

Réaliser et exploiter l'annotation syntaxique de
textes médiévaux
Méthodes et outils du *Syntactic Reference Corpus
of Medieval French*

Nicolas Mazziotta

Université de Liège/Universität Stuttgart, ILR

COST Workshop, Cracovie, 25-26 avril 2013

Plan

Exposé en deux parties :

1. SRCMF

Le projet en quelques mots

Principes généraux de l'analyse

Structures particulières

2. Mise en pratique

Démonstration

1^{re} partie : Syntactic Reference Corpus of Medieval French

Le projet en quelques mots

Principes généraux de l'analyse

Structures particulières

Le projet en quelques mots

Objectif et matériaux

Principes généraux de l'analyse

Un modèle dépendancier

Principes généraux

Délimiter l'unité maximale de traitement

Structures particulières

Prépositions, conjonctions de subordination

Amalgames

Pronom relatif

Coordinations

Objectif et matériaux

Objectif

- ▶ Enrichir les principaux corpus d'afr. avec une **annotation syntaxique**

Objectif et matériaux

Objectif

- ▶ Enrichir les principaux corpus d'afr. avec une **annotation syntaxique**
- ▶ ⇒ Rend possible l'examen des **interactions** syntaxe/autres domaines de la grammaire

Objectif et matériaux

Objectif

- ▶ Enrichir les principaux corpus d'afr. avec une **annotation syntaxique**
- ▶ ⇒ Rend possible l'examen des **interactions** syntaxe/autres domaines de la grammaire

Matériaux

- ▶ Base de Français Médiéval (BFM, Lyon, dir. Céline Guillot) ;
- ▶ Nouveau Corpus d'Amsterdam (NCA, Stuttgart, dir. Achim Stein).

Objectif et matériaux

Objectif

- ▶ Enrichir les principaux corpus d'afr. avec une **annotation syntaxique**
- ▶ ⇒ Rend possible l'examen des **interactions** syntaxe/autres domaines de la grammaire

Matériaux

- ▶ Base de Français Médiéval (BFM, Lyon, dir. Céline Guillot) ;
- ▶ Nouveau Corpus d'Amsterdam (NCA, Stuttgart, dir. Achim Stein).

Annotations présentes :

- ▶ Métadonnées
- ▶ Morphosyntaxe
- ▶ Lemmatisation automatique (NCA)
- ▶ Identification de l'oral représenté (BFM)

Objectif et matériaux

Une annotation à plusieurs niveaux

- ▶ Compléments relevant du verbe (syntaxe propositionnelle)
- ▶ Compléments ne relevant pas d'un verbe

Objectif et matériaux

Une annotation à plusieurs niveaux

- ▶ Compléments relevant du verbe (syntaxe propositionnelle)
- ▶ Compléments ne relevant pas d'un verbe

Compléments relevant du verbe

- ▶ Permet l'étude lexicale et syntaxique des verbes
- ▶ Tous les textes

Objectif et matériaux

Une annotation à plusieurs niveaux

- ▶ Compléments relevant du verbe (syntaxe propositionnelle)
- ▶ Compléments ne relevant pas d'un verbe

Compléments relevant du verbe

- ▶ Permet l'étude lexicale et syntaxique des verbes
- ▶ Tous les textes

Compléments ne relevant pas d'un verbe

- ▶ Description approfondie du SN, p. ex.
- ▶ Certains textes

Le projet en quelques mots

Objectif et matériaux

Principes généraux de l'analyse

Un modèle dépendancier

Principes généraux

Délimiter l'unité maximale de traitement

Structures particulières

Prépositions, conjonctions de subordination

Amalgames

Pronom relatif

Coordinations

Un modèle dépendancier

Positionnement

- ▶ Suivant Tesnière et Mel'čuk
- ▶ Un modèle relativement souple, fondé sur les **relations** unissant les unités lexicales

Un modèle dépendancier

Positionnement

- ▶ Suivant Tesnière et Mel'čuk
- ▶ Un modèle relativement souple, fondé sur les **relations** unissant les unités lexicales

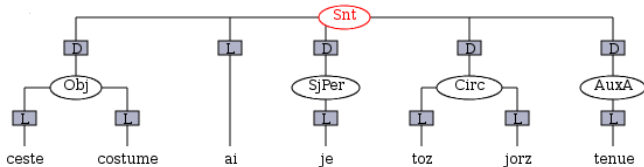
Approche concrète

Un modèle dépendanciel

Positionnement

- ▶ Suivant Tesnière et Mel'čuk
- ▶ Un modèle relativement souple, fondé sur les **relations** unissant les unités lexicales

Approche concrète

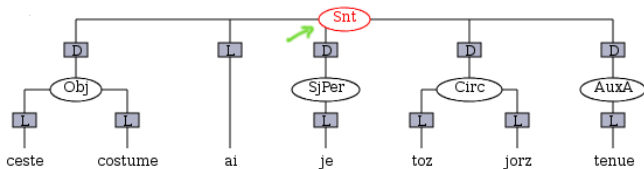


Un modèle dépendancier

Positionnement

- ▶ Suivant Tesnière et Mel'čuk
- ▶ Un modèle relativement souple, fondé sur les **relations** unissant les unités lexicales

Approche concrète



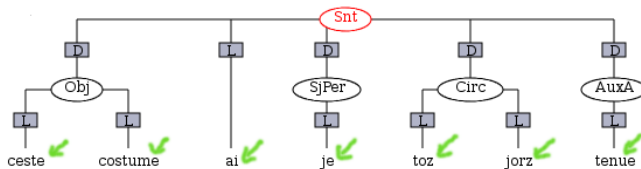
- ▶ Une phrase (Snt), unité maximale de traitement

Un modèle dépendanciel

Positionnement

- ▶ Suivant Tesnière et Mel'čuk
- ▶ Un modèle relativement souple, fondé sur les **relations** unissant les unités lexicales

Approche concrète



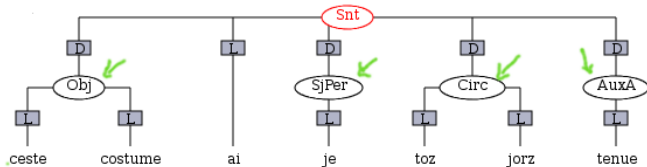
- ▶ Mots, unité minimale de traitement

Un modèle dépendanciel

Positionnement

- ▶ Suivant Tesnière et Mel'čuk
- ▶ Un modèle relativement souple, fondé sur les **relations** unissant les unités lexicales

Approche concrète



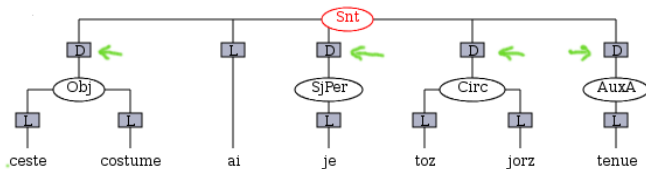
- ▶ Fonctions et des structures grammaticales (ellipses), reliant et groupant les mots

Un modèle dépendanciel

Positionnement

- ▶ Suivant Tesnière et Mel'čuk
- ▶ Un modèle relativement souple, fondé sur les **relations** unissant les unités lexicales

Approche concrète



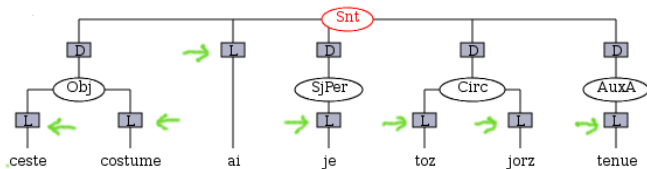
- ▶ Relations de dépendance [D] entre les mots et les structures

Un modèle dépendanciel

Positionnement

- ▶ Suivant Tesnière et Mel'čuk
- ▶ Un modèle relativement souple, fondé sur les **relations** unissant les unités lexicales

Approche concrète



- ▶ Réalisation lexicale [L] des structures

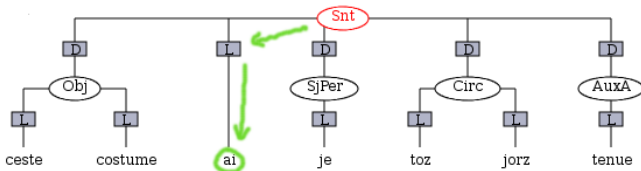
Un modèle dépendancier

Principes généraux

Le verbe dit « fini » est le nœud principal

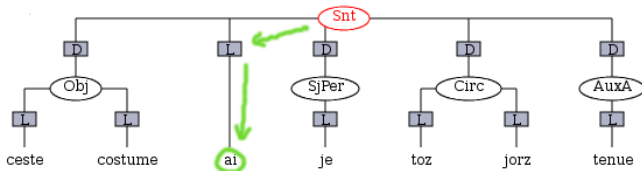
Principes généraux

Le verbe dit « fini » est le nœud principal



Principes généraux

Le verbe dit « fini » est le nœud principal



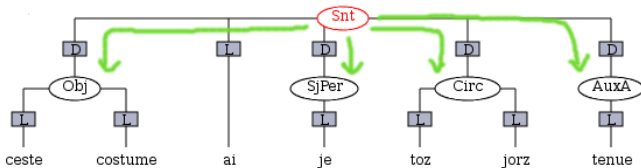
- ▶ Il réalise lexicalement [L] la phrase (Snt)
- ▶ Il en est le nœud principal (ou la « tête »)

Principes généraux

Les autres structures qui constituent Snt dépendent de ce verbe

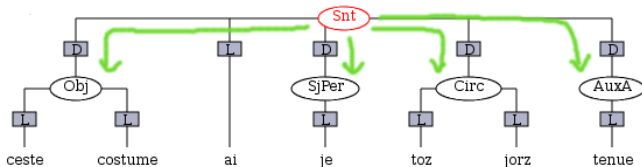
Principes généraux

Les autres structures qui constituent Snt dépendent de ce verbe



Principes généraux

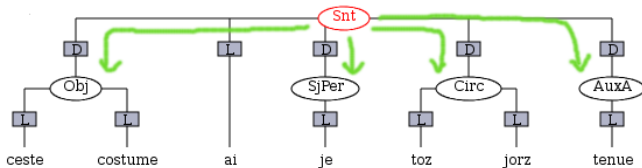
Les autres structures qui constituent Snt dépendent de ce verbe



Chaque structure de **même niveau** a une fonction par rapport au verbe :

Principes généraux

Les autres structures qui constituent Snt dépendent de ce verbe

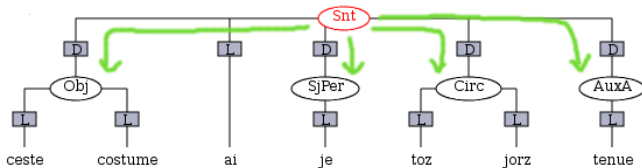


Chaque structure de **même niveau** a une fonction par rapport au verbe :

- ▶ dans sa valence (SjPer, Obj)

Principes généraux

Les autres structures qui constituent Snt dépendent de ce verbe

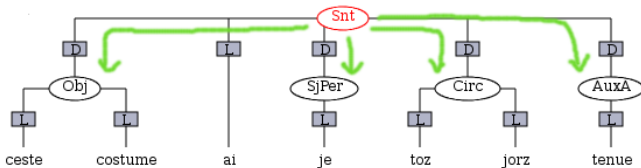


Chaque structure de **même niveau** a une fonction par rapport au verbe :

- ▶ dans sa valence (SjPer, Obj)
- ▶ hors de sa valence (Circ)

Principes généraux

Les autres structures qui constituent Snt dépendent de ce verbe



Chaque structure de **même niveau** a une fonction par rapport au verbe :

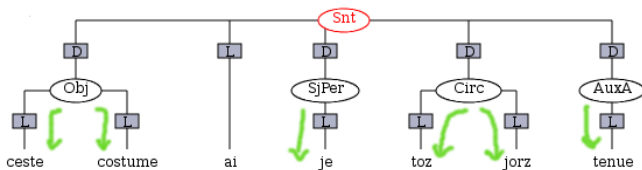
- ▶ dans sa valence (SjPer, Obj)
- ▶ hors de sa valence (Circ)
- ▶ liée à sa morphologie et modifiant éventuellement la valence (AuxA)

Principes généraux

Les dépendants du verbe sont rérealisés lexicalement

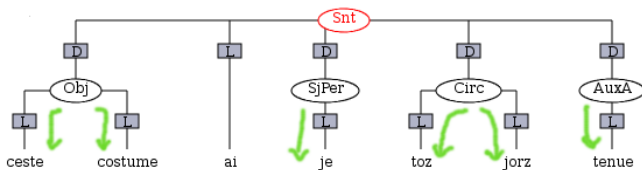
Principes généraux

Les dépendants du verbe sont réalisés lexicalement



Principes généraux

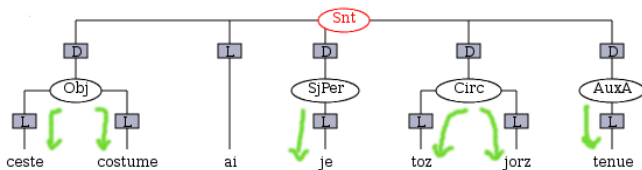
Les dépendants du verbe sont réalisés lexicalement



- ... Ou bien ils constituent une structure propositionnelle (récursivité)

Principes généraux

Les dépendants du verbe sont réalisés lexicalement



- ▶ ... Ou bien ils constituent une structure propositionnelle (récursivité)
- ▶ Si l'annotation n'est pas complète (ici), plusieurs mots réalisent simultanément la structure

Principes généraux

Fonctions principales des dépendants du verbe (hiérarchisées)

- ▶ Actant

Principes généraux

Fonctions principales des dépendants du verbe (hiérarchisées)

- ▶ Actant
 - ▶ Sujet (SjPer/SjImp)

Principes généraux

Fonctions principales des dépendants du verbe (hiérarchisées)

- ▶ Actant
 - ▶ Sujet (SjPer/SjImp)
 - ▶ Régime (Regim)
 - ▶ Objet (Obj)
 - ▶ Complément (Cmpl)

Principes généraux

Fonctions principales des dépendants du verbe (hiérarchisées)

- ▶ Actant
 - ▶ Sujet (SjPer/SjImp)
 - ▶ Régime (Regim)
 - ▶ Objet (Obj)
 - ▶ Complément (Cmpl)
- ▶ Attribut
 - ▶ Attribut du sujet (AtSj)
 - ▶ Attribut de l'objet (AtObj)

Principes généraux

Fonctions principales des dépendants du verbe (hiérarchisées)

- ▶ Actant
 - ▶ Sujet (SjPer/SjImp)
 - ▶ Régime (Regim)
 - ▶ Objet (Obj)
 - ▶ Complément (Cmpl)
- ▶ Attribut
 - ▶ Attribut du sujet (AtSj)
 - ▶ Attribut de l'objet (AtObj)
- ▶ Auxilié (Aux)
 - ▶ Auxilié actif (AuxA)
 - ▶ Auxilié passif (AuxP)

Principes généraux

Fonctions principales des dépendants du verbe (hiérarchisées)

- ▶ Actant
 - ▶ Sujet (SjPer/SjImp)
 - ▶ Régime (Regim)
 - ▶ Objet (Obj)
 - ▶ Complément (Cmpl)
- ▶ Attribut
 - ▶ Attribut du sujet (AtSj)
 - ▶ Attribut de l'objet (AtObj)
- ▶ Auxilié (Aux)
 - ▶ Auxilié actif (AuxA)
 - ▶ Auxilié passif (AuxP)
- ▶ Circonstant (Circ)

Délimiter l'unité maximale de traitement

Difficulté

Délimiter l'unité maximale de traitement

Difficulté

- ▶ Le terme *phrase* est tributaire d'une longue tradition grammaticale

Délimiter l'unité maximale de traitement

Difficulté

- ▶ Le terme *phrase* est tributaire d'une longue tradition grammaticale
- ▶ La distinction entre *phrase* et *énoncé* est loin d'être consensuelle

Délimiter l'unité maximale de traitement

Difficulté

- ▶ Le terme *phrase* est tributaire d'une longue tradition grammaticale
- ▶ La distinction entre *phrase* et *énoncé* est loin d'être consensuelle
- ▶ Certaines relations de coordination de phrase sont difficile à démêler

Délimiter l'unité maximale de traitement

Difficulté

- ▶ Le terme *phrase* est tributaire d'une longue tradition grammaticale
- ▶ La distinction entre *phrase* et *énoncé* est loin d'être consensuelle
- ▶ Certaines relations de coordination de phrase sont difficile à démêler

Une série de règles

Délimiter l'unité maximale de traitement

Difficulté

- ▶ Le terme *phrase* est tributaire d'une longue tradition grammaticale
- ▶ La distinction entre *phrase* et *énoncé* est loin d'être consensuelle
- ▶ Certaines relations de coordination de phrase sont difficile à démêler

Une série de règles

- ▶ 1 verbe fini indépendant = 1 phrase

Délimiter l'unité maximale de traitement

Difficulté

- ▶ Le terme *phrase* est tributaire d'une longue tradition grammaticale
- ▶ La distinction entre *phrase* et *énoncé* est loin d'être consensuelle
- ▶ Certaines relations de coordination de phrase sont difficile à démêler

Une série de règles

- ▶ 1 verbe fini indépendant = 1 phrase
- ▶ Pas de coordination de phrases

Délimiter l'unité maximale de traitement

Difficulté

- ▶ Le terme *phrase* est tributaire d'une longue tradition grammaticale
- ▶ La distinction entre *phrase* et *énoncé* est loin d'être consensuelle
- ▶ Certaines relations de coordination de phrase sont difficile à démêler

Une série de règles

- ▶ 1 verbe fini indépendant = 1 phrase
- ▶ Pas de coordination de phrases
- ▶ Pas de partage de compléments entre les verbes (rattachement au plus proche)

Délimiter l'unité maximale de traitement

Difficulté

- ▶ Le terme *phrase* est tributaire d'une longue tradition grammaticale
- ▶ La distinction entre *phrase* et *énoncé* est loin d'être consensuelle
- ▶ Certaines relations de coordination de phrase sont difficile à démêler

Une série de règles

- ▶ 1 verbe fini indépendant = 1 phrase
- ▶ Pas de coordination de phrases
- ▶ Pas de partage de compléments entre les verbes (rattachement au plus proche)

Conséquences

- ▶ Des postulats qui affectent toute l'analyse et refusent toute exception
- ▶ Une analyse limitée

Le projet en quelques mots

Objectif et matériaux

Principes généraux de l'analyse

Un modèle dépendancier

Principes généraux

Délimiter l'unité maximale de traitement

Structures particulières

Prépositions, conjonctions de subordination

Amalgames

Pronom relatif

Coordinations

Prépositions, conjonctions de subordination

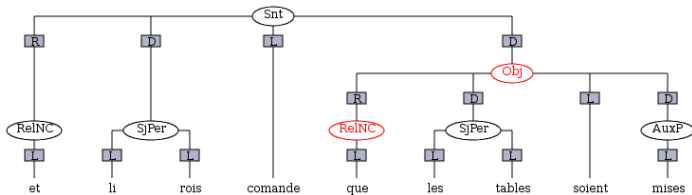
Unités de relation nommées *relateurs*

- ▶ Relateurs coordonnants (RelC, cf. infra)
- ▶ Relateurs non coordonnants (RelNC)

Prépositions, conjonctions de subordination

Unités de relation nommées *relateurs*

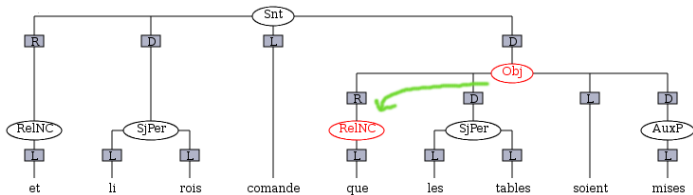
- ▶ Relateurs coordonnants (RelC, cf. infra)
- ▶ Relateurs non coordonnants (RelNC)



Prépositions, conjonctions de subordination

Unités de relation nommées *relateurs*

- ▶ Relateurs coordonnants (RelC, cf. infra)
- ▶ Relateurs non coordonnants (RelNC)

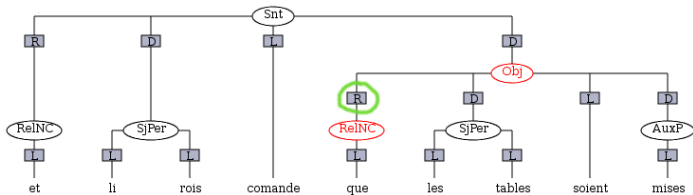


- ▶ Le RelNC « subordonnant » dépend du verbe comme les compléments (pas de « translation » ou de rection)

Prépositions, conjonctions de subordination

Unités de relation nommées *relateurs*

- ▶ Relateurs coordonnants (RelC, cf. infra)
- ▶ Relateurs non coordonnants (RelNC)

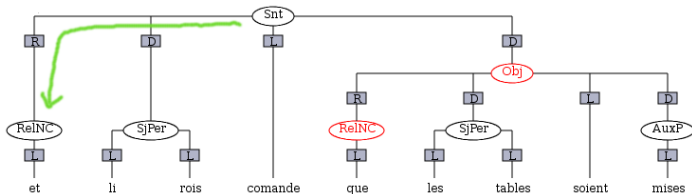


- ▶ Le RelNC « subordonnant » dépend du verbe comme les compléments (pas de « translation » ou de rection)
- ▶ Le RelNC entretient une relation particulière [R]

Prépositions, conjonctions de subordination

Unités de relation nommées *relateurs*

- ▶ Relateurs coordonnants (RelC, cf. infra)
- ▶ Relateurs non coordonnants (RelNC)



- ▶ Le RelNC « subordonnant » dépend du verbe comme les compléments (pas de « translation » ou de rection)
- ▶ Le RelNC entretient une relation particulière [R]
- ▶ Le RelNC peut introduire une phrase

Amalgames

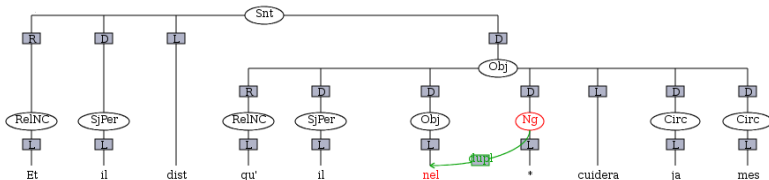
Fusion de deux fonctions

- ▶ Une seule forme indécomposable graphiquement
- ▶ Plusieurs fonctions
- ▶ L'unité est dupliquée dans la représentation

Amalgames

Fusion de deux fonctions

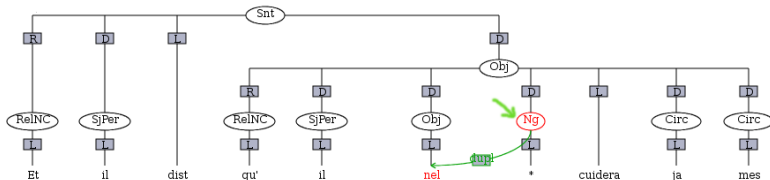
- ▶ Une seule forme indécomposable graphiquement
- ▶ Plusieurs fonctions
- ▶ L'unité est dupliquée dans la représentation



Amalgames

Fusion de deux fonctions

- ▶ Une seule forme indécomposable graphiquement
- ▶ Plusieurs fonctions
- ▶ L'unité est dupliquée dans la représentation

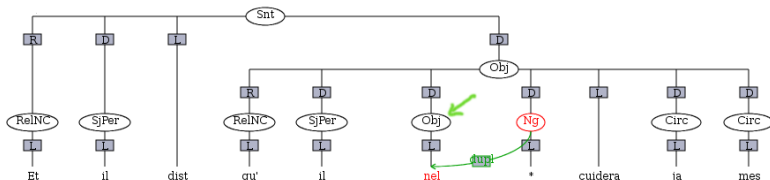


- ▶ *nel* contient l'expression de la négation

Amalgames

Fusion de deux fonctions

- ▶ Une seule forme indécomposable graphiquement
- ▶ Plusieurs fonctions
- ▶ L'unité est dupliquée dans la représentation

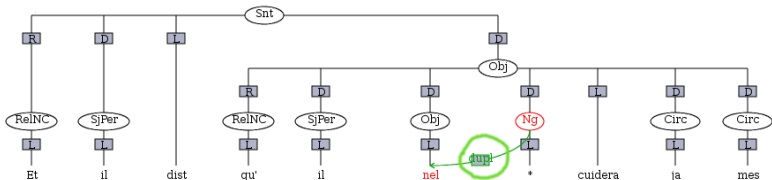


- ▶ *nel* contient l'expression de la négation **et** l'expression de l'objet

Amalgames

Fusion de deux fonctions

- ▶ Une seule forme indécomposable graphiquement
- ▶ Plusieurs fonctions
- ▶ L'unité est dupliquée dans la représentation



- ▶ *nel* contient l'expression de la négation **et** l'expression de l'objet
- ▶ La relation de duplication [dup] est explicite

Pronom relatif

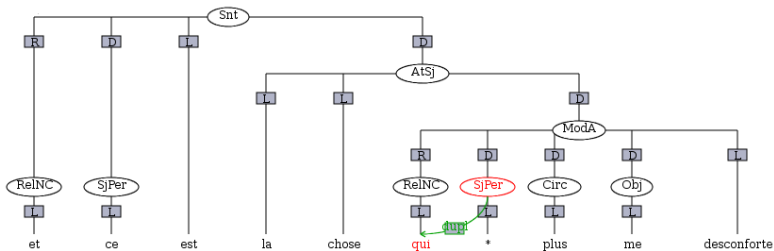
Traitement similaire aux amalgames

- ▶ Une seule forme indécomposable graphiquement
- ▶ Plusieurs fonctions
- ▶ L'unité est dupliquée dans la représentation

Pronom relatif

Traitement similaire aux amalgames

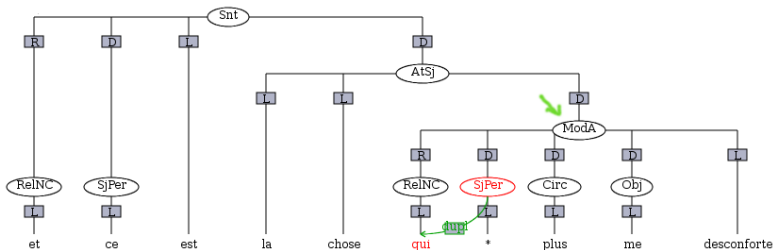
- ▶ Une seule forme indécomposable graphiquement
- ▶ Plusieurs fonctions
- ▶ L'unité est dupliquée dans la représentation



Pronom relatif

Traitement similaire aux amalgames

- ▶ Une seule forme indécomposable graphiquement
- ▶ Plusieurs fonctions
- ▶ L'unité est dupliquée dans la représentation

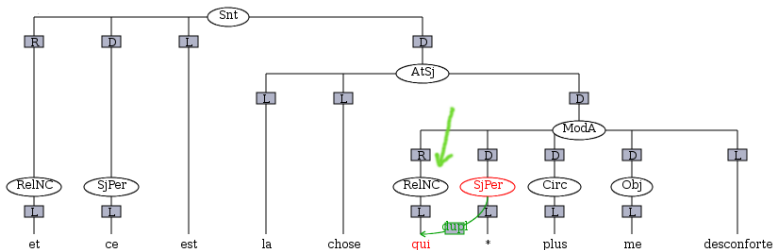


- ▶ La relative *modifie* un nom (ModA)
- ▶ En tant que proposition, sa structure est analysée

Pronom relatif

Traitement similaire aux amalgames

- ▶ Une seule forme indécomposable graphiquement
- ▶ Plusieurs fonctions
- ▶ L'unité est dupliquée dans la représentation

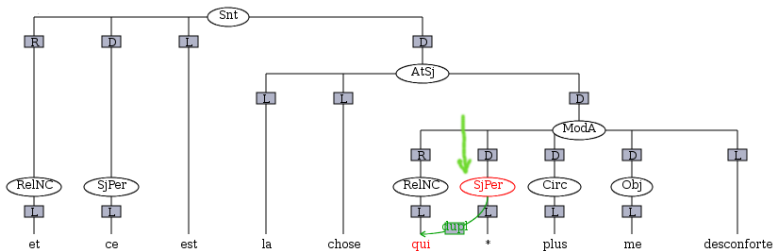


- ▶ La relative *modifie* un nom (ModA)
- ▶ En tant que proposition, sa structure est analysée
- ▶ *qui* fonctionne comme un RelNC

Pronom relatif

Traitement similaire aux amalgames

- ▶ Une seule forme indécomposable graphiquement
- ▶ Plusieurs fonctions
- ▶ L'unité est dupliquée dans la représentation

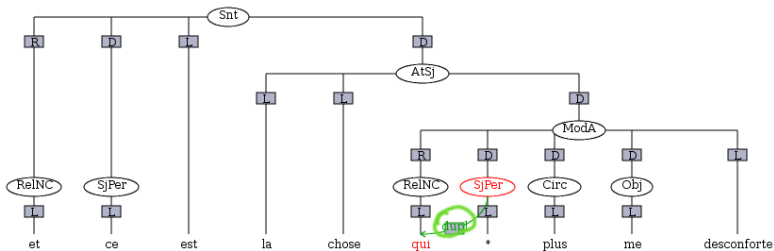


- ▶ La relative *modifie* un nom (ModA)
- ▶ En tant que proposition, sa structure est analysée
- ▶ *qui* fonctionne comme un RelNC **et** comme un SjPer

Pronom relatif

Traitement similaire aux amalgames

- ▶ Une seule forme indécomposable graphiquement
- ▶ Plusieurs fonctions
- ▶ L'unité est dupliquée dans la représentation



- ▶ La relative *modifie* un nom (ModA)
- ▶ En tant que proposition, sa structure est analysée
- ▶ *qui* fonctionne comme un RelNC **et** comme un SjPer
- ▶ La relation de duplication [dupl] est explicite

Coordinations

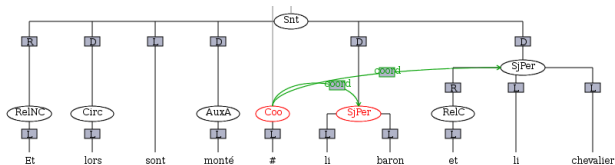
Spécificité de la relation

- ▶ Sur un autre plan que la dépendance (orthogonal)
- ▶ Lie deux unités qui dépendent d'une même autre

Coordinations

Spécificité de la relation

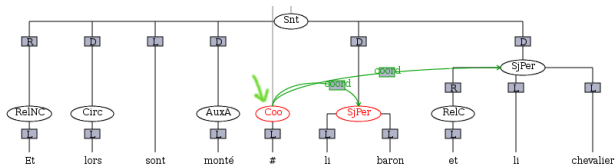
- ▶ Sur un autre plan que la dépendance (orthogonal)
- ▶ Lie deux unités qui dépendent d'une même autre



Coordinations

Spécificité de la relation

- ▶ Sur un autre plan que la dépendance (orthogonal)
- ▶ Lie deux unités qui dépendent d'une même autre

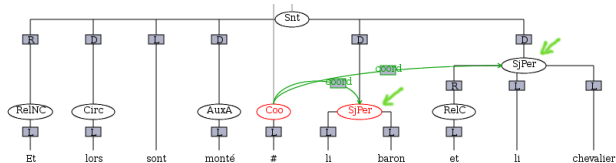


- ▶ La coordination est une structure parallèle qui n'est pas dans le réseau de dépendance de Snt. Elle n'a pas de réalisation lexicale

Coordinations

Spécificité de la relation

- ▶ Sur un autre plan que la dépendance (orthogonal)
- ▶ Lie deux unités qui dépendent d'une même autre

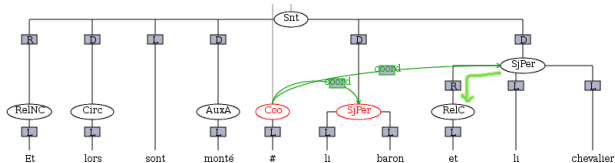


- ▶ La coordination est une structure parallèle qui n'est pas dans le réseau de dépendance de Snt. Elle n'a pas de réalisation lexicale
- ▶ Les termes coordonnés ne sont pas dépendants de la coordination

Coordinations

Spécificité de la relation

- ▶ Sur un autre plan que la dépendance (orthogonal)
- ▶ Lie deux unités qui dépendent d'une même autre

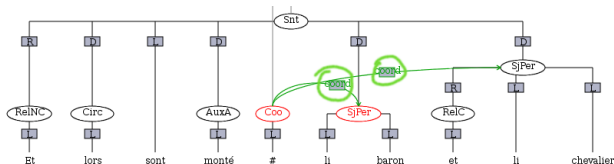


- ▶ La coordination est une structure parallèle qui n'est pas dans le réseau de dépendance de Snt. Elle n'a pas de réalisation lexicale
- ▶ Les termes coordonnés ne sont pas dépendants de la coordination
- ▶ Le relateur coordonnant dépend du terme suivant

Coordinations

Spécificité de la relation

- ▶ Sur un autre plan que la dépendance (orthogonal)
- ▶ Lie deux unités qui dépendent d'une même autre



- ▶ La coordination est une structure parallèle qui n'est pas dans le réseau de dépendance de Snt. Elle n'a pas de réalisation lexicale
- ▶ Les termes coordonnés ne sont pas dépendants de la coordination
- ▶ Le relateur coordonnant dépend du terme suivant
- ▶ La relation de coordination [dupl] est explicite

2^e partie : Démonstration

Format des textes en deux mots

NotaBene

Exportation

TigerSearch

Format des textes en deux mots

NotaBene

Exportation

TigerSearch

Format des textes en deux mots

NotaBene

Exportation

TigerSearch

Format des textes en deux mots

NotaBene

Exportation

TigerSearch

Format des textes en deux mots

NotaBene

Exportation

TigerSearch