Analyse du flux de dépendance dans un corpus de français oral annoté en microsyntaxe

Marie-Amélie Botalla

Mémoire dirigé par Sylvain Kahane

Master d'Ingénierie Linguistique, parcours Recherche et Développement
Université Paris III Sorbonne Nouvelle

Table des matières

Chapitre 1

Introduction

Ce mémoire a pour sujet l'étude du flux de dépendance sur un corpus de français parlé annoté manuellement en dépendance. Le flux de dépendance est défini comme étant l'ensemble des dépendances qui relient un mot à gauche à un mot à droite d'une position inter-mots donnée (Kahane 2001, Jardonnet 2009).

Nous nous pencherons sur la complexité du flux de dépendance et sur ce qui génère cette complexité. Nous nous intéresserons aussi aux limites que rencontre le flux, ainsi qu'aux causes de ces limites.

1.1. Flux, compétence et performance

La notion de "compétence" (Chomsky 1965) désigne tout ce qu'un locuteur d'une langue est capable de produire dans cette langue, son savoir linguistique. De part la nature récursive de certaines constructions syntaxiques, le locuteur a la possibilité d'enchâsser un nombre infini de propositions.

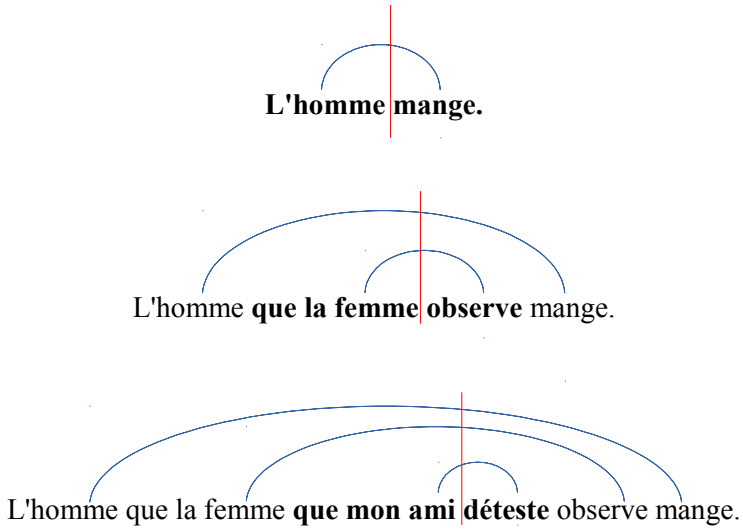

Fig.1.1. : Exemple d'enchâssements consécutifs

En figure 1.1, on a représenté la dépendance entre le verbe et son sujet, puis ajouté successivement des propositions subordonnées relatives, dont l'antécédent est inclus dans la proposition ajoutée à l'étape précédente. On remarque qu'à la position symbolisée par la ligne rouge, la taille du flux est en constante augmentation : une dépendance pour la première phrase, deux pour la deuxième, trois pour la troisième. La compétence n'entraîne donc pas de limitation du flux.

La notion de "performance" (Chomsky 1965) désigne quant à elle tout ce qu'un locuteur produit réellement. Elle se différencie de la compétence par la possibilité qu'a le locuteur de faire des erreurs et par une limitation due aux capacités cognitives du locuteur, dont la mémoire de travail ne pourrait traiter que 7±2 éléments à la fois (Miller 1956, Yngve 1960).

L'hypothèse de Miller est vérifiable en observant le nombre maximal de dépendances passant par une position. En effet, l'analyse d'une phrase par un récepteur se fait au fur et à mesure de la réception des mots. Le flux représentant a priori ce que le récepteur garde en mémoire pour analyser la suite de la phrase, on peut penser que les contraintes mémorielles portent sur le flux. Si on observe que le flux est borné, on peut en déduire qu'il existe très certainement une limitation mémorielle sur la taille maximale du flux.

Les premières grammaires formelles ont tenté de modéliser la compétence, et ce n'est qu'à partir du développement des corpus annotés en syntaxe, ou treebanks, que sont apparues les premières grammaires de la performance. Se pose alors la question suivante : une grammaire de la performance n'est-elle rien d'autre qu'une grammaire de la compétence augmentée par les paramètres de la limitation mémorielle et des erreurs potentielles, ou est-elle une grammaire fondamentalement différente ?

1.2. Définitions

Ce mémoire fait appel à différentes notions, qui sont définies ci-après.

Flux : le flux de dépendance est défini comme étant l'ensemble des dépendances syntaxiques qui relient un mot à gauche à un mot à droite d'une position inter-mots donnée.

Taille du flux : la taille du flux en une position donnée désigne le nombre de dépendances passant par cette position.

Token : un token est un segment orthographique délimité par deux blancs, ou par un blanc et un signe de ponctuation.

Mot : le mot désigne ici la plus petite unité syntaxique traitée dans le corpus. Un mot est la plupart du temps constitué d'un seul token, mais les nombres, les noms propres, les noms composés et certains mots grammaticaux constitués de plusieurs tokens sont considérés comme un seul mot.

Dépendance/Lien : on désigne par "dépendance" ou "lien" une dépendance syntaxique entre deux mots, l'un désigné par le terme de "gouverneur", l'autre par le terme de "dépendant". Visuellement, une dépendance est représentée par une flèche partant du gouverneur et pointant sur le dépendant.

Gouverneur et dépendant : dans une dépendance, le gouverneur est le mot qui légitime la présence du dépendant.

Projectivité : le flux est dit projectif en une position dans deux cas :

- aucune dépendance n'en croise d'autre : pour deux dépendances A-B et C-D telles que A est avant B et C est avant D, si A est avant C, alors D est avant B.

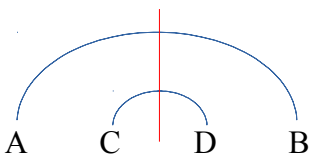

Fig. 1.2. : Projectivité

Dans la figure 1.2, le flux est projectif pour la position représentée par la ligne rouge.

- tous les descendants d'un mot, c'est-à-dire les mots qui sont gouvernés par ses dépendants, se trouvent dans sa projection : pour trois mots A, B et C tels que A gouverne B et B gouverne C, les mots B et C apparaissent soit tous les deux avant, soit tous les deux après A.

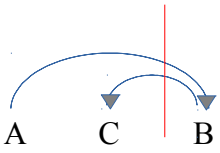

Fig. 1.3. : Projectivité

Dans la figure 1.3, les mots B et C apparaissent après A : tous les descendants du mot A se trouvent dans sa projection. Le flux entre C et B, indiqué par la ligne rouge, est projectif.

Croisement : deux dépendances A-B et C-D se croisent lorsque les mots apparaissent dans l'ordre A, C, B, D. S'il y a croisement, alors le premier cas de projectivité du flux n'est pas respecté.

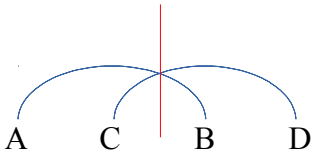

Fig. 1.4. : Croisement

En figure 1.4, on observe un croisement à la position représentée par la ligne rouge.

Insubordination : il y a insubordination lorsque une dépendance couvre le gouverneur de son gouverneur, c'est-à-dire si, pour trois mots A, B et C tels que A gouverne B et B gouverne C, le mot A se trouve entre les mots B et C. S'il y a insubordination, alors le second cas de projectivité du flux n'est pas respecté.

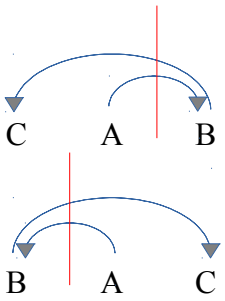

Fig. 1.5. : Insubordinations

La figure 1.5 présente les deux cas d'insubordination possibles dans le corpus : le flux entre A et B, représenté par la ligne rouge, est non projectif.

Disjonction : le flux (ou un sous-ensemble du flux) est dit disjoint en une position si chaque mot à gauche de cette position n'entretient de dépendance qu'avec un seul mot à droite de cette position, et vice versa.

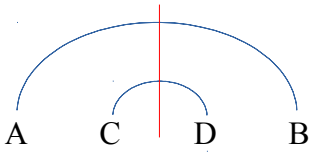

Fig. 1.6. : Disjonction

Dans la figure 1.6, le flux est disjoint à la position représentée par la ligne rouge : chaque mot entretient une dépendance exclusive. Celui de la figure 1.3 n'est pas disjoint.

Dépendance exclusive : pour un flux donné, une dépendance est dite exclusive si aucune de ses extrémités n'appartient à une autre dépendance du flux. Un flux disjoint n'a que des dépendances exclusives. Dans la figure 1.6, le mot A entretient une dépendance exclusive avec le mot B, et le mot C avec le mot D. Dans le cas contraire, on parle de conjonctivité.

1.3. État de l'art

1.3.1. Transition-Based Dependency Parsing (TBDP)

Le flux de dépendance peut être rapproché de la pile S du Transition-Based Dependency Parsing développé par Joakim Nivre (2004).

Dans cet algorithme de parsing en dépendance, on dispose d'une pile S de mots partiellement traités, d'une pile B de mots en attente de traitement, et d'un ensemble A de liens de dépendance. On ajoute des liens à l'ensemble A en effectuant différentes opérations.

On désignera respectivement par *b* et *s* les mots aux sommets des piles B et S.

- Chaque mot de la pile B est déplacé au fur et à mesure de la lecture vers la pile S. Cette opération est nommée "shift".
- Si *b* est le gouverneur de *s*, alors on ajoute un lien $s \leftarrow b$ à l'ensemble A et on supprime *s* de la pile S. Cette opération est nommée "left-arc".
- Si *s* est le gouverneur de *b*, alors on ajoute un lien $s \rightarrow b$ à l'ensemble A, et on déplace *b* vers la pile S. Cette opération est nommée "right-arc".
- Si *s* et *b* n'entretiennent pas de lien de dépendance, mais que *b* entretient un lien avec s_{-1}, le prédécesseur de *s* dans S, et que *s* a déjà un gouverneur, on supprime *s* de la pile S. Cette opération est nommée "reduce". Si *b* n'entretient pas de lien avec s_{-1}, on a recours à l'opération "shift" décrite précédemment.

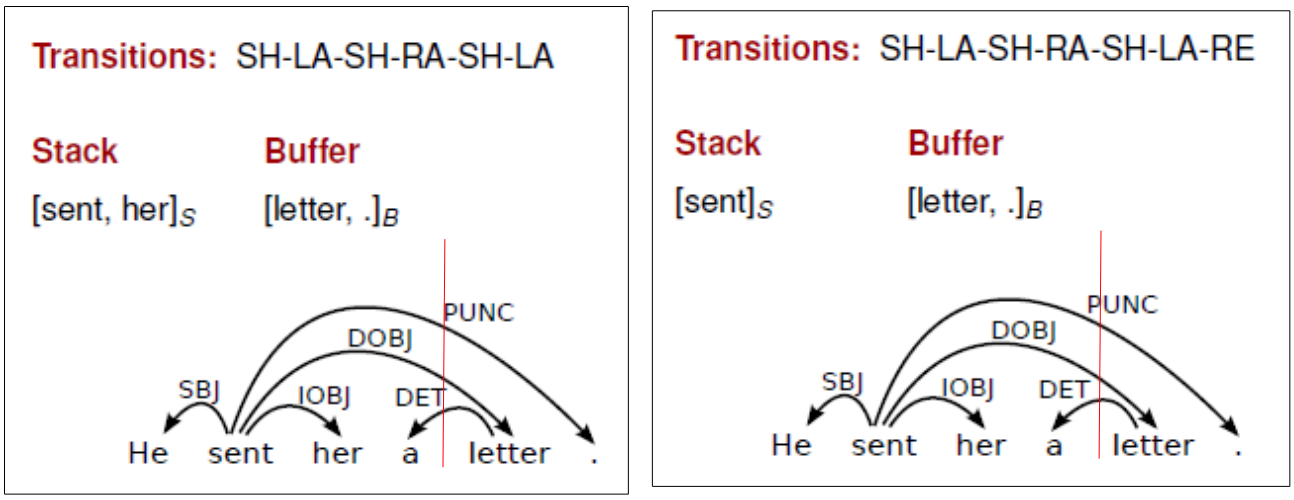

Fig. 1.7. : Exemple d'utilisation de l'opération "reduce" (Nivre 2013, pp. 14-15)

Dans le cas présenté en figure 1.7, la ligne rouge représente la position de la tête de lecture de l'algorithme : b = "letter", s = "her" et s_{-1} = "sent". Comme indiqué par l'arbre de dépendance, s et b n'entretiennent pas de lien, mais b entretient un lien avec s_{-1} et s a déjà un gouverneur. On a donc recours à l'opération "reduce".

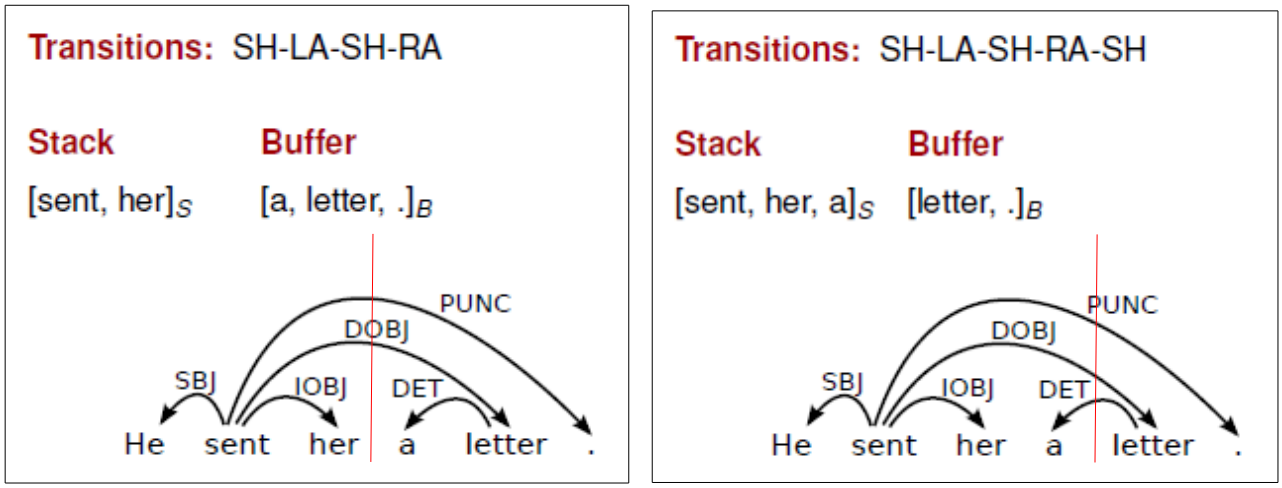

Fig. 1.8. : Exemple d'utilisation de l'opération "shift" (Nivre 2013, pp. 12-13)

Dans le cas présenté en figure 1.8, la ligne rouge représente la position de la tête de lecture de l'algorithme : b = "a", s = "her" et s_{-1} = "sent". Comme indiqué par l'arbre de dépendance, s et b n'entretiennent pas de lien, et b n'entretient pas de lien avec s_{-1}. On a donc recours à l'opération "shift".

Les éléments présents dans la pile S à un moment donné sont ceux que le récepteur de la phrase garde en mémoire : ils sont soit en attente de leur gouverneur, soit susceptibles de gouverner un ou plusieurs des mots à venir. De façon similaire, les dépendances présentes dans le flux à une position donnée sont celles dont une des deux extrémités a déjà été énoncée. La pile S à un moment donné peut donc être comparée au flux à la position correspondante. Mais à la différence du flux, la pile S contient a priori tous les mots susceptibles d'entretenir à une dépendance avec un mot à la suite, alors que le flux ne contient que les dépendances effectivement réalisées.

Le transition-based parsing a néanmoins des limites : il ne permet pas d'analyser les cas où le flux est non projectif sans l'ajout de transitions adéquates, lesquelles complexifient l'algorithme et

rendent la procédure nettement moins déterministe et donc moins efficace (Nivre 2006). En effet, l'analyse de cas non projectifs nécessite de relier *b* à s_{-1} sans que *s* n'ait trouvé son gouverneur.

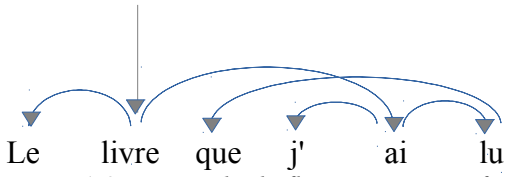

Fig. 1.9 : Exemple de flux non projectif

L'analyse du syntagme en figure 1.9 se déroule ainsi :

0.
S = [] ; B = [le, livre, que, j', ai, lu]

1. SH
S = [le] ; B = [livre, que, j', ai, lu]

2. SH-LA
S = [] ; B = [livre, que, j', ai, lu]

3. SH-LA-SH
S = [livre] ; B = [que, j', ai, lu]
4. SH-LA-SH-SH
S = [livre, que] ; B = [j', ai, lu]

5. SH-LA-SH-SH-SH
S = [livre, que, j'] ; B = [ai, lu]

6. SH-LA-SH-SH-SH-LA
S = [livre, que] : B = [ai, lu]

À cette étape, *b* = "ai", *s* = "que" et s_{-1} = "livre". L'algorithme est bloqué, car *b* et s_{-1} entretiennent un lien, mais l'opération "reduce" est impossible car *s* n'a pas encore trouvé son gouverneur.

1.3.2. Le flux de dépendance dans un corpus de français écrit

Une première analyse du flux de dépendance en français écrit a été réalisée par (Jardonnet 2009). Le French Treebank, corpus constitué de phrases issues du quotidien *Le Monde* et développé par le Laboratoire de Linguistique Formelle de l'université Paris 7, a servi de base à cette analyse, qui ne s'est toutefois concentrée que sur la nature bornée de la taille du flux.

La limite de 7±2 définie par (Miller 1956) a pu être validée après quelques modifications, telles que l'élimination des dépendances pour les signes de ponctuation et l'aplatissement de certaines dépendances dites "en bouquet" utilisés dans ce treebank pour l'analyse des coordinations. Les dépendances en bouquet sont une succession de dépendances ayant le même gouverneur : si les dépendants remplissent des fonctions similaires, on peut considérer que chaque dépendant est en fait lié au dépendant précédent par un lien paradigmatique, ce qui diminue parfois considérablement

la taille du flux.

Toutefois, le français oral présente des marques caractéristiques (hésitations, corrections, ...) qu'on ne retrouve pas en français écrit. L'étude du flux de dépendance dans un corpus de français oral fournira donc vraisemblablement des résultats différents de ceux présentés dans (Jardonnet 2009). De plus, le flux de dépendance dépend en partie de la méthode d'analyse choisie, comme le montre les deux méthodes d'analyse de la coordination au paragraphe précédent (voir les différentes analyses des coordinations dans (Pretkalniņa et al. 2014)) : certaines caractéristiques du flux peuvent inciter à réviser l'analyse initialement choisie.

1.4. Problématique

Quelles sont les limitations que rencontre le flux ? À quoi sont dues ces limitations ? Qu'est-ce qui génère de la complexité dans le flux ? Qu'est-ce qui influence la taille du flux ? Quelles sont les conditions d'apparition d'un flux non projectif ? Croisement et insubordination sont-ils cooccurrents ? Nous tenterons au cours de ce mémoire de répondre à la plupart de ces questions : les hypothèses que nous énoncerons seront validées ou invalidées par les résultats issus de l'analyse du corpus Rhapsodie.

Nous commencerons pour cela par présenter la syntaxe de dépendance et les treebanks en dépendance, ainsi que le treebank Rhapsodie sur lequel s'appuie ce mémoire (partie 2). Nous introduirons ensuite un moyen de représenter le flux de dépendance (partie 3), qui nous aidera à aborder le cas du flux non projectif (partie 4) et de la taille du flux disjoint (partie 5). Enfin, nous aborderons les critiques et perspectives que cette étude aura mises en évidence.

Chapitre 2

Le treebank Rhapsodie

Le treebank Rhapsodie est un corpus de français parlé annoté en dépendance (ainsi qu'en prosodie et en macrosyntaxe). Afin de mieux comprendre son fonctionnement, nous aborderons pour commencer la syntaxe de dépendance, ainsi que les treebanks en dépendance. Nous présenterons ensuite le projet Rhapsodie, ainsi que le projet Orfeo dans lequel il s'inscrit. Nous terminerons en présentant les choix d'analyse syntaxique faits lors de la construction du treebank Rhapsodie.

2.1. La syntaxe de dépendance

La syntaxe de dépendance considère les phrases sous l'angle des relations hiérarchiques que les mots entretiennent entre eux. Chaque mot est relié à au moins un autre mot, à l'exception d'un mot central, qui est le plus souvent le verbe principal de la phrase (Bérard 2012). La présence de chaque mot est donc légitimée par la présence d'un autre mot, qui remplit par là le rôle de gouverneur syntaxique. Si un mot *x* gouverne un mot *y* et que ce mot *y* gouverne un mot *z*, alors *x* légitime la présence de *y* mais aussi celle de *z*. Par transitivité, un mot qui remplit le rôle de gouverneur syntaxique légitime la présence de tout un groupe de mots (Kahane 2001).

L'analyse en dépendance d'une phrase peut être représentée à plat ou sous la forme d'un arbre.

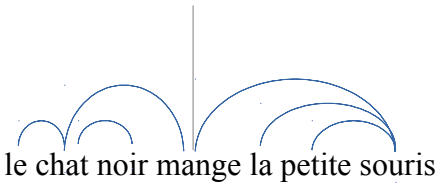

Fig. 2.1. : Dépendance à plat

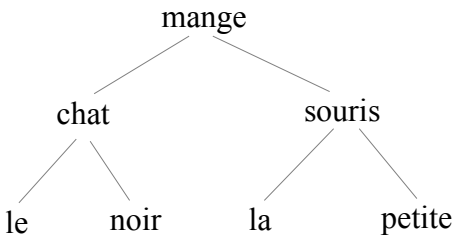

Fig. 2.2. : Dépendance en arbre

Dans les figures 2.1 et 2.2, on a représenté la phrase "le chat noir mange la petite souris" premièrement en indiquant les dépendances que les mots entretiennent entre eux, puis en organisant les mots et les liens sous la forme d'un arbre. Dans cet arbre, les nœuds sont directement représentés par les mots, et non pas par des catégories grammaticales ; il est aussi possible d'étiqueter les liens avec la fonction syntaxique du dépendant.

2.2. Les treebanks en dépendance

Un treebank est un corpus enrichi par une analyse syntaxique et/ou sémantique, voire prosodique dans le cas d'un corpus oral. Le premier treebank publié, et sans doute le plus connu, est le Penn Treebank de l'université de Pennsylvanie, qui propose une annotation syntaxique et sémantique d'un corpus en anglais (Marcus et al. 1993). Ce treebank a fait des émules et il existe maintenant de nombreux treebanks couvrant une grande variété de langues.

Dans le cas d'un treebank dit "en dépendance", l'analyse syntaxique présentée est une analyse en dépendance. Le premier et principal treebank en dépendance est le Prague Dependency Treebank pour le tchèque (Hajič et al. 2001).

Les treebanks permettent aux linguistes d'étudier différentes constructions syntaxiques à partir de données attestées, mais aussi d'entraîner et de tester des analyseurs syntaxiques automatiques, comme le transition-based parser présenté plus haut. A l'inverse, les parsers permettent de développer plus rapidement des treebanks en proposant une analyse automatique des données, que l'on pourra ensuite corriger ou non manuellement.

2.3. Présentation des projets Rhapsodie et Orfeo

Le corpus utilisé pour ce mémoire est issu du projet ANR Rhapsodie, dirigé par Anne Lacheret entre 2008 et 2012. L'objectif de ce projet était de proposer un corpus de différentes variétés de français oral (discours politiques, interviews, émissions telles que nouvelles télévisées, télé-achat, match de foot, etc) annoté en prosodie, macrosyntaxe et microsyntaxe. La constitution du treebank, ainsi que les premiers résultats, sont précisés en section 2.4.

Le corpus Rhapsodie est lui-même intégré au projet ANR Orfeo, dont l'objectif est de rassembler un Corpus d'Étude pour le Français contemporain. Couvrant une large variété des genres en français écrit et oral, et enrichi par des annotations en syntaxe, morphologie, ou encore prosodie, ce corpus fournira une base de données utile à des chercheurs ou ingénieurs en linguistique de corpus.

2.4. Présentation de la microsyntaxe du treebank Rhapsodie

2.4.1. Construction du treebank

Le corpus Rhapsodie est constitué de trois heures d'enregistrement de productions en français oral. Il permet de modéliser la performance de locuteurs de français oral dans des situations variées, telles que des émissions de radio ou des allocutions politiques.

L'analyse en dépendance a d'abord été faite par un parser automatique (Villemonte de la Clergerie 2010), puis a été corrigée manuellement par une équipe d'annotateurs, dont j'ai fait partie (Bawden et al. 2014).

Les mots ont été étiquetés selon une liste de parties du discours :

- **V :** les verbes
- **N :** les noms
- **Adj :** les adjectifs
- **Adv :** les adverbes
- **Pre :** les prépositions
- **CS :** les conjonctions de subordination
- **J :** les joncteurs (les conjonctions de coordinations et les mots liant ou clôturant un entassement)
- **D :** les déterminants
- **I :** les interjections
- **Q :** les pronoms relatifs et interrogatifs
- **Cl :** les clitiques, les clitiques sujets et l'adverbe de négation "ne"
- **Pro :** les autres pronoms
- **X :** les éléments à la catégorie indéterminable (mot inaudible, amorce de mot, position non instanciée)

Les liens ont été étiquetés selon une liste de fonctions syntaxiques :

- **root :** les racines, soit les éléments sans gouverneur
- **sub :** le sujet du verbe et des constructions prédicatives
- **obj** : le complément d'objet direct
- **pred :** les éléments qui forment un prédicat complexe avec le verbe (l'attribut du sujet ou de l'objet, le participe passé dans les formes verbales composées, l'infinitif dans les constructions avec un verbe modal)
- **obl :** les compléments obliques (le complément du verbe qui n'est pas complément d'objet direct, le complément prépositionnel qui fait partie d'une forme figée, la construction locative avec le verbe "être")
- **ad :** les ajouts
- **dep :** les dépendants des formes non verbales, les clitiques des formes figées, les constructions clivées
- **junc :** les liens entre les éléments d'un entassement et les joncteurs qui les lient
- **para :** les liens entres les éléments d'un entassement, avec précision du type d'entassement :
  - **para_coord :** la coordination, l'entassement de deux éléments à la dénotation différente
  - **para_hyper :** l'hyperonymie
  - **para_intens :** l'intensification
  - **para_disfl :** la répétition du même élément (amorce de mot, mot, groupe de mots) sans changement lexical (sauf dans le cas des mots grammaticaux)
  - **para_reform :** la reformulation d'un élément (peut être précédé par "je veux dire")
  - **para_dform :** la double formulation d'un élément (peut être précédé par "c'est-à-dire")
  - **para_negot :** la négociation (demande de confirmation, confirmation, réfutation, correction)

Le premier élément d'une construction paradigmatique est relié au gouverneur par une dépendance dite "principale", les autres le sont par des dépendances dites "héritées". La liste des dépendances héritées est la même que celle des dépendances principales, moins les liens paradigmatiques, et on ajoute *_inherited* à l'étiquette de la dépendance.

Certains éléments de constructions paradigmatiques se retrouvent donc avec deux gouverneurs : celui de la dépendance hérité et celui du lien paradigmatique. Dans ce mémoire, seul le lien paradigmatique a été pris en compte.

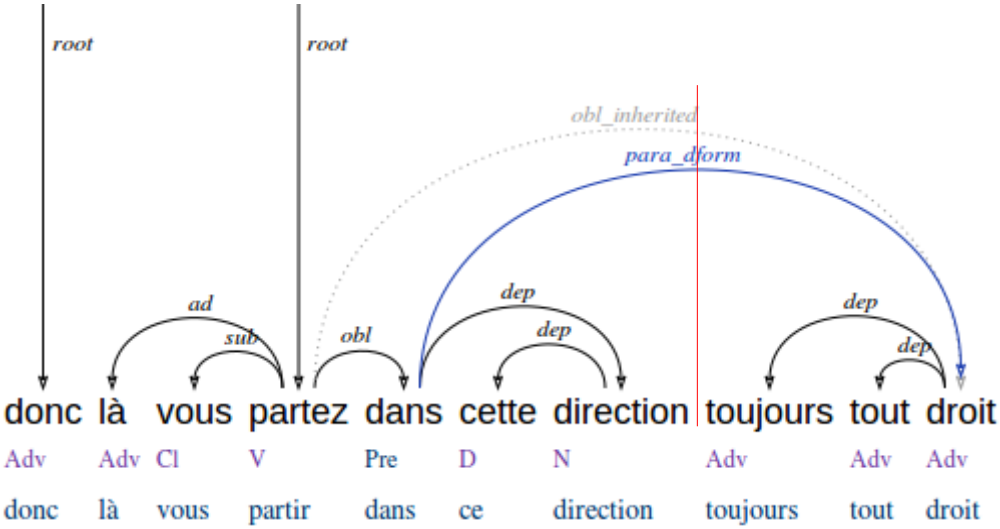

Fig. 2.3. : Lien hérité et lien paradigmatique

Dans la figure 2.3, le mot "droit" a deux gouverneurs : le mot "partez" pour la dépendance hérité, et le mot "dans" pour le lien paradigmatique. Puisque seul le lien paradigmatique est pris en compte, on considérera qu'une seule dépendance passe par la position représentée par la ligne rouge.

Ce choix peut être rapproché de l'aplatissement des dépendances en bouquet dans (Jardonnet 2009) : le second complément, introduit par "droit", a été lié au premier, introduit par "dans", au lieu d'être rattaché au verbe "partez". Cela a pour conséquence une diminution de la taille du flux entre les mots "partez" et "dans".

De même, certains liens *para_coord* doublent deux liens de type *junc* : dans ce cas, il n'a pas été tenu compte du lien *para_coord*.

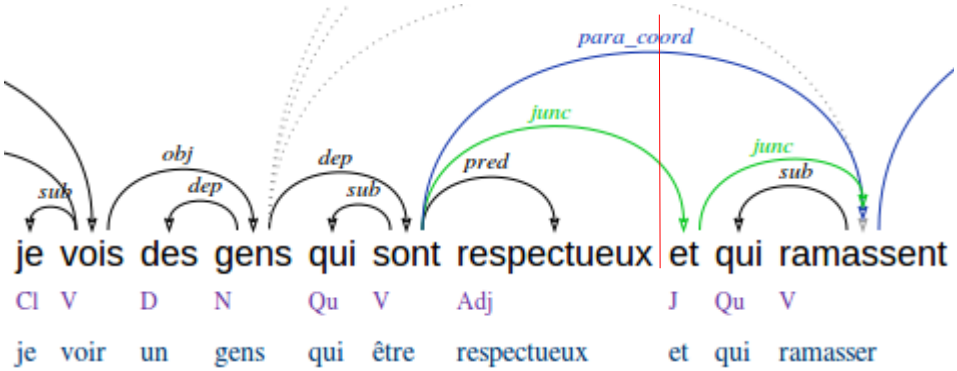

Fig. 2.4. : Lien para_coord *doublant deux liens* junc

Dans la figure 2.4, à la position représentée par la ligne rouge, on ne comptera qu'une dépendance : le lien *para_coord* entre "sont" et "ramasse" double les liens *junc* entre "sont" et "et" et "et" et "ramassent", il est donc ignoré.

2.4.2. Premières caractéristiques du treebank

Les caractéristiques du treebank Rhapsodie ont été précisées dans (Bawden et al. 2014).

Néanmoins, des corrections ont été apportées a posteriori sur le corpus. Les valeurs indiquées ci-après sont donc plus récentes que celles indiquées dans l'article susnommé.

	dep	**root**	**sub**	**ad**	**pred**	**obj**	**obl**	**junc**	**total**
principale	14291	6163	4062	2686	2166	2128	932	917	33345
inherited	1161	328	202	307	268	303	132	29	2730
total	15452	6491	4264	2993	2434	2431	1064	946	36075

Fig 2.5. : Décompte des liens de dépendance par fonction syntaxique

	disfl	**coord**	**reform**	**dform**	**intens**	**negot**	**hyper**	**total**
para_	749	556	255	191	126	117	77	2071

Fig 2.6. Décompte des liens paradigmatiques par type

	N	**V**	**Cl**	**D**	**Pre**	**Adv**	**I**	**Adj**	**J**	**Qu**	**CS**	**Pro**	**X**	**total**
POS	6311	5990	4177	4584	3944	2786	1978	1613	1141	802	727	718	198	34969

Fig 2.7. Décompte des parties du discours

Les figures 2.5, 2.6 et 2.7 détaillent la répartition des étiquettes syntaxiques dans le corpus Rhapsodie.

On remarque que le nombre de liens paradigmatiques ne correspond pas au nombre de liens hérités. En effet, dans certains cas, un lien paradigmatique reliant un gouverneur A à un dépendant B entraîne l'apparition de plusieurs liens hérités : un reliant B au gouverneur de A, et un ou plusieurs reliant A aux dépendants de B.

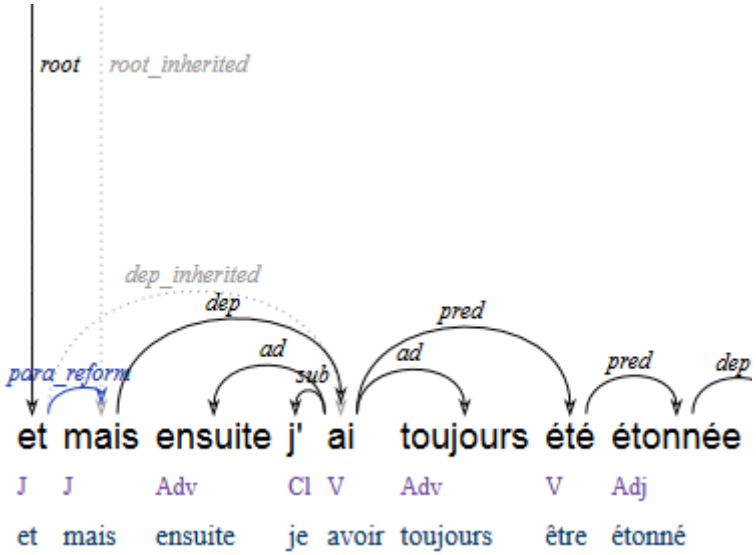

Fig. 2.8. : Lien paradigmatique entraînant deux liens hérités

Dans la figure 2.6, le lien *para_reform* entre "et" et "mais" entraîne deux liens hérités : un reliant "mais" au gouverneur de "et" (ici la racine de la phrase), et un reliant "et" à "ai", le dépendant de "mais".

Chapitre 3

Représentation du flux

Ensembles de dépendances passant par une position donnée, le flux peut être représenté de différentes façons selon les caractéristiques que l'on veut mettre en valeur. Nous introduirons ici une représentation du flux sous la forme d'une structure de traits, puis sous la forme d'une matrice. Nous approfondirons ensuite cette dernière représentation en en présentant les propriétés.

3.1. Représentation du flux sous la forme d'une structure de traits

Pour chaque position dans le corpus, le flux de dépendance peut être représenté par une structure de traits, dont le patron est présenté ci-dessous en figure 3.1.

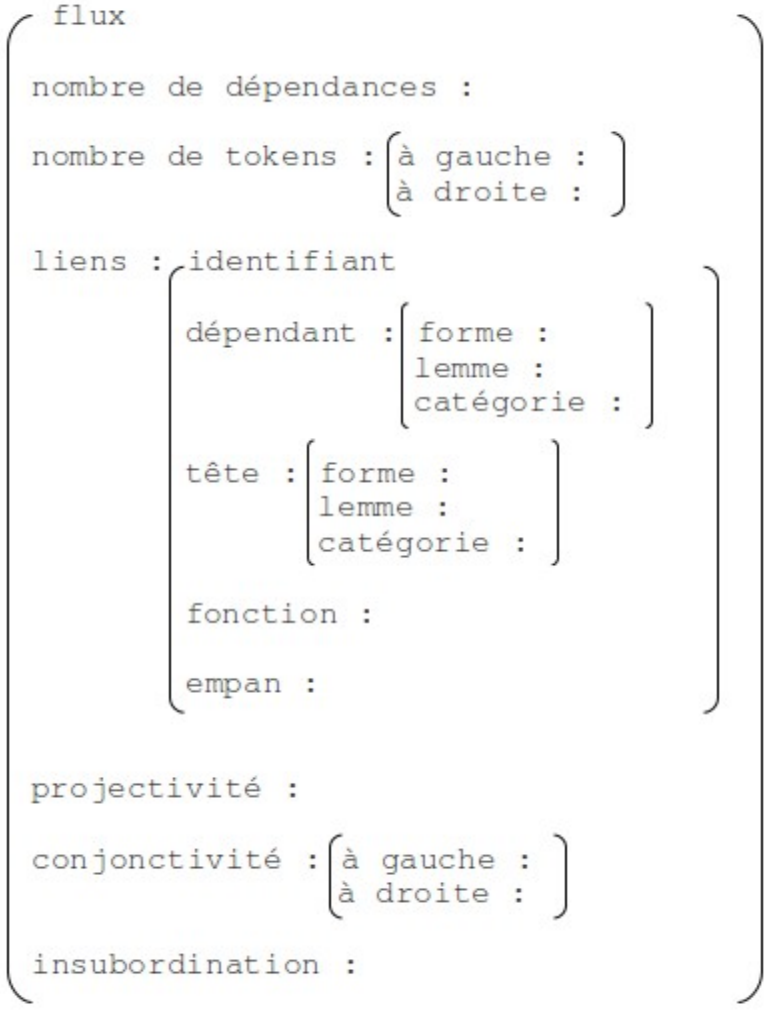

Fig. 3.1. : Représentation du flux sous la forme d'une structure de traits

Cette structure permet ensuite de rechercher des positions du corpus où le flux remplit un certain nombre de critères : elle sera donc utile à un programme de requêtes.

Néanmoins, même si cette représentation est très complète, elle ne permet pas de repérer au premier coup d'œil certaines caractéristiques du flux.

3.2. Représentation du flux sous la forme d'une liste

Le flux de dépendance peut être représenté par une liste de dépendances ordonnées (Kahane 2001). Cette liste est moins précise que la structure de traits, mais plus lisible.

Toutefois, cette représentation n'est pas adaptée à un flux non projectif, dont les dépendances n'ont pas d'ordre canonique.

3.3. Représentation du flux sous la forme d'une matrice

Une représentation du flux sous la forme d'une matrice permet de repérer rapidement certaines caractéristiques du flux, telles que les dépendances conjointes, les insubordinations ou encore les croisements.

Nous allons présenter étape par étape la construction de la matrice, puis nous en pointerons certaines propriétés.

3.3.1. Construction de la matrice

La matrice se présente sous la forme d'un tableau, dans lequel les lignes représentent la suite des mots à gauche de la position et les colonnes la suite des mots à droite. Ne sont pris en compte que les mots qui rentrent en jeu dans une dépendance passant par la position étudiée. Les dépendances sont ensuite placées dans le tableau en fonction de leurs extrémités.

Nous allons construire la matrice représentant le flux indiqué par la ligne rouge dans la figure 3.2 présentée ci-dessous.

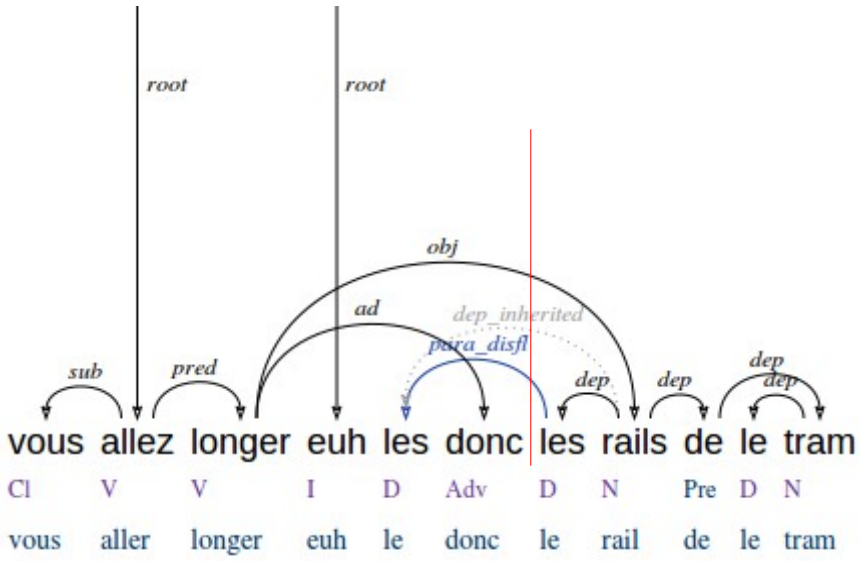

Fig. 3.2. : Exemple

Deux dépendances passent par cette position :
- une entre les mots "longer" et "rails"
- une entre les mots "les" et "les"

Les dépendances sont représentées dans le tableau par une structure de traits simplifiée :

[catégorie du gouverneur, catégorie du dépendant, fonction syntaxique, sens de la dépendance]

Les deux dépendances de l'exemple seront donc représentées respectivement par :
- [V, N, obj, +]
- [D, D, para_disfl, -]

- Numérotation des extrémités des dépendances

On numérote les mots rentrant en jeu dans des dépendances passant par la position étudiée. La numérotation part des mots les plus éloignés de la position du flux. Les mots de gauche sont donc numérotés dans l'ordre de lecture, et ceux de droite dans l'ordre inverse.

Dans l'image 3.3 ci-après, la numérotation est indiquée en rouge, en-dessous de la catégorie grammaticale du mot.

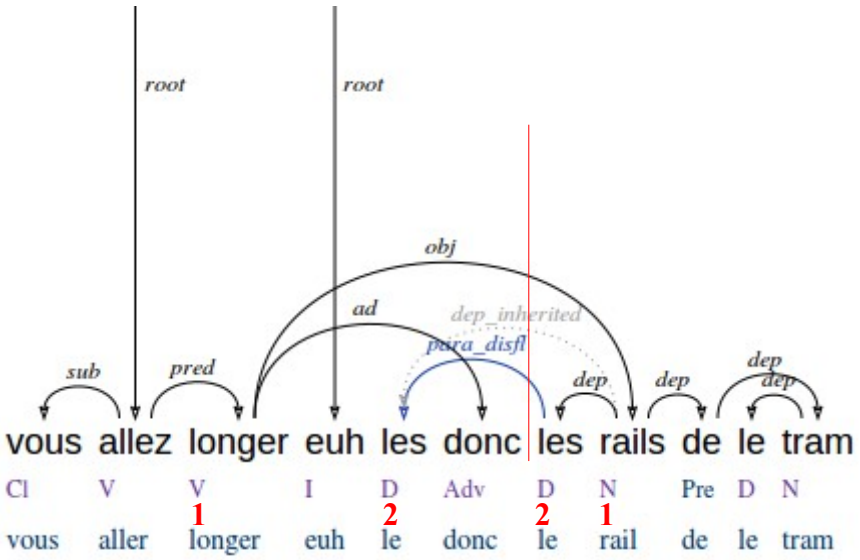

Fig. 3.3. : Numérotation des mots

- Construction du tableau vide

Dans le cas étudié ici, on a deux mots à gauche de la position, et deux mots à droite. Comme énoncé précédemment, les lignes du tableau représentent les mots à gauche de la position et les colonnes les mots à droite. Nous aurons donc un tableau (figure 3.4) composé de deux lignes numérotées 1 et 2, et deux colonnes numérotées 1 et 2.

	1	**2**
1		
2		

Fig. 3.4. : Tableau vide

- Remplissage du tableau

On place ensuite les dépendances dans le tableau, en fonction des numéros attribués à leurs extrémités. La dépendance "longer"-"rails" (notée [V, N, obj, +]) va donc en ligne 1, colonne 1 et la dépendance "les"-"les" (notée [D, D, para_disfl, -]) en ligne 2, colonne 2.

Le tableau rempli est présenté ci-dessous en figure 3.5.

	1	**2**
1	[V, N, obj, +]	
2		[D, D, para_disfl, -]

Fig. 3.5. : Tableau rempli

3.3.2. Propriétés de la matrice

La représentation du flux sous la forme d'une matrice permet de repérer rapidement certaines caractéristiques du flux.

- Flux disjoint

Propriété : À une position donnée, le flux est disjoint et projectif si et seulement si seules les cases de la diagonale principale de la matrice sont remplies.

Propriété : Deux dépendances sont disjointes si et seulement si elles ne sont placées ni sur la même ligne, ni sur la même colonne de la matrice.

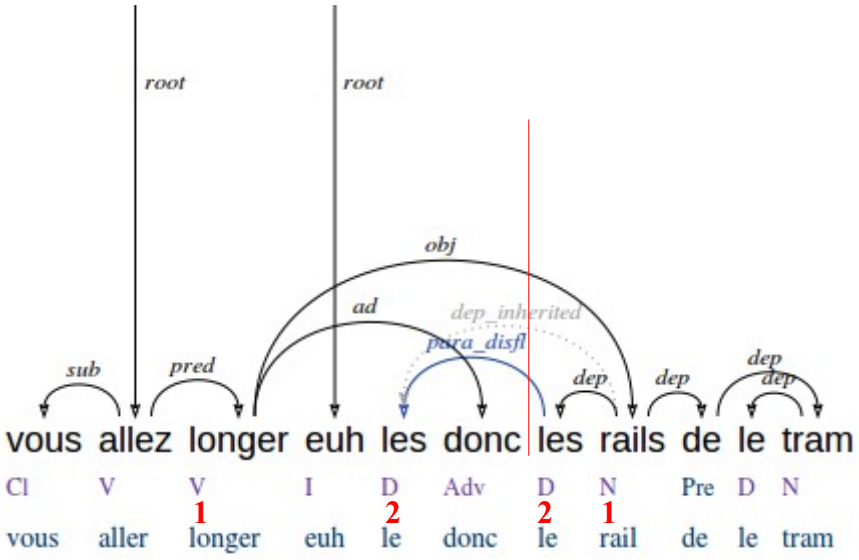

Fig. 3.6. : Flux disjoint

	1	2
1	[V, N, obj, +]	
2		[D, D, para_disfl, -]

Fig. 3.7. : Matrice du flux disjoint

La ligne rouge sur la figure 3.6 représente une position où le flux est disjoint. Les mots ont été numérotés en rouge, et le tableau en figure 3.7 est la représentation de ce flux sous la forme d'une matrice.

- Conjonctivité

Propriété : Deux dépendances sont conjointes à gauche (respectivement à droite) si et seulement si elles se trouvent sur la même ligne de la matrice (respectivement la même colonne).

La ligne rouge sur la figure 3.8 ci-après représente une position où deux dépendances sont conjointes à gauche. Les mots ont été numérotés en rouge, et le tableau en figure 3.9 est la représentation de ce flux sous la forme d'une matrice.

De même, la ligne rouge sur la figure 3.10 ci-après représente une position où deux dépendances sont conjointes à droite. Les mots ont été numérotés en rouge, et le tableau en figure 3.11 est la représentation de ce flux sous la forme d'une matrice.

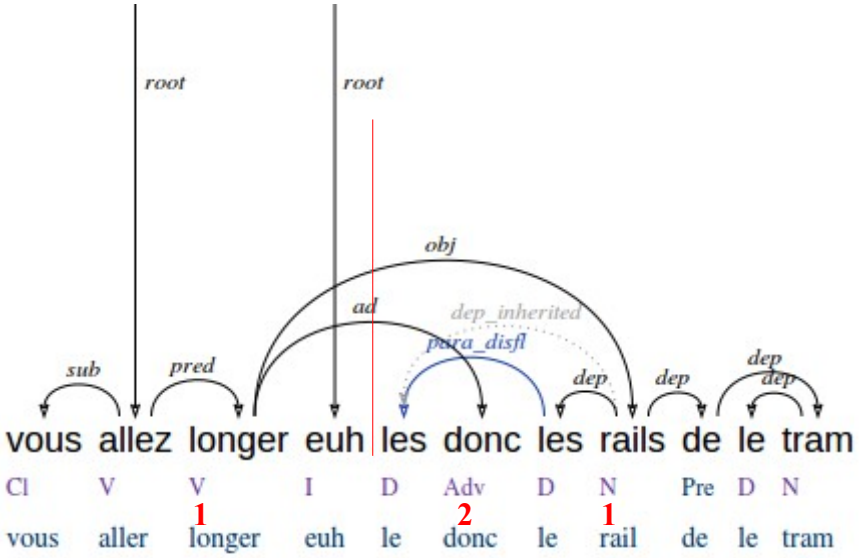

Fig. 3.8. : Dépendances conjointes à gauche

	1	2
1	[V, N, obj, +]	[V, Adv, ad, +]

Fig. 3.9. Matrice du flux avec dépendances conjointes à gauche

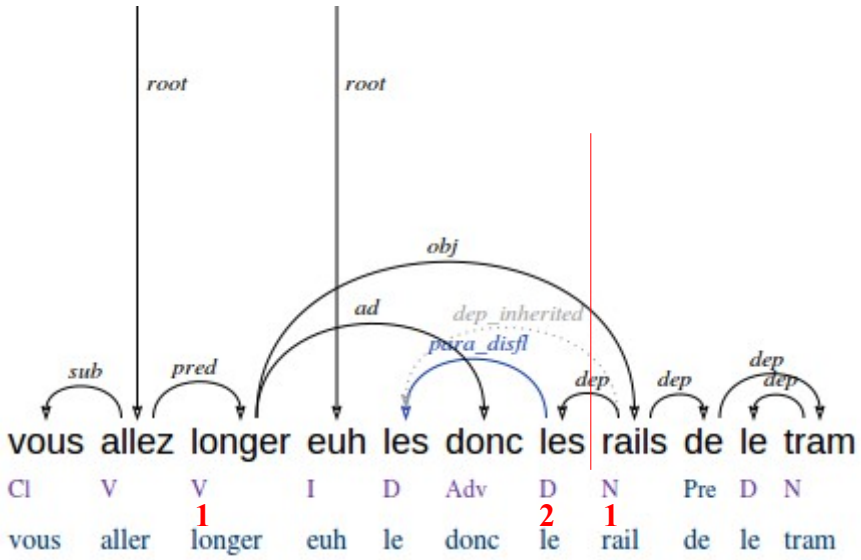

Fig. 3.10. : Dépendances conjointes à droite

	1
1	[V, N, obj, +]
2	[N, D, dep, -]

Fig. 3.11. : Matrice du flux avec dépendances conjointes à droite

- Croisements

Jusqu'à maintenant, on pouvait parcourir toutes les cases remplies du tableau en se déplaçant soit vers la droite, soit vers le bas, soit en combinant ces deux opérations. Lorsque deux dépendances se croisent, leur placement dans le tableau ne respecte pas ces règles de lecture.

Propriété : Le flux contient deux dépendances qui se croisent si et seulement si on ne peut pas parcourir la matrice du flux par une combinaison de mouvements vers le bas et vers la droite.

Preuve : Soit les deux propositions :

- P : deux dépendances se croisent
- Q : on peut passer d'une case à l'autre de la matrice en combinant des déplacements vers la droite et vers le bas.

On veut montrer que P est équivalent à ¬Q (ce qui revient à dire que Q est équivalent à ¬P).

On commence par supposer Q, et on montre ¬P.

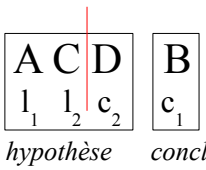

Fig. 3.12. : Illustration de Q → ¬P

On considère deux dépendances disjointes A-B et C-D telles que A < B et A < C < D. Nous allons montrer que ces dépendances ne se croisent pas, c'est-à-dire que D < B. On considère la place de ces deux dépendances dans la matrice de flux en une position entre C et D (représentée dans la figure 3.12 par la ligne rouge). On nomme :

- l_1 l'indice de A
- l_2 l'indice de C
- c_1 l'indice de B
- c_2 l'indice de D

Comme A < C, A est plus extérieur : $l_1 < l_2$.

On veut montrer que D < B, c'est-à-dire que B est plus extérieur, donc que $c_1 < c_2$.

On se déplace de la case l_1-c_1 à la case l_2-c_2 par un ou plusieurs mouvements vers le bas, car $l_1 < l_2$. D'après l'hypothèse Q, on doit aussi opérer un ou plusieurs mouvements vers la droite, donc $c_1 < c_2$. On a bien montré que D < B et donc que jamais deux dépendances ne se croisent.

Si l'hypothèse Q est vraie, alors les dépendances ne se croisent pas : $Q \rightarrow \neg P$.

On montre maintenant la réciproque : on suppose $\neg P$ et on montre Q.

Si aucune dépendance n'en croise d'autre, alors pour toutes les positions le flux possède la propriété suivante : les dépendances peuvent être totalement ordonnées, de telle façon à ce que chaque dépendance couvre la dépendance suivante, c'est-à-dire que les indices des extrémités d'une dépendance sont inférieurs ou égaux à ceux de la dépendance suivante.

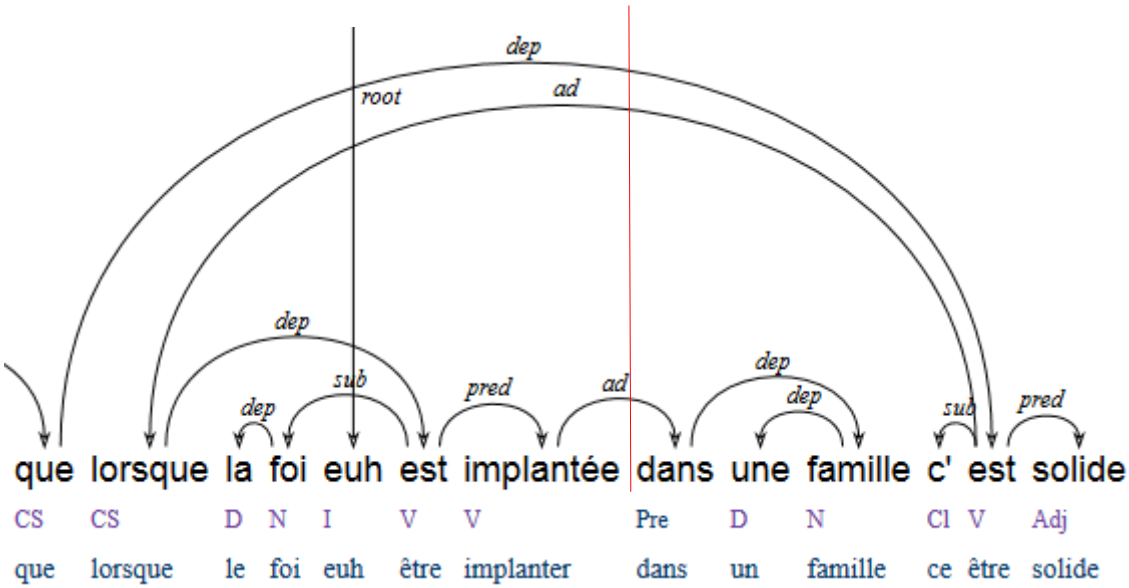

Fig. 3.13. : Dépendances ordonnées

Dans la figure 3.13, à la position représentée par la ligne rouge, la dépendance "que"-"est" couvre la dépendance "est"-"lorsque", qui elle couvre la dépendance "implantée"-"dans".

En parcourant la matrice du flux suivant l'ordre total des dépendances, on n'opérera donc que des mouvements vers la droite ou vers le bas.

Si l'hypothèse P est fausse, on passe donc d'une case de la matrice à une autre en combinant des déplacements vers la droite et vers le bas : ¬P → Q.
■

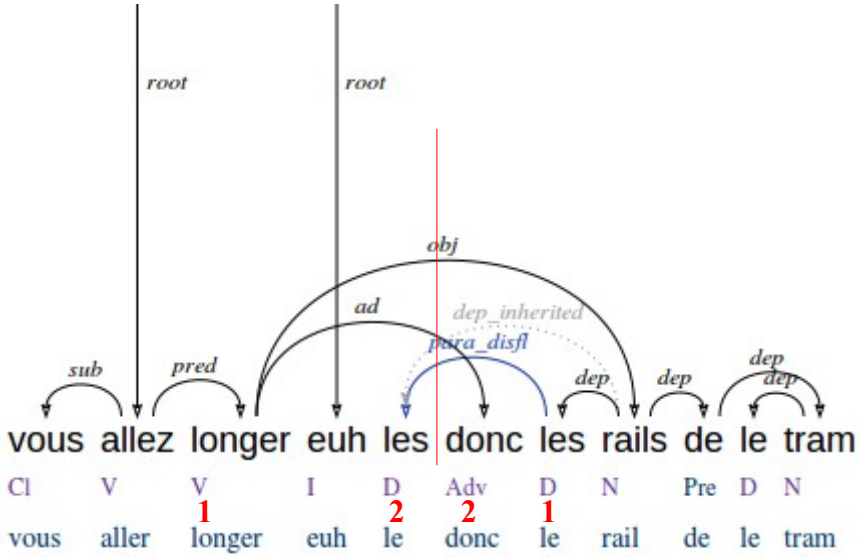

Fig. 3.14. : Flux non projectif présentant un croisement

	1	**2**
1		[V, Adv, ad, +]
2	[D, D, para_disfl, -]	

Fig. 3.15. : Matrice du flux non projectif avec croisement

La ligne rouge sur la figure 3.14 représente une position où deux dépendances se croisent. Les mots ont été numérotés en rouge, et le tableau en figure 3.15 est la représentation de ce flux sous la forme d'une matrice. Pour mettre en évidence le phénomène étudié, seules les dépendances entrant en jeu dans le croisement ont été représentées.

- Insubordination

Les cas d'insubordination se repèrent en observant les matrices présentant des dépendances conjointes : on prêtera attention au sens des dépendances. En effet, on remarque dans les matrices représentant des dépendances conjointes sans insubordination, que la dépendance la plus à droite a un sens positif (+) (voir figure 3.9), et que la dépendance la plus en bas a un sens négatif (-) (voir figure 3.11). Ce principe est valable quel que soit le sens de l'autre dépendance. Dans le cas d'une insubordination, ce principe ne sera pas respecté.

Propriété : Lorsque deux dépendances sont sur la même ligne (respectivement la même colonne), il y a insubordination de la dépendance de gauche (respectivement du haut) si et seulement si la dépendance la plus à droite (respectivement la plus en bas) a un sens négatif (-) (respectivement positif (+)).

Preuve : Soit les deux propositions :

- P : dans une paire de dépendances situées sur la même ligne, la dépendance de gauche est insubordonnée à la dépendance de droite
- Q : dans une paire de dépendances situées sur la même ligne, la dépendance de droite a

un sens négatif (-)

On veut montrer que P est équivalent à Q.

On commence par supposer P, et on montre Q.

On considère deux cases remplies de la matrice situées sur la même ligne, dont les coordonnées sont l-c_1 et l-c_2 telles que $c_1 < c_2$. Nous allons montrer que la dépendance de droite, c'est-à-dire celle de la case l-c_2, a un sens négatif (-). On nomme :

- A le mot à l'indice l
- B le mot à l'indice c_1
- C le mot à l'indice c_2

On a donc A < C < B, une dépendance entre A et B dans la case de gauche et une dépendance entre A et C dans celle de droite.

D'après l'hypothèse P, la dépendance de gauche est insubordonnée à celle de droite. On a donc C gouverneur de A et A gouverneur de B. Puisque A < C, la dépendance C-A a un sens négatif. On a donc bien montré que la dépendance de droite a un sens négatif.

Si l'hypothèse P est vraie, alors la dépendance de droite a un sens négatif : P → Q.

On montre maintenant la réciproque : on suppose Q et on montre P.

On considère deux cases remplies de la matrice situées sur la même ligne, dont les coordonnées sont l-c_1 et l-c_2, telles que $c_1<c_2$. Nous allons montrer que la dépendance de gauche (case l-c_1) est insubordonnée à celle de droite (case l-c_2). On nomme :

- A le mot à l'indice l
- B le mot à l'indice c_1
- C le mot à l'indice c_2

On a donc A < C < B, une dépendance entre A et B dans la case de gauche et une dépendance entre A et C dans celle de droite.

D'après l'hypothèse Q, la dépendance de droite a un sens négatif (-). Puisque A < C, cela signifie que C est le gouverneur de A. A entre aussi en jeu dans la dépendance de gauche, mais a déjà son gouverneur : c'est donc lui qui gouverne B. Or, il y a insubordination lorsque, pour trois mots C, A et B tels que C gouverne A et A gouverne B, le mot C se trouve entre les mots A et B (voir définition en partie 1.2). Il y a donc bien une insubordination entre les dépendances A-B et C-A.

Puisque C < B, la dépendance A-B recouvre la dépendance C-A. C'est donc bien la dépendance de gauche qui est insubordonnée à celle de droite.

Si l'hypothèse Q est vraie, alors la dépendance de gauche est insubordonnée à celle de droite : Q → P.

De manière similaire, on pourra démontrer que, pour les deux propositions :

- P : dans une paire de dépendances situées sur la même colonne, la dépendance du haut est insubordonnée à la dépendance du bas
- Q : dans une paire de dépendances situées sur la même colonne, la dépendance du bas a un sens positif (+)

P et Q sont équivalents.
■

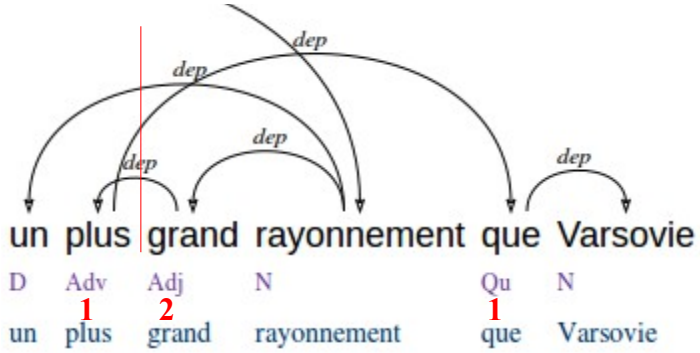

Fig. 3.16. : Dépendances conjointes à gauche et insubordination

	1	**2**
1	[Adv, Qu, dep, +]	[Adj, Adv, dep, -]

Fig. 3.17. : Matrice du flux avec dépendances conjointes à gauche et insubordination

La ligne rouge sur la figure 3.16 représente une position où deux dépendances sont conjointes à gauche et où l'une est insubordonnée à l'autre. Les mots ont été numérotés en rouge, et le tableau en figure 3.17 est la représentation de ce flux sous la forme d'une matrice. Pour mettre en évidence le phénomène étudié, seules les dépendances entrant en jeu ont été représentées.

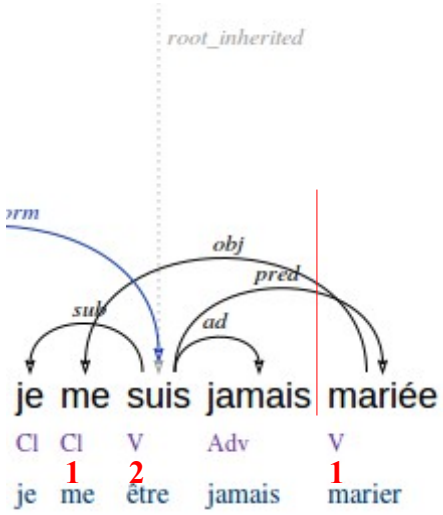

Fig. 3.18. : Dépendances conjointes à droite et insubordination

	1
1	[V, Cl, obj, -]
2	[V, V, pred, +]

Fig. 3.19. : Matrice du flux avec dépendances conjointes à droite et insubordination

La ligne rouge sur la figure 3.18 représente une position où deux dépendances sont conjointes à droite et où l'une est insubordonnée à l'autre. Les mots ont été numérotés en rouge, et le tableau en figure 3.19 est la représentation de ce flux sous la forme d'une matrice. Pour mettre en évidence le phénomène étudié, seules les dépendances entrant en jeu ont été représentées.

- Récapitulatif des propriétés de la matrice

À une position donnée
  - le flux est disjoint et projectif si et seulement si seules les cases de la diagonale principale de la matrice sont remplies
  - deux dépendances sont disjointes si et seulement si elles ne sont placées ni sur la même ligne, ni sur la même colonne de la matrice
  - deux dépendances sont conjointes à gauche (respectivement à droite) si et seulement si elles se trouvent sur la même ligne de la matrice (respectivement la même colonne)
  - le flux contient deux dépendances qui se croisent si et seulement si on ne peut pas parcourir la matrice du flux par une combinaison de mouvements vers le bas et vers la droite
  - lorsque deux dépendances sont sur la même ligne de la matrice (respectivement la même colonne), il y a insubordination de la dépendance de gauche (respectivement du haut) si et seulement si la dépendance de droite (respectivement du bas) a un sens négatif (-) (respectivement positif (+))

3.4. Matrice la plus fréquente

Le corpus Rhapsodie présente 6456 matrices du flux différentes, la matrice vide n'ayant pas été comptée. Parmi ces matrices, 4562 sont des hapax, c'est-à-dire qu'elles ne sont présentent qu'une seule fois dans le corpus. La matrice la plus fréquente, présentée ci-dessous en figure 3.20, a 1465 occurrences.

	1
1	[Pre, N, dep, +]

Fig. 3.20. : Matrice la plus fréquente

3.5. Transitions

Avancer dans une phrase, c'est passer d'une position inter-mots à une autre, et donc à une matrice du flux à une autre. Pour passer d'une matrice à la matrice de la position suivante, il faut parfois réaliser certaines opérations. Ces opérations se résument à :
  - l'ajout d'une ligne
  - l'ajout d'une colonne
  - l'ajout d'une dépendance
  - la suppression d'une ligne
  - la suppression d'une colonne
  - la suppression d'une dépendance

Chacune de ces opérations peut se combiner avec les autres, et ces combinaisons auront une influence sur les caractéristiques du flux de la position d'arrivée. On notera que, parfois, de nombreuses opérations sont nécessaires pour passer de la matrice du flux à une position à la matrice du flux à la position suivante : le flux peut radicalement changer d'une position à une autre.

3.5.1. Ajout d'une ligne ou d'une colonne

L'ajout d'une ligne ou d'une colonne est toujours combiné avec l'ajout d'une dépendance, car ni une ligne ni une colonne ne peuvent être vides. Cette opération entraîne donc une augmentation de la taille du flux, ainsi qu'une éventuelle conjonctivité de la dépendance à venir (voir partie 3.5.2 ci-après).

	1	**2**
1	[V, Pre, ad, +]	
2		[Pre, N, dep, +]

	1	**2**
1	[V, Pre, ad, +]	
2		[Pre, N, dep, +]
3		[N, D, dep, -]

Fig. 3.21. : Ajout d'une ligne

En figure 3.21, une ligne a été ajoutée en bas du tableau : elle s'accompagne de l'apparition d'une nouvelle dépendance [N, D, dep, -].

3.5.2. Ajout d'une dépendance

L'ajout d'une dépendance peut être associé soit à l'ajout d'une ligne, soit à l'ajout d'une colonne, soit aux deux. Dans le cas où il n'est associé qu'à l'ajout d'une ligne, la dépendance sera conjointe à droite, et dans le cas où il n'est associé qu'à l'ajout d'une colonne, la dépendance sera conjointe à gauche.

Dans la figure 3.21 présentée en partie 3.5.1, il y a eu ajout d'une dépendance [N, D, dep, -] sans l'ajout d'une colonne : la dépendance [N, D, dep, -] est donc conjointe à droite avec la dépendance [Pre, N, dep, +].

3.5.3. Suppression d'une ligne ou d'une colonne

La suppression d'une ligne ou d'une colonne est forcément associée à la suppression d'une dépendance, car ni une ligne ni une colonne ne peuvent être vides. À l'inverse de l'ajout, cette opération diminue la taille du flux et peut dans certains cas annuler le caractère conjoint de deux dépendances (voir partie 3.5.4 ci-après).

	1	**2**
1	[V, V, dep, +]	
2		[Pre, Pro, dep, +]

	1
1	[V, V, dep, +]

Fig. 3.22. : Suppression d'une ligne et d'une colonne

En figure 3.22, la colonne de droite et la ligne du bas ont été supprimées du tableau, et la dépendance [Pre, Pro, dep, +] a pour cette raison été aussi supprimée.

3.5.4. Suppression d'une dépendance

Similairement à l'ajout, la suppression d'une dépendance peut être associée soit à la suppression d'une ligne, soit à la suppression d'une colonne, soit aux deux. Néanmoins, si on supprime une dépendance sans supprimer la ligne dans laquelle elle se trouvait, c'est que cette suppression n'entraîne pas le vide de la ligne en question et donc qu'une autre dépendance s'y trouve (idem dans le cas d'une suppression de dépendance sans suppression de la colonne correspondante). Il y avait donc avant transition un cas de conjonctivité, qui a été supprimé par la suppression de la dépendance.

	1	**2**
1	[V, V, dep, +]	[V, Adv, ad, +]

	1
1	[V, V, dep, +]

Fig. 3.22. : Suppression d'une dépendance sans suppression de ligne

Dans la figure 3.22, la suppression de la dépendance [V, Adv, ad, +] n'est pas associée à la suppression de la ligne dans laquelle elle se trouvait. La conjonctivité entre cette dépendance et la dépendance [V, V, dep, +] a donc été supprimée.

3.5.5. Le cas des insertions

Les ajouts de lignes, colonnes et dépendances se font principalement en bas et à droite du tableau : les plus petites dépendances sont celles qui changent le plus fréquemment. Néanmoins, il est possible que lignes et colonnes s'insèrent à gauche et en haut : ce phénomène entraîne un croisement entre la dépendance à venir et une dépendance déjà présente.

	1
1	[V, Cl, sub, -]

	1	**2**
1		[V, Cl, sub, -]
2	[V, Cl, dep, -]	

Fig. 3.23. : Insertion d'une colonne

Dans la figure 3.23, la ligne a bien été ajoutée en bas, mais la colonne a été insérée à gauche : la nouvelle dépendance, [V, Cl, dep, -] croise la dépendance déjà présente, [V, Cl, sub, -].

Chapitre 4

Projectivité du flux

Deux phénomènes sont responsables de la non projectivité du flux (voir définitions en partie 1.2) : les croisements et les insubordinations. Nous présenterons tout d'abord les croisements dans le corpus Rhapsodie, puis les insubordinations. Nous étudierons ensuite la possibilité pour ces deux phénomènes d'être cooccurrents.

4.1. Croisement de deux dépendances

4.1.1. Définition

Introduite par (Ihm et Lecerf, 1960), la notion de projectivité qualifie les arbres pour lesquels, en ordre de lecture, aucune dépendance n'en croise d'autre ni aucune dépendance ne couvre la racine de l'arbre.

Ici, on appellera "non projectif" un flux qui contient deux dépendances qui se croisent : si ces dépendances sont A-B et C-D, alors les mots apparaissent dans l'ordre A, C, B, D.

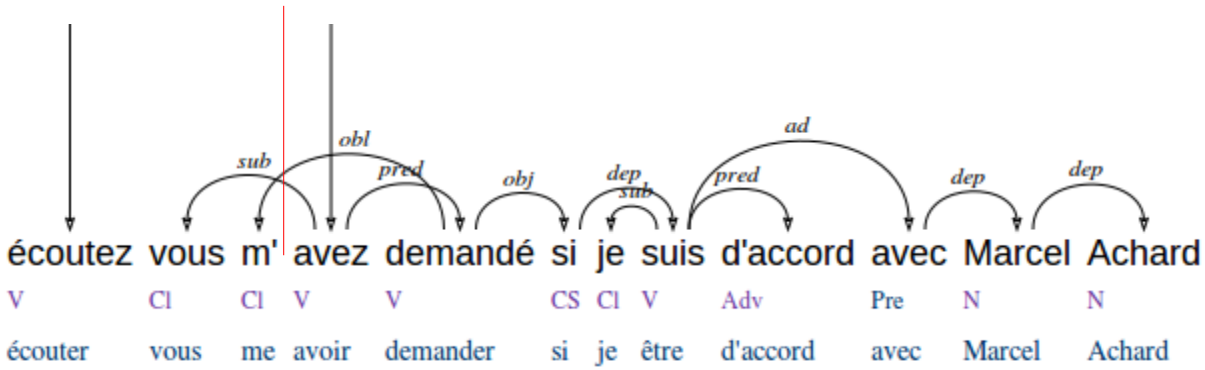

Fig. 4.1. : Arbre présentant un flux non-projectif

Dans l'arbre présenté en figure 4.1., le flux est non projectif à la position indiquée par la ligne rouge : les dépendances *sub* et *obl* se croisent.

Dans le corpus Rhapsodie, le flux est non projectif en 937 positions (soit environ 2,9% du total des positions du corpus), réparties dans 346 arbres différents (soit un peu moins de 12% du total des arbres du corpus). On dénombre 581 croisements, qui peuvent être dus à différents phénomènes.

Les flux des 937 positions où le flux est non-projectif peuvent être réparties sur 659 matrices différentes (voir en 3.2). Sur ces 659 matrices, 550 ne correspondent qu'à un seul flux et une matrice correspond au maximum à 13 flux. Cette dernière matrice est présentée ci-dessous en figure 4.2, et un exemple d'occurrence de ce flux est présenté en figure 4.3 : la position pour laquelle la matrice du flux correspond à celle présentée en figure 4.2 est représentée par la ligne rouge.

	1	2
1		[N, V, dep, +]
2	[V, Qu, obj, -]	

Fig. 4.2. Matrice la plus fréquente dans les cas de croisement

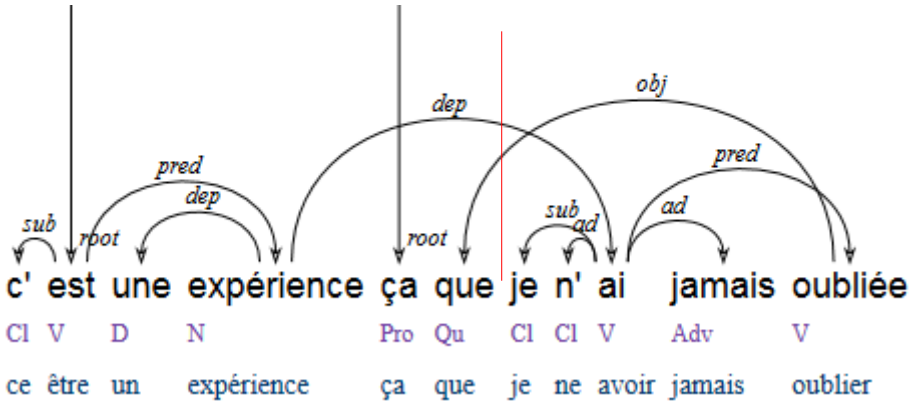

Fig. 4.3. : Position où la matrice du flux correspond à la matrice la plus fréquente

Il est intéressant de noter que ces 13 flux sont tous répartis sur un arbre différent.

4.1.2. Typologie de quelques croisements dans le corpus Rhapsodie

- Les liens paradigmatiques

Sur la totalité des croisements du corpus Rhapsodie, 114 font intervenir un ou plusieurs lien(s) *para*, soit un peu moins de 20%.

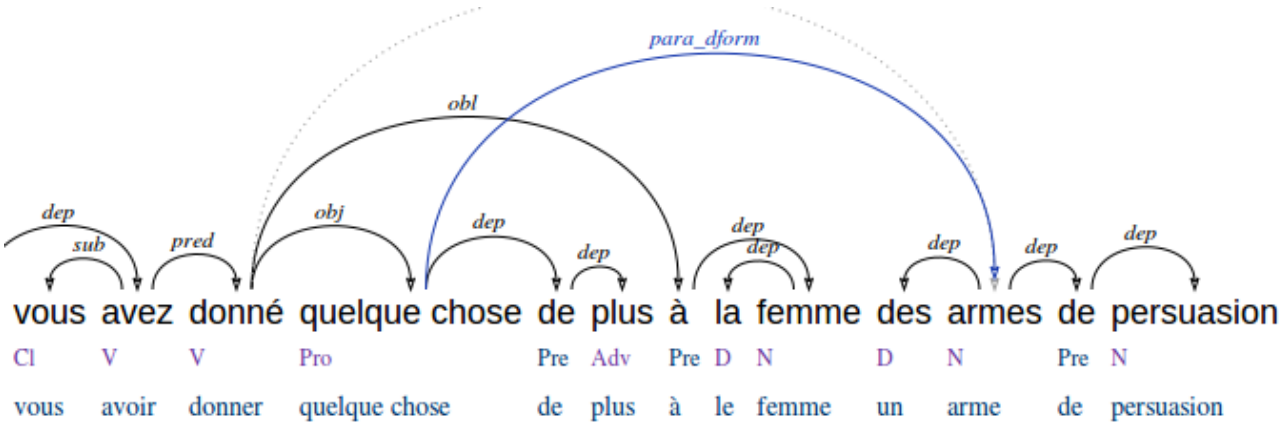

Fig. 4.4. : Croisement dû à un lien paradigmatique

Sur la figure 4.4 le lien paradigmatique entre "quelque chose" et "armes" croise le lien entre "donné" et "à".

- Le passé composé

Parmi les croisements ne faisant pas intervenir de lien paradigmatique, certains sont dus à la présence du passé composé. En effet, le passé composé a été analysé d'une façon telle que le sujet dépend de l'auxiliaire, mais les compléments dépendent du participe.

Dans le cas où l'un des compléments est pronominalisé et se retrouve donc avant l'auxiliaire, on observe un croisement entre la dépendance *auxiliaire-sujet* et la dépendance *participe-complément*.

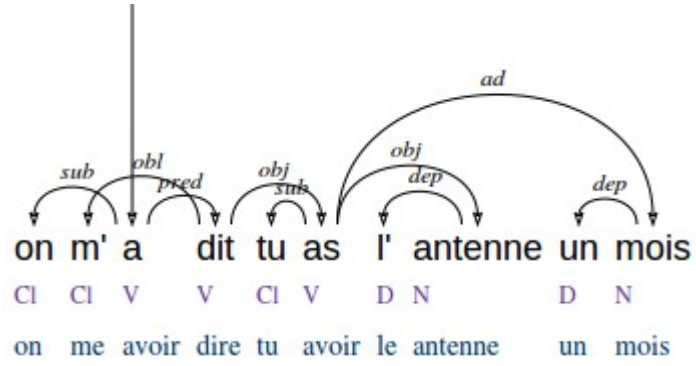

Fig. 4.5. : Complément pronominalisé et passé composé

Sur la figure 4.5, la dépendance entre "a" et "on", soit celle entre l'auxiliaire et le sujet, croise la dépendance entre "dit" et "m'", soit celle entre le participe et le complément.

- Les subordonnées relatives avec verbe de modalité

Dans les subordonnées relatives, l'antécédent est analysé comme gouverneur du verbe fini. Dans le cas d'un verbe de modalité suivi d'un infinitif, il arrive que l'un des compléments de l'infinitif soit réalisé par le pronom relatif : il y a donc croisement entre la dépendance *antécédent-modal* et la dépendance *infinitif-complément*.

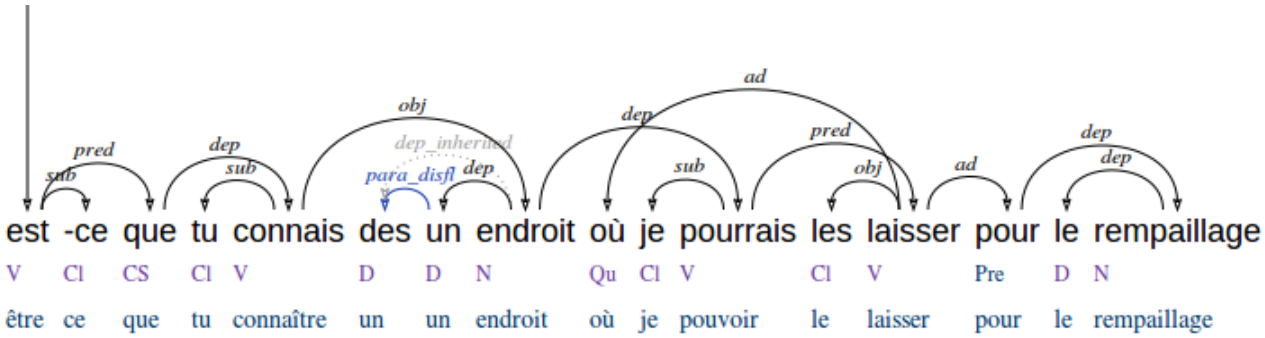

Fig 4.6. : Subordonnée relative avec verbe de modalité

Sur la figure 4.6, la dépendance entre "endroit" et "pourrais", soit celle entre l'antécédent et le modal, croise la dépendance entre "laisser" et "où", soit celle entre l'infinitif et le complément.

On observera le même phénomène dans le cas où le verbe principal de la subordonnée est un auxiliaire : si le pronom relatif est un complément du verbe, il y aura croisement entre la dépendance *antécédent-auxiliaire* et la dépendance *participe-complément*.

- Les comparaisons

Dans une structure comparative, l'adverbe comparatif ("plus", "moins", "aussi") est un dépendant de l'élément de comparaison et est le gouverneur de la conjonction de subordination "que". L'élément de comparaison étant lui-même un dépendant du verbe, il y a croisement entre les dépendances *verbe-élément de comparaison* et *adverbe-conjonction*.

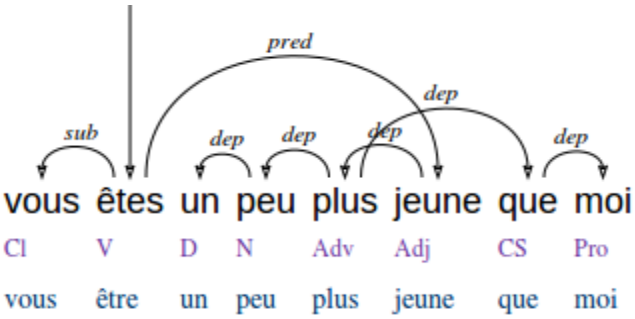

Fig. 4.7. : Comparaison

Sur la figure 4.7, la dépendance entre "êtes" et "jeune", soit celle entre le verbe et l'élément de comparaison, croise la dépendance entre "plus" et "que", soit celle entre l'adverbe et la conjonction.

- Les adjonctions

Certaines adjonctions à un verbe sont intercalées entre un complément du verbe et une dépendance gouvernée par ce complément. La dépendance *verbe-adverbe* croise donc la dépendance *nom-complément*.

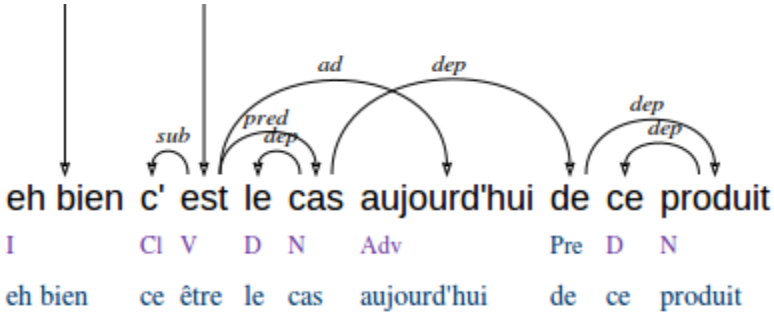

Fig. 4.8. : Croisement dû à une adjonction

Dans le cas présenté en figure 4.8, la dépendance entre "est" et "aujourd'hui", soit celle entre le verbe et l'adverbe, croise la dépendance entre "cas" et "de", soit celle entre le nom et le complément.

- Le clitique "en" antéposé

En français, le clitique "en est placé entre le verbe et son sujet, mais peut être dépendant d'un complément postérieur du verbe. Il y a donc croisement entre la dépendance *verbe-sujet* et la dépendance *complément-clitique*.

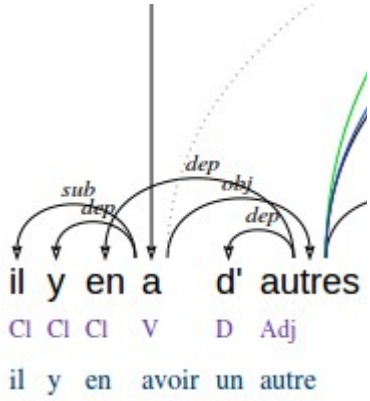

Fig. 4.9. : Clitique antéposé

Sur la figure 4.9, la dépendance entre "a" et "il", soit celle entre le verbe et son sujet, croise la dépendance entre "autres" et "en", soit celle entre le complément et le clitique antéposé.

- Autres phénomènes

On trouve dans le corpus Rhapsodie des occurrences d'une structure considérée comme rare en français : la relative extraposée.

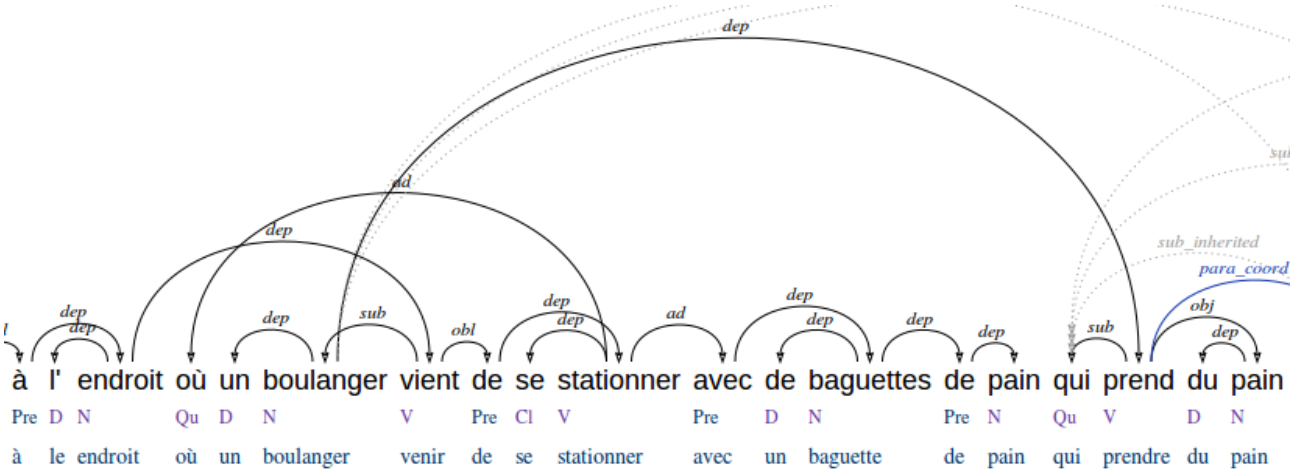

Fig. 4.10. : Relative extraposée

Dans la figure 4.10 la relative extraposée "qui prend du pain", est de plus à une autre subordonnée relative, introduite par le complément relatif "où". L'antécédent de la relative extraposée est ici le sujet de l'autre relative. La dépendance entre "boulanger" et "prend", soit la dépendance *antécédent-verbe* de la relative extraposée, croise la dépendance entre "endroit" et "vient, soit une dépendance *antécédent-modal*, et celle entre "stationner et "où", soit une dépendance *verbe-adjonction*, de l'autre relative.

4.2. Insubordination d'une dépendance

4.2.1. Définition

La projectivité peut aussi être définie comme le caractère de toute dépendance D dont les extrémités sont séparées par des descendants du gouverneur de D (Nasr 1996).

On parlera d'insubordination d'une dépendance lorsque cette projectivité n'est pas respectée, c'est-à-dire lorsque, pour trois mots A, B et C tels que A gouverne B et B gouverne C, le mot A se trouve entre les mots B et C (voir définition en partie 1.2).

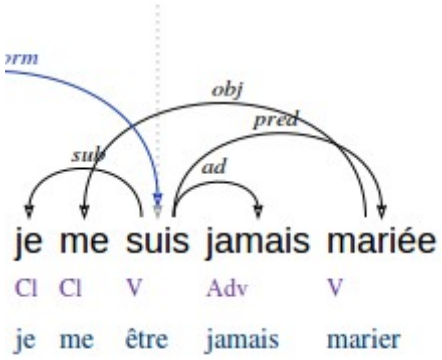

Fig. 4.10. : Dépendance insubordonnée

Dans l'exemple présenté en figure 4.10, la dépendance *prédicat-clitique* entre "mariée" et "me" n'est pas subordonnée à la dépendance *verbe-prédicat* entre "suis" et "mariée", alors qu'elle est censée être dans la projection de sa tête.

Dans le corpus Rhapsodie, on trouve 355 dépendances insubordonnées, réparties sur 302 arbres différents. Ces insubordinations peuvent être dues à différents phénomènes.

Ces insubordinations recouvrent 433 positions dans le corpus. Les flux de ces positions peuvent être réparties sur 221 matrices différentes (voir en partie 3.2). Sur ces 221 matrices, 176 ne correspondent qu'à un seul flux et une matrice correspond au maximum à 43 flux. Cette dernière matrice est présentée ci-dessous en figure 4.11.

	1
1	[V, Cl, obj, -]
2	[V, V, pred, +]

Fig. 4.11. : Matrice la plus fréquente dans les cas d'insubordination

Ces 43 flux sont répartis sur pas moins de 40 arbres différents.

4.2.2. Typologie de quelques insubordinations dans le corpus Rhapsodie

- Les liens paradigmatiques

Contrairement au cas des croisements, on ne dénombre que treize d'insubordinations contenant un lien paradigmatique.

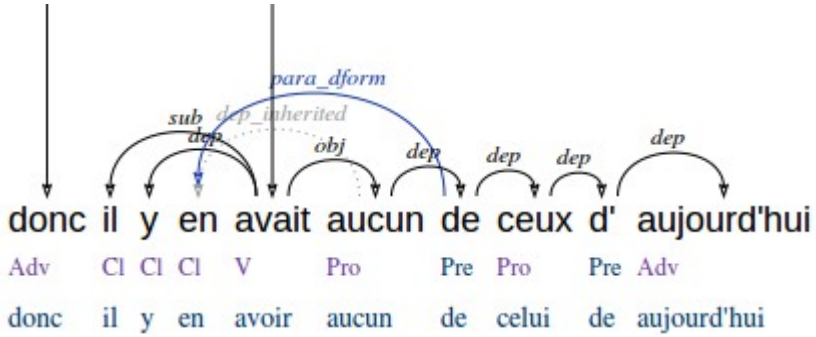

Fig 4.12 : Insubordination contenant un lien paradigmatique

Dans la figure 4.12, le lien paradigmatique entre "de" et "en" est insubordonné au lien entre "aucun" et "de".

- Le passé composé

Tout comme dans le cas des croisements, le passé composé couplé à un complément pronominalisé est une des causes de l'insubordination des dépendances. Le verbe principal, tête de la dépendance *verbe-prédicat*, se retrouve entre la tête et le dépendant de la dépendance *prédicat-clitique*. C'est le déplacement du complément pronominalisé, qui se retrouve avant le verbe en français, qui entraîne l'insubordination de la dépendance.

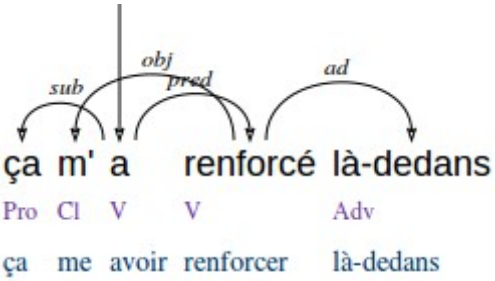

Fig. 4.13. : Passé composé et complément pronominalisé

Sur la figure 4.13, la dépendance entre "renforcé" et "m'", soit celle entre le prédicat et le clitique, est insubordonnée à la dépendance entre "a" et "renforcé", soit celle entre le verbe et le prédicat.

- Les subordonnées relatives avec verbe de modalité

Lorsque le pronom relatif est un dépendant du verbe infinitif, il y a insubordination de la dépendance *infinitif-complément* par rapport à la dépendance *modal-infinitif*.

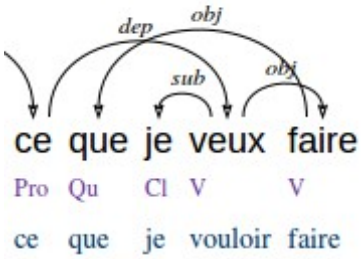

Fig. 4.14. : Subordonnée relative avec verbe de modalité

Sur la figure 4.14, la dépendance entre "faire" et "que", soit celle entre l'infinitif et le complément, est insubordonnée à la dépendance entre "veux" et "faire", soit celle entre le modal et l'infinitif.

Similairement, il y a insubordination dans le cas où le verbe de la proposition subordonnée relative est au passé composé.

- Les comparaisons

Comme nous l'avons énoncé précédemment en partie 4.1.2, dans une structure comparative, l'adverbe comparatif ("plus", "moins", "aussi") est un dépendant de l'élément de comparaison et est le gouverneur de la conjonction de subordination "que". L'élément de comparaison se trouvant entre l'adverbe comparatif et la conjonction de subordination en français, la dépendance *adverbe-conjonction* est insubordonnée à la dépendance *adjectif-adverbe*.

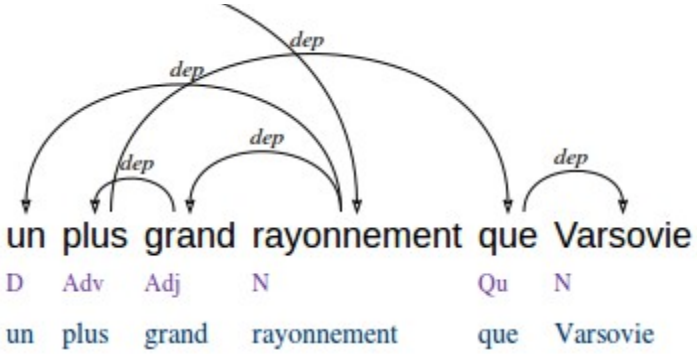

Fig. 4.15. : Comparaison

Dans la figure 4.15, la dépendance entre "plus" et "que", soit celle entre l'adverbe de comparaison et la conjonction de subordination, est insubordonnée à la dépendance entre "grand" et "plus", soit celle entre l'adjectif et l'adverbe.

- Les adjonctions

Une adjonction antéposée peut entraîner une insubordination de la dépendance dont elle fait partie.

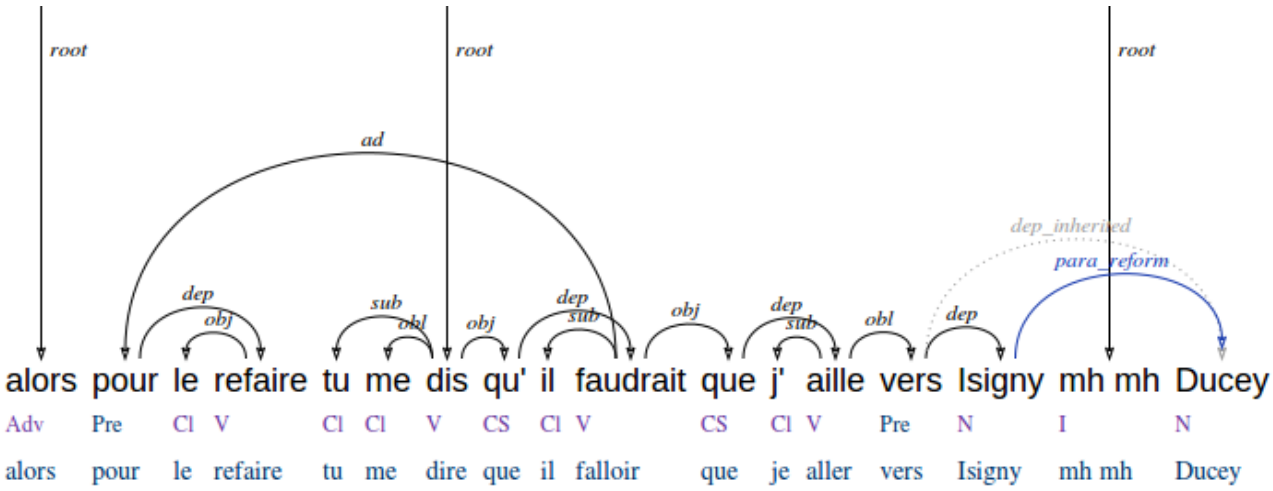

Fig. 4.16. : Adjonction antéposée

Dans la figure 4.16, la dépendance entre "faudrait" et l'adjonction antéposée "pour" est insubordonnée à la dépendance entre "qu'" et "faudrait".

- Le clitique "en" antéposé

En français, le clitique "en" est placé entre le verbe et son sujet, mais peut être dépendant d'un complément postérieur du verbe. Il y a donc insubordination de la dépendance *complément-clitique* par rapport à la dépendance *verbe-complément*.

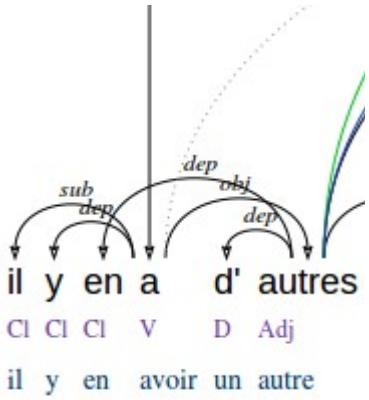

Fig. 4.17. : Clitique antéposé

Dans la figure 4.17, la dépendance entre "autres" et "en", soit celle entre le complément et le clitique antéposé, est insubordonnée à la dépendance entre "a" et "autres", soit celle entre le verbe et le complément.

4.3. Corrélation entre les deux types de non projectivité

4.3.1. Hypothèse

On remarque que les deux types de non projectivité (voir croisements en partie 4.1 et insubordinations en partie 4.2) occurrent dans des situations très similaires. On pourrait donc émettre l'hypothèse qu'un croisement est cooccurrent à une insubordination, et vice versa. Cette

hypothèse n'est pourtant pas toujours vraie.

4.3.2. Résultat

Une comparaison de la liste des arbres présentant un ou plusieurs croisements et de celle des arbres contenant une ou plusieurs insubordinations nous permet de constater que :

- 78 arbres présentent au moins un croisement mais aucune insubordination
- 34 arbres présentent au moins une insubordination mais aucun croisement

Nous présenterons quelques unes des caractéristiques de ces arbres.

4.3.3. Typologie de quelques croisements n'entraînant pas d'insubordination

- Construction en "faire + infinitif"

Dans cette construction, l'éventuel agent de l'infinitif est dépendant du verbe principal : si l'infinif prend aussi un complément et que ce complément est placé après "quelqu'un", alors il y aura croisement sans insubordination.

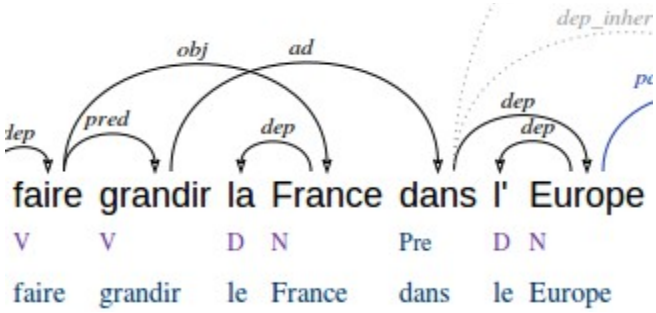

Fig 4.18. : Construction en "faire + infinitif"[1]

Sur la figure 4.18, le complément de "grandir", introduit par "dans", se trouve après le complément du verbe "faire" : il y a croisement entre les liens "faire"-"France" et "grandir"-"dans", mais on ne constate pas d'insubordination.

On ne dénombre néanmoins que deux occurrences de ce phénomène dans le corpus.

1 Le syntagme "la France" a été rattaché à "faire" dans le syntagme "faire grandir la France", car, bien qu'agent de "grandir", il se pronominalise en pronom objet : "la faire grandir".

- Adjonction intercalée entre un mot et son complément

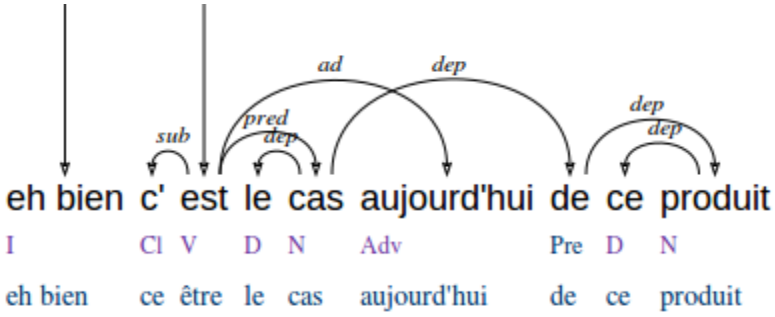

Fig. 4.19. : Adjonction intercalée entre un mot et son complément

Dans la figure 4.19, le croisement entre les liens "est"-"aujourd'hui" et "cas"-"de" n'est pas accompagné d'une insubordination.

4.3.4. Typologie de quelques insubordinations n'entraînant pas de croisement

- Adjonction antéposée

Les adjonctions antéposées constituent la grande majorité des cas d'insubordination sans croisement. On trouve de même quelques compléments obliques antéposés.

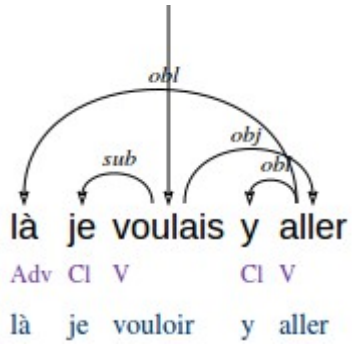

Fig 4.20. : Exemple de complément oblique antéposé

Dans la figure 4.20, la dépendance entre "aller" et "là" est insubordonnée à la dépendance entre "voulais" et "aller", mais cette insubordination n'est pas associée à un croisement.

- Comparaison

On note que l'unique occurrence d'une comparaison qui n'est pas associée à un croisement est une comparaison dont l'adjectif n'est pas prédicat d'un verbe.

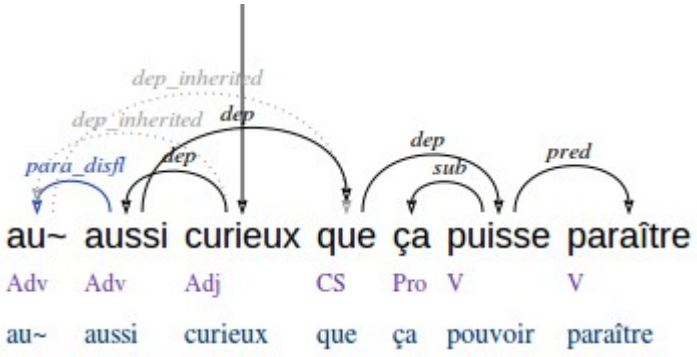

Fig. 4.21. : Comparaison

Dans la figure 4.21., la dépendance entre "aussi" et "que" est insubordonnée à celle entre "curieux" et "aussi", sans qu'on observe pour autant de croisement dans la même zone.

4.3.5. Conclusion

Bien qu'une majorité des cas de croisement entraînent une insubordination, et vice versa, l'existence de contre-exemples nous contraint à réfuter l'hypothèse émise : croisement et insubordination ne sont pas systématiquement cooccurrents.

Chapitre 5

Taille du flux disjoint

L'augmentation de la taille du flux peut être attribuée à deux facteurs : les mots qui ont un grand nombre de dépendants, c'est-à-dire les bouquets de dépendances, et les enchâssements. Nous nous intéresserons ici aux enchâssements, et plus particulièrement aux cas où le nombre de dépendances exclusives est supérieur à 3.

5.1. Définition

On rappelle que le flux est dit "disjoint" en une position donnée lorsque chaque mot à gauche de cette position n'entretient de dépendance qu'avec un seul mot à droite de cette position, et vice versa.

De même, on appellera "dépendance exclusive" une dépendance pour laquelle aucun des mots en jeu n'entretient d'autre dépendance passant par la position étudiée.

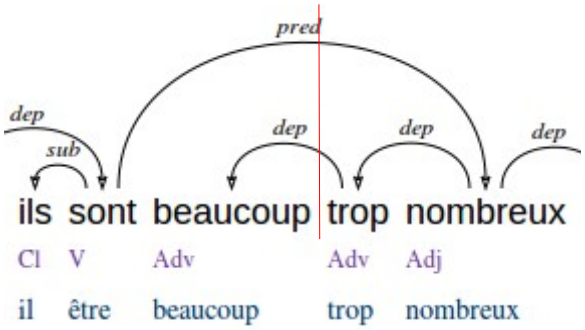

Fig. 5.1.: Exemple de flux disjoint

La figure ci-dessus présente un flux disjoint à la position représentée par la ligne rouge, entre les mots "beaucoup" et "trop". En effet, les mots à gauche de la position ("sont" et "beaucoup") ainsi que ceux à droite ("trop" et "nombreux") n'entrent chacun en jeu que dans une seule dépendance passant par cette position.

Dans le cas présenté ici, on dit que le flux disjoint a une taille de 2, car nous sommes en présence de deux dépendances exclusives.

5.2. Hypothèse

On fait l'hypothèse que la taille du flux disjoint ne peut dépasser 3, c'est-à-dire que pour une position donnée, le nombre de dépendances exclusives ne sera jamais supérieur à 3. Il est néanmoins possible que d'autres dépendances passent par cette position.

Par exemple, dans la figure 5.2, à la position représentée par la ligne rouge, on observe une dépendance exclusive entre les mots "la" et "foi", ainsi que trois dépendances non exclusives reliant les mots "que" et "lorsque" aux mots "est" et "est". On considérera donc que la taille du flux disjoint

est de 1.

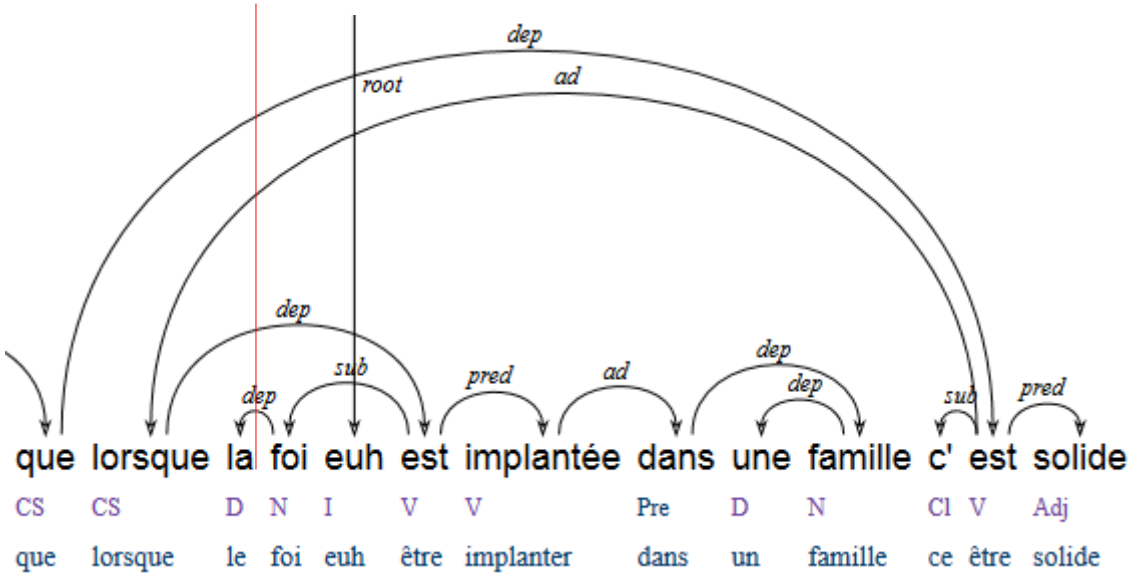

Fig. 5.2. : Exemple de dépendance exclusive accompagnée de dépendances non exclusives

5.3. Résultats

Liens	**1**	**2**	**3**	**4**
Positions	16031	3590	362	38

Fig. 5.3. : Nombre de positions présentant des liens exclusifs

La figure 5.3. indique le nombre de positions dans le corpus Rhapsodie pour lesquelles un, deux, trois ou quatre mots entretiennent une relation exclusive avec un autre mot. La proportion de positions où plus de trois mots entretiennent une relation exclusive avec un autre mot est extrêmement faible (moins de 1%) mais pourrait néanmoins invalider l'hypothèse énoncée.

Une observation des positions pour lesquelles on compte plus de trois dépendances exclusives nous permet de remarquer qu'aucune de ces positions ne présente de dépendance non exclusive. Nous rappelons que, comme indiqué précédemment en partie 2.4.1, ni les liens hérités ni les liens *para_coord* qui doublent deux liens *junc* n'ont été pris en compte.

5.4. Caractéristiques des flux disjoints à plus de trois dépendances exclusives

5.4.1. Matrices des flux

Les flux des 38 positions où on dénombre plus de trois dépendances exclusives peuvent être répartis sur 30 matrices différentes (voir en 3.2). Sur ces 30 matrices, 24 correspondent à un seul flux, 5 correspondent à 2 flux et 1 correspond à 4 flux. Cette dernière matrice est présentée ci-dessous en figure 5.4.

	1	2	3	4
1	[V, CS, obj, +]			
2		[Pre, J, junc, +]		
3			[V, Pre, ad, +]	
4				[Pre, Pre, para_disfl, +]

Fig. 5.4. : Matrice du flux disjoint à 4 dépendances exclusives ou plus le plus fréquent

Il faut néanmoins noter que les quatre occurrences de ce flux sont des positions consécutives du même arbre.

5.4.2. Des dépendances en commun

Les 38 positions où on dénombre plus de trois dépendances exclusives sont réparties sur 16 arbres, donc 6 contiennent plus d'une de ces positions (jusqu'à 7 positions dans un seul arbre). On remarque que, pour tous ces arbres, au moins deux des dépendances exclusives passent par toutes les positions comptabilisées.

5.4.3. Des dépendances à grand empan

Lorsqu'un arbre présente plusieurs positions pour lesquelles la taille du flux disjoint est supérieure à 3, ces positions peuvent être soit proches les unes des autres, soit éloignées les unes de autres.

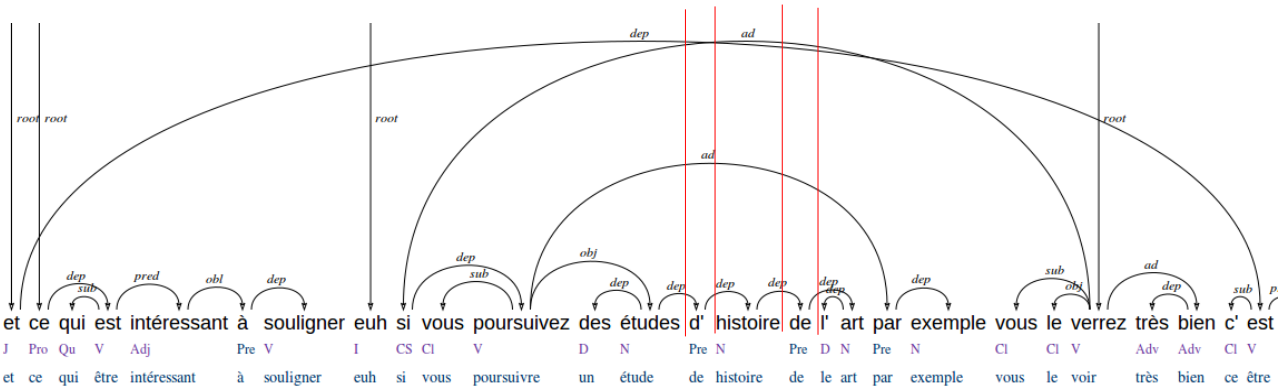

Fig. 5.5. : Exemple de positions proches

Dans la figure 5.5, les positions où la taille du flux disjoint est supérieure à 3 sont représentées par des lignes rouges. On remarque que ces positions sont proches : elles sont même consécutives.

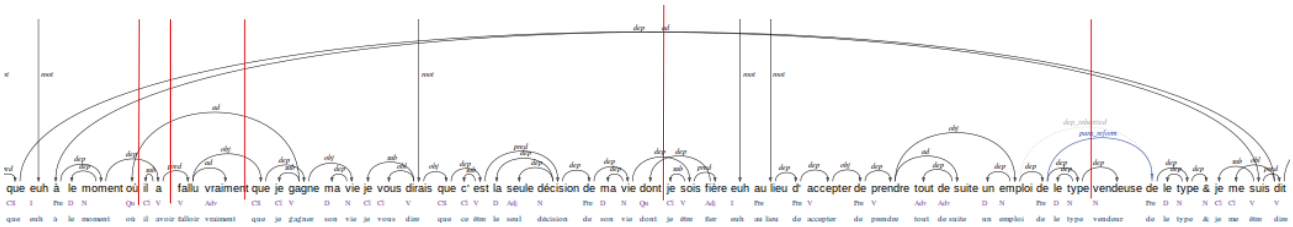

Fig. 5.6. : Exemple de positions éloignées

Dans la figure 5.6, les positions où la taille du flux disjoint est supérieure à 3 sont aussi représentées par des lignes rouges. Si les trois premières peuvent être considérées comme proches, on remarque que les deux dernières sont éloignées des autres positions.

Néanmoins, comme énoncé précédemment, si un arbre regroupe plusieurs positions où la taille du flux disjoint est supérieure à 3, alors il y a au moins deux dépendances qui passent par toutes ces positions. Dans le cas de positions éloignées (figure 5.6), cela implique que ces dépendances aient un grand empan, c'est-à-dire que les deux extrémités de la dépendance sont éloignées, séparées par un nombre important de mots.

Les dépendances à grand empan s'additionnent donc aux autres dépendances plus courtes, augmentant ainsi la taille du flux. Pour expliquer la taille du flux disjoint, on peut donc se pencher sur ces dépendances à grand empan.

5.5. Les dépendances à grand empan

Nous venons d'affirmer que l'étude des dépendances à grand empan peut être un moyen d'expliquer la taille du flux disjoint. Nous nous intéresserons donc ici à deux choses : tout d'abord, les caractéristiques des dépendances à grand empan ; enfin, les causes de la longueur de ces dépendances.

Nous ne prendrons néanmoins en compte que les dépendances qui passent par des positions où la taille du flux disjoint est supérieure à 3.

5.5.1. Typage des dépendances à grand empan

On peut extraire des catégories de dépendances à grand empan en observant les dépendances situées dans la case en haut à gauche de la matrice du flux : ce sont les dépendances qui relient les mots les plus éloignés. Dans le cas d'un croisement au sommet du flux, on observera les dépendances en ligne 1, colonne 2 et ligne 2, colonne 1.

Ces dépendances ainsi que leurs occurrences dans les flux disjoints étudiés sont présentées dans le tableau 5.7 ci-après.

Dépendance	Occurrences
[J, V, dep, +]	15
[V, Pre, ad, -]	9
[V, CS, obj, +]	7
[CS, V, dep, +]	5
[N, J, junc, +]	4
[N, N, para_coord, +]	4
[V, CS, ad, -]	1
[V, N, sub, -]	1
[V, Pre, ad, +]	1
[V, Adv, ad, +]	1
[N, N, para_dform, +]	1
[V, V, para_coord, +]	1

Fig. 5.7. : Dépendances les plus longues pour chaque position

Si les types de dépendances présentés dans le tableau précédent sont assez variés, on peut néanmoins en extraire cinq grandes catégories :

- les dépendances entre une conjonction de coordination et un verbe
- les dépendances entre un verbe et son sujet ou un de ses compléments
- les dépendances entre une conjonction de subordination et le verbe de la subordonnée
- les liens paradigmatiques
- les jonctions entre un mot et un joncteur ou entre deux mots coordonnés sans joncteur

5.5.2. Causes de la longueur des dépendances à grand empan

La longueur d'une dépendance est bien évidemment due au nombre de mots se trouvant entre ses extrémités. Mais la présence de ces mots peut être due à différents phénomènes, qui sont susceptibles de se combiner. Parmi eux, on trouve :

- une autre dépendance à grand empan
- une incise ou une parenthèse
- un autre dépendant du gouverneur dans le cas d'une dépendance postposée
- un dépendant du dépendant dans le cas d'une dépendance antéposée
- un lien paradigmatique

Dans les figures qui suivront, les positions pour lesquelles la taille du flux disjoint est supérieure à 3 seront représentées à titre indicatif par des lignes rouges.

- Autre dépendance à grand empan

La présence d'une dépendance à grand empan est parfois due à une autre dépendance à grand empan. La présence de cette autre dépendance sera elle due à un autre phénomène.

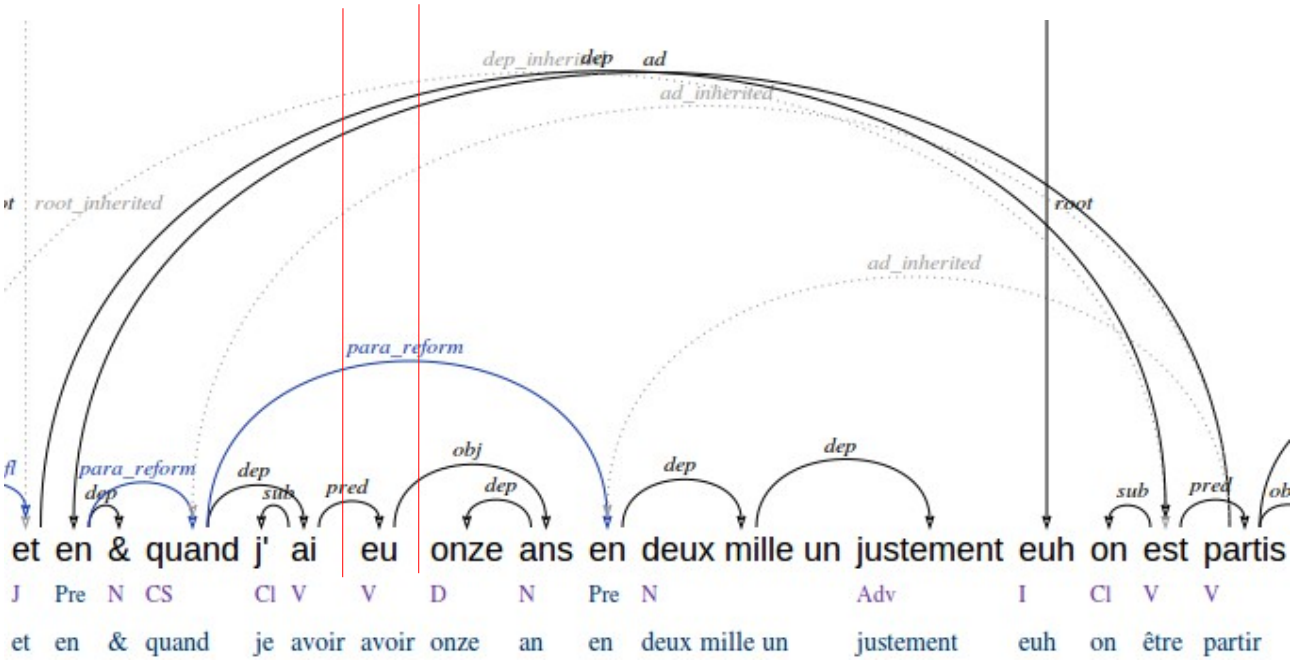

Fig. 5.8. : Autre dépendance à grand empan

Dans la figure 5.8, la longueur de la dépendance entre "partis" et "en", étiquetée *ad*, entraîne un allongement de la dépendance entre "et" et "est", étiquetée *dep*. La longueur de cette seconde dépendance est donc due à la longueur de la première.

- Incise ou parenthèse

Les incises et les parenthèses peuvent être placées entre les deux extrémités d'une dépendance, ce qui implique un allongement de cette dépendance.

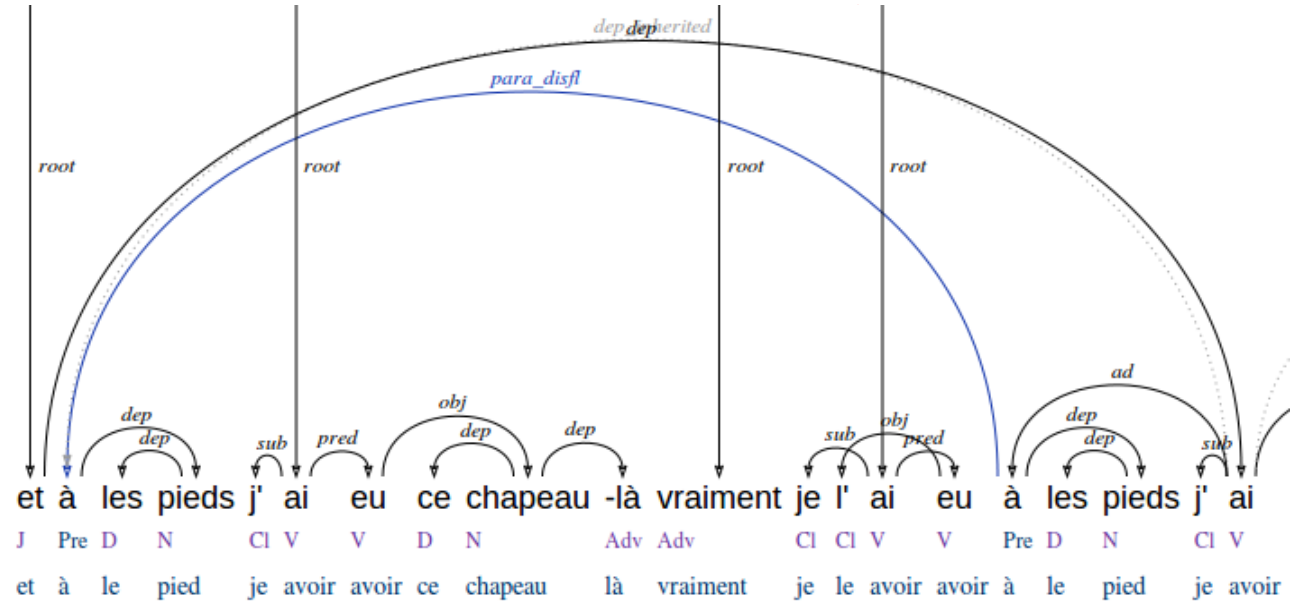

Fig. 5.9. : Parenthèses

Dans la figure 5.9, on trouve trois parenthèses ("j'ai eu ce chapeau-là", "vraiment" et "je l'ai eu"), qui entraînent un allongement de la dépendance *para_disfl* entre "à" et "à". Ici, on peut même les considérer comme la cause de la présence de cette dépendance *para_disfl*. Elles entraînent aussi un allongement de la dépendance entre "et" et "ai", étiquetée *dep*.

- Autre dépendant du gouverneur dans le cas d'une dépendance à tête initiale

Soit trois mots A, B et C tels que A gouverne B, A gouverne C et A < B < C. Si B gouverne par projection toute une suite de mots, alors la dépendance A-C peut être une dépendance à grand empan.

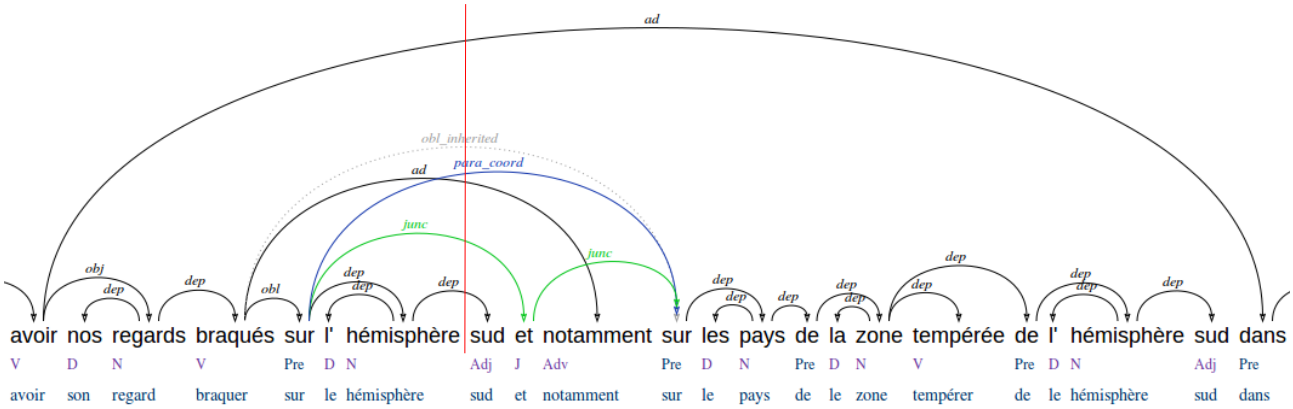

Fig. 5.10. : Dépendant du gouverneur

Dans la figure 5.10, le verbe "avoir", gouverneur de la dépendance à grand empan "avoir"-"dans", est aussi gouverneur de la dépendance "avoir"-"regards", qui s'intercale entre les deux extrémités de la première dépendance. Les descendants du mot "regards", et donc par extension les descendants du mot "avoir", sont donc la cause de l'allongement de la dépendance "avoir"-"dans".

En principe, plus un dépendant est lourd (c'est-à-dire qu'il a beaucoup de mots dans sa projection), plus il est éloigné de son gouverneur, et cela justement pour éviter les dépendances trop longues et les contraintes prosodiques que cela génère. On peut donc s'attendre à ce que le mot "dans" gouverne une suite de dépendants de longueur plus importante que "regards".

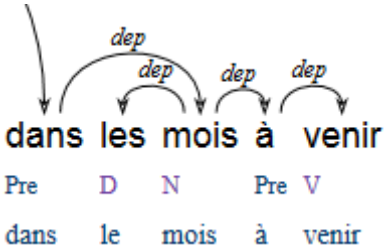

Fig. 5.11. Poids du mot "dans"

La figure 5.11 présente la suite de dépendants du mot "dans". Sa longueur est nettement moins importante que celle des dépendants du mot "regards", ce qui transgresse le principe énoncé précédemment. Il sera intéressant de se pencher sur la prosodie de cette phrase, afin de mieux comprendre comment l'ordre inhabituel de ces deux dépendants a été produit par le locuteur.

- Lien paradigmatique

Le corpus étudié étant un corpus oral, de nombreuses hésitations ou reformulations sont présentes.

Dans la figure 5.12 ci-après, le dépendant antéposé du participe passé "parti" entre en jeu dans une dépendance paradigmatique, ce qui entraîne un allongement de la dépendance entre "parti"

et "en", et donc de celle entre "et" et "est", comme indiqué précédemment.

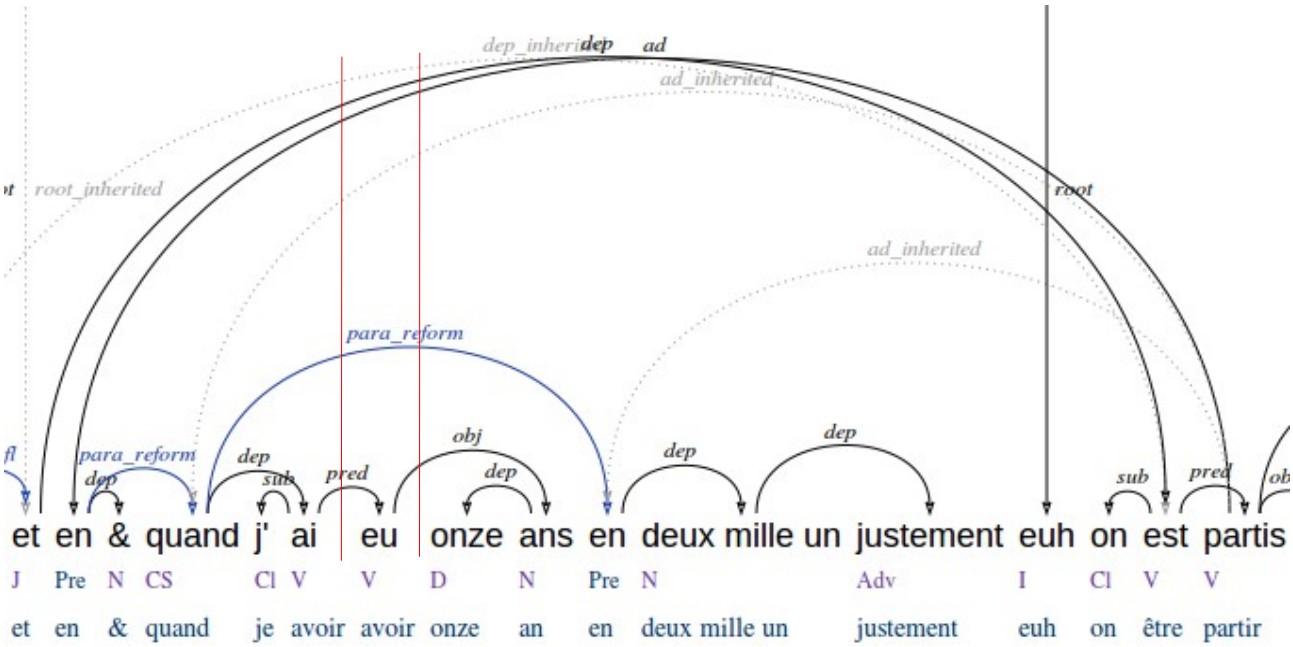

Fig. 5.12. : Liens paradigmatiques

- Dépendant du dépendant dans le cas d'une dépendance à tête finale

Si le dépendant d'une dépendance à tête finale D1 gouverne par projection toute une suite de mots, alors D1 peut être une dépendance à grand empan.

Dans la figure 5.13 ci-après, la longueur de la dépendance *sub* entre le verbe "dispose" et le nom "parti" est uniquement due à la succession de mots dans la projection du dépendant.

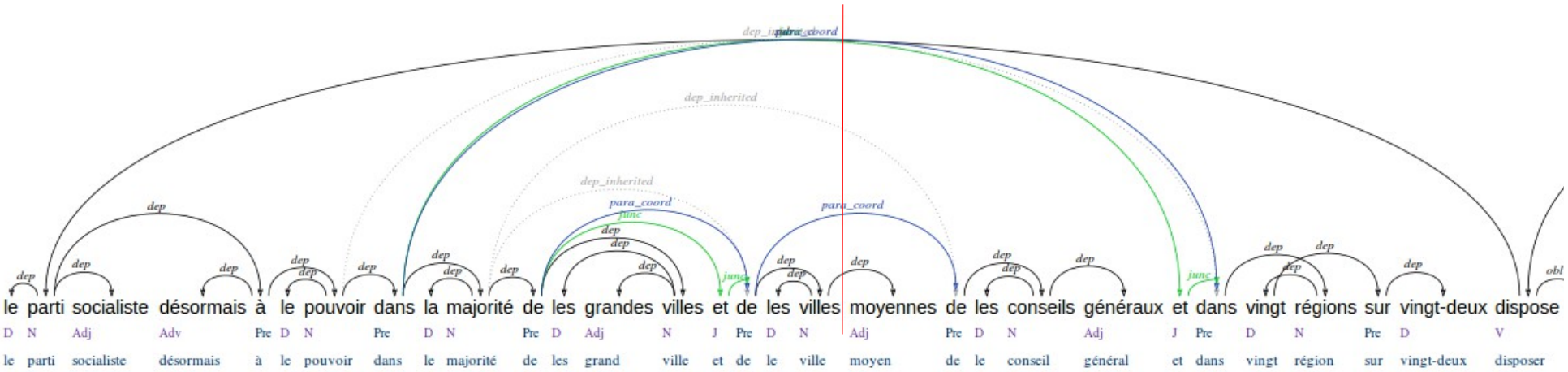

Fig. 5.13. : Dépendant du dépendant

Chapitre 6

Conclusion

L'analyse que nous avons effectuée au cours de ce mémoire apporte une réponse à certaines des questions initialement posées. D'autres restent en suspens, et d'autres encore sont apparues au cours de notre étude.

6.1. Conclusions sur le flux de dépendance

Quelles sont les conditions d'apparition d'un flux non projectif ? Croisement et insubordination sont-ils cooccurrents ? Quelles sont les limitations que rencontre le flux ? Qu'est-ce qui influence la taille du flux ?

Nous avons déterminé les conditions d'apparition du flux non projectif (partie 4), et avons établi une typologie des croisements et des insubordinations. Celle-ci nous a ensuite permis de montrer que ces deux phénomènes étaient fréquemment cooccurrents. L'étude de la pseudo-projectivité a ensuite mis en valeur l'influence des liens paradigmatiques dans les caractéristiques du flux.

Nous nous sommes ensuite penchés sur le nombre maximal de dépendances exclusives (partie 5). Notre hypothèse de départ était qu'à une position donnée ce nombre ne dépassait jamais 3, soit un total de 6 mots (3 de chaque côté de la position), valeur proche de celle de 7±2 avancée par (Miller 1956). La proportion de positions pour lesquelles le nombre de dépendances exclusives ne correspond pas à ce critère étant inférieure à 1%, l'hypothèse est considérée comme validée.

6.2. Critiques et remarques

Le corpus Rhapsodie ne rassemble que trois heures d'enregistrement : s'il couvre un large spectre des variétés du français oral, il reste trop peu fourni pour servir d'appui à une étude approfondie du flux de dépendance. Il a permis d'extraire quelques principes sur le flux de dépendance, mais ceux-ci devront être vérifiés sur un corpus plus conséquent.

Les choix pour l'analyse microsyntaxique font l'objet de questionnements. Dans une construction au passé composé, par exemple, l'auxiliaire est considéré comme gouverneur du sujet, mais c'est le participe passé qui gouverne les compléments : dans le cas d'un complément pronominalisé, cela entraîne un croisement (partie 4.1.2). Une autre analyse du passé composé considère que les clitiques dépendent tous de l'auxiliaire (Mel'čuk 1988) : il n'y a alors plus de croisement. Il en va de même pour le traitement des liens paradigmatiques : lors de la répétition d'un syntagme, on peut soit considérer qu'il n'y a qu'un seul lien paradigmatique, qui relie les deux têtes, soit considérer que chaque élément du syntagme entretient un lien avec sa répétition dans l'autre syntagme. Choisir l'une ou l'autre des analyses aura une influence sur la taille du flux.

6.3. Perspectives

Comme précisé précédemment, il sera intéressant de valider les résultats énoncés dans ce mémoire par une analyse sur un corpus de français oral plus conséquent. On pourra ensuite comparer ces résultats avec ceux obtenus à partir d'un corpus de français écrit : (Jardonnet 2009) avait par exemple montré que la taille du flux dans le French Treebank de l'Université Paris 7, constitué de phrases extraites du quotidien *Le Monde*, pouvait dépasser 5, valeur qu'elle n'atteint qu'une seule fois dans le corpus Rhapsodie. Cela permettra de mettre en valeur certaines caractéristiques du français oral.

De même, le corpus Rhapsodie couvre différentes variétés de français oral : discours présidentiels, interviews, voire simple demande de chemin dans la rue. Ces variétés pourraient être réparties en catégorie selon certains critères prédéfinis (par exemple contexte formel/informel, discours préparé/spontané, interlocuteur connu/inconnu...) : les résultats obtenus varieraient probablement.

En ce qui concerne les choix d'analyse microsyntaxique, réétudier le flux en s'appuyant sur le même corpus annoté avec d'autres choix permettra de relativiser certaines des valeurs obtenues, ou au contraire de mettre l'accent sur certains phénomènes passés inaperçus.

En partie 3, nous avions présenté succinctement une représentation du flux sous la forme d'une structure de traits. Cette structure pourra servir de base à un programme de requêtes sur le corpus, qui permettra de rechercher dans le corpus des positions pour lesquelles le flux correspond à un certain patron.

La représentation du flux de dépendance sous la forme d'une matrice entraîne aussi de nouvelles questions. Peut-on prévoir certaines transitions en fonction de la matrice de départ ? Existe-t-il des combinaisons d'opérations de transition impossibles ? De même, on pourra étudier les triplets formés par les matrices de deux positions successives et le mot situé entre ces deux positions. Les réponses à ces questions pourront servir de point de départ au développement de nouveaux parseurs syntaxiques.

On pourra de plus se pencher sur la corrélation entre le flux de dépendance et la prosodie, en partant de l'hypothèse que plus le flux est important, moins une frontière prosodique est souhaitable, et inversement.

L'étude réalisée au cours de ce mémoire a mis en évidence certains grands principes du flux de dépendance. Chacun de ces principes pourra faire l'objet d'études plus approfondies, qui permettront de mieux appréhender la richesse et la complexité du français oral.

Bibliographie

Bawden, R., Botalla, M.-A., Gerdes, K. et Kahane, S. (2014). Correcting and Validating Syntactic Dependency in the Spoken French Treebank Rhapsodie. In *Proceedings of the 9th Language Resources and Evaluation Conference* (*LREC*), Reykjavik

Bérard, L. (2012). *Dépendances à distance en français contemporain - Etude sur corpus : "c'est ce qu'on pense qui devrait être fait"*. Thèse de doctorat, Université de Lorraine

Chomsky, N. (1965). *Aspects of the Theory of Syntax.* Cambridge MA: MIT Press

Hajič, J., Hladká, B. et Petr Pajas, P. (2001). The Prague Dependency Treebank: Annotation Structure and Support. In *Proceedings of the IRCS Workshop on Linguistic Databases*, pp. 105-114, Philadelphia, December 2001. University of Pennsylvania.

Ihm, P. et Lecerf, Y. (1963). *Eléments pour une grammaire générale des langues projectives.* Rapport EUR 210.f, CETIS

Jardonnet, U. (2009). Analyse du flux de dépendance. Mémoire de master, Université Paris Ouest Nanterre La Défense

Kahane, S., Nasr, A. et Rambow, O. (1998). Pseudo-Projectivity: A Polynomially Parsable Non-Projective Dependency Grammar. In *Proceedings of ACL/COLING*, Montréal, 646-52

Kahane, S. (2001). Grammaires de dépendance formelles et théorie Sens-Texte. *TALN 2001*

Kahane S. (avec la participation de K. Gerdes, P. Pietrandrea, C. Benzitoun, R. Bawden) (2012). Protocole de codage microsyntaxique. http://www.projet-rhapsodie.fr/

Marcus, M., Santorini, B. et Marcinkiewicz, M. A. (1993). Building a Large Annotated Corpus of English: The Penn Treebank. In *Computational Linguistics*, volume 19, n° 2, pp. 313-330.

Mel'čuk, I. (1988). *Dependency Syntax: Theory and Practice.* Albany, N.Y.: The SUNY Press

Miller, G. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. In *The Psychological Review, vol. 63, pp. 81-97*

Nasr, A. (1996). *Un modèle de reformulation automatique fondé sur la Théorie Sens-Texte : Applications aux langues contrôlées.* Thèse de doctorat, Université Paris 7

Nivre, J., Hall, J. et Nilsson, J. (2004). Memory-based dependency parsing. In Ng, H. T. and Riloff, E. (eds.) *Proceedings of the Eighth Conference on Computational Natural Language Learning (CoNLL)*, May 6-7, 2004, Boston, Massachusetts, pp. 49-56

Nivre, J. (2005). Dependency grammar and dependency parsing. MSI report 05133, Växjö University: School of Mathematics and Systems Engineering.

Nivre, J. (2006). Constraints on Non-Projective Dependency Parsing. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pp. 73-80

Nivre, J. (2013). Training Deterministic Parsers Using Non-Deterministic Oracles. Diaporama de la conférence au séminaire Alpage, juin 2013

Pretkalniņa, L., Znotiņš, A., Rituma, L. et Goško, D. (2014). Dependency parsing representation effects on the accuracy of semantic applications — an example of an inflective language. In *Proceedings of the 9th Language Resources and Evaluation Conference* (*LREC*), Reykjavik

Villemonte de la Clergerie, E. (2010). Building factorized TAGs with meta-grammars. In *Proceedings of the 10th International Conference on Tree Adjoining Grammars and Related Formalisms - TAG+ 10*, pp. 111-118

Yngve, V. (1960). A Model and an Hypothesis for Language Structure. In *Proceedings of the American Philosophical Society*, Vol. 104, n° 5, pp. 444-466

Annexe

Liste des principaux scripts Python rédigés et utilisés au cours de ce mémoire

bdd_trees.py : Transforme les arbres de la base de données de Rhapsodie en dictionnaire exploitable par un script Python

toutesmatrices.py : Fournit les matrices du flux pour chaque position (utilisé dans la partie 3)

proj_type1.py : Fournit pour chaque arbre la liste des paires de dépendances qui se croisent (utilisé dans la partie 4)

proj_type2.py : Fournit pour chaque arbre la liste des paires de dépendances qui présentent un insubordination (utilisé dans la partie 4)

elts_gd.py : Produit un dictionnaire indiquant, pour chaque position, l'indice absolu des mots qui entrent en jeu dans une dépendance passant par cette position, en précisant s'ils sont placés à gauche ou à droite (utilisé dans la partie 5)

l_sup3.py : À partir du dictionnaire produit par le programme elts_gd.py, fournit les positions pour lesquelles il y a plus de trois dépendances exclusives (utilisé dans la partie 5)

Code source du script de générations des matrices du flux

toutesmatrices.py

```
#!/usr/bin/python
# -*- coding: utf-8 -*-

import pickle

# Récupération des dictionnaires

id_dep = pickle.load(open("id_dep.p","rb"))
# contient toutes les dépendances sous la forme {n° arbre: [[indice du dep,
indice du gov, fonction, sens+empan, cat du dep, cat du gov, mot dep, mot
gov],...]}
dep_ord = pickle.load(open("dep_ord.p","rb"))
#contient pour chaque arbre les dépendances sous la forme {n° arbre : [[dep1,
gov1, fct], [dep2, gov2, fct], ...]}, avec dep1 < dep2

# Production du dictionnaire contenant la liste des dépendances pour chaque
position

id_flux_temp = {}

for t in dep_ord.keys():
      id_flux_temp[t] = {}
      l = []
```

```
        l_fin = []
        posmax = 0
        for dep in dep_ord[t]:
                if dep[1] > posmax:
                        posmax = dep[1]
        pos = 1
        while pos < posmax:
                l.append([])
                if len(l) != 1:
                        l[-1].extend(l[-2])
                for dep in dep_ord[t]:
                        if dep[0] == pos:
                                l_fin.append(dep[1])
                                l[-1].append(dep)
                f = 0
                while f < len(l_fin):
                        if l_fin[f] == pos:
                                l_fin.pop(f)
                                l[-1].pop(f)
                                f -= 1
                        f += 1
                id_flux_temp[t][pos] = l[-1]
                pos += 1

# à ce moment, id_flux_temp contient, pour chaque position, la liste des
dépendances qui y passent, sous la forme [indice du mot de gauche, indice du mot
de droite]

id_flux = {}

for t in id_flux_temp.keys():
        id_flux[t] = {}
        for p in id_flux_temp[t].keys():
                id_flux[t][p] = []
                for dep in id_flux_temp[t][p]:
                        for elt in id_dep[t]:
                                a = [elt[0],elt[1]]
                                a.sort()
                                if dep[0] == a[0] and dep[1] == a[1]:
                                        id_flux[t][p].append(list(elt)[:-2])

for t in id_flux.keys():
        for p in id_flux[t].keys():
                entree = []
                sortie = []
                for dep in id_flux[t][p]:
                        if min(dep[0],dep[1]) not in entree:
                                entree.append(min(dep[0],dep[1]))
                        if max(dep[0],dep[1]) not in sortie:
                                sortie.append(max(dep[0],dep[1]))
                entree.sort()
                sortie.sort()
                sortie.reverse()
                for d in range(len(id_flux[t][p])):
                        dep = id_flux[t][p][d]
                        ajout = []
                        pl_in = 1
                        for i in range(len(entree)):
                                if entree[i]==min(dep[0],dep[1]):
                                        ajout.append(pl_in)
                                pl_in += 1
```

```
                    pl_out = 1
                    for i in range(len(sortie)):
                          if sortie[i]==max(dep[0],dep[1]):
                                ajout.append(pl_out)
                          pl_out += 1
                    id_flux[t][p][d] = id_flux[t][p][d]+ajout

# Constitution des matrices sous la forme d'un dictionnaire

mat_flux = {}

for t in id_flux.keys():
      mat_flux[t] = {}
      for p in id_flux[t].keys():
            mat_flux[t][p] = {}
            lin = 0
            col = 0
            for dep in id_flux[t][p]:
                  if dep[-2] > lin:
                        lin = dep[-2]
                  if dep[-1] > col:
                        col = dep[-1]
            for i in range(lin):
                  mat_flux[t][p][i+1] = []
                  for j in range(col):
                        mat_flux[t][p][i+1].append("")
            for dep in id_flux[t][p]:
                  pl_in = dep[-2]
                  pl_out = dep[-1]-1
                  mat_flux[t][p][pl_in][pl_out] = dep

# Exportation des données

pickle.dump(mat_flux,open("mat_flux.p","wb"))

# Impression html

fichier = open("toutesmatrices.html","w")

print >> fichier, "<html><head><meta http-equiv=\"Content-Type\"
content=\"text/html; charset=utf-8\"></head><body>"

for t in mat_flux.keys():
      print t
      for p in mat_flux[t]:
            print >> fichier, "<table border=\"1\">"
            print >> fichier, "<caption>",t,"-",p,"</caption>"

            print >> fichier, "<tr>"
            print >> fichier, "<th></th>"
            if len(mat_flux[t][p]) > 0:
                  for col in range(len(mat_flux[t][p][1])):
                        print >> fichier, "<th>",col+1,"</th>"
            print >> fichier, "</tr>"

            for lin in mat_flux[t][p].keys():
                  print >> fichier, "<tr>"
                  print >> fichier, "<th>",lin,"</th>"
                  for col in mat_flux[t][p][lin]:
                        if len(col) > 0:
                              print >> fichier, "<td>["
```

```
                                print >> fichier, col[5].encode("utf-8"),","
                                print >> fichier, col[4].encode("utf-8"),","
                                print >> fichier, col[2].encode("utf-8"),","
                                if col[3] > 0 :
                                      print >> fichier, "+"
                                if col[3] < 0:
                                      print >> fichier, "-"
                                print >> fichier, "]</td>"
                          else:
                                print >> fichier, "<td>",col,"</td>"
                    print >> fichier, "</tr>"
              print >> fichier, "</table> <br/>"

        print >> fichier, "<br/> ---------------- <br/>"

print >> fichier, "</body></html>"

fichier.close()
```