

**Problèmes d'annotation de corpus**

**au niveau des**

**constructions sémantiques**

**du discours**

István CSÜRY

Université de Debrecen (H),

Département de Français

# 1. Introduction / 1

- Les orientations de l'étude des structurants de discours (SD: connecteurs, marqueurs)
  - Pragmatique > linguistique
  - Théorique > empirique
- La dimension textuelle / discursive de l'étude des SD
  - La nature des entités mises en relation
  - Les traits et les relations lexicologiques des SD
- L'exploration de corpus et la structure du discours
  - Difficultés à identifier les items (*3 cas de figure*) et les structures
  - Repères combinatoires

# 1. Introduction / 2

- Travaux précédents
  - Le champ lexical de *mais* (2001)
  - KKK (2005)
- Les objectifs de la communication
  - Envisager les possibilités et les avantages de l'emploi de NooJ dans le traitement multidimensionnel des discours
  - Passer en revue les problèmes empiriques du repérage des constituants mis en relation par les SD
  - Mettre en évidence le rôle des traits lexicaux, des cooccurrences et de la position syntaxique
  - Évoquer les relations (hiérarchiques et croisées) établies par les SD

## 2. Perspectives de l'emploi de NooJ dans le traitement des relations de discours (RD)

- L'inadaptation des outils d'analyse de corpus à l'étude de l'organisation discursive
- Dans mes travaux antérieurs:
  - Concordances simples de connecteurs
  - Dispositif d'annotation en XML et son usage en Cl@rk System (*KKK*)
- Exigences (minimales) pour aller plus loin:
  - Emploi d'outils compatibles (XML)
  - Possibilité de modification / développement de codages et annotations
  - Combinaison de critères lexicaux et grammaticaux dans l'analyse

# 3. Les problèmes du repérage des constituants des RD

## 1. Entités de nature différente

- Verbales explicites
- Inférables
- Dialogales
- Extralinguistiques

## 2. Constituants de niveaux différents

- Intrapropositionnels

- Intraphrastiques (propositions)
- Phrases / énoncés
- Ensembles multiphrases

## 1. Constituants situés différemment

- Contigus ou éloignés
- « bi- ou unilatéraux »

## 2. Enchevêtrement des relations

## 3./2. Les problèmes du repérage des constituants des RD: exemples

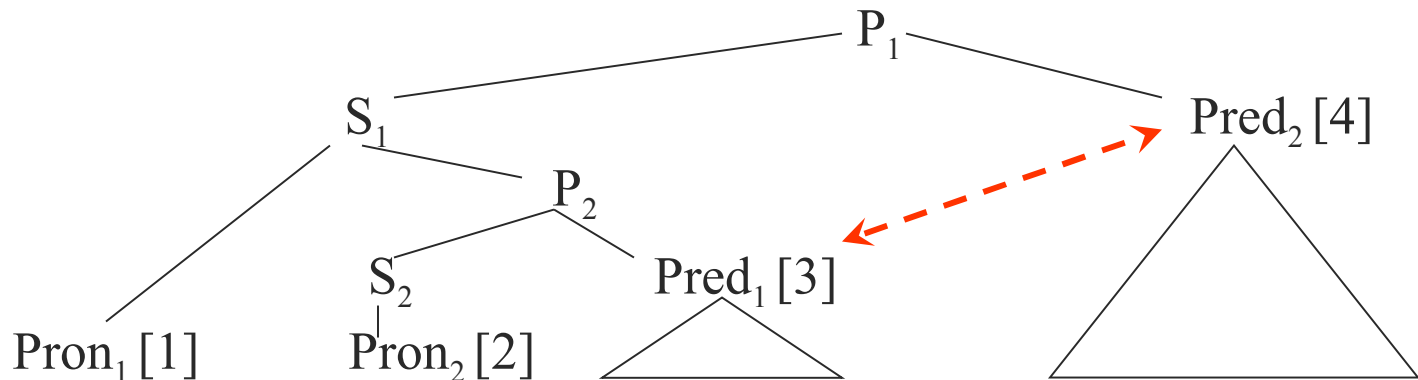
- (1) Cependant, si l'on se place dans le cadre du Régime d'aide du Canada, cela veut dire qu'il s'agit d'argent consacré au bien-être. Dans ce cas, la question continuera de se poser de savoir si ce sont les gens qui ont des salaires moyens ou de petits salaires qui peuvent en bénéficier. J'aimerais voir fonctionner des services de garde qui ne soient pas placés sous la rubrique de bien-être, de façon à ce que [1]ceux d'entre nous [2]qui [3]n'ont pas le droit de bénéficier ou ne veulent pas passer le test permettant de les faire bénéficier de ce service [4]puissent **MALGRÉ TOUT** y accéder. (*Hansard*)

### 3./3. La relation marquée par le connecteur dans (1)

{[ceux d'entre nous] n'ont pas le droit de bénéficiaire ou ne veulent pas passer le test permettant de les faire bénéficiaire de ce service}

**malgré tout**

{[ceux d'entre nous] puissent accéder [aux services de garde ]})



# [ 4. Les caractéristiques des SD ]

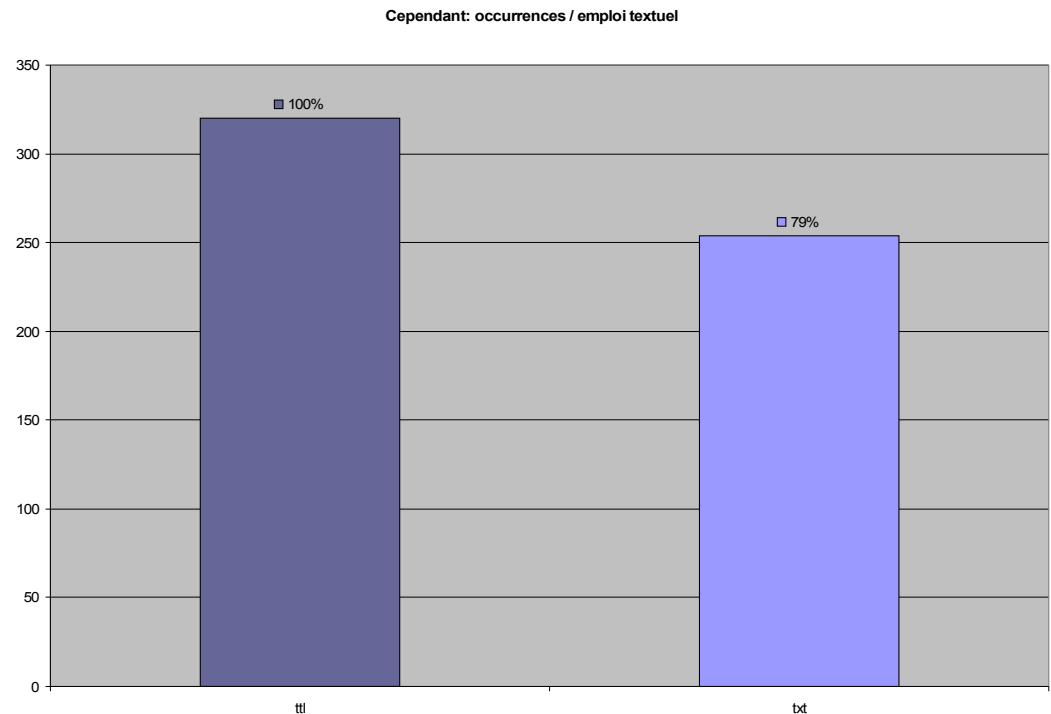
1. Traits lexicaux
  - Sens référentiel ou procédural
  - Types de relation et réalisations
  - Nature et place des termes mis en relation
  - Composantes sémantiques particulières
  - Types de discours; registre
2. Cooccurrences
  - Séquences de connecteurs
3. Position syntaxique
  - Relation intra- ou transphrastique
  - Les structures phrastique et textuelle
  - La position des adverbiaux en français

## 4.3.1. La position des connecteurs adverbiaux en français

- Rubattel (1982 : 54): “aucun principe connu n’explique ces différences d’acceptabilité”
- Weinrich (1989: 324) à propos de la position préverbale des Adv:  
*[cette variante] demeure cependant rare dans l’ensemble et n’est employé que dans le langage soigné, et avec plusieurs restrictions. Il ne peut en particulier s’agir d’adverbes dont le poids phonique ou graphique est trop important. Conviennent particulièrement à cette position les adverbes d’assertion [...] et, parmi les adverbes de séquence [...], ceux qui marquent dans le texte une séquence marquée (conséquence logique, opposition...) [...] Cette place de l’adverbe a aussi pour effet de désigner le sujet comme “thème” en le séparant du verbe. Le verbe arrivant à la suite est ainsi caractérisé comme une information nouvelle (“rhème”).*
- Ouvrages plus récents: classement des adverbes, positions relatives, mais pas grand-chose sur les particularités des connecteurs

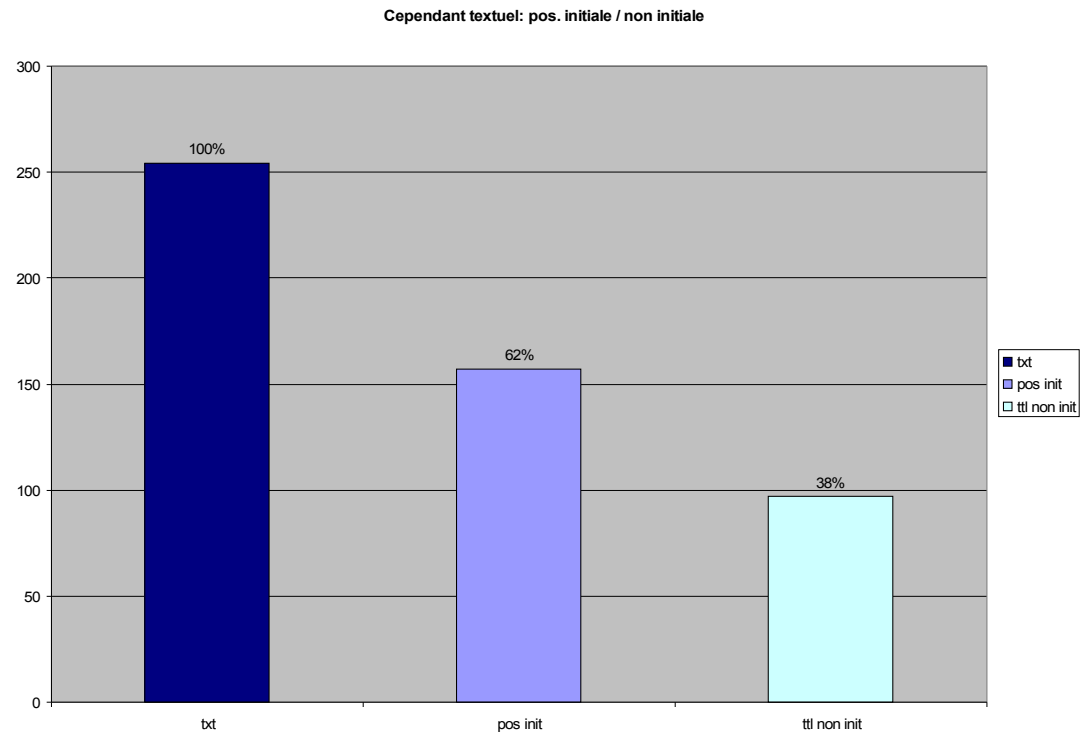
## 4.3.1. / 2. Un exemple: *cependant*

- corpus: 4 romans + 7 pages d'Internet aléatoirement choisies
- 320 occurrences
- **emploi « transphrastique » 254 fois**



## 4.3.1. / 3. La position de *cependant*

- initiale: 157, soit 62%
- préverbale: 32
- périphérie gauche au total: 189, soit 74%
- groupe verbal: 41
- post-GV ou dans compléments: 24



## 4.3.1. / 4. La position préverbale

- séquence avec un autre connecteur (présence d'une conjonction)
- GN sujet topique
- constituant adverbial (cadratif; marqueur d'univers de discours ou d'intégration linéaire)

## 4.3.1. / 5. La position enchâssée

- le véritable choix: entre la position initiale et non-initiale, plutôt qu'entre les différentes positions non-initiales: contraintes grammaticales (ou, du moins, tendances)
  - sur 32 cas de position préverbale, forme verbale simple en 28 cas (surtout imp.). Les 4: 1 passif à l'imp., 1 question avec *peut-être* + inversion, 2 pqp. du même texte
  - S pronominal et/ou de faible volume
  - phrase interrogative, impérative
  - phrase introduite par *si*
  - *il faut* (ou autre constr. impersonnelle) - tendance; moins régulièrement observé
- cas rares de « variantes libres » (Weinrich!)

## 4.3.1. / 6. D'autres exemples: Le connecteur signale un changement de cadre

(2) Et une inquiétude le mordait au coeur. S'il allait échouer ? Mais que pouvait-il craindre ? Il se tirerait toujours d'affaire ! **POURTANT** c'était une grosse partie qu'il jouait, ce soir-là ! (*Bel-Ami*)

## 4.3.1. / 7. Le connecteur tombe sous la portée d'un cadratif

(3) Charles n'était point de complexion facétieuse, il n'avait pas brillé *pendant la noce*. Il répondit médiocrement aux pointes, calembours, mots à double entente, compliments et paillardises que l'on se fit un devoir de lui décocher dès le potage. *Le lendemain*, **EN REVANCHE**, il semblait un autre homme. C'est lui plutôt que l'on eût pris pour la vierge de la veille, tandis que la mariée ne laissait rien découvrir où l'on pût deviner quelque chose. (*Madame Bovary*)

NB: Bonami, Godard & Kampers-Mahne (2004): à la différence des Adv intégrés, la portée des Adv incidents n'est pas restreinte syntaxiquement

## 4.3.1. / 8. Le connecteur suit un topique contrastif

(4) Deux ou trois fois par semaine, *la jeune femme* arrivait avant qu'il fût levé, se déshabillait en une minute et se glissait dans le lit, toute frémissante du froid du dehors. *Duroy*, **PAR CONTRE**, dînait tous les jeudis dans le ménage et faisait la cour au mari en lui parlant agriculture ; et comme il aimait lui-même les choses de la terre, ils s'intéressaient parfois tellement tous les deux à la causerie qu'ils oubliaient tout à fait leur femme sommeillant sur le canapé. (*Bel-Ami*)

## 4.3.1. / 9. Le connecteur suit le topique: l'effet sur les inférences (1)

- (5) L'élargissement ne devrait pas poser de problèmes graves à la PCP, et les nouveaux États membres devraient être en mesure d'adopter celle-ci sans difficulté. *La mise en oeuvre effective de l'acquis* pourrait **NÉANMOINS** souffrir de l'insuffisance, dans certains pays candidats, de structures administratives spécialisées dans des domaines tels que la collecte des données de base nécessaires. (*Docus de l'UE*)

## 4.3.1. / 10. Le connecteur suit le topique: l'effet sur les inférences (2)

- (6) La Roumanie doit affronter un défi particulier en matière d'affaires intérieures et de justice. Elle a pour l'instant fait des progrès limités en matière de reprise de l'acquis en ce domaine. *Le nouveau gouvernement a **CEPENDANT** entrepris un programme de grande ampleur pour mener à bien les réformes institutionnelles indispensables. (Docus de l'UE)*

# [ 5. La problématique des interrelations des RD ]

1. Problèmes de délimitation
  - Ouverture et fermeture
  - Réalisations linguistiques différentes
  - Contenus non explicites
2. Hiérarchie et entrecroisement
  - Entrecroisement des RD d'ordre différent
  - Dépendance du même constituant de plusieurs parents
3. Traitement incrémentiel
  - Situation de discours: interprétation progressive avec retours sur le contexte précédent; prise en compte incrémentielle de l'ensemble des SD /RD
  - Situation d'analyse: démarches régressives; prise en compte d'« îlots » de RD
4. Repères (à part des SD)
  - Réseaux de coréférence; champs lexicaux / sémantiques
  - Phénomènes morfo-grammaticaux

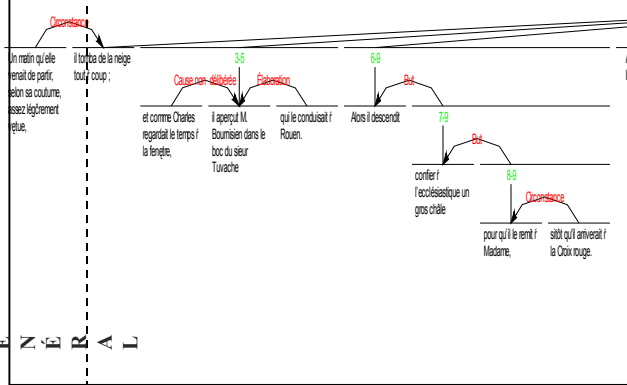
## 5.1.1. Interrelations de RD: un exemple

(7) Un matin qu'elle venait de partir, selon sa coutume, assez légèrement vêtue, il tomba de la neige tout à coup ; et comme Charles regardait le temps à la fenêtre, il aperçut M. Bournisien dans le boc du sieur Tuvache qui le conduisait à Rouen. Alors il descendit confier à l'ecclésiastique un gros châte pour qu'il le remit à Madame, sitôt qu'il arriverait à la *Croix rouge*. A peine fut-il à l'auberge que Bournisien demanda où était la femme du médecin d'Yonville. L'hôtelière répondit qu'elle fréquentait fort peu son établissement. Aussi, le soir, en reconnaissant madame Bovary dans l'Hirondelle, le curé lui conta son embarras, sans paraître, du reste y attacher de l'importance ; car il entama l'éloge d'un prédicateur qui pour lors faisait merveille à la cathédrale, et que toutes les dames couraient entendre. N'importe s'il n'avait point demandé d'explications, d'autres plus tard pourraient se montrer moins discrets. Aussi jugea-t-elle utile de descendre chaque fois à la *Croix rouge*, de sorte que les bonnes gens de son village qui la voyaient dans l'escalier ne se doutaient de rien. (*Madame Bovary*)

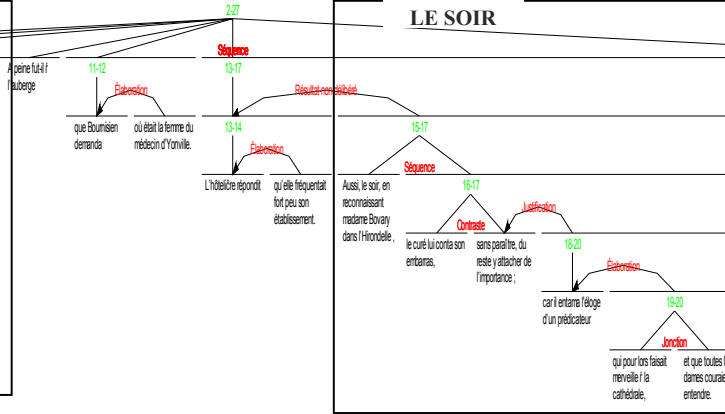
# 5.1.2. Problèmes de délimitation de cadres temporels

UN JOUR

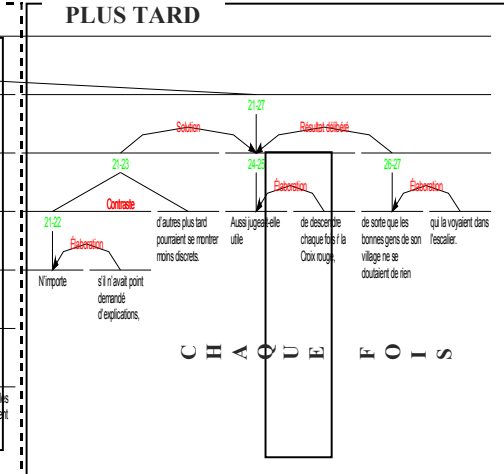
UN MATIN



LE SOIR

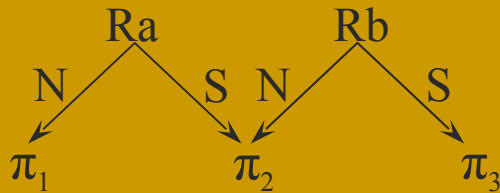


PLUS TARD



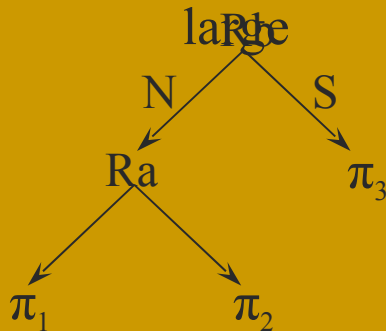
## 5.2. Les interrelations des RD - la portée des connecteurs

### Connecteur à portée étroite



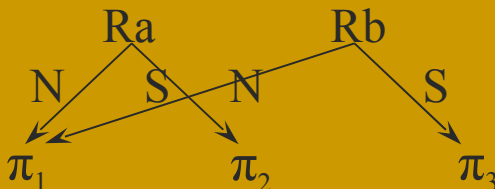
Marie est triste PARCE QUE son fils est malade. EN EFFET, il a mangé trop de choux à la crème.

### Connecteur à portée large



Je suis allé voir Paul PARCE QUE je voulais lui demander un service MAIS il n'était pas à la maison.

### Connecteur à grande envergure



Je vais voir Paul PARCE QUE je veux lui demander un service, je ne peux DONC pas te laisser la voiture.

# 6. Conclusion

- Impératifs et difficultés du traitement des structures discursives en linguistique de corpus

*[...] les études sur le discours se caractérisent actuellement par une approche qualitative, sur des données de faible volume, avec des méthodes manuelles et donc subjectives [...] ce qui fait obstacle [...] à la généralisation de leurs résultats. [...] On a besoin de techniques permettant d'appréhender et d'articuler des fonctionnements souvent enchevêtrés, à différents niveaux de granularité (Péry-Woodley)*

- Perspectives (immédiates) de l'emploi de NooJ
  - Perfectionnement de l'analyse lexicale des SD
  - Combinaison de critères lexicaux et grammaticaux dans l'analyse des RD – coopération entre recherches différentes
  - Annotation structurale et sémantique de corpus
  - Prolongements? (traduction...)

[ Merci de votre attention ]

István CSÜRY

[csury@delfin.unideb.hu](mailto:csury@delfin.unideb.hu)

Université de Debrecen, Département de  
Français

[www.francia.unideb.hu](http://www.francia.unideb.hu)

