

**UNE HYPOTHÈSE SUR L'ORDRE LINÉAIRE:  
L'ARBRE DE DÉPENDANCE ET SA LINÉARISATION,  
OPTIMISÉE SOUS DES CONTRAINTES  
TOPOLOGIQUES, MÉTRIQUES ET MÉMORIELLES**

JACQUES VERGNE

*GREYC URA 1526 Université de Caen*

## 0. Introduction

Comme point de départ, nous nous plaçons dans le cadre théorique de Lucien Tesnière, en reprenant sa définition de la connexion comme processus; mais nous prenons cette définition au mot en la précisant et en la fondant sur un processus qui utilise la mémoire à court terme de la personne qui écoute ou lit; de plus nous prenons comme segment relié non pas le mot comme le fait Tesnière, mais une suite de quelques mots contigus: le syntagme noyau (sections 1 et 2). Nous obtenons ainsi un arbre de dépendance assez différent du stemma de Tesnière (section 3).

En ce qui concerne la correspondance entre ordre linéaire et ordre structural, nous abandonnons le concept de projection de l'arbre sur l'axe linéaire (le résultat de la projection est dépendant du dessin de l'arbre), et nous posons une hypothèse qui utilise des concepts issus de l'algorithmique, concepts sur les modes d'énumération des nœuds d'un arbre quelconque (les énumérations possibles dépendent de la structure de l'arbre et non pas de son dessin): parmi toutes les linéarisations possibles, ne sont réalisées que celles qui minimisent la distance entre deux segments reliés. Nous donnons une interprétation mémorielle de cette hypothèse: minimiser la distance entre deux segments reliés permet de minimiser l'effort de mémoire (section 4).

Nous confrontons ensuite cette hypothèse à un corpus de phrases en français où un verbe a un objet et un autre complément, et, au sujet des ordres linéaires de ces trois segments, nous comparons l'ordre attesté avec l'ordre possible, ce qui nous permet d'expliquer pourquoi l'objet est ou n'est pas contigu au verbe (section 5).

Ensuite, nous étudions la linéarisation de la proposition relative en français, et les mêmes concepts nous permettent d'expliquer pourquoi le sujet est postposé au verbe dans certaines relatives (section 6).

## 1. Connexion et segments connectés

### 1.1. La connexion: un retour à la définition originale de Tesnière

La connexion est le concept central de Tesnière, c'est une relation entre deux mots:

"entre un mot et ses voisins, l'esprit aperçoit des connexions"

(voir Tesnière, 1959: page 11, en fait la première page de Tesnière)

Cette définition de la connexion ne concerne les formes qu'indirectement, car il ajoute: "ces connexions ne sont indiquées par rien"; ceci est à tempérer par: "deux mots en connexion sont généralement placés l'un à côté de l'autre sur la chaîne linéaire" (voir dans l'esquisse, Tesnière, 1953: page 4).

Relevons trois points dans sa définition de la connexion:

- (1) c'est un processus humain ("aperçoit"),
- (2) un processus mental ("l'esprit"),
- (3) un processus lié à la perception d'une personne qui lit ou entend ("aperçoit").

Nous allons proposer une définition fondée sur un processus, qui reprend ces trois caractéristiques en précisant le détail de ce processus (jusqu'à rendre possible sa modélisation dans un processus informatisé).

### 1.2. Les segments connectés: non pas des mots mais des groupes de quelques mots contigus

Tesnière présente la phrase comme un ensemble organisé de mots: il propose donc une hiérarchie à 2 niveaux, le mot et la phrase:

**phrases**                      (composées de mots contigus et **connectés**)  
**mots**

Il introduit le concept de connexion (qui est traditionnellement nommée dépendance) dès la première page; c'est dire qu'il introduit juste le strict nécessaire de segmentation pour présenter une relation.

Il introduit d'abord un segment non fondé sur une relation (le mot), puis une relation, définissant ainsi séparément et successivement la segmentation, puis la relation, alors que le syntagme (tel qu'il est défini par Chomsky dans la théorie X barre) est un segment dont la définition s'appuie sur une relation, car un syntagme inclut ses compléments, qui sont eux mêmes des syntagmes.

Nous appelons **segment primaire** un segment dont la définition ne fait pas appel à une relation de dépendance - complémentation (mot, syntagme noyau), et **segment secondaire** un segment fondé sur une relation de dépendance - complémentation (syntagme, proposition).

Nous proposons une hiérarchie de segments ayant deux niveaux intermédiaires supplémentaires:

- le syntagme noyau (nominal ou verbal), syntagme sans ses compléments, segment définissable par ses constituants et par ses frontières, sans utiliser une relation de complémentation, segment non récursif, en résumé segment primaire; dans la suite de ce travail, nous appelons ce segment **séquence**, séquence nominale ou verbale; une séquence est constituée d'un mot central (un nom ou un verbe) et de ses satellites contigus (déterminant, adjectif autour du nom, et négation, pronom atone, adverbe autour du verbe). Les mots situés hors des séquences sont les prépositions, les conjonctions, les pronoms relatifs, et la ponctuation, appelons-les **clips**, pour évoquer leur rôle d'accrocheur des séquences dans la structure de la phrase.

- le **bloc**, composé de quelques séquences contiguës, ou séparées par des blocs insérés entre deux séquences du même bloc, avec un clip éventuel en début de bloc:

<b>phrases</b>	(composées de <b>séquences</b> contiguës et <b>connectées</b> )
	(composées de <b>blocs</b> contigus ou insérés)
<b>blocs</b>	(composés de <b>séquences</b> )
<b>séquences</b>	(composées de mots contigus)
<b>mots</b>	

Le concept de segment primaire permet de fonder une première segmentation sans faire appel à un concept de relation de dépendance, puis de fonder un concept de relation de dépendance sur le concept de segment primaire: une relation entre segments primaires; on a ainsi dissocié segmentation et relation de dépendance.

D'autre part, le concept de segment primaire permet de construire une hiérarchie de segments où chaque niveau est caractérisé par un type particulier de segment (d'où un nombre de niveaux fixé au départ), et où un segment du niveau **n** est constitué de segments du niveau **n - 1** : un tout est d'un type différent de ses parties (un segment primaire est défini non récursivement). Ceci a pour conséquence qu'on peut s'abstraire du niveau le plus fin (le mot) pour exprimer synthétiquement les structures des niveaux supérieurs. Une structure est intelligible si on porte le regard sur des niveaux voisins de la hiérarchie: ainsi, la question de l'ordre linéaire se pose confusément en termes d'ordre linéaire des mots dans les phrases, mais elle se clarifie, quand on la pose en termes d'ordre linéaire des mots dans les séquences, et en termes d'ordre linéaire des séquences dans les phrases.

Voici un exemple de phrase segmentée en séquences:

(conventions: séquence nominale **séquence verbale** clip)  
La mesure de concentration de chlorophylle est utilisée  
pour **estimer** la biomasse phytoplanctonique .

Voici un exemple de la même phrase segmentée en blocs, avec deux blocs insérés entre sujet et verbe (conventions: limites de bloc marquées par []):

[ La mesure [ de concentration ] [ de chlorophylle ] **est utilisée** ]  
[ pour **estimer** la biomasse phytoplanctonique ] .

## 2. Dépendance mémorielle

### 2.1. La dépendance entre séquences dans la phrase

Dans la suite de cette étude, nous considérerons les phrases comme des suites de séquences contiguës et connectées, en abandonnant ainsi le niveau hiérarchique des mots, et nous ne nous intéresserons qu'aux connexions entre séquences. Quant à ces connexions, nous allons les nommer dépendances selon l'usage actuel. Quand nous écrirons "verbe", il faudra entendre

"séquence verbale", quand nous écrivons "sujet" ou "objet", il faudra entendre "séquence nominale sujet" ou "séquence nominale objet".

La dépendance entre séquences est définie comme résultat d'un processus de calcul à partir des positions relatives des séquences, calcul effectué par le lecteur - auditeur:

une dépendance entre 2 séquences est une relation que le lecteur - auditeur établit entre:

une séquence qu'il est en train de lire - entendre  
et une autre séquence qu'il a déjà en mémoire.

Voici comme exemple le processus de mise en relation d'une séquence verbe avec une séquence sujet:

- 1- le lecteur - auditeur mémorise une séquence A comme étant un sujet en attente de verbe
- 2- il s'en souvient en recevant une séquence B, séquence verbale conjuguée
- 3- il relie B à A (la séquence nominale sujet possible la plus récente): sujet <-- verbe
- 4- puis il oublie A comme sujet n'attendant plus de verbe
- 5- et il oublie toutes les séquences situées entre A et B comme compléments du sujet n'attendant plus de compléments (ce dernier oubli est fondé sur une propriété topologique de l'arbre de dépendance, qui sera développée ci-dessous).

Nous abandonnons ainsi une vision statique de la dépendance pour la conception dynamique d'un processus exécuté par le lecteur - auditeur.

Tesnière reprend le concept traditionnel de subordination, sans plus rattacher la dépendance à sa définition première de la connexion comme processus mental. Il ne distingue pas les deux types de dépendance suivants:

- la dépendance d'un "actant" ou d'un "circonstant" à un verbe, qu'on pourrait appeler dépendance "actancielle":

"Quant aux actants et aux circonstants, ce sont les subordonnés immédiats du verbe."  
(voir Tesnière, 1959: page 103, §13)

- et la dépendance d'un adjectif à un nom, qu'on pourrait appeler dépendance de "détermination":

"De même que le verbe régit un certain nombre de subordonnés, qui sont les actants et les circonstants, de même ces subordonnés du verbe ont à leur tour des subordonnés de différentes natures." (voir Tesnière, 1959: page 144, §1)

Il nous semble que le fait de mélanger dans le même arbre deux types différents de dépendance crée un arbre hétérogène, et masque certaines propriétés.

Pour la distinguer de la dépendance de Tesnière, nous proposons d'appeler la nouvelle dépendance définie ci-dessus **dépendance mémorielle**.

Le principe linguistique général de ce calcul est la saturation de valence à distance minimale: il faut comprendre "saturation" non comme un état mais encore comme un

processus effectué par le lecteur - auditeur. Ce calcul est un travail de la mémoire à court terme du lecteur - auditeur, plus exactement un cycle: mémoriser, se souvenir, relier, oublier, au moment de la lecture - audition de chaque séquence. Ces opérations se font sur des mémoires spécialisées: par exemple, la mémoire des séquences nominales sujets en attente de séquence verbale conjuguée.

Voici un exemple d'arbre de dépendance à la fin du calcul des relations:

phrase	représentation en séquences
<u>la mesure</u>	N
. de <u>concentration</u>	. pN
. . de <u>chlorophylle</u>	. . pN
. <b>est utilisée</b>	. V
. . pour <b>estimer</b>	. . qI
. . . <u>la biomasse phytoplanctonique</u>	. . . N

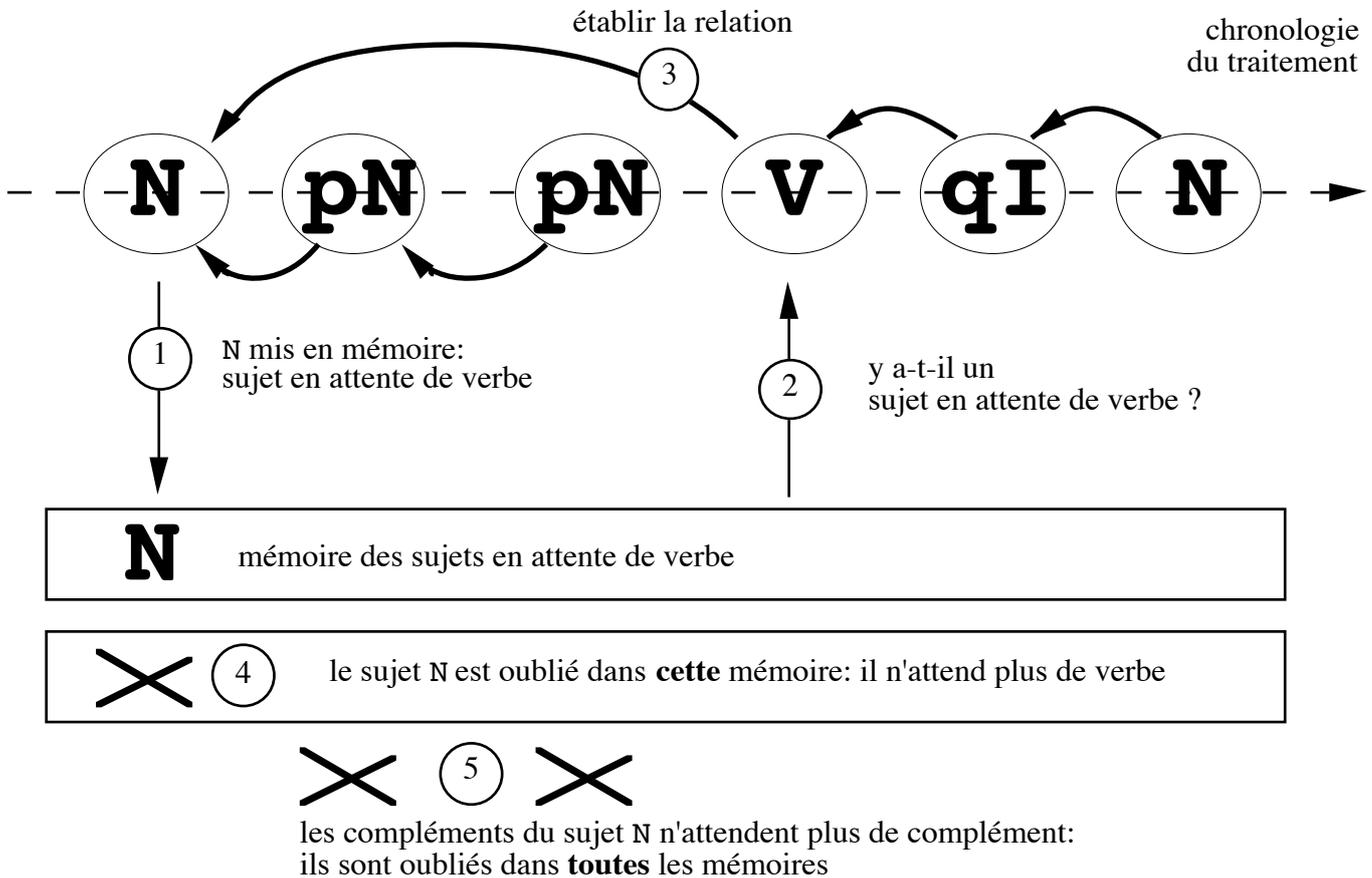
Étudions comment est établie la relation sujet - verbe: la mesure **est utilisée**

Voici la représentation en séquences dans l'ordre linéaire au début du calcul des relations:

(codes: N = séquence nominale, p = préposition nominale, q = préposition verbale infinitive, V = séquence verbale conjuguée, I = séquence verbale infinitive)

-- **N** -- **pN** -- **pN** -- **V** -- **qI** -- **N** -- ➔

Voici un schéma illustrant le processus de construction de l'arbre des dépendances entre séquences (les nœuds sont des séquences, les branches sont des dépendances), et plus précisément le processus de mise en relation de la séquence verbale V avec sa séquence nominale sujet N, en 5 opérations numérotées comme dans la définition ci-dessus:



la branche 1 de N sujet (ses compléments) est close, et V est le début de sa branche 2

Figure 1: Processus de construction de l'arbre des dépendances

Les séquences de la phrase sont traitées de gauche à droite, dans l'ordre dans lequel un auditeur - lecteur la reçoit.

Les oublis des opérations 4 et 5:

- contraignent, restreignent toutes les relations qui s'établiront ensuite,
- concrétisent une propriété de l'arbre linéarisé: une branche close est close définitivement.

L'idée de base de ce processus de mise en relation est qu'il ne s'agit pas de rechercher dans le **passé** (à gauche dans la structure) un sujet pour un verbe en cours d'examen, mais d'observer l'état **présent** des mémoires: les séquences en attente d'une relation sont mises en attente au moment de leur traitement. Une mise en relation n'est pas faite en une opération au moment du traitement du verbe, mais en 3 opérations (1, 2 et 3) par la médiation de la mémoire: mettre en mémoire, lire la mémoire, relier.

Généralisation du processus de mise en relation des séquences:

Toutes les relations de dépendance et de coordination sont ainsi calculées (on peut sans doute pousser encore plus loin la généralisation en reliant ainsi les coréférents anaphoriques, les mots tels que: "ne que", "plus que", etc.).

Dans le modèle informatisé <sup>1</sup>, des mémoires spécialisées (actuellement 8) gèrent les types de relations différents:

- relations de dépendance internes au bloc:
  - mémoire des sujets en attente de verbe
  - des verbes en attente d'objet
  - des clips en attente de séquence
  - mémoire des sujets avec partitif "de" en attente de verbe
- relations de dépendance entre deux blocs:
  - mémoire des régissants en attente de dépendant postposé
  - mémoire des dépendants antéposés en attente de régissant
- relations de coordination entre deux blocs:
  - mémoire des séquences clippées en attente de coordonné
  - mémoire des séquences N sujets en attente de coordonné
  - mémoire des séquences non clippées, et non sujets en attente de coordonné
  - mémoire des relatives en attente de coordonné.

## 2.2. Dépendance mémorielle et contiguïté au niveau hiérarchique des séquences

Si nous observons le même type de dépendance (dépendance mémorielle) au même niveau hiérarchique (celui des séquences), on a un regard homogène: on observe alors (sur les 3 corpus étudiés, en français, espagnol, anglais <sup>2</sup>) que la plupart des séquences (70%) dépendent de la séquence précédente (leur "régissant", selon le terme et concept de Tesnière). Autrement dit, l'indice le plus fréquent de dépendance mémorielle entre deux séquences est la contiguïté entre ces deux séquences.

Définissons la **longueur d'une dépendance** comme le nombre positif ou négatif de séquences d'une séquence à sa séquence "régissant" dans la phrase: contiguïté  $\Leftrightarrow$  longueur de la dépendance = -1 (ou +1, si la séquence dépendante est antéposée à la séquence régissante).

Voici la distribution des longueurs des dépendances sur le corpus français:

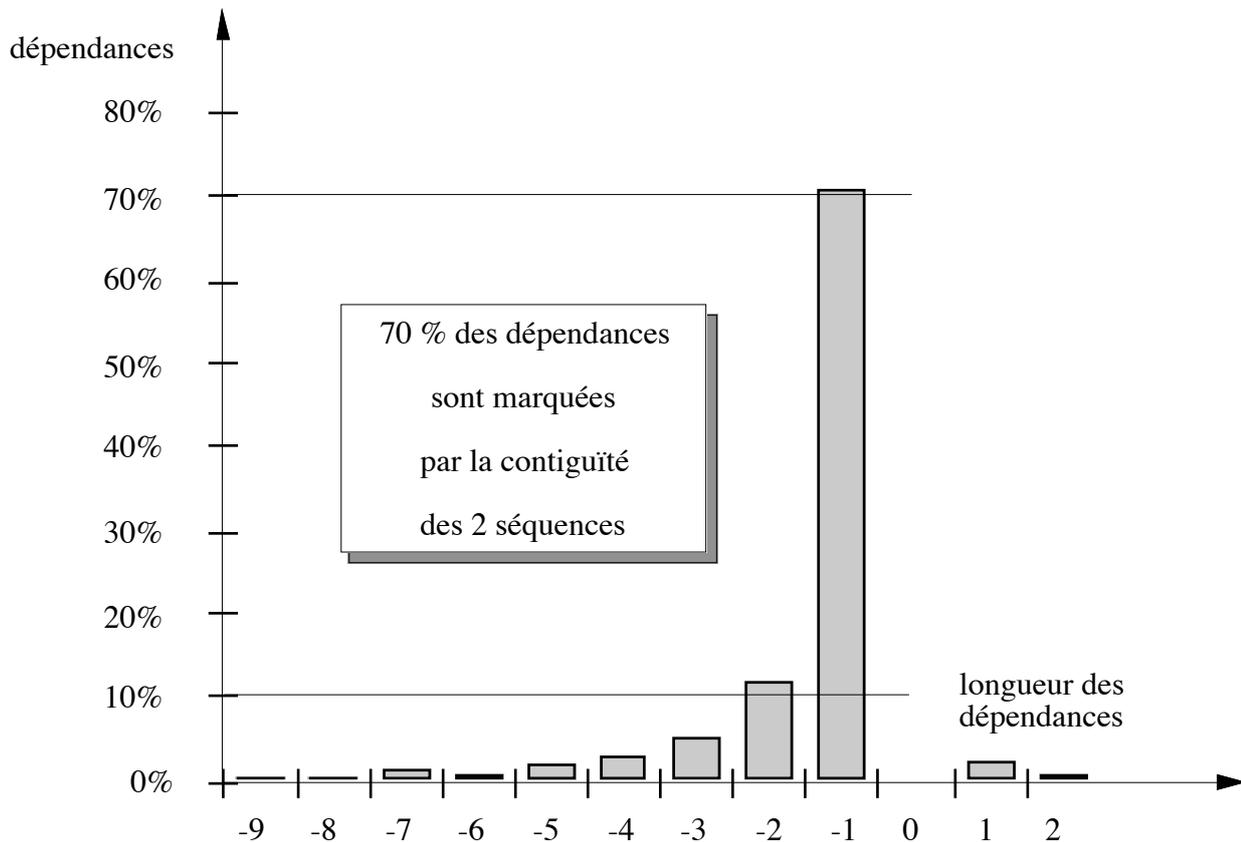


Figure 2: Distribution des longueurs des dépendances

Quelles dépendances ne sont pas marquées par la contiguïté?

En fait **toute** dépendance est marquée par la contiguïté **sauf** en cas d'impossibilité topologique: quand une séquence régit plus d'une séquence, un dépendant est contigu, les autres sont non contigus et placés le plus souvent après le premier dépendant (cas courant de 2 dépendants postposés).

Pourquoi la contiguïté est-elle favorisée? On rejoint là notre hypothèse générale annoncée dans l'introduction: la contiguïté est la distance minimale entre deux séquences reliées, car minimiser la distance entre deux séquences reliées permet de minimiser l'effort de mémoire du lecteur - auditeur.

### 2.3. L'arbre de dépendance des séquences

Définition: le graphe des relations de dépendance entre séquences est un arbre, ce qui signifie que toute séquence a une seule séquence régissante, sauf la séquence racine, qui n'a aucun régissant dans la phrase.

Dans un tel arbre de dépendances, en reprenant la terminologie de l'arbre en informatique <sup>3</sup>: racine, nœuds et feuilles sont des séquences: •

(NB: nœud d'un arbre  $\neq$  nœud au sens de Tesnière; dans cet exposé, nous l'emploierons toujours dans le sens "nœud d'un arbre")

la racine est la séquence nominale (sujet, s'il y a une verbale conjuguée après) du bloc principal (ou à défaut, son unique séquence verbale);

les branches sont des dépendances mémorielles entre séquences:

•<-- \_1--• ou •<-- \_2--•

un nœud à 1 branche implique une dépendance marquée par la contiguïté: •<-- \_1--•

un nœud à 2 branches implique:

une dépendance marquée par contiguïté (première branche): •<-- \_1--•

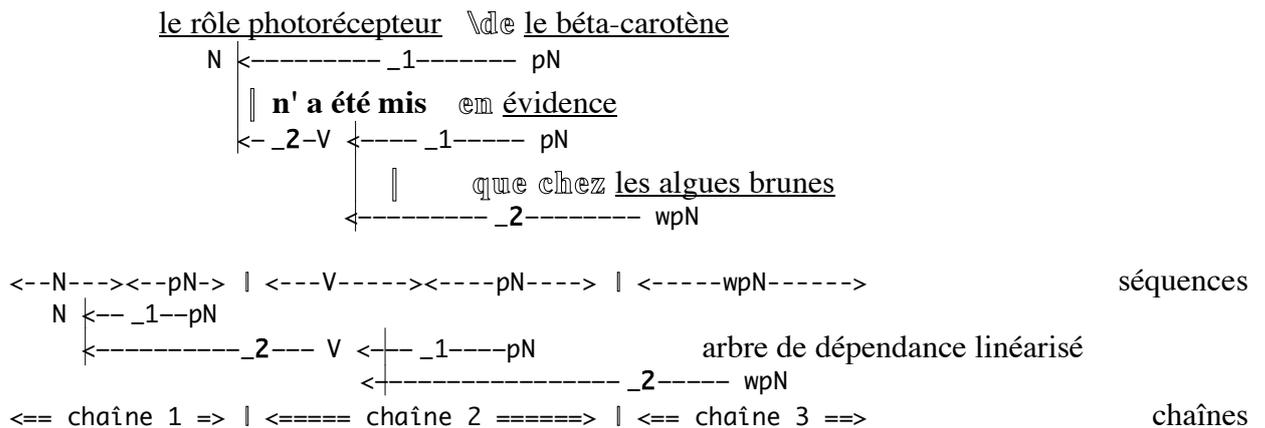
et l'autre non marquée par contiguïté: la seconde branche continue (est linéarisée)

quand la première est terminée (à sa feuille): •<----- \_2-----•

• |<----- \_1----- •  
 |<----- \_2----- || ----•

Le signe || marque (dans l'ordre linéaire) une coupure entre la feuille d'une branche 1 et le premier nœud d'une branche 2, et donc précède une séquence qui ne dépend pas de la séquence précédente. Ce signe marque aussi la coupure entre deux chaînes de séquences dépendantes contiguës (segment secondaire, et segment prosodique, car la prosodie permet à l'auditeur de reconstruire l'arbre de dépendance).

Les phrases ci-dessous donne un exemple d'un tel arbre de dépendances:



Conventions:

- \de est la préposition "de" issue de la décomposition de l'amalgame "du";

- typage des séquences par la typographie:

séquence nominale    **séquence verbale**    clip

- codes:    N = séquence nominale

          V = séquence verbale conjuguée

          I = séquence verbale infinitive

          p = préposition

          w = adverbe de bloc

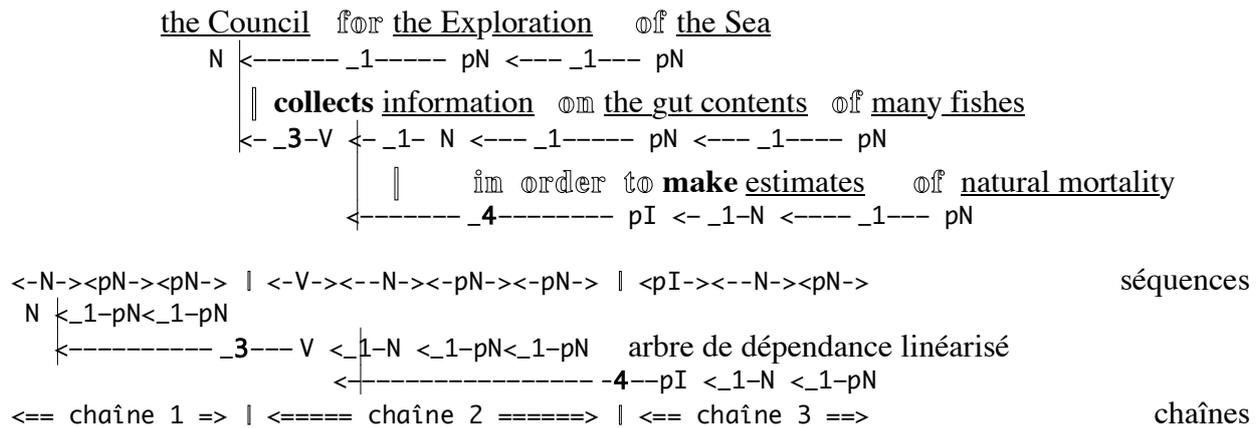
Dans cette phrase, chaque séquence dépend de la séquence précédente (longueur de dépendance = -1), sauf celles qui constituent une deuxième branche d'un nœud:

**n' a été mis** dépend de le rôle photorécepteur 2 séquences avant (longueur = -2),

les algues brunes dépend de **n' a été mis** 2 séquences avant (longueur = -2).

Remarque: un lecteur familier de Tesnière sera peut-être surpris qu'une séquence nominale (le sujet) est la racine de l'arbre; c'est la conséquence de la définition de la dépendance mémorielle: ce sujet a deux dépendants, son complément de nom, et son verbe.

Voici une phrase en anglais, ayant un arbre de même structure que celui de la phrase précédente:



Dans cette phrase, chaque séquence dépend de la séquence précédente (longueur de dépendance = -1), sauf celles qui constituent une deuxième branche d'un nœud:

**collects** dépend de the Council 3 séquences avant (longueur de dépendance = -3),

**make** dépend de **collects** 4 séquences avant (longueur de dépendance = -4).

## 2.4. Valence et saturation de valence

### 2.4.1. Définition

Pour Tesnière, **la** valence est une propriété du verbe: c'est le nombre d'actants qu'un verbe est capable de régir (voir Tesnière, 1959: page 238); mais sa définition n'est pas stable: il parle (même page) d'**une** valence soit inemployée, libre, soit pourvue, saturée; ici une valence est **une** relation potentielle.

Nous proposons de définir **une valence** d'une séquence (quelque soit son type) comme étant une relation potentielle que cette séquence est capable d'établir avec une autre séquence.

Par exemple, une séquence nominale sans clip peut avoir (liste non exhaustive):

- un verbe conjugué dont elle est le sujet
- un complément bloc prépositionnel nominal ou verbal
- un complément bloc proposition relative
- une séquence nominale coordonnée.

On peut donc caractériser les valences d'une séquence par l'arbre de dépendance potentiel ayant cette séquence pour racine, et les séquences qu'elle est capable de régir pour feuilles.

La saturation d'une valence est la réalisation de la relation potentielle (comme pour Tesnière).

La définition proposée ici est différente et plus large que celle de Tesnière:

- différente, car nous définissons **une** valence comme une relation possible, et non **la** valence comme nombre d'actants, ou l'ensemble des relations possibles avec les actants, comme synonyme de structure actancielle (terme aussi employé par Tesnière).

- plus large, car nous l'appliquons non pas à un mot, mais à une séquence (à tout nœud de l'arbre de dépendance), non seulement à la séquence verbale, mais aussi aux séquences nominales, et adjectivales, et adverbiales.

#### 2.4.2. Une application du concept de saturation de valence: "de", préposition ou partitif?

Comment savoir si "de" est préposition ou partitif?

Un cas évident est le cas où "de" est précédé d'une préposition: c'est alors un partitif:

ils sont souvent liés à d'autres molécules

Cas général:

Hypothèse: "de" est partitif si la séquence nominale qui le suit peut être sujet ou objet d'un verbe ("de" fait alors partie de la séquence nominale).

Cas du sujet:

dans: N1 de N2 V N1 est sujet de V

Les molécules de nombreux pigments caroténoïdes **n'échappent pas** ...

dans: de N2 V N2 est sujet de V

de nombreuses molécules énergétiques **sont transférées** ...

donc, s'il n'y a aucun N1 attendant un V, N2 peut attendre un V

Cas de l'objet:

dans: V N1 de N2 N1 est objet de V transitif

... **semblent jouer un rôle** de photoprotection

dans: V de N2 N2 est objet de V transitif

... **a suscité** de très nombreux travaux

donc, si V attend un objet, et s'il n'en a pas quand N2 est lu, alors N2 est objet de V.

Cette hypothèse est confirmée sans exception sur le corpus français.

(NB: le "de" des amalgames "du" et "des" a les mêmes propriétés)

### 3. Structures associées: l'arbre des dépendances <=> la phrase linéaire

L'arbre et la phrase linéaire forment deux faces d'un même objet: l'arbre est la face abstraite, la phrase linéaire est la face concrète; l'arbre est codé dans la phrase, complètement, car la phrase est intelligible: on peut reconstituer l'arbre à partir de la phrase. Voici un tableau présentant les correspondances entre les caractéristiques de l'arbre et celles de la phrase:

arbre de dépendance	phrase
racine, nœud, feuille	↔ <u>séquence</u> nominale ou verbale (pas le clip)
racine	↔ <u>séquence</u> nominale (sujet) du bloc principal
branche	↔ <u>dépendance mémorielle</u> entre 2 séquences
nœud à 1 branche	↔ dépendance marquée par <u>contiguïté</u>
nœud à 2 branches	↔ 1 dépendance marquée par <u>contiguïté</u> , l'autre pas
espace à 2 dimensions	espace à 1 dimension

#### 4. Linéarisation de l'arbre dans la phrase

Comment l'arbre est-il codé dans la phrase?

Ce problème n'est pas seulement linguistique, mais d'abord topologique: il est surdéterminé par les propriétés des arbres, et les propriétés des espaces à une dimension. Ce problème n'est pas seulement linguistique ou topologique, mais aussi mémoriel: il est surdéterminé par les limites de mémorisation des séquences dans leur attente de mise en relation; la mémorisation est facilitée si les séquences reliées sont proches, ou mieux, contiguës.

Appellons **optimisation** de la linéarisation, le fait que les limites de mémorisation sont le plus possible respectées, que les efforts de mémorisation sont les plus faibles possibles, ce qui se concrétise principalement par une minimisation des longueurs des dépendances.

Le terme "linéarisation" peut s'entendre de deux manières:

- 1- les propriétés statiques conjointes de l'arbre et de la phrase,
- 2- le processus qui permet au scripteur - locuteur de passer de l'arbre à la phrase (on pourrait appeler "délinéarisation", ou "arborisation" le processus inverse qui permet au lecteur - auditeur de reconstruire la phrase à partir de l'arbre).

La suite de l'étude est dans le sens -1-.

##### 4.1. Linéarisation de l'arbre dans la phrase ou comment l'arbre est codé dans la phrase

Un arbre peut en théorie être codé intégralement dans une phrase - espace à 1 dimension (par exemple avec une structure parenthésée, comme dans le langage de programmation Lisp<sup>4</sup>). "Codé intégralement" signifie qu'on peut restituer l'arbre à partir de la phrase.

Pour que la restitution complète de l'arbre soit possible à partir de la phrase, deux types de singularités doivent être **marquées** directement ou indirectement (c'est-à-dire calculables d'après d'autres indices):

singularité (1): toute séquence régissant plus d'une séquence (nœuds à 2 branches ou plus),

singularité (2): tout couple: feuille || séquence ne dépendant pas de la précédente (la coupure || est marquée par une coupure prosodique à l'oral et souvent par une virgule à l'écrit).

Ces deux types de singularités sont entourées d'un ovale dans le schéma suivant:

(le nœud en gris a 2 branches, les nœuds en blanc sont des feuilles, les autres sont noirs)

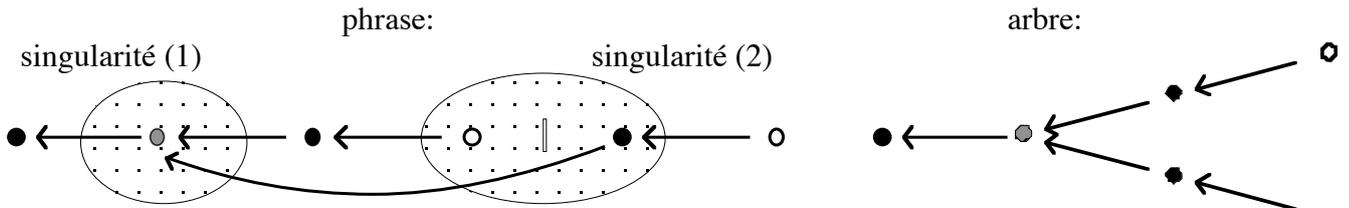


Figure 3: Singularités marquant l'arbre dans la phrase

Les cas les plus courants de nœuds à 2 branches sont:

- une séquence nominale sujet avec son expansion (en reprenant terme et concept de Martinet):

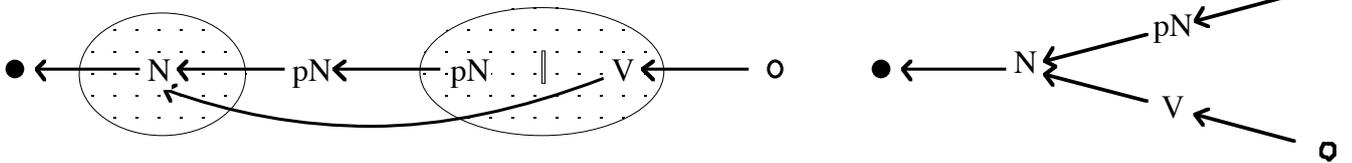


Figure 3 bis: Cas de la séquence nominale sujet

- ou une séquence verbale avec objet et circonstant:

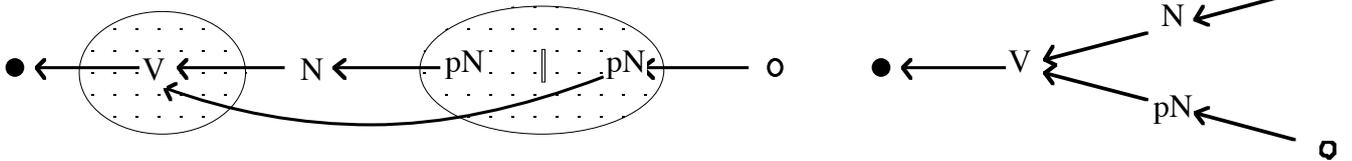


Figure 3 ter: Cas de la séquence verbale avec objet et circonstant

(voir ci-dessous en 4.3 l'inventaire des nœuds à 2 branches en français)

#### 4.2. Linéarisation de l'arbre dans la phrase, ou dans quelle chronologie les nœuds de l'arbre sont énumérés dans la phrase

C'est exactement le problème (classique en informatique) de définir **un ordre linéaire des nœuds d'un arbre**, c'est-à-dire le problème de l'énumération des nœuds d'un arbre, ou le problème du parcours d'un arbre avec relevé de ses nœuds.

Divisons ce problème en deux questions:

- 1) quel est l'ordre linéaire régissant - dépendants?
- 2) quel est l'ordre linéaire des dépendants entre eux?

##### 4.2.1. Ordre linéaire régissant - dépendants

Dans un arbre on distingue classiquement deux dimensions: la largeur et la profondeur.

La profondeur d'un nœud dans un arbre est sa distance à la racine en nombre de nœuds: la racine est à la profondeur 0, les nœuds situés à la fin des branches de la racine (appelés ses fils, par analogie à un arbre généalogique) sont à profondeur 1, etc.

Parcourir un arbre en largeur d'abord revient à passer par le nœud de profondeur 0 (la racine), puis par tous les nœuds de profondeur 1 (les fils de la racine), puis profondeur 2, etc. , jusqu'aux feuilles.

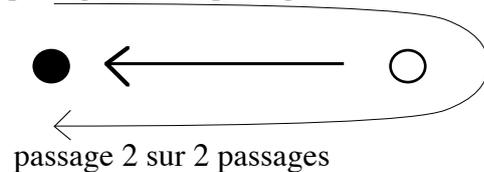
On peut aussi parcourir un arbre en profondeur d'abord: racine -> feuilles -> racine. À partir de chaque nœud, on passe si possible (= si ce n'est pas une feuille) à un de ses fils, ainsi une dépendance est marquée par contiguïté.

La linéarisation de l'arbre de dépendances consiste en un parcours de l'arbre en profondeur d'abord, en partant de la racine: parmi les deux parcours possibles, c'est le parcours qui conserve le plus les contiguïtés (contrainte mémorielle), car chaque branche est parcourue entièrement avant de parcourir une autre branche.

Dans un parcours d'arbre en profondeur d'abord, on passe sur chaque nœud de **b** branches **b + 1** fois:

nœud à 1 branche => 2 passages

passage 1 sur 2 passages



nœud à 2 branches => 3 passages

passage 1 sur 3 passages

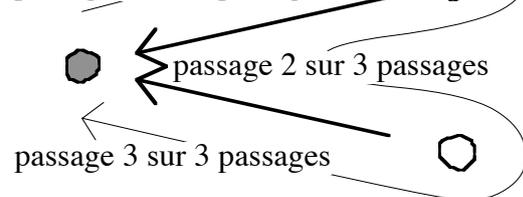


Figure 4: Parcours d'arbre en profondeur d'abord

Le type de **relevé** spécifie à quel passage le nœud est relevé:

premier passage, ..., dernier passage

- au premier passage, le relevé est dit **préfixé**, car le nœud est relevé **avant** ses branches
- au deuxième passage, le relevé est dit **infixé**, car le nœud est relevé **entre** ses branches
- au dernier passage, le relevé est dit **postfixé**, car le nœud est relevé **après** ses branches.

- relevé préfixé: le nœud est relevé au premier passage

=> ordre linéaire: nœud régissant, puis ses branches dépendantes

c'est le relevé normal en français (comme dans les figures 3 ci-dessus, en 4.1)

exemple: séquence verbale conjuguée avec objet et complément du verbe

phrase: ces poissons **peuvent transporter** tous ces pigments vers leurs tissus

N1

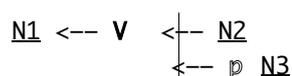
**V**

N2

P

N3

arbre:



=> ordre linéaire: nœud régissant, puis ses 2 branches dépendantes

$V$  ,  $\underline{N2}$  ,  $P \underline{N3}$

- relevé infixé: le nœud est relevé après le premier passage et avant le dernier

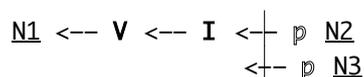
=> ordre linéaire: au moins 1 branche dépendante, puis son nœud régissant,  
puis au moins 1 branche dépendante

ceci est possible pour un nœud de 2 branches ou plus; ce relevé est moins fréquent ou ancien en français (voir aussi les linéarisations 3 et 4, figure 8, ci-dessous en 4.3.3.)

phrase: une distance **pourra** , dans un tel espace , être calculée par une formule

$\underline{N1}$   $V$  ,  $P$   $\underline{N2}$  ,  $I$   $P$   $\underline{N3}$

arbre:



=> ordre linéaire:

1 branche dépendante, puis son nœud régissant, puis l'autre branche dépendante

$P \underline{N2}$  ,  $I$  ,  $P \underline{N3}$

- relevé postfixé: le nœud est relevé au dernier passage

=> ordre linéaire: les branches dépendantes, puis leur nœud régissant

c'est le relevé de certaines séquences verbales en allemand:

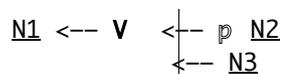
- séquence verbale conjuguée dans une proposition subordonnée:

phrase: **Düngte** er die Felder mit Mist , sorgte er dafür

, daß es im Zukunft eine Ernte geben würde .

$P \underline{N1}$   $P \underline{N2}$   $\underline{N3}$   $V$

arbre:



=> ordre linéaire: 2 branches dépendantes, puis leur nœud régissant

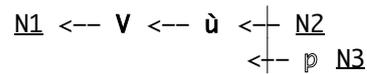
$P \underline{N2}$  ,  $\underline{N3}$  ,  $V$

- séquence verbale participe passé dans une proposition avec auxiliaire et participe passé:

phrase: Bauern **haben** das Land nach überlieferten Regeln bestellt .

$\underline{N1}$   $V$   $\underline{N2}$   $P$   $\underline{N3}$   $\grave{u}$

arbre:



=> ordre linéaire: 2 branches dépendantes, puis leur nœud régissant



- séquence verbale infinitive précédée de "zu"

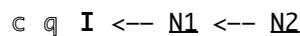
phrase: Manche vom ihnen sind im der Lage,

sogar die harten Chitinpanzer von Insekten amzuknabbern

oder die ansonsten unverdauliche Zellulose der Pflanzen zu verwerten .



arbre:



=> ordre linéaire: 1 branche dépendante, puis son nœud régissant



Choix du type de relevé:

Le choix du type de relevé est propre à un type particulier de nœud dans un contexte particulier pour une langue particulière.

Il faut distinguer le type de relevé global de l'arbre, type normal, propre à une langue, du type de relevé local propre à un nœud, qui est le plus souvent le même que le type global car l'optimisation maximale se produit quand on a le même type de relevé pour tous les nœuds de l'arbre. On a donc un défaut d'optimisation par exemple quand un nœud a un relevé postfixé, ou infixé dans un arbre globalement relevé préfixé: par exemple, certains nœuds verbaux en allemand, un bloc antéposé, la place obligatoire du pronom relatif en début de bloc.

La racine de l'arbre de dépendance est la séquence nominale du bloc principal = le dit ou écrit en premier, sauf nécessité supérieure (le bloc antéposé comme clip de paragraphe, ou la position imposée du pronom relatif en début de bloc).

Fonction syntaxique:

Une dépendance entre deux séquences, le type des séquences reliées, la présence ou l'absence de clip avant la séquence dépendante impliquent la fonction d'une séquence par rapport à l'autre (et réciproquement); par exemple une dépendance  $N \leftarrow V$  ou  $V \rightarrow N$  est une relation sujet - verbe.

(le nom d'une fonction est indépendant de l'ordre linéaire des séquences reliées)

Terminologie:

Les termes: préfixé, infixé, postfixé dénomment la position d'un régissant par rapport à l'ensemble de ses dépendants, selon qu'il est situé avant, entre, ou après ses dépendants (nous reprenons ainsi les termes et les concepts venant de l'informatique).

Les termes: antéposé, postposé dénomment la position d'un dépendant par rapport à son régissant, selon qu'il est situé avant, ou après son régissant (nous reprenons ainsi les termes et les concepts venant de la linguistique).

4.2.2. Ordre linéaire des dépendants entre eux

Nous nous situons dans le cadre du relevé global préfixé, relevé normal en français, dans lequel le régissant précède ses dépendants; 3 cas sont alors possibles:

- tout nœud a 1 branche, sauf le dernier, l'unique feuille:



Figure 5: Ordre linéaire des dépendants entre eux (tout nœud a 1 branche)

Arbre et phrase ont même structure, toute dépendance est marquée par contiguïté; la phrase est donc minimale et optimale (la plus simple possible) du point de l'effort de mémoire minimal. C'est une unique chaîne de séquences dépendantes contiguës, un unique segment prosodique.

- un nœud a 2 branches de même poids:  
poids d'une branche = nombre total de nœuds de la branche, quelque soit sa structure

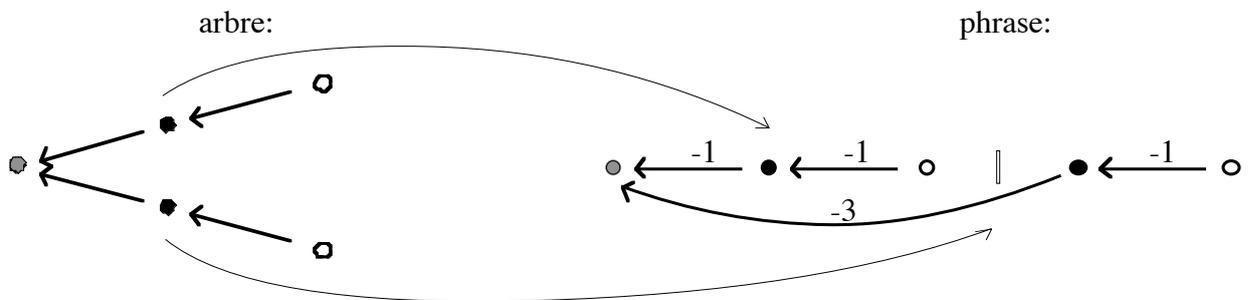


Figure 5 bis: Ordre linéaire des dépendants entre eux (nœud à 2 branches de même poids)

Les 2 branches sont linéarisées (dites - écrites) successivement:

- la première branche est contiguë au nœud régissant,
- et la deuxième branche est reliée au nœud régissant par une dépendance distendue, de longueur égale au poids de la première branche + 1.

La phrase est constituée de 2 chaînes de séquences dépendantes contiguës (de 3 séquences puis 2 séquences), 2 segments prosodiques séparés par une coupure prosodique.

Comme les 2 branches ont le même poids, les deux solutions d'ordre linéaire des dépendants sont identiques du point de vue de la longueur de la dépendance distendue; cette structure impose au lecteur - auditeur de mémoriser le nœud régissant à 2 branches R comme

pouvant régir un deuxième dépendant D2, premier nœud de la branche 2, pour pouvoir relier D2 à R, après la feuille de la branche 1.

Dans ces deux premiers cas, la question du choix d'un meilleur (par minimisation de l'effort de mémoire) ordre linéaire des dépendants entre eux ne se pose pas; elle se pose seulement dans ce troisième cas:

- un nœud a 2 branches de poids différents:

voici les deux linéarisations possibles, selon la branche linéarisée en premier:

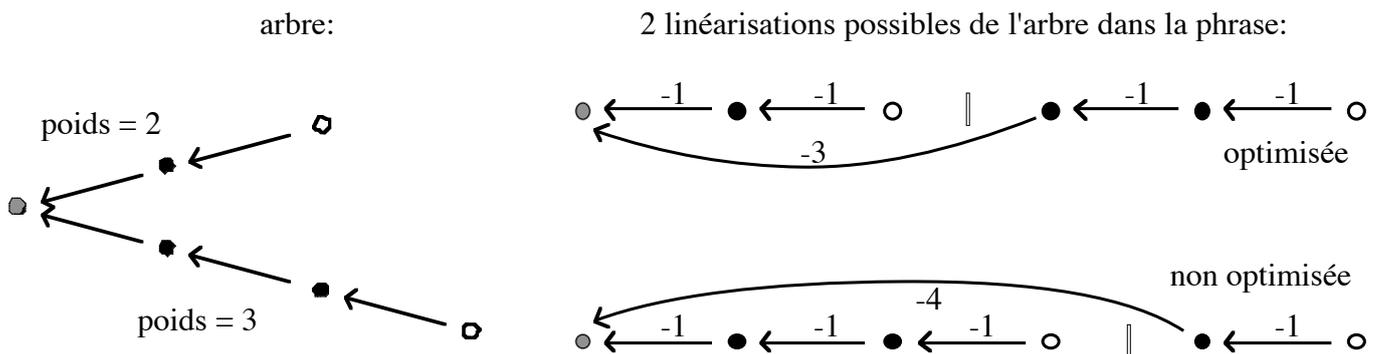


Figure 5 ter: Ordre linéaire des dépendants entre eux (nœud à 2 branches de poids différents)

La linéarisation est alors optimisée, surdéterminée par une **contrainte de minimisation de l'effort de mémoire** pour le lecteur - auditeur: les séquences reliées sont les plus proches possibles dans la phrase; dans la linéarisation d'un nœud à 2 branches, la branche de poids inférieur est écrite - dite d'abord pour distendre le moins possible la dépendance: régissant <--- branche de poids supérieur, dont la longueur est alors: poids de la première branche + 1 = poids de la branche de poids inférieur + 1; en généralisant à un nœud ayant un nombre quelconque de branches, les branches sont écrites - dites dans l'ordre des poids croissants. Ce principe est plus faible que certaines contraintes linguistiques; par exemple, les expansions du sujet lui restent contiguës, même si les expansions du verbe sont plus courtes (sauf la relative en "qui").

La phrase est toujours constituée de 2 chaînes de séquences dépendantes contiguës, 2 segments prosodiques séparés par une coupure prosodique: remarquons que l'optimisation tend à égaliser les longueurs des chaînes: la linéarisation optimisée donne 2 chaînes de 3 séquences, et la linéarisation non optimisée donnerait 2 chaînes de 4 et 2 séquences.

En résumé:

L'ordre linéaire des séquences dans la phrase découle directement de la structure de l'arbre des dépendances mémorielles entre les séquences; il se décrit en deux temps:

ordre linéaire régissant - dépendants:

l'arbre est parcouru en profondeur d'abord;

3 types de relevé sont possibles: préfixé, infixé, postfixé, d'où 3 positions possibles d'un régissant par rapport ses dépendants: avant, entre, après;

la question du type de relevé se pose globalement pour tout l'arbre, et localement pour un nœud et ses branches

ordre linéaire des dépendants entre eux:

les branches sont linéarisées par ordre de poids croissants, sous la contrainte de minimisation de l'effort de mémoire pour le lecteur - auditeur.

Remarque:

Tesnière explique les relations entre l'ordre structural et l'ordre linéaire en précisant qu'on parle en produisant l'ordre linéaire à partir de l'ordre structural, et qu'on comprend en produisant l'ordre structural à partir de l'ordre linéaire (voir Tesnière, 1959: pages 16 et suivantes) mais il évoque très brièvement comment cela se passe, quelles sont les relations de structure entre les deux, uniquement dans le paragraphe suivant:

"... syntaxiquement, la vraie phrase, c'est la **phrase structurale** dont la phrase linéaire n'est que l'image projetée tant bien que mal, et avec tous les inconvénients d'aplatissement que comporte cette projection sur la chaîne parlée." (page 20, § 10)

La linéarisation de l'arbre par parcours en profondeur d'abord est différente de la projection évoquée par Tesnière et précisée par David Hays, dont le résultat linéaire dépend de la manière dont l'arbre est dessiné, et de la direction de projection, alors que la linéarisation ici décrite donne un résultat linéaire qui ne dépend que de la **structure** de l'arbre et des choix de branches et de type de relevé faits à chaque nœud durant le parcours de l'arbre.

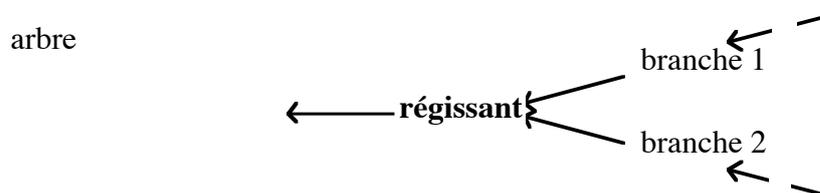
#### 4.3. Inventaire des séquences à 2 dépendants ou plus, et leur ordre linéaire

Toute séquence dépend de la séquence précédente sauf en cas de séquences à 2 dépendants ou plus: seul le premier dépendant est contigu. Pour mieux comprendre la linéarisation de l'arbre, il faut donc faire l'inventaire exhaustif des cas de séquences à 2 dépendants ou plus.

Dans les langues étudiées (français, anglais, espagnol, allemand), le type de relevé global est **préfixé** (ordre linéaire régissant - dépendants), le type de relevé local est aussi le plus souvent **préfixé** (optimisation maximale avec le même type de relevé pour tous les nœuds) mais il y a des nœuds à relevé localement **infixé** en français, et le nœud verbal subordonné en allemand est relevé localement **postfixé**; on observe que les branches restent en relevé préfixé.

##### 4.3.1. Cas général (dans un relevé global préfixé)

Voici le cas général d'un nœud régissant à 2 branches, avec ses 3 linéarisations possibles: régissant placé avant, entre ou après ses dépendants:



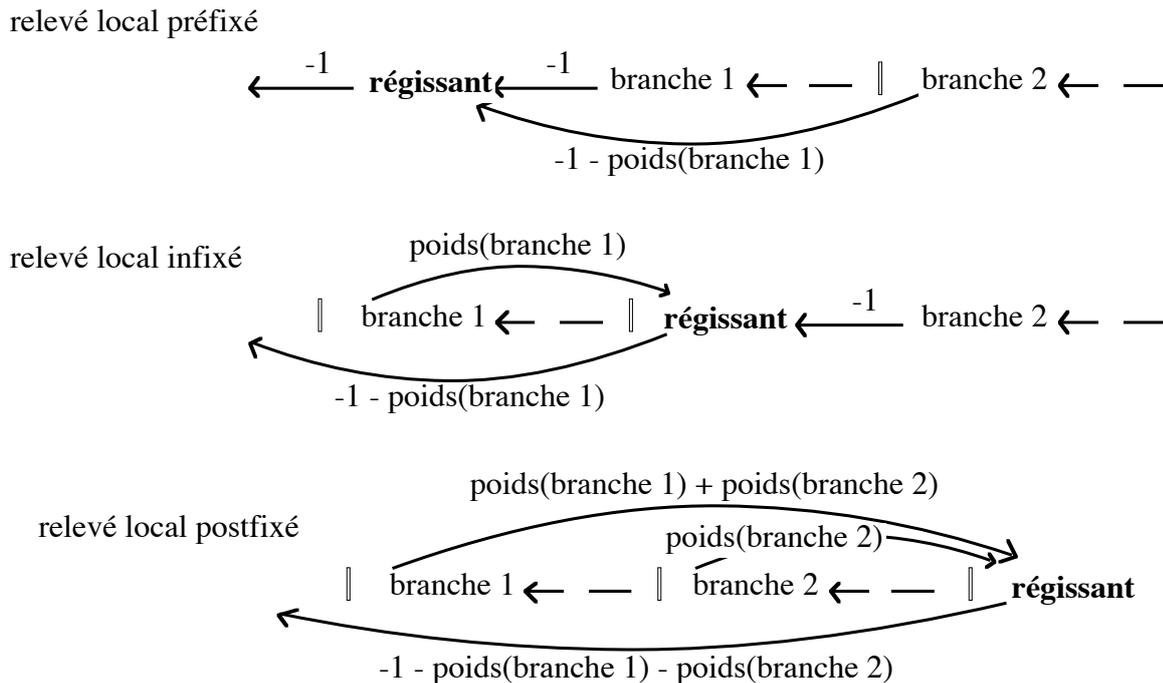


Figure 6: Linéarisation d'un nœud à 2 branches (les 3 relevés locaux)

La linéarisation optimisée est la linéarisation pour laquelle la somme des longueurs des dépendances est minimale parmi les linéarisations possibles.

Supposons arbitrairement que la linéarisation optimisée est celle représentée par le schéma:  
ordre linéaire: branche 1 - branche 2.

La linéarisation optimisée équivaut alors à l'expression ( $\sum$  signifie "somme"):

$\sum$  des valeurs absolues des longueurs des dépendances si branche 1 linéarisée en premier  
 $\leq$   $\sum$  des valeurs absolues des longueurs des dépendances si branche 2 linéarisée en premier

que l'on développe alors pour chaque relevé:

relevé local préfixé:  $3 + \text{poids}(\text{branche 1}) \leq 3 + \text{poids}(\text{branche 2})$   
 $\Leftrightarrow \text{poids}(\text{branche 1}) \leq \text{poids}(\text{branche 2})$

relevé local infixé:  $2 + 2 * \text{poids}(\text{branche 1}) \leq 2 + 2 * \text{poids}(\text{branche 2})$   
 $\Leftrightarrow \text{poids}(\text{branche 1}) \leq \text{poids}(\text{branche 2})$

relevé local postfixé:  $1 + 2 * \text{poids}(\text{branche 1}) + 3 * \text{poids}(\text{branche 2})$   
 $\leq 1 + 2 * \text{poids}(\text{branche 2}) + 3 * \text{poids}(\text{branche 1})$   
 $\Leftrightarrow \text{poids}(\text{branche 2}) \leq \text{poids}(\text{branche 1})$

Donc le choix de la linéarisation optimisée d'un nœud à 2 branches se fait sur la simple comparaison des poids de ces 2 branches.

Étudions ensuite les 3 réalisations les plus courantes d'un nœud à 2 branches, selon qu'il est une séquence nominale sujet, une séquence verbale avec 2 dépendants, ou un régissant de 2 branches coordonnées.

#### 4.3.2. Nœud à 2 branches = séquence nominale sujet

Une branche est l'expansion du sujet, l'autre est le verbe conjugué avec ses propres expansions.

Convention dans la suite de l'exposé: les limites de bloc sont marquées par [ ] , ce qui permet de marquer des blocs contigus: [NV] [pN] , ou des blocs insérés: [N [pN] V] ; les constituants des blocs sont des séquences nominales ou verbales, ou des clips.

L'expansion du sujet est constituée de 0, 1 ou plusieurs blocs contigus, insérés entre sujet et verbe:

- => la partie 2 du bloc entourant (séquence verbale conjuguée) dépend de la partie 1 (sa séquence nominale sujet), qui précède le ou les blocs insérés
- => la coupure ( || ) est entre le bloc inséré et la partie 2 du bloc entourant:

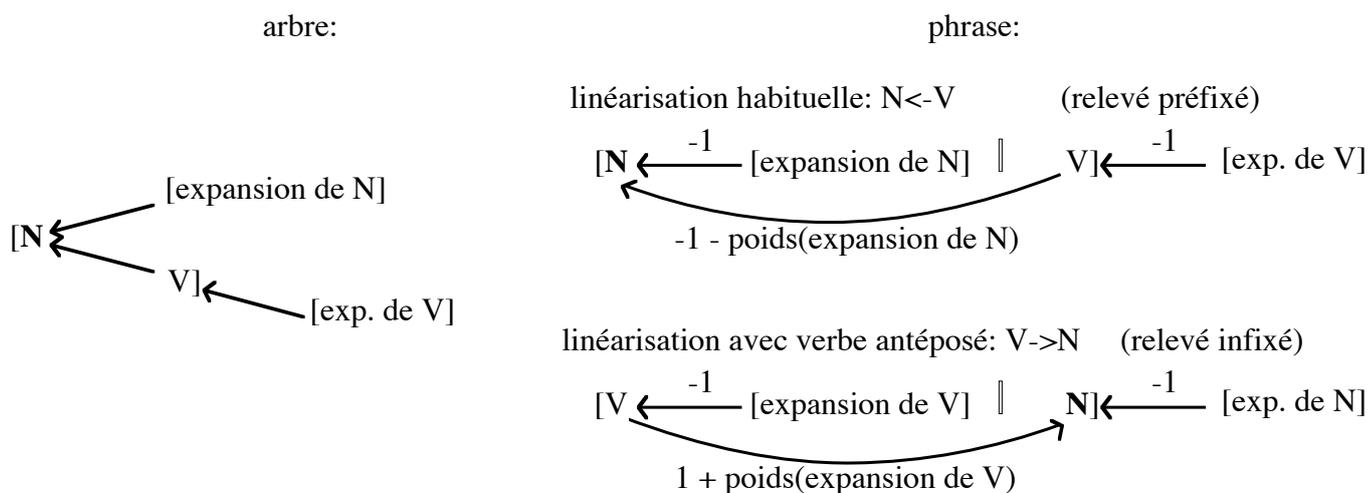


Figure 7: Linéarisation d'un nœud séquence nominale sujet à 2 branches

Linéarisation habituelle: N sujet suivi de ses expansions, puis V suivi de ses expansions  
 ordre régissant - dépendants = relevé préfixé du nœud sujet  
 ordre des dépendants entre eux:     branche 1 = expansions de N,  
   puis    branche 2 = V, et ses expansions  
 optimisée si  $\text{poids}(\text{expansion de N}) \leq 1 + \text{poids}(\text{expansion de V})$  (le plus souvent)

Une linéarisation plus rare est celle avec le verbe antéposé, choisie quand  $1 + \text{poids}(\text{expansion de V}) < \text{poids}(\text{expansion de N})$  ce qui permet l'optimisation:

ordre dépendant 1 - régissant - dépendant 2 = relevé local infixé du nœud sujet

avec V et ses expansions comme dépendant 1, et les expansions de N comme dépendant 2.

Le seul cas en français où l'expansion du sujet ne lui est pas contiguë est le cas où cette expansion est une relative en "qui", placée après le verbe:

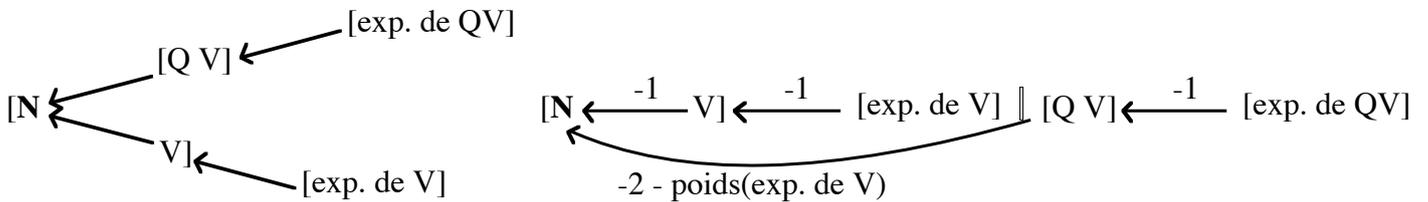
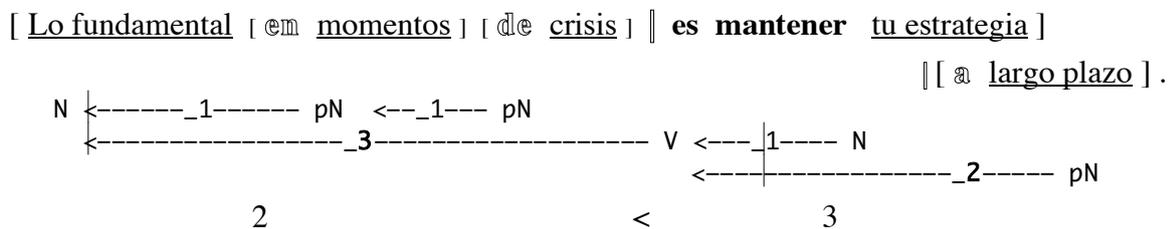
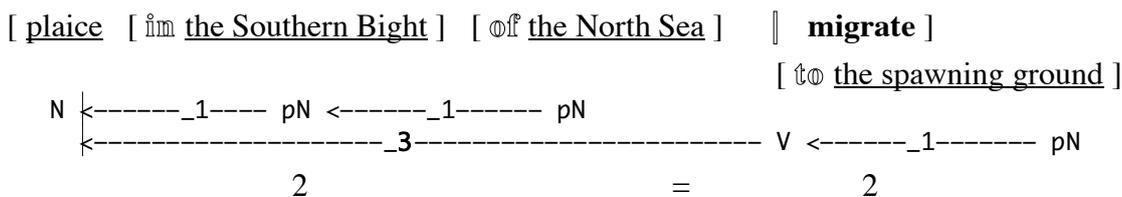
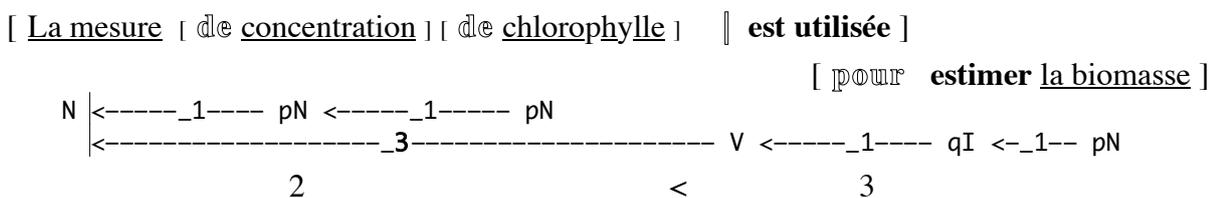


Figure 7 bis: Linéarisation d'un nœud séquence nominale sujet à 2 branches avec une relative postposée au verbe

Cette linéarisation est optimisée si:  $\text{poids(branche 1)} \leq \text{poids(branche 2)}$   
 ou:  $1 + \text{poids(expansion de V)} \leq 1 + \text{poids(expansion de [QV])}$   
 ou:  $\text{poids(expansion de V)} \leq \text{poids(expansion de [QV])}$

Exemples de linéarisation habituelle, avec nœud sujet préfixé et verbe postposé:



Dans les 3 exemples, la linéarisation est optimisée:

$$\text{poids(expansion de N)} \leq 1 + \text{poids(expansion de V)}$$

Remarquons aussi que ces 3 phrases, ayant le même arbre de dépendance, ont aussi même structure prosodique, avec une coupure située avant le verbe conjugué.

La structure d'insertion des blocs implique les dépendances suivantes



Exemples de linéarisation avec nœud sujet infixé et verbe antéposé:

[ Pour le premier exemple ] [ , sont opérationnels des lecteurs ]  
 [ de caractères ] [ dactylographiés ]  
 pN -----1-----> V -----1-----> N <---\_1--- pN <---\_1--- ù  
 [ ( [ lecture ] [ de codes postaux ] [ , aide ] [ \à les aveugles ] ) ]  
 ( N <---\_1--- pN , N <---\_1--- pN )

avec branche verbe de poids 2, et branche des expansions du sujet de poids 6.

Cette linéarisation est magnifiquement optimisée: les 3 dépendances sont marquées par contiguïté, et l'insertion de bloc est évitée; une première antéposition (pN > V) en entraîne une deuxième (V > N), comme dans la relative avec pronom relatif objet (voir ci-dessous en 6.2.), où l'antéposition du pronom relatif (O > V) entraîne l'antéposition du verbe (V > N); la postposition du verbe avec nœud sujet préfixé n'aurait pas été optimisée (1+1+1+1<1+1+7+8):

[ Pour le premier exemple ] [ , des lecteurs [ de caractères ] [ dactylographiés ]  
 (NpN , NpN) sont opérationnels ]  
 N <---\_1--- pN <---\_1--- ù  
 <-----\_7-----> V  
 pN -----8----->

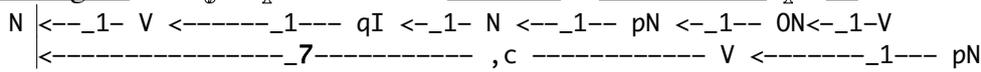
Exemple de linéarisation avec relative du sujet placée après le verbe (de La Bruyère, cité par Grevisse, dans "le bon usage"):

Bien des gens vont jusques à sentir le mérite d'un manuscrit qu'on leur lit  
 N <---\_1- V <-----\_1--- qI <-\_1- N <---\_1-- pN <-\_1-- ON<-\_1-V  
 <-----\_7-----> , Q <-----\_1-- V <-----\_1--- pN  
 , qui ne peuvent se déclarer en sa faveur.

cette linéarisation n'est pas optimisée: 1+ poids(expansion de V) ≤ poids(expansion de N)  
 (6>3)

Aujourd'hui, la relative placée après le verbe n'est plus utilisée, même à l'écrit; on remplacerait la subordination par une coordination, "qui" étant remplacé par "et ":

Bien des gens vont jusques à sentir le mérite d'un manuscrit qu'on leur lit



, et ne peuvent se déclarer en sa faveur.

avec presque le même arbre, et toujours une linéarisation non optimisée.

4.3.3. Nœud à 2 branches = séquence verbale avec 2 dépendants ou plus

Les dépendants du verbe prennent place dans la structure actancielle de ce verbe (conjugué, infinitif, participe présent ou passé).

Détaillons le cas le plus courant: un dépendant objet et un dépendant circonstant.

Le bloc dépendant circonstant [pN] peut prendre 4 positions par rapport aux 3 séquences du bloc [NVN]:

$$[4] [N] [3] V [2] N [1]$$

(l'ordre d'exposé des 4 cas est à la fois par complexité croissante, et approximativement par fréquence décroissante)

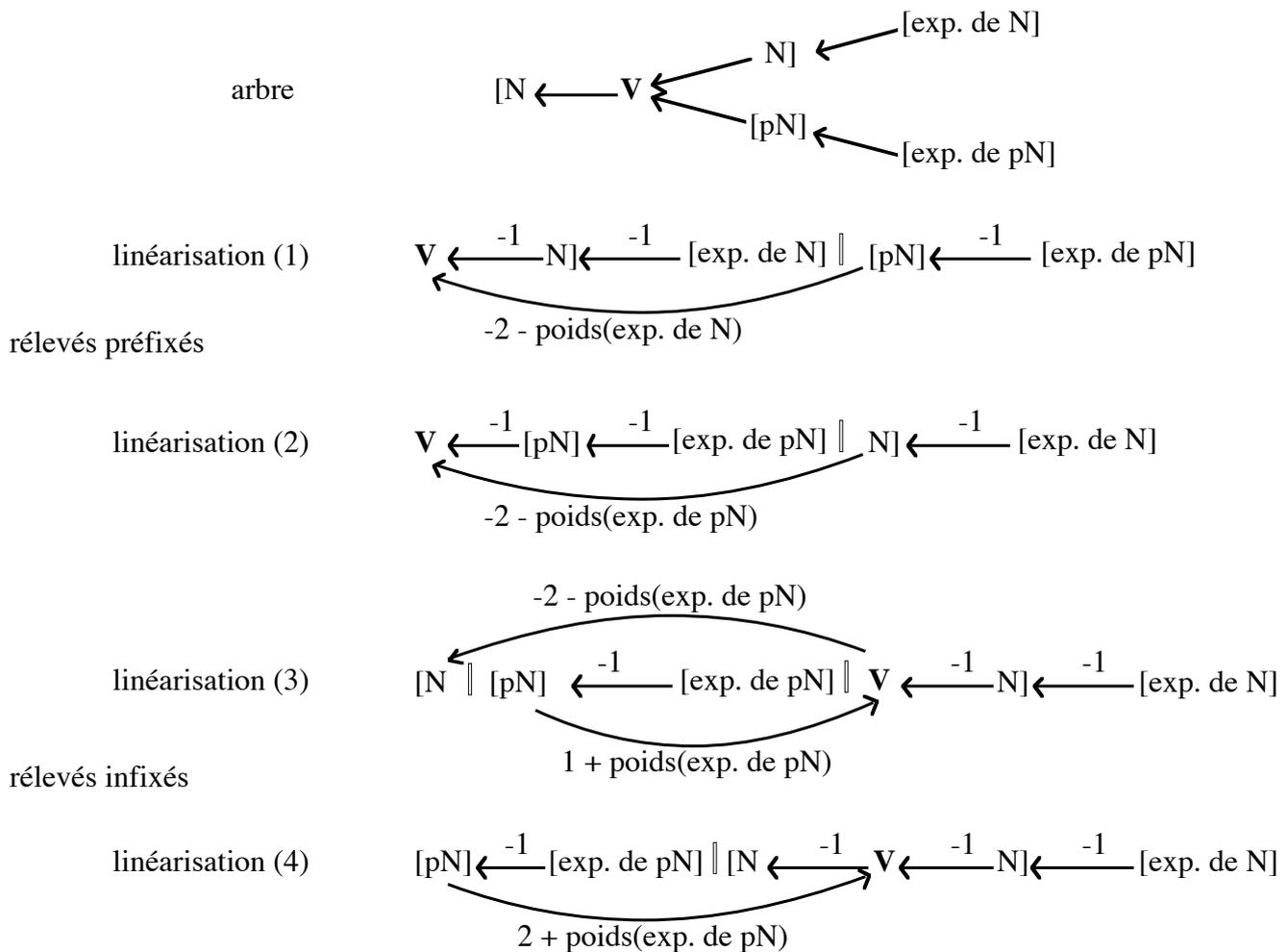


Figure 8: Linéarisation d'un nœud séquence verbale à 2 branches

Les linéarisations (1) et (2) ont toutes deux un ordre linéaire régissant - dépendants (un relevé préfixé); elles diffèrent en ce qui concerne l'ordre des dépendants entre eux:

(1) linéarisation la plus fréquente:

ordre des dépendants entre eux: on trouve cette linéarisation (toujours observée optimisée) quand:

$$\text{poids}(\text{expansion de N}) \leq \text{poids}(\text{expansion de pN})$$

[pN] est placé après les expansions de l'objet de [NVN] : [NVN] [] [pN]

(2) linéarisation plus rare:

ordre des dépendants entre eux: on trouve cette linéarisation (toujours observée optimisée) quand:

$$\text{poids}(\text{expansion de pN}) \leq \text{poids}(\text{expansion de N})$$

[pN] est inséré dans [NV.N] entre verbe et objet: [NV[pN]N]

les linéarisations (3) et (4) ont toutes deux un ordre linéaire dépendants - régissant - dépendants (un relevé infixé); elles diffèrent en ce qui concerne la définition du segment infixé:

(3) le segment infixé est une séquence verbale seule

séquence verbale conjuguée: pN est inséré dans [N.VN], antéposé à V: [N[pN]VN]

séquence verbale auxiliée: pN est inséré entre auxiliaire et auxilié: [NV[pN]ùN] ou [NV[pN]IN]

séquence verbale infinitive ou participe présent clippée: [c[,pN],qIN] ou [c[,pN],sRN]

(4) le segment infixé est une séquence verbale conjuguée avec son sujet et l'expansion de ce sujet.

Exemples de linéarisation (1): [NVN] [pN] verbe, branche objet, branche complément

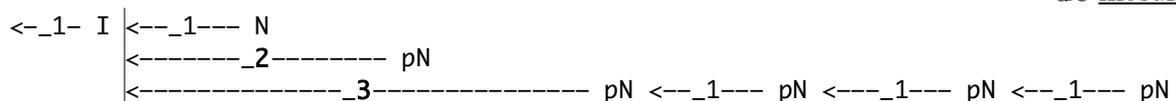
[ on rencontre les pigments caroténoïdes ] [ sous diverses formes ]



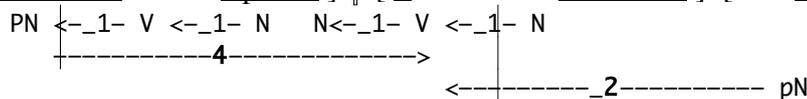
une approche statistique consisterait

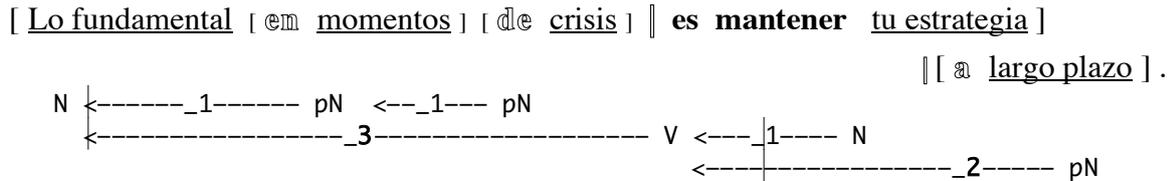
à caractériser les objets sur un point dans un espace de dimension égale à le nombre

de mesures



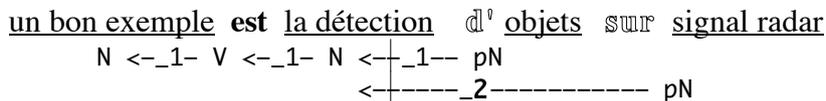
[ if the animal were a plaice ] || [ it switches behaviour ] [ on the spawning ground ]



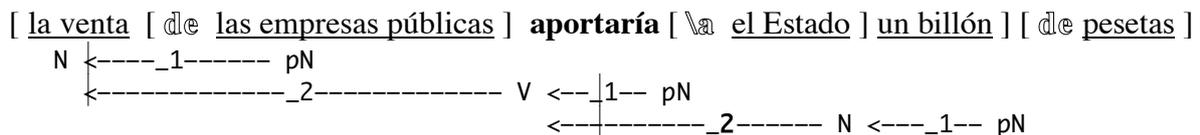
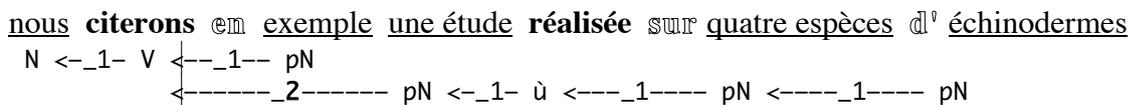


Dans ces 4 exemples, une séquence verbale a 1 branche objet et une branche complément toutes deux de poids 1, et est linéarisée de manière optimisée dans l'ordre: séquence verbale, branche objet, branche complément; on peut faire l'hypothèse que, dans le cas de branches de poids égaux, cette linéarisation est préférée à [NV [pN] N] , la linéarisation de type 2, qui donne 2 séquences nominales contiguës, ce qui entraînerait une difficulté de segmentation entre 2 séquences de même type; d'une manière générale, on observe que le locuteur - scripteur tente d'éviter de placer 2 séquences de même type en contiguïté.

Cas du verbe nominalisé, avec le même arbre:



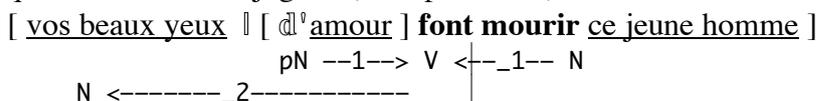
Exemples de linéarisation (2): [NV [pN] N] verbe, branche complément, branche objet



Dans ces deux exemples, une séquence verbale a 1 branche objet de poids supérieur à 1 et une branche complément de poids 1, et est linéarisée de manière optimisée dans l'ordre: séquence verbale, branche complément, branche objet; aucun exemple n'a été trouvé dans le corpus anglais. Le cas particulier de l'objet placé après le complément est exposé plus en détails ci-dessous en section 5.

Exemples de linéarisation (3): [N [pN] VN] [NV[pN]IN] [c[,pN],qIN]

Séquence verbale conjuguée (n'est plus usité):

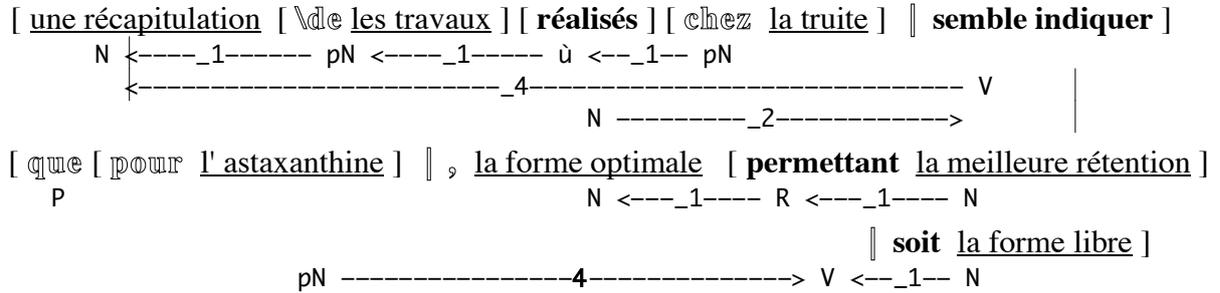


Remarque: une coupure prosodique est nécessaire pour marquer que [ d' amour ] dépend de la séquence suivante et non précédente.

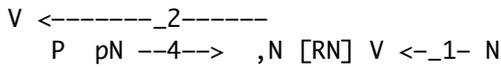




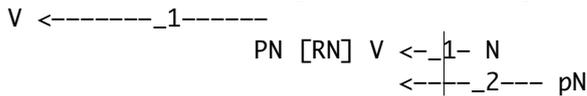
L'antéposition d'un bloc avant sujet - verbe ne se produit pas seulement avant le bloc principal, mais aussi dans un bloc subordonné: [pN] est alors inséré entre clip (P = conjonction de subordination) et sujet: [P[pN]NVN] :



on a un relevé infixé:



un relevé normalement préfixé aurait été plus optimisé (2+4+1>1+1+2):



Ceci montre que d'autres contraintes sont plus fortes que la contrainte d'optimisation de la linéarisation, des contraintes probablement dues à la nécessité de marquer la structure du niveau supérieur: la structure du paragraphe en termes de phrases (voir à ce sujet les travaux de Nadine Lucas).

4.3.4. Nœud à 2 branches = régissant de 2 branches coordonnées (ou énumérées)

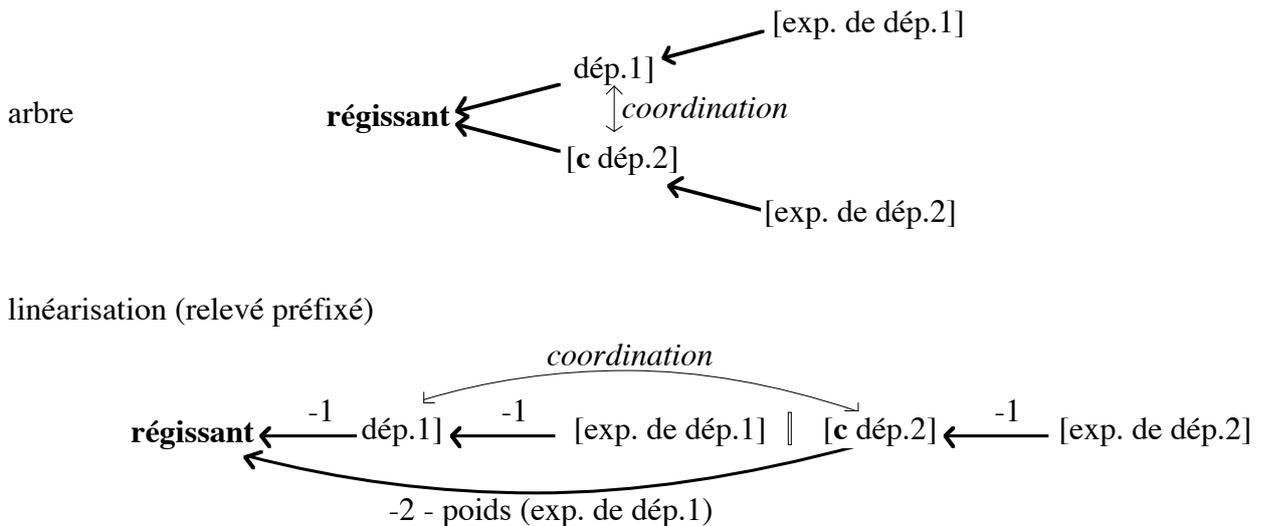


Figure 9: Linéarisation d'un nœud à 2 branches coordonnées

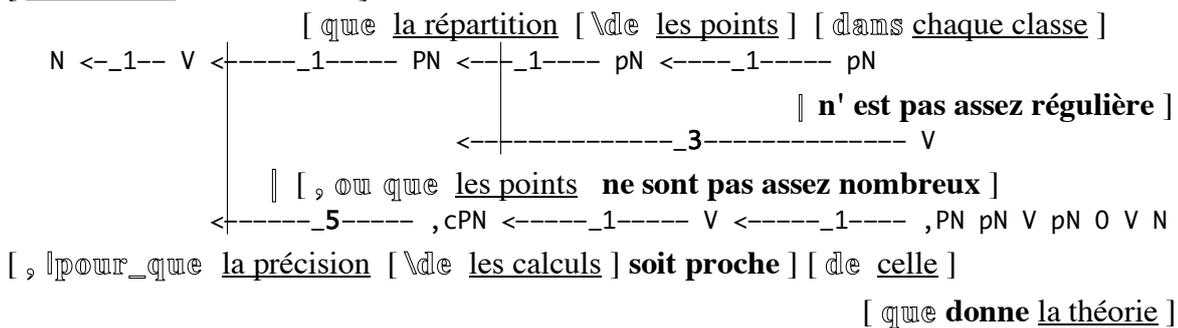
L'ordre linéaire observé est toujours le même: régissant - dépendant 1 - dépendant 2 coordonné.

linéarisation optimisée  $\Leftrightarrow$  poids(exp. de dép.1)  $\leq$  poids(exp. de dép.2)

Des séquences coordonnées sont isodépendantes: la seconde séquence coordonnée dépend du régissant de la première séquence coordonnée; réciproquement, les dépendants d'un même régissant, s'ils sont isofonctionnels, et mieux, isomorphes, peuvent être considérés comme coordonnés, même s'il n'y a pas de conjonction de coordination, comme dans le cas d'une énumération. La coupure ( | ) est avant la seconde séquence coordonnée.

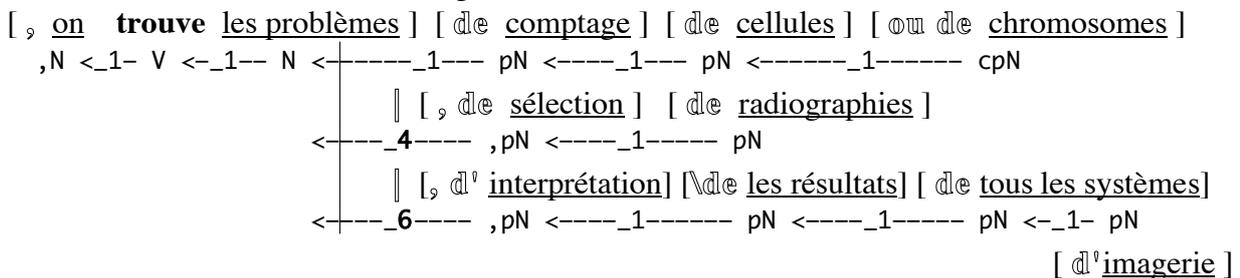
Exemples:

[ la difficulté **est souvent** ]



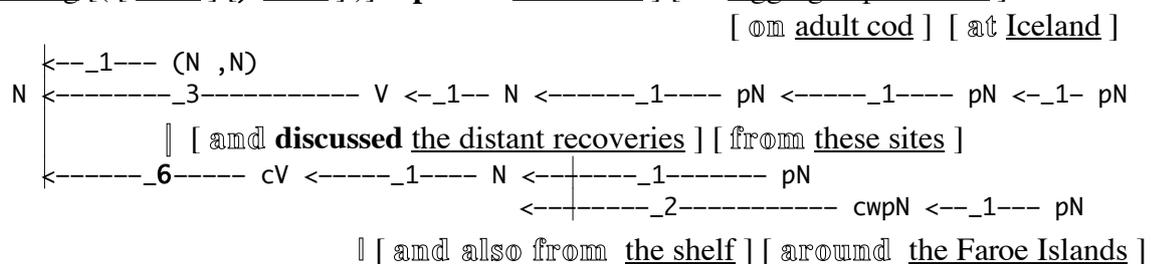
cas optimisé: branche 1 de poids 4, branche 2 de poids 8.

[ dans le domaine ] [ de les images médicales ]

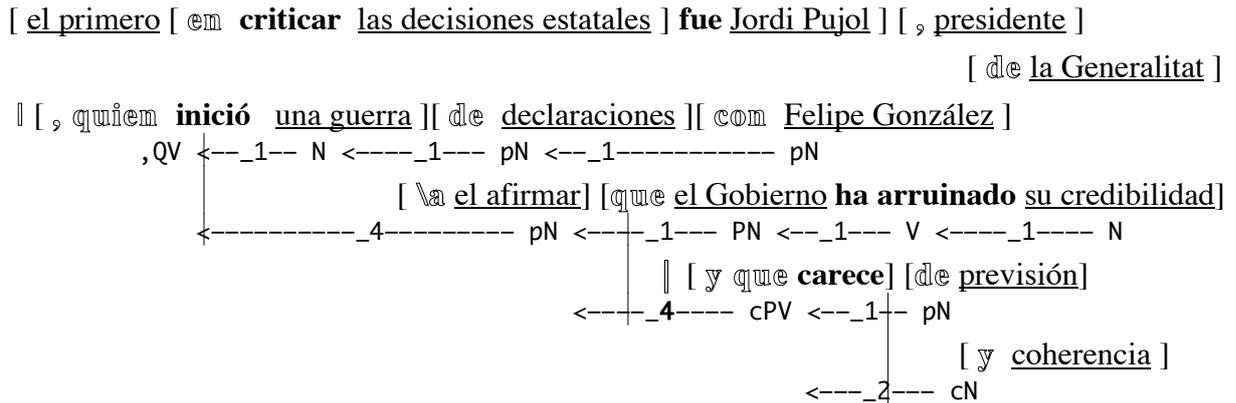


cas partiellement optimisé: branche 1 de poids 3, branche 2 de poids 2, branche 3 de poids 4.

[ Taning [( [ 1934 ] [ , 1935 ] )] **reported the results** ] [ of tagging experiments ]



cas optimisé et équilibré: branche 1 de poids 5, branche 2 de poids 5.



cas optimisé et équilibré: branche 1 de poids 3, branche 2 de poids 3.

5. Cas particulier de la linéarisation de l'objet

Étudiés des exemples d'ordre linéaire d'un nœud à 2 branches tirés du corpus français:  
 un verbe, son objet et son complément

Le régissant, le verbe, est toujours préfixé; les 2 dépendants s'observent dans les deux ordres possibles, toujours dans l'ordre de la linéarisation optimisée, ce qui en fait une illustration intéressante.

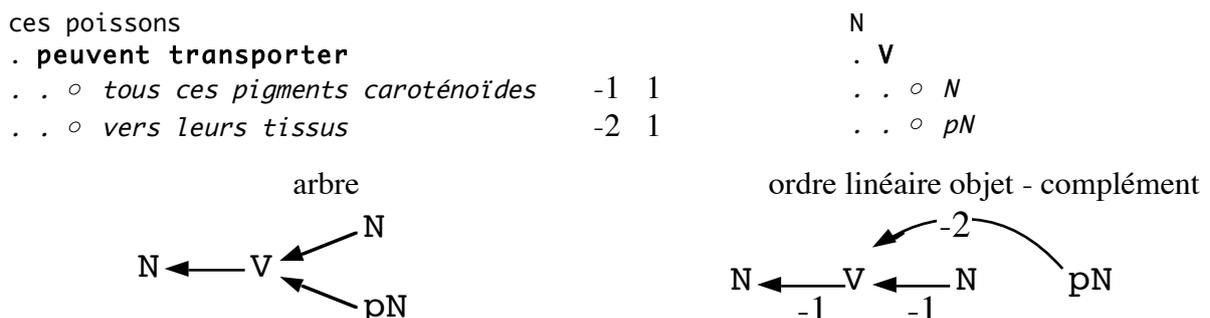
- On observe toujours l'ordre branche objet - branche complément
- quand poids(expansion de l'objet) ≤ poids(expansion du complément)
- et on observe toujours l'ordre branche complément - branche objet
- quand poids(expansion du complément) ≤ poids(expansion de l'objet)

Les exemples qui suivent illustrent tous les cas possibles en compliquant progressivement.

- Conventions:
- le **régissant verbal** est en gras
  - les *branches dépendantes* sont en italiques, leur début est marqué par ◦
  - pour chaque branche les deux nombres sont respectivement:
    - la longueur de la dépendance
    - le poids de la branche

5.1. Branches de poids égaux

(préférence à placer l'objet contre le verbe, pour éviter deux séquences nominales contiguës)



5.2. Branches de poids différents

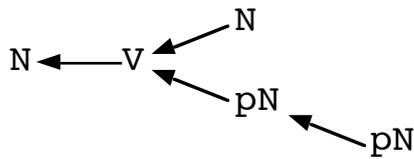
- branche objet de poids =1
- branche complément du verbe de poids >1

ces caroténolipoprotéines

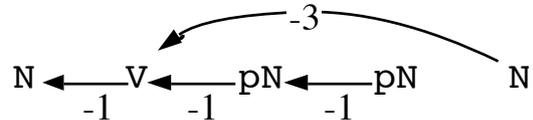
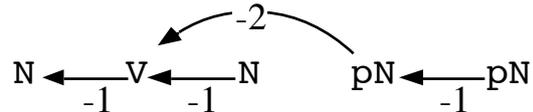
. **semblent jouer**  
 . . ◦ *un rôle important* -1 1  
 . . ◦ *dans les phénomènes* -2 2  
 . . . ◦ *de maturation sexuelle*

N  
 . V  
 . . ◦ N  
 . . ◦ pN  
 . . . pN

arbre



ordre linéaire observé objet - complément => optimisé



si ordre linéaire complément - objet => non

optimisé

on

. **rencontre**  
 . . ◦ *les pigments caroténoïdes* -1 1  
 . . ◦ *sous diverses formes* -2 >3  
 . . . ◦ *qui*  
 . . . . ◦ *sont essentiellement*

N  
 . V  
 . . ◦ N  
 . . ◦ pN  
 . . . Q  
 . . . . V

l' auteur

. **remercie**  
 . . ◦ *le Professeur J.Ceccaldi* -1 1  
 . . ◦ *pour l' intérêt* -2 6  
 . . . ◦ *soutenu*  
 . . . . ◦ *qu'*  
 . . . . ◦ *il*  
 . . . . ◦ *a manifesté*  
 . . . . ◦ *l\à\_le\_cours\_de ce travail*

N  
 . V  
 . . ◦ N  
 . . ◦ pN  
 . . . ù  
 . . . . ◦  
 . . . . ◦ N  
 . . . . ◦ V  
 . . . . ◦ pN

5.3. Branches de poids différents

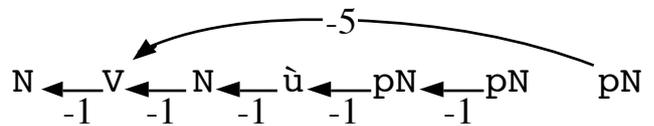
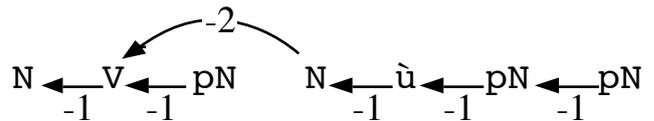
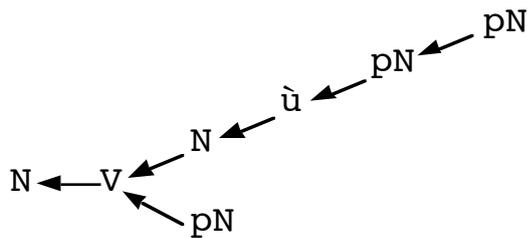
- branche complément du verbe de poids =1
- branche objet de poids >1

nous  
 . **citerons**  
 . . ◦ *en exemple* -1 1  
 . . ◦ *une étude* -2 4  
 . . . *réalisée*  
 . . . . *sur quatre espèces*  
 . . . . . *d' échinodermes*

N  
 . V  
 . . ◦ pN  
 . . ◦ N  
 . . . ù  
 . . . . pN  
 . . . . . pN

arbre

ordre linéaire observé complément - objet => optimisé



si ordre linéaire objet - complément, alors non optimisé

elle  
 . **définit**  
 . . ◦ *en fait* -1 1  
 . . ◦ *toutes les formes* -2 4  
 . . . *qui*  
 . . . . *ne sont pas*  
 . . . . . *\de les esters*

N  
 . V  
 . . ◦ pN  
 . . ◦ N  
 . . . Q  
 . . . . V  
 . . . . . N

la liaison  
 . entre le pigment caroténoïde  
 . et la protéine  
 . **confère**  
 . . ◦ *\à le complexe* -1 1  
 . . ◦ *un caractère hydrosoluble particulier* -2 1 ou 3

N  
 . pN  
 . cN  
 . V  
 . . ◦ pN  
 . . ◦ N

les pigments caroténoïdes  
 . absorbent  
 . . l' énergie lumineuse  
 . . lorsque la chlorophylle  
 . . . est présente  
 . . . . en faible quantité  
 . , puis **transmettent**  
 . . ◦ *à celle-ci* -1 1  
 . . ◦ *l' énergie* -2 2  
 . . . *ainsi captée*

N  
 . V  
 . . N  
 . . PN  
 . . . U  
 . . . . pN  
 . ,cV  
 . . ◦ pN  
 . . ◦ N  
 . . . ù

le mot		N
. d'interprétation		. pN
. <b>signifie</b>		. V
. . ◦ <i>dans le cas</i>	-1 3	. . ◦ pN
. . . <i>qui</i>		. . . Q
. . . . <i>nous occupe</i>		. . . . V
. . ◦ <i>une simple catégorisation</i>	-4 6	. . ◦ N
. . . <i>du phénomène</i>		. . . pN
. . . . <i>perçu</i>		. . . . ù
. . ◦ <i>,c'est-à-dire son affectation</i>		. . ◦ ,N
. . . <i>à une famille</i>		. . . pN
. . . . <i>de phénomènes ressemblants ;</i>		. . . . pN

Cette linéarisation est optimisée, car les branches objets 2 et 3 (de poids 3 chacune) sont coordonnées et doivent rester contiguës, et donc sont équivalentes à une branche objet de poids 6.

...		...
les travaux		N
. de Kuhn		. pN
. qui		. Q
. . <b>décrivaient</b>	-1	. . V
. . . ◦ <i>pour la première fois</i>	-1 1	. . . ◦ pN
. . . ◦ <i>la présence</i>	-2 3	. . . ◦ N
. . . . <i>d' astacine</i>		. . . . pN
. . . . <i>chez le homard</i>		. . . . pN
. . . ◦ <i>comme caroténoïde</i>	-5 4	. . . ◦ pN
. . . . <i>différent</i>		. . . . ù
. . . . . <i>de ceux</i>		. . . . . pN
. . . . . . <i>\de les végétaux</i>		. . . . . . pN

Cette linéarisation d'un nœud à 3 branches est optimisée, avec les branches linéarisées par poids croissants: 1, 3, et 4.

## 6. Cas particulier de la linéarisation des relatives en français

Les relatives en français fournissent un champ d'expérimentation intéressant pour des concepts qui tentent d'expliquer l'ordre linéaire, car on trouve le sujet et le verbe de la relative linéarisés dans les deux ordres possibles. Mais le problème se complique par rapport au cas de l'ordre objet - complément étudié en section précédente, car on étudie la linéarisation de deux nœuds au lieu d'un, et la position du pronom relatif est fixée en début de relative.

Une relative est un bloc (quelquefois deux) régi par une séquence nominale, l'antécédent; ce bloc contient toujours une séquence verbale conjuguée.

Un pronom relatif est un amalgame constitué:

- d'un pronom tonique (= pouvant être dissocié une séquence verbale), séquence nominale nœud de l'arbre, pouvant être:
  - sujet: il régit alors le verbe de la relative
  - objet ou circonstant: c'est alors une feuille de l'arbre (il ne peut régir),

- et d'un clip de subordination qui subordonne la nominale sujet et la verbale conjuguée de la relative, et dont la position est obligatoire en début de bloc;
- et dans l'ordre linéaire pronom] [clip, car le pronom est extérieur au bloc qui commence au clip, et ce pronom est régi le plus souvent par une séquence du bloc qui suit, mais peut être régi par une séquence du bloc suivant (voir ci-dessous le pronom relatif objet ou dépendant d'un bloc subordonné de la relative)

Ce clip impose la position du pronom relatif en début de bloc, ce qui rend la linéarisation particulière.

Il y a 3 types de pronoms relatifs selon leur fonction dans la relative:

- sujet (Q = "qui"),
- objet (O = "que"),
- dépendant du verbe (pn = "dont", "où", préposition + "lequel"), du sujet ou de l'objet ("dont")

"dont" et "où" sont des amalgames à 3 composantes: préposition - pronom relatif, donc préposition - pronom - clip, dans cet ordre linéaire: [préposition pronom] [clip];

le pronom Q dépend de son coréférent;

le pronom [O] ou [pn] dépend du **premier** candidat (V, I pour les deux, N pour "dont" seulement): on a saturation au plus tôt.

L'antécédent (coréférent anaphorique du pronom relatif) régit le sujet de la relative; il n'est pas toujours contigu à la relative, car il peut avoir une autre expansion, linéarisée avant la relative.

L'antécédent est le plus souvent déterminé (partitif, déterminant, ou quantificateur, ou nom propre), mais ce n'est pas toujours la séquence nominale déterminée la plus récente:

*Il est ici intéressant de noter le cas des crustacés des zones très profondes **qui** ont généralement une teinte rouge très marquée.*

(ce sont les crustacés qui ont une teinte rouge, alors que "zones" est déterminé)

### 6.1. Pronom relatif sujet

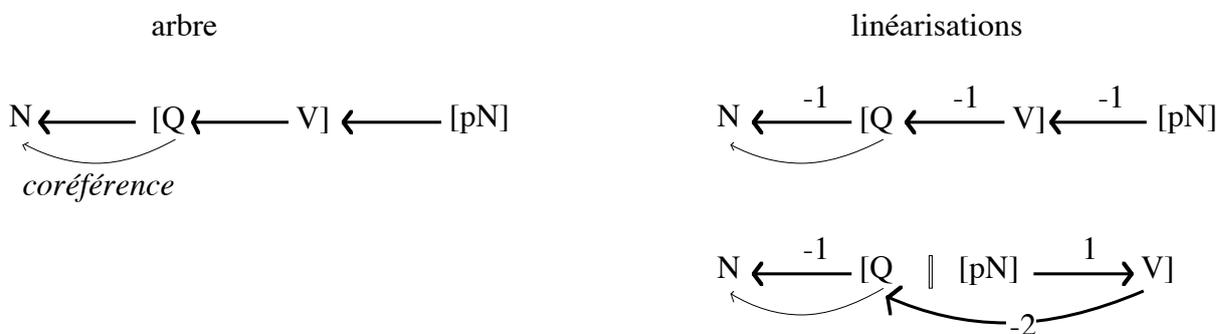


Figure 10: Linéarisation d'une relative avec pronom relatif sujet

La séquence verbale est accordée en genre, nombre, personne avec l'antécédent, comme s'il était son sujet; comme la place du sujet est fixe en début de relative (c'est le pronom relatif), la branche sujet est toujours la première; mais l'insertion d'un bloc antéposé au verbe est possible, mais rare (pas d'exemple dans le corpus).

Reprenons la dernière phrase exemple, et antéposons la branche 1:

...		...
les travaux		N
. de Kuhn		. pN
. qui		. Q
. . . ◦ <i>pour la première fois</i>	1 1	. . . ◦ pN
. . <b>décrivaient</b>	-2	. . V
. . . ◦ <i>la présence</i>	-1 3	. . . ◦ N
. . . . <i>d' astacine</i>		. . . . pN
. . . . <i>chez le homard</i>		. . . . pN
. . . ◦ <i>comme caroténoïde</i>	-4 4	. . . ◦ pN
. . . . <i>différent</i>		. . . . ù
. . . . . <i>de ceux</i>		. . . . . pN
. . . . . \ <i>de les végétaux</i>		. . . . . pN
relation antécédent - relative:	travaux décrivaient	

On obtient une linéarisation possible (par relevé infixé), légèrement plus optimisée:

$$1+1+2+5 > 1+2+1+4 \quad (9 > 8)$$

qui n'a pas été choisie, pour une raison que nous ne pouvons expliquer.

## 6.2. Pronom relatif objet

Dans l'arbre, on défait l'amalgame du pronom relatif en P clip et O pronom objet; dans la phrase ils se retrouvent amalgamés en O, en début de bloc; ici, le problème est plus complexe que les cas étudiés ci-dessus: on n'étudie pas l'optimisation de la linéarisation d'un seul nœud, mais l'optimisation simultanée de la linéarisation de 2 nœuds de la relative: le nœud sujet et le nœud verbe:

- le nœud verbe a toujours un relevé infixé, impliqué par la place obligatoire antéposée du pronom relatif objet;

- le nœud sujet peut être relevé de deux manières: préfixé ou infixé, comme dans le cas général (voir ci-dessus, nœud à 2 branches = séquence nominale sujet, en 4.3.2); quand les deux nœuds ont un relevé infixé, on a deux contiguïtés autour du verbe: l'antéposition de l'objet a tendance à entraîner l'antéposition du verbe, car l'optimisation est meilleure quand tous les nœuds ont le même type de relevé (comme dans une interrogative).

On a donc deux linéarisations possibles:

l'ordre linéaire: sujet et ses expansions - verbe et ses expansions

l'ordre linéaire: verbe et ses expansions - sujet et ses expansions

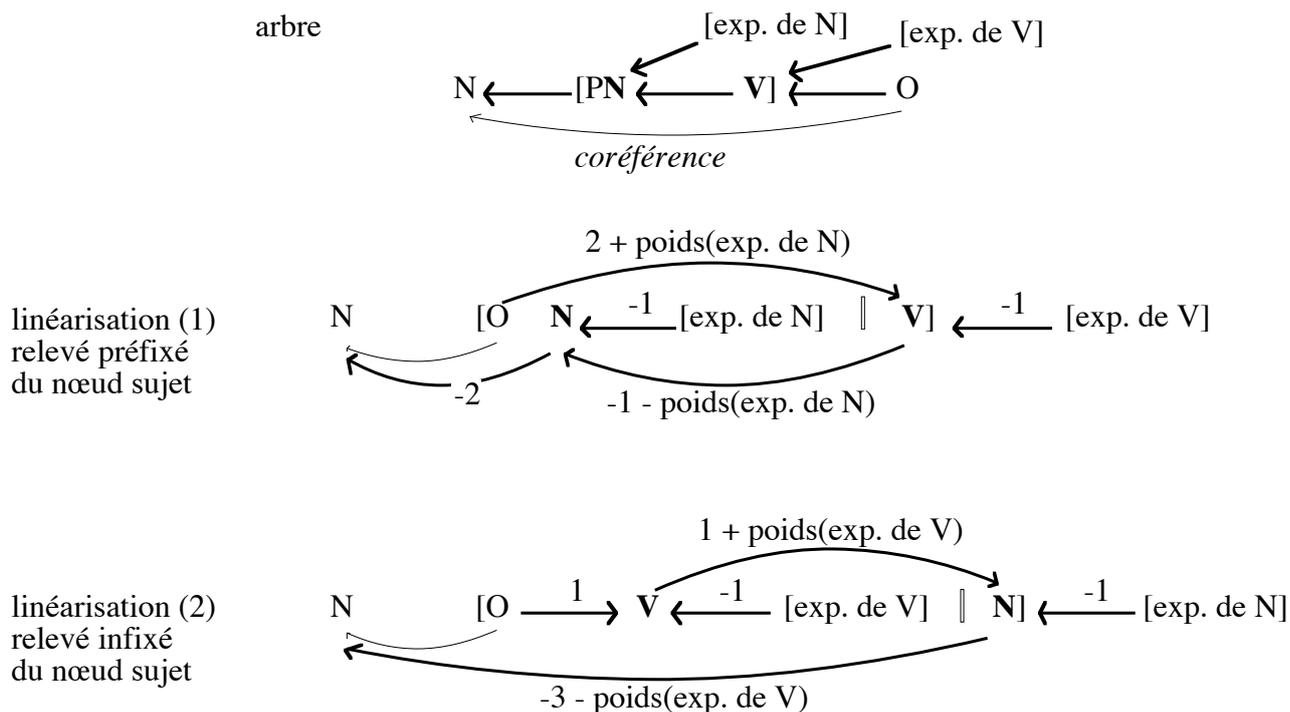


Figure 11: Linéarisation d'une relative avec pronom relatif objet

Choix théorique entre linéarisations (1) ou (2): quelle est la linéarisation optimisée?

On choisit le minimum de la somme des valeurs absolues des longueurs des dépendances:

linéarisation (1):

$$\sum = 2 + \text{poids}(\text{exp. de N}) + 1 + 1 + 2 + 1 + \text{poids}(\text{exp. de N}) = 7 + 2 * \text{poids}(\text{exp. de N})$$

linéarisation (2):

$$\sum = 1 + \text{poids}(\text{exp. de V}) + 1 + 1 + 1 + 3 + \text{poids}(\text{exp. de V}) = 7 + 2 * \text{poids}(\text{exp. de V})$$

Donc la linéarisation (1) est optimisée si on a:  $\text{poids}(\text{exp. de N}) \leq \text{poids}(\text{exp. de V})$

et la linéarisation (2) est optimisée si on a:  $\text{poids}(\text{exp. de V}) \leq \text{poids}(\text{exp. de N})$ .

On retrouve presque le même résultat que dans le cas général de la linéarisation optimisée du nœud sujet (ci-dessus, en 4.3.2), mais on ne compare plus les branches d'un même nœud, mais les expansions des deux nœuds.

On observe que les linéarisations sont presque toujours optimisées, mais pas toujours, car d'autres phénomènes sont en jeu; il n'y a pas de coupure prosodique entre O et N ou O et V comme entre tout clip et la séquence clippée: l'élision est possible et fréquente: *qu'il, qu'un, qu'élèvent*.

Le pronom [O] dépend du **premier** candidat régissant séquence verbale transitive (V I).

Exemple de linéarisation (1), par relevé préfixé du nœud sujet :

conventions: le **pronom relatif** est en gras

*sujet ou verbe avec leur expansion* sont en italiques, leur début est marqué par ◦ pour chacun des deux nœuds, les deux nombres sont respectivement:

- la longueur de la dépendance
- le poids du nœud avec son expansion

Mais on		wN
. verra aussi		. V
. . ce		. . N
. . . . . <b>que</b>	3 1	. . . . . <b>0</b>
. . . . . ◦ <i>l' utilisation</i>	-2 2	. . . . . ◦ <i>N</i>
. . . . . <i>de ces modèles</i>		. . . . . <i>pN</i>
. . . . . ◦ <i>apporte</i>	-2 3	. . . . . ◦ <i>V</i>
. . . . . <i>dans la résolution</i>		. . . . . <i>pN</i>
. . . . . . <i>de certains problèmes complexes</i>		. . . . . . <i>pN</i>

relation antécédent - relative: apporte ce

linéarisation optimisée: poids(expansion de N) < poids(expansion de V) (1<2)

Exemples de linéarisation (2), par relevé infixé du nœud sujet :

La précision		N
. des calculs		. pN
. est proche		. V
. . de celle		. . pN
. . . . . <b>que</b>	1 1	. . . . . <b>0</b>
. . . . . ◦ <i>donne</i>	1 1	. . . . . ◦ <i>V</i>
. . . . . ◦ <i>la théorie</i>	-3 1	. . . . . ◦ <i>N</i>

relation antécédent - relative: donne celle

linéarisation optimisée: poids(expansion de V) = poids(expansion de N) (0=0)

La linéarisation (1) aurait donné:

La précision		N
. des calculs		. pN
. est proche		. V
. . de celle		. . pN
. . . . . <b>que</b>	2 1	. . . . . <b>0</b>
. . . . . ◦ <i>la théorie</i>	-2 1	. . . . . ◦ <i>N</i>
. . . . . ◦ <i>donne</i>	-1 1	. . . . . ◦ <i>V</i>

qui semble également optimisée (1+1+3=2+2+1, et égalité des poids des expansions de N et V), mais qui paraît plus lourde; pourquoi?

Hypothèse: dans la comparaison entre 2 linéarisations équivalentes sur le plan de la somme des valeurs absolues des longueurs des dépendances, la plus optimisée est celle qui a le plus de dépendances marquées par contiguïté: 2 pour la linéarisation (2), 1 pour la linéarisation (1); si cette hypothèse est confirmée, toute relative dont le nœud verbal est composé de la seule séquence verbale (et dont le sujet est une séquence nominale et non un pronom personnel atone) correspond à la linéarisation (2).

Ce deuxième exemple confirme cette hypothèse:

L'ergonomie		N
. . . <b>que</b>	1 1	. . . <b>0</b>
. . . ◦ <i>peut représenter</i>	1 1	. . . ◦ <i>V</i>
. . . ◦ <i>un système</i>	-3 3	. . . ◦ <i>N</i>
. . . <i>à interrogation</i>		. . . <i>pN</i>
. . . <i>en langage naturel</i>		. . . <i>pN</i>
. ne devra pas être infirmée		. <i>V</i>
. . par une édition peu pratique		. . <i>pN</i>

relation antécédent - relative:    représenter ergonomie

linéarisation optimisée: poids(expansion de V) < poids(expansion de N)    (0<2)

La linéarisation (1) aurait donné:

L'ergonomie		N
. . . <b>qu'</b>	4 1	. . . <b>0</b>
. . . ◦ <i>un système</i>	-2 3	. . . ◦ <i>N</i>
. . . <i>à interrogation</i>		. . . <i>pN</i>
. . . <i>en langage naturel</i>		. . . <i>pN</i>
. . . ◦ <i>peut représenter</i>	-3 1	. . . ◦ <i>V</i>
. ne devra pas être infirmée		. <i>V</i>
. . par une édition peu pratique		. . <i>pN</i>

qui est loin de l'optimisation (1+1+3 < 4+2+3), et qui en plus introduit une insertion au niveau 2, avec deux séquences verbales contiguës:

linéarisation (2) observée:	linéarisation (1) non observée:
[ L'ergonomie	[ L'ergonomie
[ que peut représenter un système ]	[ qu' un système
[ à interrogation ]	[ à interrogation ]
[ en langage naturel ]	[ en langage naturel ]
ne devra pas être infirmée ]	peut représenter ]
[ par une édition peu pratique ] .	ne devra pas être infirmée ]
	[ par une édition peu pratique ] .

Cas du pronom personnel sujet atone:

Quand N est un pronom personnel sujet atone, il ne peut avoir d'expansion (poids(exp. de N)=0), car un pronom personnel sujet atone ne peut régir, et il fait un tout indissociable avec V (on peut même dire qu'il fait partie de V, comme en espagnol ou en latin), et il n'y a qu'une seule linéarisation possible:

N [O NV] [exp. de V] qui est une fusion des deux linéarisations possibles

On		N
. étudie ensuite		. <i>V</i>
. . un aspect différent		. . <i>N</i>
. . . . . , <b>que</b>	2 1	. . . . . , <b>0</b>
. . . . . ◦ <i>l'on</i>	-2 1	. . . . . ◦ <i>N</i>
. . . . . ◦ <i>pourra appliquer pratiquement</i>	-1 3	. . . . . ◦ <i>V</i>
. . . . . <i>à tous les formalismes</i>		. . . . . <i>pN</i>
. . . . . <i>décrits précédemment</i>		. . . . . <i>ù</i>

relation antécédent - relative:    appliquer aspect

linéarisation optimisée: poids(expansion de N) < poids(expansion de V) (0<2)

### 6.3. Pronom relatif dépendant du verbe

Ce cas est presque identique au cas précédent, sinon que le verbe peut avoir un objet; dans l'arbre on défait l'amalgame du pronom relatif en P clip et [pn] pronom dépendant du verbe avec sa préposition; dans la phrase ils se retrouvent amalgamés en pn, en début de bloc; ici, on étudie encore l'optimisation simultanée de la linéarisation de deux nœuds de la relative: le nœud sujet et le nœud verbe:

- le nœud verbe a toujours un relevé infixé, impliqué par la place obligatoire antéposée de son pronom relatif complément,

- le nœud sujet peut être relevé de deux manières: préfixé ou infixé, comme dans le cas général.

On a encore deux linéarisations possibles:

l'ordre linéaire: sujet et ses expansions - verbe et ses expansions

l'ordre linéaire: verbe et ses expansions - sujet et ses expansions

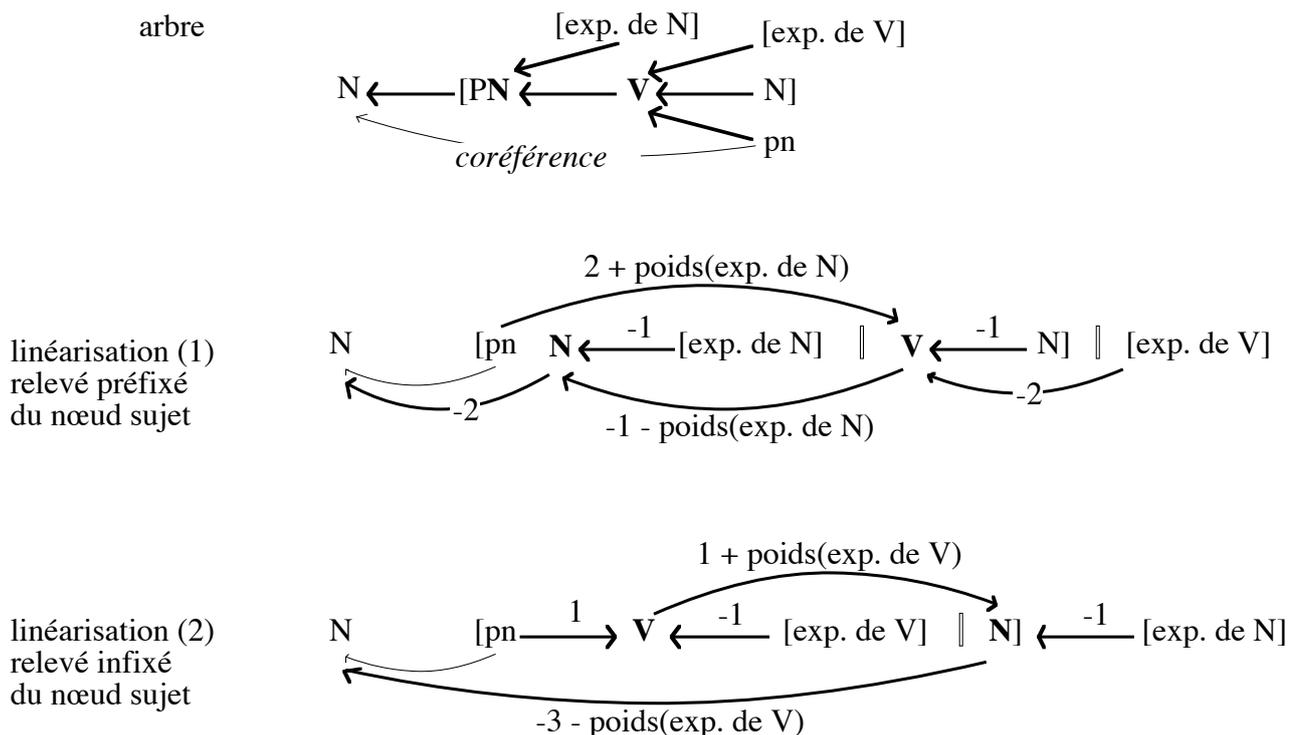


Figure 12: Linéarisation d'une relative avec pronom relatif dépendant du verbe

Choix théorique entre linéarisations (1) ou (2): quelle est la linéarisation optimisée?

On choisit la linéarisation pour laquelle la somme des valeurs absolues des longueurs des dépendances est minimale, en calculant en l'absence d'objet, qu'on ne trouve jamais en linéarisation (2), car elle deviendrait trop peu optimisée:

linéarisation (1):  $\sum = 7 + 2 * \text{poids}(\text{exp. de N})$

linéarisation (2):  $\Sigma = 7 + 2 * \text{poids}(\text{exp. de V})$

Donc la linéarisation (1) est optimisée si on a:  $\text{poids}(\text{exp. de N}) \leq \text{poids}(\text{exp. de V})$

et la linéarisation (2) est optimisée si on a:  $\text{poids}(\text{exp. de V}) \leq \text{poids}(\text{exp. de N})$ .

On retrouve le même résultat qu'avec le pronom relatif objet.

On observe que les linéarisations sont presque toujours optimisées, mais pas toujours, car d'autres phénomènes sont en jeu; il n'y a pas de coupure prosodique entre pn et N ou pn et V comme entre tout clip et la séquence clippée: la liaison est possible et fréquente: *dont il, dont un*, ainsi que l'enchaînement: *avec lequel il, avec lequel un, où l'on*.

Les pronoms relatifs composés d'une préposition et des différentes formes fléchies de "lequel", et l'amalgame "où" ("dans" avec "lequel" non marqué en genre et nombre), dépendent du **premier** candidat verbal V ou I: saturation au plus tôt; ces pronoms sont accordés en genre et nombre avec leur antécédent.

"dont" est un amalgame de "de" et "lequel" non marqué en genre et nombre; il diffère des autres car il peut dépendre de V ou de N (sujet ou objet, ou autre): "dont" dépend du **premier** candidat régissant (N ou V, ou I) possible qui lui est postérieur dans l'ordre linéaire: toujours saturation au plus tôt. Donc, si le sujet n'est pas un pronom (atone ou démonstratif):

ordre linéaire: sujet et ses expansions - verbe et ses expansions => "dont" dépend de N sujet

ordre linéaire: verbe et ses expansions - sujet et ses expansions => "dont" dépend de V

(on verra que ce principe souffre des exceptions)

Exemples de linéarisation (1), par relevé préfixé du nœud sujet :

Le choix		N
. de l'espace		. pN
. . de représentation		. . pN
. . . . <b>dans lequel</b>	2 1	. . . . <b>pn</b>
. . . <i>o les formes</i>	-3 1	. . . <i>o N</i>
. . . . <i>o vont être codées</i>	-1 1	. . . . <i>o V</i>
. est		. V
. . un		. . N
. . . des facteurs les plus importants		. . . pN
. . . . dans la réalisation		. . . . pN
. . . . . d'un tel système		. . . . . pN

relation antécédent - relative: codées dans espace (accord masculin singulier)

linéarisation optimisée:  $\text{poids}(\text{expansion de N}) = \text{poids}(\text{expansion de V})$  (0=0)

La linéarisation (2) aurait été possible (meilleure car 2 contiguïtés):

Le choix		N
. de l'espace		. pN
. . de représentation		. . pN
. . . . <b>dans lequel</b>	1 1	. . . . <b>pn</b>
. . . . <i>o vont être codées</i>	1 1	. . . . <i>o V</i>
. . . . <i>o les formes</i>	-4 1	. . . . <i>o N</i>
. est		. V
. . un		. . N
. . . des facteurs les plus importants		. . . pN
. . . . dans la réalisation		. . . . pN
. . . . . d'un tel système		. . . . . pN

Elle est malgré tout moins heureuse, peut-être parce qu'on préfère "codées" après "formes".

Il		N
. est ici intéressant		. V
. . de noter		. . qI
. . . le cas		. . . N
. . . . des crustacés		. . . . pN
. . . . . des zones très profondes		. . . . . pN
. . . . . . <b>chez lesquels</b>	4 1	. . . . . . <b>pn</b>
. . . . . . <i>o les conditions particulières</i>	-3 3	. . . . . . <i>o N</i>
. . . . . . <i>de température</i>		. . . . . . <i>pN</i>
. . . . . . <i>et de hautes pressions</i>		. . . . . . <i>cpN</i>
. . . . . . <i>o paraissent inhiber</i>	-3 3	. . . . . . <i>o V</i>
. . . . . . <i>la formation</i>		. . . . . . <i>N</i>
. . . . . . . <i>des caroténoprotéines</i>		. . . . . . . <i>pN</i>

relation antécédent - relative: inhiber chez crustacés (accord masculin pluriel)

linéarisation optimisée: poids(expansion de N) = poids(expansion de V) (2=2)

Ici, la linéarisation (2) n'est pas possible car la présence de l'objet donnerait un ordre linéaire:

pronom relatif - verbe - objet et ses expansions - sujet et ses expansions

ordre inexistant en français, peut-être pour éviter la contiguïté de deux séquences nominales, et pour éviter un déplacement si important du sujet par rapport à sa position normale.

Hypothèse: un verbe antéposé à son sujet ne peut avoir d'objet (objet séquence nominale, car un pronom objet atone reste possible); on en déduit: la linéarisation (2) ne peut se produire que si le verbe de la relative n'a pas d'objet séquence nominale.

La linéarisation (1) n'existe pas avec "dont" dépendant du verbe (voir ci-dessus), quand le sujet n'est pas un pronom, car dans ce cas, "dont" dépend du sujet.

Exemples de linéarisation (2), par relevé infixé du nœud sujet :

on			N
. ne peut pas exprimer			. V
. . dans un tel modèle			. . pN
. . certains types			. . N
. . . de contraintes			. . . pN
. . . . . <b>auxquelles</b>	1 1		. . . . . <b>pn</b>
. . . . . ◦ <i>pourtant obéissent</i>	1 1		. . . . . ◦ V
. . . . . ◦ <i>les formes</i>	-3 2		. . . . . ◦ N
. . . . . <i>étudiées</i>			. . . . . ù

relation antécédent - relative: obéissent à contraintes (accord féminin pluriel)

linéarisation optimisée: poids(expansion de V) < poids(expansion de N) (0<1)

Remarque: "contraintes" n'est pas déterminé directement, mais indirectement par "certains types de" qui joue un rôle de quantificateur.

La linéarisation (1) aurait donné:

on			N
. ne peut pas exprimer			. V
. . dans un tel modèle			. . pN
. . certains types			. . N
. . . de contraintes			. . . pN
. . . . . <b>auxquelles</b>	3 1		. . . . . <b>pn</b>
. . . . . ◦ <i>les formes</i>	-2 2		. . . . . ◦ N
. . . . . <i>étudiées</i>			. . . . . ù
. . . . . ◦ <i>pourtant obéissent</i>	-2 1		. . . . . ◦ V

qui n'est pas optimisée (1+1+3<3+2+2).

Remarque: l'adverbe "pourtant" est antéposé à son verbe, contrairement à la postposition habituelle: encore une fois, l'optimisation est meilleure si tout est postposé, ou tout antéposé: c'est pour cela qu'on a "pourtant obéissent" dans la linéarisation (2), et qu'on préférerait "obéissent pourtant" dans la linéarisation (1).

Exemple avec "dont" (Saint Exupéry, Courrier sud, cité par Grevisse, dans "le bon usage"):

. . Quelle douceur			. . N
. aujourd'hui répandait			. V
cette lampe			N
. . . <b>dont</b>	1 1		. . . <b>pn</b>
. . ◦ <i>coulait</i>	1 1		. . ◦ V
. ◦ <i>une lumière</i>	-3 2		. ◦ N
. . <i>d'huile !</i>			. . pN

relation antécédent - relative: coulait de lampe (pas d'accord)

linéarisation optimisée: poids(expansion de V) < poids(expansion de N) (0<1)

Remarque: encore une phrase où la postposition des sujets est généralisée: *répandait lampe, coulait lumière*, ce qui permet une meilleure optimisation globale.

La linéarisation (1) aurait donné:

. . Quelle douceur		. . N
. aujourd'hui répandait		. V
cette lampe		N
. . . <b>dont</b>	1 1	. . . <b>pn</b>
. ◦ <i>une lumière</i>	-2 2	. ◦ <i>N</i>
. . <i>d'huile</i>		. . <i>pN</i>
. . ◦ <i>coulait !</i>	-2 1	. . ◦ <i>V</i>

relation antécédent - relative **différente**: lumière de lampe

"dont" dépend du premier candidat, car les deux relations "coulait de lampe" et "lumière de lampe" sont possibles, donc la linéarisation est contrainte par l'optimisation, et aussi par la relation antécédent - relative choisie par le locuteur.

Mais, quand une seule relation est possible, les deux linéarisations impliquent cette même relation, comme dans l'exemple suivant (extrait de: Les linguistiques contemporaines, repères théoriques, de Catherine Fuchs et Pierre Le Goffic, Hachette 1992):

. Est ensuite récusée		. V
la distinction traditionnelle		N
. entre une morphologie		. pN
. . conçue		. . ù
. . . comme l'étude		. . . pN
. . . . de la façon		. . . . pN
. . . . . <b>dont</b>	1 1	. . . . . <b>pn</b>
. . . . . ◦ <i>sont constitués</i>	1 1	. . . . . ◦ <i>V</i>
. . . . . ◦ <i>les mots</i>	-3 1	. . . . . ◦ <i>N</i>
. et une syntaxe		. cN
. . conçue		. . ù
. . . comme l'étude		. . . pN
. . . . de la combinaison		. . . . pN
. . . . . des mots		. . . . . pN
. . . . . en phrases		. . . . . pN

relation antécédent - relative: constitués de façon

linéarisation optimisée: poids(expansion de V) = poids(expansion de N) (0=0)

La linéarisation (1) aurait donné:

. . . comme l'étude		. . . pN
. . . . de la façon		. . . . pN
. . . . . <b>dont</b>	2 1	. . . . . <b>pn</b>
. . . . . ◦ <i>les mots</i>	-2 1	. . . . . ◦ <i>N</i>
. . . . . ◦ <i>sont constitués</i>	-1 1	. . . . . ◦ <i>V</i>

relation antécédent - relative **identique**: constitués de façon

car "constituer" régit avec la préposition "de".

Cas du pronom personnel sujet atone:

Quand N est un pronom personnel sujet atone ou démonstratif, comme pour un pronom relatif objet, il n'y a qu'une seule linéarisation possible: N [pn NV] [exp. de V] qui est une fusion des deux linéarisations possibles

Un autre exemple		N
. est		. V
. . l'utilisation		. . N
. . . de critères		. . . pN
. . . . de type structurel		. . . . pN
. . . pour organiser		. . . qI
. . . . un dictionnaire		. . . . N
. . . . . de représentants		. . . . . pN
. . . . . . de l'apprentissage		. . . . . . pN
. . . . . , afin d'accéder rapidement		. . . . . ,qI
. . . . . à un sous-ensemble		. . . . . pN
. . . . . . <b>sur lequel</b>	2 1	. . . . . . <b>pn</b>
. . . . . . ◦ <i>on</i>	-2 1	. . . . . . ◦ <i>N</i>
. . . . . . ◦ <i>fera</i>	-1 2	. . . . . . ◦ <i>V</i>
. . . . . . . <i>une décision statistique</i>		. . . . . . . <i>N</i>

relation antécédent - relative: fera sur sous-ensemble (accord masculin singulier)

linéarisation optimisée: poids(expansion de N) < poids(expansion de V) (0<1)

Cas particulier de "dont":

Si le sujet est un pronom atone, "dont" dépend du premier candidat possible, qui est toujours le verbe (sauf s'il y a un objet) car le sujet ne peut régir:

cette lampe **dont** *une lumière d'huile coulait*

relation antécédent - relative : *lumière* de lampe

mais: cette lampe **dont** *elle coulait*

relation antécédent - relative : *coulait* de lampe

Si la relative comporte un objet séquence nominale, la question se pose de savoir si le pronom relatif dépend du verbe ou de l'objet, comme pour un bloc prépositionnel nominal introduit par "de" et situé après l'objet. Voici un exemple où la valence complément de nom de l'objet est saturée, et où "dont" dépend du verbe:

. Fabuleuse est		. V
la manière		N
. . . <b>dont</b>	2 1	. . . <b>pn</b>
. ◦ <i>il</i>	-2 1	. ◦ <i>N</i>
. . ◦ <i>a forgé</i>	-1 5	. . ◦ <i>V</i>
. . . <i>la vie</i>		. . . <i>N</i>
. . . . <i>de milliers</i>		. . . . <i>pN</i>
. . . . . <i>de filles</i>		. . . . . <i>pN</i>
. . . . . <i>et de garçons</i>		. . . . . <i>cpN</i>

relation antécédent - relative: forgé de manière

linéarisation optimisée: poids(expansion de N) < poids(expansion de V) (0<4)

(voir ci-dessous le cas où "dont" dépend de l'objet en 6.5)

6.4. Pronom relatif dépendant du sujet : dont

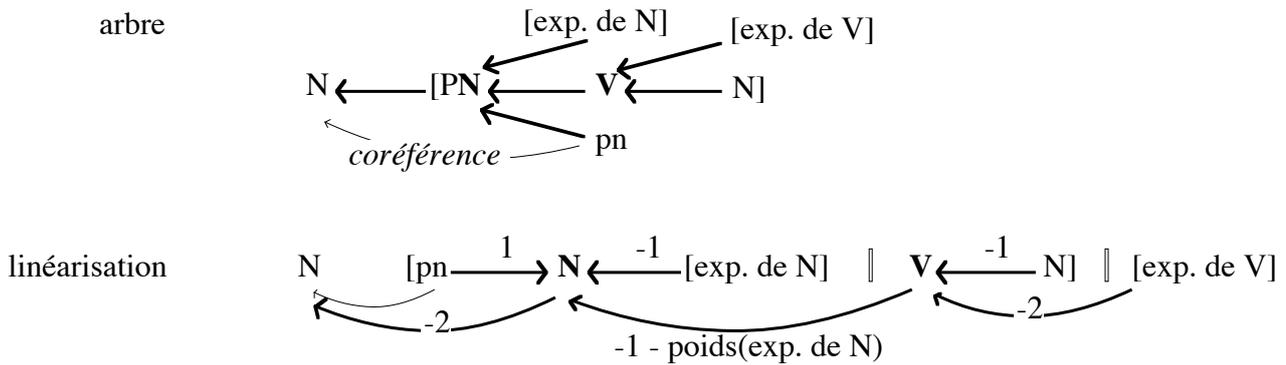


Figure 13: Linéarisation d'une relative avec pronom relatif dépendant du sujet

Une seule linéarisation est possible, car l'inversion sujet verbe impliquerait une dépendance de "dont" au verbe. "dont" dépend du **premier** candidat régissant (N ou V) possible qui lui est postérieur dans l'ordre linéaire:

L'emploi		N
. de l'ordinateur		. pN
. autorise		. V
. . des manipulations formelles		. . N
. . . sur des symboles		. . . pN
. . . . abstraits		. . . . ù
. . . . . <b>dont</b>	1 1	. . . <b>pn</b>
. . . . . ◦ <i>la signification réelle</i>	-2 1	. ◦ N
. . . . . ◦ <i>n'importe plus</i>	-1 1	. . ◦ V

relation antécédent - relative: signification de symboles

linéarisation optimisée: poids(expansion de N) = poids(expansion de V) (0=0)

Cas du pronom personnel sujet atone:

Quand N est un pronom personnel sujet atone, il ne peut régir, et en particulier pas de pronom relatif, donc "dont" ne dépend pas du sujet quand ce dernier est pronom personnel atone; il dépend alors le plus souvent de V, ou plus rarement de N objet.

6.5. Pronom relatif dépendant de l'objet: dont

La relative comporte un objet séquence nominale (et pas un pronom objet atone):

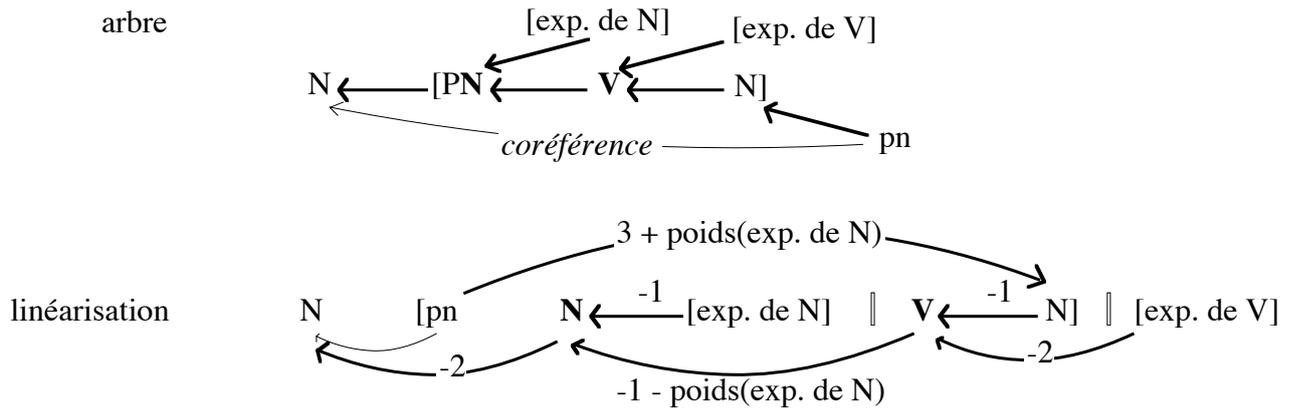


Figure 14: Linéarisation d'une relative avec pronom relatif dépendant de l'objet

On cherche à éviter cette structure, car très peu optimisée, et l'inversion objet verbe n'existe pas en français, car elle est ambiguë avec sujet verbe (quand on n'a pas d'accusatif, différent du nominatif). Cette structure comporte toujours un sujet pronom personnel atone, sinon le sujet serait le régissant du pronom relatif:

On		N
. dispose		. V
. . d'un ensemble		. . pN
. . . d'apprentissage		. . . pN
. . , c'est-à-dire d'un jeu		. . ,cpN
. . . de tels vecteurs		. . . pN
. . . . . <b>dont</b>	4 1	. . . . . <b>pn</b>
. . . . . <i>o on</i>	-2 1	. . . . . <i>o N</i>
. . . . . <i>o connaît</i>	-1 4	. . . . . <i>o V</i>
. . . . . <i>en plus</i>		. . . . . pN
. . . . . <i>la classe</i>		. . . . . N
. . . . . <i>d'appartenance</i>		. . . . . pN

relation antécédent - relative: classe (d'appartenance) de vecteurs  
(ou: connaît de vecteurs?)

linéarisation optimisée: poids(expansion de N) < poids(expansion de V) (0<3)

Si on considère que "classe d'appartenance" est lexicalisé comme "pomme de terre" (pas de déterminant), sa valence complément de nom n'est pas saturée et "classe" peut régir "dont".

### 6.6. Pronom relatif dépendant d'une séquence d'un bloc subordonné de la relative

Il arrive que le pronom relatif dépende non pas du sujet, du verbe ou de l'objet de la relative, mais d'une séquence qui dépend du sujet, du verbe ou de l'objet de la relative; en voici deux cas:

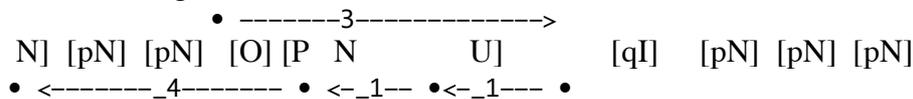
- pronom relatif objet d'un bloc subordonné de la relative

...		...
une technique		N
. de suivi		. pN
. . de contour		. . pN
. . . . <b>qu'</b>	3 1	. . . . <b>0</b>
. o <i>il</i>	-4 1	. o N
. . o <i>est facile</i>	-1 5	. . o U (verbale intransitive)
. . . <i>d'imaginer</i>		. . . qI
. . . . <i>dans des images</i>		. . . . pN
. . . . . <i>à deux niveaux</i>		. . . . . pN
. . . . . . <i>de gris</i>		. . . . . . pN

relation antécédent - relative: imaginer technique

linéarisation optimisée: poids(expansion de N) < poids(expansion de V) (0<4)

Examinons la segmentation en blocs:



[O], facette pronom du pronom relatif fait bloc à lui seul, ne fait pas partie du bloc de la relative [P N U] dont le verbe est intransitif, et est objet de I de [qI]; [O] est objet du **premier** (dans l'ordre linéaire) verbe dont la valence objet n'est pas saturée.

Le seul candidat régissant précédant le régissant est "est facile" verbale intransitive.

Autre exemple: *le bâtiment qu'il a souhaité que vous visitiez*  
relation antécédent - relative: *visitez bâtiment*

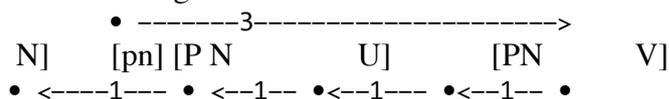
- pronom relatif dépendant du verbe d'une subordonnée de la relative

...		...
le rendez-vous		N
. . . . . <b>auquel</b>	4 1	. . . . . <b>pn</b>
. o <i>il</i>	-2 1	. o N
. . o <i>a souhaité</i>	-1 3	. . o V (verbale transitive)
. . . <i>que vous</i>		. . . PN
. . . . <i>soyez associé</i>		. . . . V

relation antécédent - relative: associé à rendez-vous (et pas: souhaité à rendez-vous)

linéarisation optimisée: poids(expansion de N) < poids(expansion de V) (0<3)

Examinons la segmentation en blocs:



[pn] facette pronom du pronom relatif fait bloc à lui seul, ne fait pas partie du bloc de la relative [P N V] dont le verbe ne peut le régir, et dépend de V de [PNV]; [pn] dépend de la **première** (dans l'ordre linéaire) séquence nominale ou verbale capable de le régir.

Chaque candidat régissant précédant le régissant a une "bonne raison" de ne pouvoir régir [pn]: les deux nominales sujets sont des pronoms personnels atones, et "souhaité" ne régir pas avec la préposition "à".

### 6.7. Quelques remarques globales sur la linéarisation des relatives

Nous avons proposé des concepts syntaxiques qui permettent d'expliquer l'ordre linéaire dans les relatives: à partir des concepts de segment primaire (la séquence), de dépendance mémorielle, et de linéarisation optimisée de l'arbre de dépendance sous des contraintes topologiques, métriques et mémorielles, nous avons tenté de montrer que cet ordre dépend non pas du verbe ou du sujet indépendamment, mais presque uniquement de la comparaison métrique des expansions du verbe et du sujet.

On remarque que ces deux expansions sont le plus souvent de poids très différents, et dans ce cas le choix de l'ordre linéaire est clair: vient en premier celui du sujet ou du verbe dont l'expansion est de poids le plus faible (souvent de poids nul); en particulier, quand le nœud verbal est linéarisé en premier, alors, le plus souvent, il n'a aucune expansion.

Quand les expansions ont des poids égaux, elles sont alors le plus souvent de poids nuls, et on observe alors les deux linéarisations.

### 7. La phrase définie comme arbre linéarisé

On pourrait maintenant définir la phrase comme un arbre de dépendances mémorielles linéarisé, un graphe connexe, donc comme une suite de séquences reliées. Un graphe connexe est un graphe dont tous les nœuds sont reliés, directement ou indirectement.

La coupure entre deux phrases est alors manifestée par l'absence de dépendance entre les deux phrases: deux arbres différents, deux graphes non connexes.

### 8. Conclusion

La dépendance mémorielle définie comme processus de mise en relation constitue un exemple d'étude des formes des textes non pas comme des objets statiques, mais comme des traces de processus humains (émettre, recevoir); et ces traces des processus dans les textes permettent de faire des hypothèses sur certaines propriétés de ces processus: ainsi le processus de mise en relation est aussi une hypothèse sur le processus humain de mise en relation.

*Adresse de l'auteur:*

GREYC URA 1526 Université de Caen

14032 Caen cedex

France

courrier électronique: Jacques.Vergne@info.unicaen.fr

NOTES

1) Le modèle informatisé sert à observer le corpus à travers le filtre des concepts, à déduire segments et relations à partir des concepts, et à confronter les concepts au corpus.

2) Voici quelques caractéristiques sur le corpus: cette recherche porte sur le français sur les deux aspects, étude linguistique "assistée par ordinateur" et analyse syntaxique automatique; une étude linguistique précise, avec expérimentations d'analyse automatique a été faite pour l'espagnol et pour l'anglais; des études de transposabilité des concepts ont été faites sur l'allemand, le polonais, le basque, le japonais (avec des locuteurs de chaque langue) et le latin.

Voici quelques caractéristiques quantitatives des corpus:

- corpus français: 2 textes informatifs scientifiques

l'avant-propos d'un ouvrage sur la reconnaissance des formes	140	phrases
3850 mots		

un article tiré d'une revue de biologie marine	107	phrases	3108
--	-----	---------	------

mots

total du corpus français:	247	phrases	6958	mots
---------------------------	-----	---------	------	------

- corpus anglais:

un article tiré de la même revue de biologie marine	134	phrases	2121	mots
---	-----	---------	------	------

- corpus espagnol:

un article tiré d'un hebdomadaire économique	223	phrases	4691	mots
--	-----	---------	------	------

3) En informatique, un **arbre** est un type particulier de graphe: un graphe est constitué de points (appelés **nœuds**) reliés par des segments orientés (appelés arcs). Dans un arbre, les arcs sont appelés **branches**, et il y a deux types particuliers de nœuds: la **racine**, à laquelle n'arrive aucune branche, et les **feuilles** d'où ne part aucune branche. De plus, en partant de la racine et en suivant les branches dans le sens de leur orientation, on arrive toujours à une feuille, sans jamais repasser deux fois par un même nœud. On a en somme une structure analogue à celle d'un arbre naturel. Par contre, la représentation des arbres peut se faire dans n'importe quelle direction: ainsi, dans cet article, nous représentons nos arbres avec la racine à gauche, et les feuilles à droite, pour pouvoir y placer des phrases écrites de gauche à droite.

4) Lisp signifie "list processor", c'est-à-dire langage spécialisé dans le traitement des listes. Dans ce langage de programmation, on code un arbre sous la forme d'une liste de listes, dans laquelle des parenthèses marquent lesdites singularités.

## RÉFÉRENCES

- Chomsky, Noam. 1969: *Structures syntaxiques*. Paris: Point Seuil, .
- Chomsky, Noam. 1971. *Aspects de la théorie syntaxique*. Paris: Seuil.
- Chomsky, Noam. 1980. *Essais sur la forme et le sens*. Paris: Seuil.
- Grevisse, Maurice. 1936, 1986. *Le bon usage*, Duculot.
- Lucas, Nadine. 1993a. Syntaxe du paragraphe dans les textes scientifiques en Japonais et en Français. In *Actes du colloque international: Parcours linguistiques de discours spécialisés*. Berne-Paris: éditions Moirand et alii., pages 249-261.
- Lucas, Nadine; Kikuko Nishina; Tomoyoshi Akiba; K.G. Suresh. 1993b. Discourse analysis of scientific textbooks in Japanese: a tool for producing automatic summaries. Tôkyô: Department of Computer Science, Tôkyô Institute of Technology.
- Lucas, Nadine. 1995. Le style scientifique en Japonais et en Français. In *Japon Pluriel, actes du 1<sup>er</sup> colloque de la Société française des études japonaises*. Arles: éditions Picquier, pages 393-402.
- Martinet, André. 1980. *Éléments de linguistique générale*. Paris: Armand Colin.
- Stevens, Peter. 1978. *Les formes dans la nature*. Paris: Seuil.
- Tesnière, Lucien. 1953. *Esquisse d'une syntaxe structurale*. Paris: Klincksieck.
- Tesnière, Lucien. 1982 (1<sup>ère</sup> édition: 1959). *Éléments de syntaxe structurale*. Paris: Klincksieck.
- Vergne, Jacques. 1995a. Les cadres théoriques des Traitements Automatiques des Langues syntaxiques: quelle adéquation linguistique et algorithmique? une étude et une alternative. In *Actes de "TALN'95" Conférence sur le Traitement Automatique du Langage Naturel*. Marseille, pp. 24-33.
- Vergne, Jacques. 1995b. *Esquisse d'une syntaxe des langues concrètes - Application à l'analyse syntaxique automatique*. Les cahiers du GREYC, université de Caen, numéro 11, 180 pages.

## SUMMARY

As starting point, we place ourselves in Lucien Tesnière's theoretical frame, while taking his definition of "connexion" as a process; but we take this definition at his word, while making it more precise and basing it on a process which uses the short term memory of the person who listens or reads; moreover, we do not take the word as linked segment, as Tesnière does, but a sequence of a few contiguous words: the core phrase (sections 1 et 2). In this way, we obtain a dependency tree which is quite different from Tesnière's "stemma" (section 3).

About matching between linear order and structural order, we abandon the concept of projection of the tree on the linear axis (the result of the projection is dependant on the drawing of the tree), and we make an hypothesis which uses concepts coming from algorithmics, concepts about how to enumerate nodes of any tree (possible enumerations depend on the

structure of the tree and not on its drawing): among all possible linearisations, the ones which are realised are the linearisations which minimise distance between two linked segments. We give a memory interpretation of this hypothesis: minimising distance between two linked segments minimises the memory effort (section 4).

Then we confront this hypothesis to a corpus of sentences in French where a verb has an object and another complement, and, about the linear order of these three segments, we compare the attested order with the possible order, and it drives us to explain why the object is or is not contiguous to the verb (section 5).

At last, we study the linearisation of the relative clause in French, and same concepts drive us to explain why the subject is postposed to the verb in some relative clauses (section 6).