifs de l'étude
2. Interprétations partielles et anaphores : phénomènes et caractérisation
  - ambiguïté sémantique *versus* ambiguïté TAL
  - ambiguïté pour l'interlocuteur *versus* ambiguïté assumée par le locuteur
  - une première classification des antécédents potentiellement flous
  - discussion
3. Anaphores à antécédents flous et résolution automatique des anaphores
  - annotation pour évaluer *versus* annotation pour représenter
  - quelques principes d'annotation adaptés aux antécédents flous
  - double balisage et attribut MIN
  - discussion
4. Bilan et perspectives

# Objectifs

1. Montrer qu'il n'est pas toujours indispensable de déterminer précisément quel est l'antécédent d'une anaphore
2. Apporter la notion d'anaphore à antécédent flou pour rendre compte d'un type d'interprétation partielle (sous-spécifiée, ambiguë)
3. Montrer un exemple d'impact de la sémantique des entités du discours sur l'identification des antécédents possibles d'une anaphore
4. Fournir des recommandations pour les systèmes de résolution des anaphores, à travers la proposition de quelques principes d'annotation

# Antécédent flou : cas 1

1. « Je **lui** ai dit sur un ton de plaisanterie que **son idée** était intéressante, qu'**elle** montrait les choses sous un angle auquel en effet on n'est pas habitué » Frantext K899

- antécédent possible 1 = **l'idée**, qui montre les choses sous un angle...
- antécédent possible 2 = **la personne**, qui montre les choses...
- deux antécédents dont l'un est une facette/propriété/fonction de l'autre
- les deux sont au féminin singulier, etc. → **ambiguïté sémantique**
- l'ambiguïté freine-t-elle la compréhension du texte ?
- l'ambiguïté est-elle assumée par le locuteur ?

2. « **Le travail de l'intellect** naît d'une insatisfaction profonde devant le monde, mais **il** s'avère incapable d'y porter remède »

*NB : autres exemples :* « une commune acception de la réalité [...] **elle** [...] »  
 « la nostalgie d'une innocence perdue [...] **elle** »  
 « le caractère de Marcello [...] **il** »  
 « le comportement de tout public quel qu'**il** soit »  
 « Marcello [...] son engagement [...] **il** »

# Antécédent flou : cas 2

1. « **les rapports des rapporteurs et les documents auxquels ils se réfèrent**, pour autant qu'**ils** n'ont pas été communiqués [...] » Frantext P659
  - antécédents possibles = premier coordonné ; deuxième coordonné ; tous
  - coordination, surtout avec des pluriels → **ambiguïté sémantique**
  - l'ambiguïté freine-t-elle la compréhension du texte ?
  - l'ambiguïté est-elle assumée par le locuteur ?
  
2. « **cette faille, cette malédiction** qui est comme une résurgence laïque du péché originel, est d'autant plus grave qu'**elle** laisse un vide [...] »
  - hypothèse de la correction ou reformulation (qui annule le premier terme) : l'antécédent peut être 1. le dernier terme ou, éventuellement, 2. l'ensemble
  - hypothèse de la précision : l'antécédent peut être 1. le dernier terme, 2. l'ensemble, ou, éventuellement, 3. le premier terme seul
  - hypothèse de l'apposition : l'antécédent est plutôt 1. le premier terme
  - même phénomène d'**ambiguïté sémantique**
  - *NB : « le rêve, l'évasion est nécessaire. **Il** permet de mieux supporter... »  
« le rêve, l'évasion est nécessaire. **Elle** permet de mieux supporter »*

# Antécédent flou : cas 3

1. « Voulez-vous **Jean Martin pour époux** ? – Oui, je **le** veux » ;  
 « Jean a **coupé tout le bois**. Il **l'**a fait en petits morceaux »
  - antécédents possibles = « tout le bois » ; « coupé tout le bois »
  - problème de rattachement → **ambiguïté sémantique entre une anaphore individuelle et une anaphore événementielle**
  - l'ambiguïté freine-t-elle la compréhension du texte ?
  - l'ambiguïté est-elle assumée par le locuteur ?
  
2. « **Jean s'est cassé la jambe. Il n'est pas allé au cinéma. C'**est arrivé hier » ;  
 « quand on rencontrait ma domestique, **on lui chantait sur l'escalier : "allons, Babet, un peu de complaisance"**. J'ai entendu **ça** un jour »  
 Frantext L863
  - antécédents possibles = premier événement ; l'ensemble des deux événements (avec la relation causale)
  - **ambiguïté sémantique entre deux anaphores événementielles**

# Quelles ambiguïtés ?

## 1. Ambiguïté sémantique et ambiguïté TAL

- cas présentés : ambiguïtés sémantiques car les antécédents potentiels sont des entités de discours différentes, avec des propriétés sémantiques différentes (*focalisation sur les éléments sémantiques*)
- marquage de l'antécédent d'une anaphore abstraite : l'antécédent de « **ça** » est-il la phrase complète, ou seulement sa tête verbale ?  
→ ambiguïté TAL (*focalisation sur les éléments textuels*)

## 2. Ambiguïté pour l'interlocuteur et ambiguïté assumée par le locuteur

- pour les ambiguïtés sémantiques : ambiguïtés pour l'interlocuteur et éventuellement aussi pour le locuteur
  - ambiguïté pour l'interlocuteur mais sans doute pas pour le locuteur : « Je **lui** ai dit sur un ton de plaisanterie que **son idée** était intéressante, qu'**elle** montrait les choses... »
  - ambiguïté pour les deux, assumée par le locuteur : « **cette faille, cette malédiction** est d'autant plus grave qu'**elle** laisse un vide... »

# Types d'antécédents potentiellement flous

## 1. Les possessifs, quand :

- le possesseur est un être animé et l'objet possédé est un trait de personnalité auquel il peut s'assimiler (facette, propriété, fonction...)
- le possesseur est un objet et le possédé est sa fonction principale

## 2. Les groupes complexes « le N<sub>1</sub> de le N<sub>2</sub> » avec des participants similaires

## 3. Les coordinations

- avec des participants similaires
- avec des pluriels : « **les N<sub>1</sub> et les N<sub>2</sub> ... ils** »

## 4. Les juxtapositions

- avec des participants du même genre que pour les coordinations
- lorsqu'il n'est pas possible de décider entre correction et précision, et dans des énumérations telles que « le 1<sup>er</sup> secrétaire, M. Martin et sa femme » quand le texte ne permet pas de dire si M. Martin est le 1<sup>er</sup> secrétaire ou une tierce personne

# Les antécédents flous : discussion

## 1. Compréhension et anaphore à antécédent flou

- rejoint la remarque de (Prandi, 1987) à propos des ambiguïtés syntaxiques liées à la portée des prépositions : « ce genre d'ambiguïté structurale, très fréquent dans l'expression nominale, reste d'ailleurs sans conséquences sur l'interprétation, du fait, probablement, de son caractère systématique et non annulable »
- l'existence du phénomène a été montrée avec des exemples faisant intervenir le simple pronom personnel...
- parmi tous les occurrences d'anaphores, ça reste néanmoins un phénomène peu représenté, peut-être marginal...

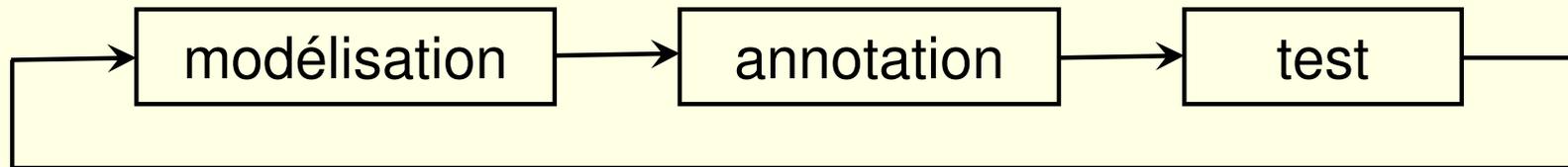
## 2. En fin de compte

- l'intérêt de la notion d'anaphore à antécédent flou, c'est de rendre compte de phénomènes d'interprétation partielle : on n'a pas besoin d'analyser précisément toutes les relations anaphoriques pour comprendre un texte
- le but visé, c'est de modéliser (et formaliser) ces interprétations partielles

# L'annotation : pourquoi ?

## 1. Annotation comme phase méthodologique

- annoter en tant qu'étape lors de la conception d'un modèle :



- annoter pour évaluer des systèmes de résolution des anaphores
- annoter pour comparer des systèmes

## 2. Annotation comme moyen commode de représenter des interprétations

- annoter permet de se forcer à délimiter les antécédents potentiels, et ceci peut aider à mieux appréhender les phénomènes
- une annotation présente de manière simple et lisible les différentes hypothèses d'interprétation
- une annotation n'est pas dépendante de la théorie et/ou formalisme sémantique du système considéré, ce qui va dans le sens de la généralité et favorise les échanges dans la communauté scientifique

# Quelques principes d'annotation

## 1. Principe de groupement

rend compte des groupements d'antécédents :

« **Jean** dormait. **Marie** lisait. **Ils** étaient heureux. » {Jean, Marie}

## 2. Principe des alternatives

ensemble d'antécédents plausibles (tous pouvant être conservés pour l'interprétation), parmi lesquels se trouve celui auquel pensait le locuteur :

« Je **lui** ai dit que **son idée** était intéressante, qu'**elle**... » {elle | son idée}

## 3. Principe des possibles

ensemble de possibilités syntaxiques, certaines étant peu plausibles sémantiquement (pourraient bloquer l'interprétation) :

« le premier secrétaire, M. Martin et sa femme » : 2 ou 3 personnes, l'une de ces possibilités étant en totale contradiction avec la réalité

## 4. Principe de double balisage

limites inférieure et supérieure de l'empan textuel contenant l'antécédent

# A propos du double balisage

1. Pour les anaphores abstraites telles que « **c'**est arrivé hier »
  - le principe des alternatives ne suffit pas, elles seraient trop nombreuses à déterminer : « **ça** » peut reprendre la phrase précédente ; une partie seulement ; les N phrases précédentes ; tout le paragraphe ; etc.
  - exemples rencontrés présentant le même type d'ambiguïté TAL :  
« dans **ce contexte** » ; « à **cet égard** » ; « pour **ce** faire »
  - il faudrait pouvoir gérer des antécédents à limites floues et progressives, ce qui s'avère incompatible avec les balises ouvrantes et fermantes XML
  
2. Attribut MIN et double balisage
  - attribut MIN de MUC-7 : permet de définir l'empan textuel minimal acceptable dans un balisage qui délimite un empan textuel maximal  
→ même entité de discours
  - balisage INF et balisage SUP : permettent de définir les limites inférieure et supérieure de l'empan textuel où se trouve l'antécédent flou  
→ couvrent plusieurs entités de discours
  - un attribut MIN peut être placé dans le balisage INF comme dans le SUP

# L'annotation : discussion

## 1. Vers un schéma d'annotation

- a priori sans difficulté en partant des principes d'annotation proposés
- mais est-ce vraiment utile ? est-ce que ça ne va pas pénaliser certains systèmes qui préfèrent ne pas voir certaines ambiguïtés au profit d'une plus grande efficacité ?

## 2. Vers un système de résolution des anaphores à antécédents flous

- algorithme a priori réalisable puisque les catégories d'antécédents flous reposent sur des critères syntaxiques
- mais est-ce vraiment utile ? les antécédents flous apparaissent très rarement dans les corpus interrogés, donc leur prise en compte n'augmenterait pas de manière significative les performances d'un système

## 3. En fin de compte

- l'intérêt, c'est surtout de donner des idées aux concepteurs de systèmes
- le but visé, c'est d'encourager les approches intégrant l'aspect partiel, sous-spécifié ou ambigu de l'interprétation linguistique

# Bilan et perspectives

## 1. Présentation d'une première étape de recueil de données et de proposition d'une caractérisation

- données pour l'instant limitées, car interroger **Frantext** sur des groupes nominaux tels que « **le N<sub>1</sub> de le N<sub>2</sub>** » ou des empan textuels tels que « **le N<sub>1</sub>, le N<sub>2</sub> ... il** » soit ramène énormément de bruit, soit prend beaucoup de temps de calcul (listes de substantifs et/ou exploitation de la catégorisation)
- caractérisation très syntaxique et pas assez sémantique

## 2. Perspectives

- continuer la collecte de données, explorer d'autres méthodes
- approfondir l'étude des rapports entre N<sub>1</sub> et N<sub>2</sub> : quelles propriétés sémantiques ? quelles contraintes et quelles conséquences sur l'interprétation ?
- modéliser et formaliser dans un cadre de sémantique sous-spécifiée