

# Guide d'annotation syntaxique de ParCoLab, v2.0

Aleksandra Miletic, Dejan Stosic, Cécile Fabre  
CLLE-ERSS, Université Toulouse - Jean Jaurès

17 avril 2018

# Table des matières

<b>Remarques préliminaires</b>	<b>3</b>
<b>1 Annotation syntaxique et le TAL</b>	<b>4</b>
1.1 Analyse en constituants et analyse en dépendances . . . . .	4
1.2 Déterminer l'existence de la dépendance et son orientation . . . . .	7
<b>2 Règles d'annotation syntaxique</b>	<b>9</b>
2.1 Présentation globale du jeu d'étiquettes . . . . .	9
2.2 Élément racine . . . . .	13
2.2.1 Verbe principal . . . . .	13
2.2.2 Autres types de tête . . . . .	14
2.2.3 Caractère unique de la tête de phrase . . . . .	15
2.3 Dépendants du verbe . . . . .	16
2.3.1 Verbe auxiliaire . . . . .	17
2.3.2 Sujet . . . . .	17
2.3.3 Sujet logique . . . . .	18
2.3.4 Objet direct . . . . .	20
2.3.5 Objet indirect . . . . .	21
2.3.6 Prédicatif nominal . . . . .	22
2.3.7 Prédicatif adverbial . . . . .	24
2.3.8 Prédicatif complémentaire . . . . .	25
2.3.9 Prédicatif optionnel . . . . .	27
2.3.10 Dépendant sous forme d'un adverbe . . . . .	28
2.3.11 Dépendant sous forme d'un nom fléchi . . . . .	29
2.3.12 Dépendant sous forme d'une préposition . . . . .	29
2.3.13 Propositions participiales . . . . .	30
2.3.14 Prédicat complexe : verbe modal ou aspectuel introduisant un infinitif	30
2.3.15 Traitement des enchaînements des dépendants . . . . .	31
2.4 Dépendants du nom . . . . .	32
2.4.1 Dépendant sous forme d'un nom fléchi . . . . .	32
2.4.2 Dépendant sous forme de préposition . . . . .	32
2.4.3 Dépendant sous forme d'adjectif . . . . .	33
2.4.4 Apposition . . . . .	34
2.5 Dépendants de l'adjectif . . . . .	36
2.5.1 Dépendant sous forme d'un adverbe . . . . .	36
2.5.2 Dépendant sous forme d'un nom fléchi . . . . .	37
2.5.3 Dépendant prépositionnel . . . . .	38
2.5.4 Construction <i>sav</i> 'tout' + Adjectif . . . . .	38
2.6 Dépendants de l'adverbe . . . . .	39
2.6.1 Dépendant sous forme d'adverbe . . . . .	39
2.6.2 Dépendant sous forme de nom fléchi . . . . .	39

2.6.3	Dépendant sous forme de préposition . . . . .	40
2.7	Relations diverses . . . . .	40
2.7.1	Complément de préposition . . . . .	40
2.7.2	Complément de numéral . . . . .	41
2.7.3	Négation . . . . .	41
2.7.4	Interrogation . . . . .	42
2.7.5	Réflexif . . . . .	44
2.7.6	Éléments extra-prédicatifs . . . . .	45
2.7.7	Emphase . . . . .	45
2.7.8	Éléments polylexicaux . . . . .	48
2.7.9	Citations et emplois métalinguistiques . . . . .	52
2.7.10	Ponctuation . . . . .	53
2.8	Subordination . . . . .	53
2.8.1	Subordonnées adverbiales à subordonnant mono-fonctionnel . . . . .	53
2.8.2	Subordonnées complétives . . . . .	54
2.8.3	Subordonnées interrogatives indirectes . . . . .	57
2.8.4	Subordonnées relatives . . . . .	59
2.8.5	Subordonnées corrélatives . . . . .	61
2.8.6	Ambiguïté des subordonnées en <i>da</i> . . . . .	62
2.9	Discours indirect . . . . .	63
2.10	Coordination . . . . .	64
2.11	Juxtaposition . . . . .	67
2.12	Ellipse . . . . .	68
	<b>Index des exemples par étiquette</b>	<b>72</b>
	<b>Bibliographie</b>	<b>74</b>

# Remarques préliminaires

Ce guide est destiné aux annotateurs qui travaillent sur l’annotation syntaxique du volet serbe du corpus *ParCoLab*. Il s’organise de manière suivante : le chapitre 1 présente le cadre théorique dans lequel s’inscrit l’annotation syntaxique de *ParCoLab* et les règles d’annotation générales adoptées dans le cadre du projet, alors que le chapitre 2 contient le jeu d’étiquettes et les règles d’annotation concrètes relatives à chaque étiquette. À la fin du document, un index des exemples disponibles pour chaque étiquette syntaxique est proposé ([Index des exemples par étiquette](#)).

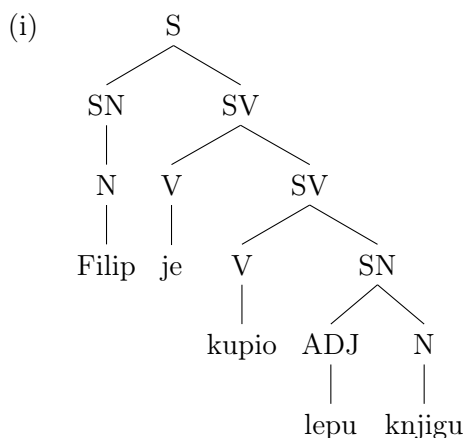
# 1. Annotation syntaxique et le TAL

Dans l'analyse syntaxique, aussi bien théorique qu'appliquée au TAL, deux approches principales existent : l'analyse en constituants et l'analyse en dépendances. Ce chapitre est dédié à une brève présentation des deux cadres, à l'argumentation du choix qui a été fait dans le cadre de ce projet, ainsi qu'à la présentation des notions de base de l'approche sélectionnée.

## 1.1 Analyse en constituants et analyse en dépendances

L'analyse en constituants aussi bien que l'analyse en dépendances ont donné naissance à de nombreuses théories syntaxiques. Parmi les représentant de la grammaire en constituants, on trouve *Government and Binding Theory* (Chomsky, 1993, 1982), *Generalized phrase structure grammar* (GPSG) (Gazdar et al., 1985) et *Head-driven phrase structure grammar* (HPSG) (Pollard and Sag, 1994). Du côté de la grammaire en dépendances on peut citer les théories suivantes : *Word Grammar* (WG) (Hudson, 1984), *Functional Generative Description* (FGD) (Sgall et al., 1986), *Dependency Unification Grammar* (DUG) (Hellwig, 1986), *Meaning-Text Theory* (MTT) (Mel'čuk, 1988), ou *Functional Dependency Grammar* (FDG) (Tapanainen and Järvinen, 1997). Le principal critère de distinction entre ces deux approches théoriques est leur vision de la structure syntaxique et des représentations syntaxiques dont elles se servent. Comme les deux visions entraînent des implications importantes aussi bien pour la linguistique que pour le TAL, nous les présentons dans la suite.

L'analyse en constituants est en général plus répandue et mieux connue que l'analyse en dépendances. Dans ce cadre théorique, on considère que l'unité de base de la structure syntaxique sont les constituants (syntagmes, groupes). Analyser une phrase consiste à la décomposer en constituants niveau par niveau, jusqu'à arriver aux mots mêmes. Les arbres syntaxiques qui en résultent se présentent comme dans l'exemple (i).



Nous pouvons remarquer plusieurs caractéristiques de l'arbre :

1. l'arbre représente une structure à **plusieurs niveaux** :
2. la **racine** de l'arbre (le nœud supérieur) est un nœud représentant **la phrase** ;

- il n'y a **pas de marquage explicite des fonctions syntaxiques** : on indique les catégories morphosyntaxiques des syntagmes, alors que la fonction syntaxique est dérivée de la structure de l'arbre.

L'application de cette approche dans le cadre du TAL consiste souvent à indiquer la structure en constituants de la phrase en ajoutant des crochets délimitant les syntagmes. L'arbre montré ci-dessus serait donc représenté sous la forme suivante :

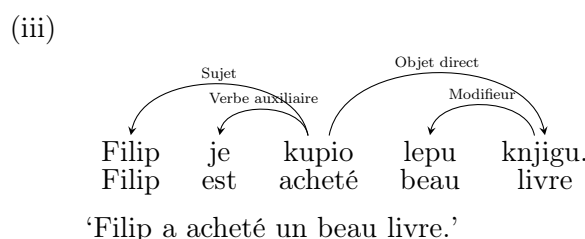
[S [SN [N Filip ] ] [SV [V je ] [SV [V kupio ] [SN [ADJ lepu ] [N knjigu ] ] ] ] ]

Cette représentation (ainsi que les fondements théoriques de l'analyse en constituants) pose cependant un problème important : l'analyse est basée sur l'ordre linéaire des mots dans la phrase. Ceci pose des difficultés dans le traitement des langues à ordre des mots flexible, tel le serbe, qui permet par ailleurs des constituants discontinus. Il est donc tout à fait possible d'avoir des phrases comme *Lepu je knjigu Filip kupio*, présentée dans l'exemple (ii).

- (ii) Lep-u                      je                      knjig-u                      Filip                      kupi-o  
beau-ACC.SG.F VAUX-3SG.PRES livre-ACC.SG Filip.NOM.SG acheté  
'C'est un beau livre que Filip a acheté'

Ici, la focalisation spécifique de la phrase entraîne un ré-ordonnement des mots, et la décomposition de la phrase en constituants devient ardue : le sujet *Filip* se trouve au milieu du syntagme verbal, l'auxiliaire *je* est séparé du verbe principal *kupio* par la tête de l'objet direct *knjigu* et le sujet *Filip*, et par ailleurs, l'adjectif modifieur de l'objet direct *lepu* est séparé de sa tête *knjigu* par l'auxiliaire. Il est donc impossible d'effectuer une analyse en constituants en respectant l'ordre linéaire des mots dans la phrase. C'est précisément pour cette raison que l'analyse syntaxique des langues comme le serbe se fait de préférence dans le cadre de l'analyse en dépendances.

Cette deuxième approche considère que la structure syntaxique d'une phrase est un ensemble de relations de dépendance qui s'établissent entre les mots individuels d'une phrase : chaque mot de la phrase a un gouverneur et peut à son tour en gouverner d'autres. Les arbres syntaxiques de cette approche se présentent sous la forme indiquée dans l'exemple (iii).



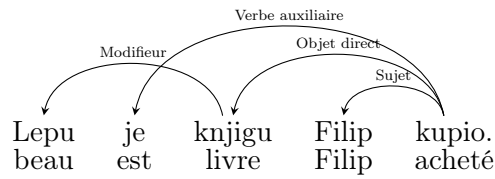
Les principales différences par rapport à l'arbre en constituants sont les suivantes :

- l'arbre a **un seul niveau** : les relations s'établissent directement entre les formes fléchies ;
- la racine** de l'arbre est représentée par **le verbe principal** de la phrase ;
- les **fonctions syntaxiques** sont **annotées explicitement**, mais les syntagmes ne le sont pas<sup>1</sup>.

L'avantage de cette approche est que l'ordre linéaire des mots dans la phrase a peu d'impact sur l'analyse. Ainsi, les mêmes relations présentes dans l'exemple (iii) se mettent en place dans l'exemple (iv), malgré le fait que les mots sont distribués différemment.

1. Ils peuvent néanmoins être récupérés en considérant les différents sous-arbres.

(iv)



‘C’est un beau livre que Filip a acheté.’

En fonction du courant de la grammaire en dépendances dans lequel on s’inscrit, les relations peuvent s’établir entre différents types d’unités : formes fléchies, lemmes, ou unités polylexicales. Également, les relations peuvent exprimer différents types de dépendances : il peut s’agir des rôles thématiques ou des relations syntaxiques (souvent dites “de surface”). Dans le cadre du projet *ParCoLab*, l’annotation portera sur des relations syntaxiques de surface, et les relations s’établiront entre les formes fléchies individuelles (les tokens) constituant la phrase.

En ce qui concerne la forme de l’arbre syntaxique créé par l’analyse en dépendances, elle est soumise à plusieurs contraintes :

1. **complétude** : tous les mots de la phrase doivent être inclus dans l’arbre (chaque mot doit avoir un gouverneur) ;
2. **acyclicité** : un mot ne peut pas gouverner son gouverneur, ni un mot hiérarchiquement supérieur à son gouverneur (il ne peut pas y avoir des cycles fermés dans l’arbre) ;
3. **nature unique du gouverneur** : un mot ne peut avoir qu’un seul gouverneur (un mot peut être la cible d’une seule dépendance).

Dans certains cadres applicatifs en TAL, on impose une dernière contrainte aux arbres créés dans l’annotation : celle de la **projectivité**. Un arbre syntaxique est considéré comme non projectif si au moins une dépendance qui le constitue est elle-même non projective. Une dépendance entre un gouverneur G et un dépendant D est considérée comme non projective s’il existe au moins un token entre G et D dans l’ordre linéaire de la phrase qui n’est pas dominé par G (autrement dit, que G n’est ni son gouverneur immédiat, ni son ancêtre). Si l’on revient encore une fois à l’exemple (iv), on remarque que la relation entre le gouverneur *knjigu* et le dépendant *lepu* est non projective, vu que l’auxiliaire *je*, situé entre ces deux formes, n’est pas dominé par *knjigu*. En revanche, la dépendance entre le verbe principal *kupio* et l’auxiliaire *je* est projective, car les deux tokens intervenants *knjigu* et *Filip* sont gouvernés par *kupio*. La non-projectivité d’une relation (et, par conséquent, d’un arbre) se traduit typiquement par le croisement des arcs désignant les dépendances dans la représentation graphique de l’arbre syntaxique.

Vu que le traitement de ce type de relations reste encore problématique dans le cadre du parsing, certains projets (et certains outils de parsing) exigent que les arbres analysés soient projectifs et imposent un traitement projectif artificiel aux constructions syntaxiques non-projectives par nature. Or, la non-projectivité est le reflet de la nature discontinue des constituants dans une phrase. Le taux de non-projectivité dans les corpus de différentes langues est souvent utilisée comme un indicateur du degré de liberté de l’ordre des constituants. Alors que le taux des relations non-projectives est relativement bas pour toutes les langues (en général, 1%-2%), le pourcentage d’arbres syntaxiques contenant au moins une relation non-projective peut atteindre (voire dépasser) 20% dans les langues à ordre de constituants flexible telles les langues slaves. Ceci signifie qu’un schéma d’annotation strictement projectif appliqué à une telle langue produirait des analyses fausses pour une phrase sur cinq dans le corpus. Par conséquent, dans le cadre de ce projet, la non-projectivité est autorisée dans la phase de l’annotation manuelle. Ceci permettra d’estimer le taux de non-projectivité

en serbe, et également de mettre à l'épreuve les méthodes de parsing proposant une solution pour ce type de relation.

## 1.2 Déterminer l'existence de la dépendance et son orientation

L'une des questions centrales dans l'analyse en dépendances est de savoir déterminer si deux formes sont reliées par une relation ou non. Pour ce faire, la Théorie Sens-Texte (Mel'čuk, 1988) propose les critères suivants :

- le critère de linéarité : la position dans la phrase de l'une des deux formes est déterminée par rapport à la position de l'autre si une relation de dépendance existe entre elles, et
- le critère d'unité prosodique : deux formes liées par une dépendance font une unité prosodique, ou bien l'une des formes peut être liée prosodiquement avec une unité prosodique dont l'autre forme est la tête. (Mel'čuk, 1988, pp. 129-132).

Il existe également plusieurs critères qui permettent de déterminer laquelle entre les deux formes reliées par une dépendance est le gouverneur, et laquelle est le dépendant. Certains d'entre eux sont donnés dans la suite. Ils font appel aux notions de gouverneur, dépendant et construction, cette dernière désignant la construction syntaxique créée par le gouverneur et le dépendant.

1. C'est le **gouverneur** qui **détermine la catégorie distributionnelle de la construction** et **peut** souvent **la remplacer**.

Par exemple, si l'on reprend la phrase *Filip je kupio lepu knjigu* et qu'on analyse la construction *lepu knjigu*, nous constatons que la construction occupe la fonction de l'objet direct dans la phrase, qui est une fonction typique du nom. Par conséquent, nous pouvons conclure que c'est *knjigu* qui est le gouverneur, et *lepu* le dépendant. Par ailleurs, on peut également remplacer la construction par le gouverneur sans compromettre la gramaticalité de la phrase : *Filip je kupio knjigu*.

2. Le **dépendant** d'une relation **peut être optionnel** (il peut être omis de la phrase), mais le **gouverneur est obligatoire** dans la phrase.

Par exemple, en considérant encore la construction *lepu knjigu*, on peut omettre le modifieur *lepu* de la phrase *Filip je kupio lepu knjigu* et obtenir une phrase grammaticale *Filip je kupio knjigu*, mais ceci n'est pas le cas du gouverneur *knjigu* : ??*Filip je kupio lepu*.

3. La **réalisation morphosyntaxique du dépendant est déterminée par le gouverneur** (à travers les règles d'accord ou de rectification).

La forme *lepu*, qui correspond à l'accusatif singulier féminin de l'adjectif *lep* 'beau', est déterminée par le fait que son gouverneur, la forme *knjigu* est un nom féminin, à l'accusatif singulier. Le cas du nom est déterminé à son tour par son gouverneur *kupio* (participe passé du verbe *kupiti* 'acheter'), dont la structure argumentale exige un objet direct à l'accusatif.

4. La **position du dépendant** dans la phrase est **déterminée par rapport au gouverneur**.

Ce critère est reflété, par exemple, dans le fait que l'adjectif *lepu* se trouve à gauche du nom *knjigu*, ce qui est la position de préférence des modifieurs adjectivaux d'un nom. En revanche, ce critère est beaucoup plus pertinent pour des langues à ordre de constituants rigide que pour le serbe.

On remarque que ces règles sont très générales et qu'elles ne font pas mention des relations syntaxiques spécifiques. Effectivement, les règles citées ci-dessus sont indépendantes de langue et servent d'outil de base dans l'analyse en dépendances de toute langue. En revanche, chaque



langue individuelle dispose d'un inventaire de relations syntaxiques de surface qui lui sont spécifiques. Certaines langues (cf. le russe, le tchèque) bénéficient de travaux théoriques décrivant en détail leur fonctionnement syntaxique dans le cadre de l'analyse en dépendants. Il est donc possible de se baser sur ces ouvrages pour dresser l'inventaire des relations syntaxiques qui seront encodées en corpus. Or, la syntaxe théorique du serbe repose traditionnellement sur l'analyse en constituants. Il n'y a donc pas de formalisme prêt à être exploité dans le cadre d'une application en corpus. Par conséquent, les relations syntaxiques proposées dans le jeu d'étiquettes décrit dans ce document ont été définies spécifiquement pour ce projet, avant le démarrage du travail sur le corpus. Elles ont été mises à l'épreuve dans le cadre d'une annotation manuelle initiale, mais le jeu d'étiquettes reste certainement perfectible. Les fonctions retenues et leurs domaines d'application respectifs sont présentés dans la suite de ce document.

## 2. Règles d’annotation syntaxique

### Avertissement

Afin de pouvoir exploiter ce guide d’annotation, il est nécessaire de maîtriser les notions suivantes :

- dépendance syntaxique ;
- arbre de dépendances syntaxiques ;
- racine de l’arbre de dépendances syntaxiques ;
- dépendances projectives et non-projectives ;
- structure argumentale d’un verbe.

Ce chapitre est dédié à la présentation du jeu d’étiquettes syntaxiques<sup>1</sup> et du schéma d’annotation associé de *ParCoLab*. Nous proposons d’abord une vue d’ensemble du jeu d’étiquettes pour présenter ensuite en détail les règles d’application de chaque étiquette.

### 2.1 Présentation globale du jeu d’étiquettes

Le jeu d’étiquettes que nous présenterons dans la suite a été construit sur base de plusieurs principes. D’abord, étant donné la nature parallèle de *ParCoLab*, le corpus est destiné à être utilisé par la communauté serbophone aussi bien que francophone. Par conséquent, le noyau des relations retenues coïncide avec les fonctions syntaxiques traditionnellement reconnues dans la grammaire serbe (cf. sujet grammatical, sujet logique, objet direct, objet indirect, prédicatifs nominal et adverbial, etc.). En revanche, ces fonctions traditionnelles ont été soumises à une analyse basée sur des critères formels de surface dans l’objectif de garantir que la distinction des fonctions syntaxiques encodées en corpus se base sur des propriétés accessibles à un parser. Il s’agit notamment des critères suivants :

- réalisations morphosyntaxiques possibles du gouverneur et du dépendant (POS, lemme) ;
- flexion du dépendant et du gouverneur si des traits morphosyntaxiques spécifiques sont liés à la fonction en question (par exemple, un cas ou une forme verbale spécifique) ;
- pour les dépendants verbaux considérés comme objets, la possibilité de pronominalisation avec un clitique et le type de clitique utilisé ;
- accord : les constituants et les traits concernés ;
- règles de linéarisation :
  - ordre canonique gouverneur - dépendant ;
  - caractère flexible ou rigide de l’ordre gouverneur - dépendant ;

---

1. Le terme *jeu d’étiquettes* désigne l’ensemble des étiquettes ou des labels utilisés à un niveau d’annotation dans un corpus. Il existe donc des jeux d’étiquettes morphosyntaxiques, qui expriment typiquement les parties du discours, des jeux d’étiquettes syntaxiques, exprimant des relations syntaxiques, des jeux d’étiquettes discursifs, servant à l’analyse du discours, etc.

- caractère obligatoire ou non de l’adjacence du gouverneur et du dépendant (possibilité que le dépendant soit séparé du gouverneur par un autre constituant) ;
- non-projectivité possible ou non de la relation (possibilité que le dépendant soit séparé du gouverneur par un mot n’ayant pas le même gouverneur, ce qui entraîne un croisement des arcs dans l’arbre syntaxique).

Pour certains phénomènes plus complexes, moins fréquemment abordés dans les travaux théoriques, des traitements déjà existants dans d’autres corpus ont été repris ou adaptés.

Cette démarche a abouti à un jeu de 48 relations syntaxiques de base, et un traitement spécifique pour l’ellipse, qui sera détaillé dans la suite. Le jeu est présenté dans le tableau 2.1. Chaque étiquette est accompagnée d’une brève définition de la relation et d’un exemple. Dans l’exemple, le **gouverneur** est indiqué en gras, et le dépendant en souligné.

**NB**

Avant de continuer, rappelons encore que dans le cadre de la syntaxe en dépendances, les relations s’établissent entre les mots individuels. Par conséquent, dans les exemples donnés ci-dessous, le dépendant correspond toujours à un seul mot. Dans le cadre de la syntaxe en constituant, ce mot correspondrait à la tête du constituant exerçant la fonction syntaxique en question. C’est pourquoi, dans la première ligne du tableau, nous annotons seulement la forme predsednik ‘président’ en tant qu’apposition, et non pas tout le syntagme predsednik Francuske ‘président de la France’.

TABLE 2.1: Jeu d’étiquettes syntaxiques ParCoLab

Etiquette	Définition	Exemple
Ap	apposition	<i>Oland</i> , <u>predsednik Francuske</u> ‘ <b>Hollande</b> , <u>président</u> de la France’
AuxV	verbe auxiliaire dans une forme verbale composée	<i>Filip je <u>stigao</u></i> ‘Filip <u>est</u> arrivé’
Cit	élément métalinguistique	<i>misao</i> « <i>budan sam</i> » ‘ <b>idée</b> « je suis réveillé »’
ComplNum	complément d’un numéral sous forme du paucal <sup>2</sup>	<i>dva čoveka</i> ‘ <b>deux</b> hommes’
ComplPrep	complément de préposition	<i>kolač od višanja</i> ‘gâteau <b>aux</b> cerises’
ConjCoord <sup>3</sup>	relation entre le coordonné précédant immédiatement la conjonction de coordination et la conjonction elle-même	<i>Filip je <b>vredan</b> i pametan</i> ‘Filip est <b>travailleur</b> <u>et</u> intelligent’
Coord	relation entre la conjonction de coordination et le dernier coordonné	<i>Filip je vredan <b>i</b> pametan</i> ‘Filip est travailleur <b>et</b> <u>intelligent</u> ’
Correl	relation entre deux éléments d’une structure corrélatrice	<i>tako vruće da peče</i> ‘ <b>si</b> chaud <u>que</u> ça brûle’

2. Le paucal est une forme casuelle spécifique, imposée aux mots qui se déclinent par les numéraux *dva* ‘deux’, *tri* ‘trois’ et *četiri* ‘quatre’. Il s’agit d’une trace de l’ancien dual.

TABLE 2.1: Jeu d'étiquettes syntaxiques ParCoLab

Etiquette	Définition	Exemple
DepAdjAdj	dépendant d'un adjectif sous forme d'un autre adjectif	<i>sav zadihan</i> 'tout <u>essoufflé</u> '
DepAdjAdv	dépendant d'un adjectif sous forme d'un adverbe	<i>Jedva vidljiv i sasvim tih</i> 'A peine <b>visible</b> et <u>complètement silencieux</u> '
DepAdjCas	dépendant d'un adjectif sous forme d'un nom fléchi	<i>sličan ocu</i> 'semblable père.DAT'
DepAdjPrep	dépendant d'un adjectif sous forme d'un groupe prépositionnel	<i>On je zaljubljen u Milicu</i> 'Il est <b>amoureux</b> <u>de</u> Milica'
DepAdvAdv	dépendant d'un adverbe sous forme d'un adverbe <sup>4</sup>	<i>još dugo</i> 'encore <b>longtemps</b> '
DepAdvCas	dépendant d'un adverbe sous forme d'un nom fléchi	<i>mnogo ljudi</i> 'beaucoup gens.GEN'
DepAdvPrep	dépendant d'un adverbe sous forme d'une préposition	<i>više od njega</i> ' <b>plus</b> <u>que</u> lui'
DepEx_	ellipse : préfixe ajouté à l'étiquette de l'élément dont le gouverneur est élidé <sup>5</sup>	(cf. section 2.12)
DepNAdj	dépendant de nom sous forme d'un adjectif	<i>dugo pismo</i> 'longue <b>lettre</b> '
DepNCas	dépendant de nom sous forme d'un nom fléchi	<i>zrak sunca</i> ' <b>rayon</b> soleil.GEN'
DepNPrep	dépendant de nom sous forme d'un groupe prépositionnel	<i>pismo od Filipa</i> ' <b>lettre</b> <u>de</u> Filip'
DepVAdv	dépendant d'un verbe sous forme d'un adverbe	<i>Filip lepo peva</i> 'Filip <b>chante</b> bien'
DepVCas	dépendant d'un verbe sous forme d'un nom fléchi et qui n'est pas un ObjDir, ObjIndir ou prédicatif	<i>Plaši se grmljavine</i> 'Il a <b>peur</b> tonnerre.GEN'
DepVInf	dépendant infinitif d'un verbe	<i>prestat i plakati</i> ' <b>arrêter</b> pleurer'
DepVPart	dépendant d'un verbe introduit par un participe présent ou passé	<i>Vratio se pevajući</i> 'Il est <b>rentré</b> <u>en chantant</u> '
DepVPrep	dépendant d'un verbe sous forme d'un groupe prépositionnel qui n'est pas un ObjIndir ou un prédicatif	<i>Učestvovao je u organizaciji</i> 'Il a <b>participé</b> <u>dans</u> l'organisation'
Emph	dépendant de la racine de la proposition exprimant l'emphasis, privé de fonction syntaxique	<i>To dolazi zima</i> lit. 'Ça <b>vient</b> hiver'
ExtraPred	dépendant de la racine de la proposition sous forme d'un adverbe extra-prédicatif	<i>On zapravo kasni</i> 'Il est <b>en fait</b> en retard'
Interrog	dépendant de la racine de la proposition sous forme d'un marqueur d'interrogation	<i>Dolazi li Filip</i> ? 'Est-ce que Filip <b>vient</b> ?'

4. Il s'agit typiquement d'intensificateurs.

5. Le traitement de l'ellipse a été repris de Prague Dependency Treebank (Hajič et al., 1999).

TABLE 2.1: Jeu d'étiquettes syntaxiques ParCoLab

Etiquette	Définition	Exemple
Juxt	juxtaposition de deux éléments de haut niveau où aucune autre relation ne s'applique	<i>Posledice su se brzo <b>osetile</b> : <u>vrto</u>glavica i <u>muč</u>nina</i> 'Les conséquences se sont vite fait <b>sentir</b> : vertige et nausée'
Neg	négation (verbale ou nominale)	<i><u>Ne</u> <b>dolazi</b></i> 'Il <u>ne</u> <b>vient pas</b> '
ObjDir	objet direct, à l'accusatif ou au génitif	<i>Filip <b>jede</b> jabuku</i> 'Filip <b>mange</b> pomme.ACC'
ObjIndirCas	objet indirect au datif	<i>Filip <b>daje</b> jabuku <u>Ani</u></i> 'Filip <b>donne</b> pomme.ACC <u>Ana</u> .DAT'
ObjIndirPrep	objet indirect réalisé comme <i>o</i> 'de' + N_locatif	<i>Filip <b>misli</b> <u>o</u> putovanju</i> 'Filip <b>pense de</b> voyage.LOC'
Polylex	relation réunissant les éléments d'une locution prépositionnelle ou adverbiale ou d'une conjonction complexe	<i>Dolazi <b>zato što</b> mora</i> 'Il vient <b>parce qu'</b> il est obligé'
PredCompletive	relation entre le subordonnant et le prédicat de la complétive	<i>Zna <b>da</b> <u>dolazim</u></i> 'Il sait <b>que</b> je <u>viens</u> '
PredPercont	relation entre le prédicat de la principale et le prédicat de la percontative	<i><b>Pitao</b> je zašto <u>dolaze</u></i> 'Il a <b>demandé</b> pourquoi ils <u>venaient</u> '
PredRap	relation entre le verbe introductif et le verbe principal du discours rapporté	« <i><u>Dolazim</u></i> », <b>kaže</b> . « <i>J'<u>arrive</u></i> », <b>dit-il</b> .
PredRel	relation entre l'antécédent d'une relative et son prédicat	<i><u>čovek koji je došao</u></i> 'l'homme qui est <u>venu</u> '
PredSub	relation entre le subordonnant et le prédicat de la subordonnée	<i>Dolazi <b>kad</b> <u>završi</u></i> 'Il vient <b>quand</b> il <u>fini</u> '
PredicAdv	prédicatif adverbial : complément adverbial du verbe <i>biti</i> 'être'	<i>Filip <b>je u</b> Beogradu</i> 'Filip <b>est à</b> Belgrade'
PredicComplObj	prédicatif complémentaire lié à l'objet direct : complément nominal, adjectival ou prépositionnel d'un verbe obligatoirement attributif autre que le verbe <i>biti</i> 'être'	<i><b>Proglasili</b> su ga <u>kraljem</u></i> 'Ils l'ont <b>proclamé</b> <u>roi</u> .INS'
PredicComplSuj	prédicatif complémentaire lié au sujet : complément nominal, adjectival ou prépositionnel d'un verbe obligatoirement attributif autre que le verbe <i>biti</i> 'être'	<i><b>Proglasio</b> se <u>kraljem</u></i> 'Il s'est <b>proclamé</b> <u>roi</u> .INS'
PredicNom	prédicatif nominal : complément nominal du verbe <i>biti</i> 'être'	<i>Filip <b>je</b> <u>profesor</u></i> 'Filip <b>est</b> professeur'
PredicOpt	prédicatif optionnel : complément nominal ou adjectival d'un verbe optionnellement attributif	<i>Filip se <b>vratio</b> <u>umoran</u></i> 'Filip est <b>rentré</b> <u>fatigué</u> '
Ref	relie le verbe au pronom réflexif	<i><b>Osvežio</b> se</i> 'Il s'est <b>rafraîchi</b> '
Root	relie la racine externe et la tête de la phrase	<i><b>ROOT</b> <u>Dolazi</u> sutra</i> ' <b>ROOT</b> Il <u>vient</u> demain'

TABLE 2.1: Jeu d’étiquettes syntaxiques ParCoLab

Etiquette	Définition	Exemple
Sub	relation entre le verbe principal et le subordonnant introduisant une proposition subordonnée	<i><b>Dolazi</b> <u>kad</u> završi</i> ‘Il <b>vient</b> <u>quand</u> il finit’
Suj	sujet grammatical, exprimé par le nominatif	<i><u>Filip</u> je stigao</i> ‘ <u>Filip</u> est <b>arrivé</b> ’
SujLog	sujet logique, exprimé par le datif, génitif ou accusatif	<i><u>Filipu</u> je dosadno</i> ‘ <u>Filip</u> s’ <b>ennuie</b> ’ (lit. ‘Filip.DAT est ennuyeux’)
Ponct	relie la ponctuation au premier token précédent qui n’en est pas une	<i><b>Razočaran</b> , vratio se <b>kući</b> ,</i> ‘ <b>Déçu</b> , il est rentré à la <b>maison</b> ,’

Les modifications les plus importantes par rapport à la tradition grammaticale serbe concernent le traitement des dépendants du nom et des dépendants du verbe autre que les objets et les prédicatifs. Comme la distinction entre les ajouts et les arguments est souvent basée sur des critères sémantiques plutôt que syntaxiques, elle est difficile à opérer pour un annotateur humain, et d’autant plus pour un parser. Par conséquent, nous adoptons une autre approche pour le traitement de ces éléments : nous mettons en place une série d’étiquettes sous-spécifiées, qui commencent par **Dep**, désignant le terme neutre de dépendant, suivi des indications de la nature morphosyntaxique du gouverneur et du dépendant. Ainsi, l’étiquette **DepVPrep** indique un dépendant d’un verbe sous forme d’une préposition (qui n’est pas un prédicatif ou un objet indirect casuel), tandis que **DepNCas** désigne un dépendant d’un nom sous forme d’un nom fléchi (cf. *padežni atribut*). Par analogie, le même type d’étiquettes a été mis en place pour les dépendants d’un adjectif et d’un adverbe. Chacune de ces étiquettes est présentée en détail dans la suite de ce document (cf. sections 2.3.10, 2.3.11, 2.3.12, 2.3.13, 2.3.14, 2.4.1, 2.4.2, 2.4.3, 2.5.1, 2.5.2, 2.5.3, 2.5.4, 2.6.1, 2.6.2 et 2.6.3).

## 2.2 Élément racine

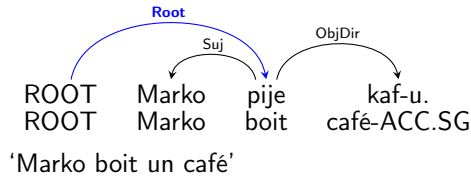
Cette section décrit la relation syntaxique **Root**, qui s’établit entre la racine externe de la phrase et la phrase elle-même. En effet, dans la section 1.1, nous avons mentionné que l’une des contraintes que doit respecter un arbre syntaxique réalisé dans le cadre de l’analyse en dépendances était la complétude : chaque mot de la phrase doit avoir un gouverneur. Nous avons également vu que la racine de la phrase était le verbe principal. Se trouvant au sommet de l’arbre, il n’était gouverné par aucun autre mot de la phrase. Pour contraindre la complétude de l’arbre et permettre au verbe principal aussi d’avoir un gouverneur, on introduit un nœud artificiel, représentant la racine “externe”. Dans le cadre de ce projet, ce nœud se trouve à gauche de la phrase et il est marqué comme *ROOT*. Selon notre schéma d’annotation, *ROOT* a un descendant unique dans la phrase. La liste des cas de figure possibles est donnée dans la suite.

### 2.2.1 Verbe principal

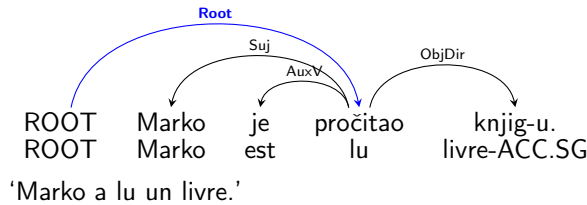
Le dépendant prototypique de la racine externe est le **verbe principal** de la phrase.

3. Le traitement de la coordination sera présenté en détail dans la section 2.10.

(1)



(2)

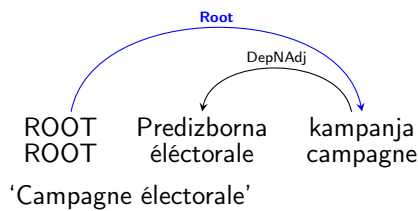


## 2.2.2 Autres types de tête

### Phrases averbales

Une phrase peut ne pas contenir de verbe fini (par exemple, en cas d'ellipse ou s'il s'agit d'un titre). Dans ce cas, on considère comme descendant de la racine externe la **tête du segment averbal**.

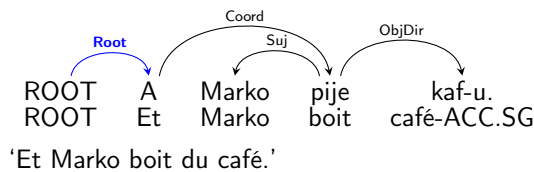
(3)



### Phrases introduites par un connecteur

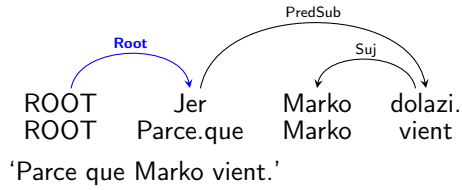
Dans le cas où une phrase est introduite par une conjonction en position initiale (ayant donc une fonction trans-phrastique, ou, plus précisément, le rôle d'un connecteur discursif), c'est la **conjonction** qui est annotée comme descendant de la racine externe. La conjonction gouverne, quant à elle, la tête de la proposition. La relation qui relie la conjonction et le verbe dépend de la nature de la conjonction : dans le cas des conjonctions de coordination (exemple 4), il s'agit de la relation **Coord** (cf. section 2.10), alors que pour les conjonctions de subordination (exemple 5) c'est **PredSub** (cf. section 2.8).

(4)

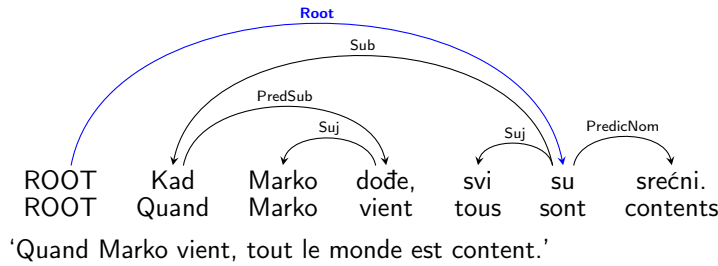


Il ne faut pas confondre ce cas de figure avec celui où la subordonnée est simplement antéposée à la principale, cf. l'exemple 6 :

(5)



(6)



Malgré le fait qu'ici aussi la phrase commence par une conjonction de subordination, la phrase contient également la proposition principale. La racine correspond donc au verbe principal de la proposition principale, qui gouverne à son tour la subordonnée.

### 2.2.3 Caractère unique de la tête de phrase

Dans le cadre du projet *ParCoLab*, nous considérons qu'il n'existe qu'un descendant de la racine externe dans une phrase. Autrement dit, nous adoptons le principe de la racine unique. Ceci est notamment important dans le traitement des cas de figure suivants.

#### Rattachement des modifieurs phrastiques

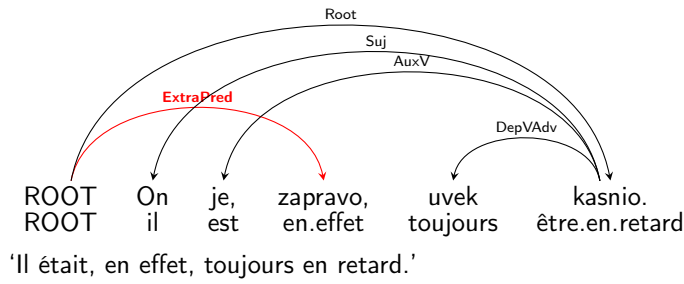
La linguistique théorique considère que certains éléments de la phrase se trouvent au même niveau de la structure syntaxique que le contenu propositionnel lui-même : il s'agit notamment des adverbes phrastiques. Logiquement, ces éléments devraient se positionner au même niveau de l'arbre syntaxique que la tête de la proposition qu'ils modifient. Autrement dit, ils devraient être considérés comme descendants de la racine externe. Or, dans le cadre de ce travail, ces éléments ne sont pas rattachés à la racine externe, mais à la tête interne de la proposition (typiquement au verbe principal, cf. 2.7.6). Ceci est fait dans un effort d'éviter la sur-production des relations non-projectives dans l'arbre syntaxique : les modifieurs phrastiques pouvant intervenir au milieu de la phrase aussi bien qu'en début, on risquerait d'avoir des arcs qui se croisent dans la représentation de la structure syntaxique, comme dans l'exemple 7.

Pour éviter cette situation, nous adoptons le traitement dans lequel le modifieur phrastique est gouverné par le verbe principal de la proposition (cf. exemple 8).

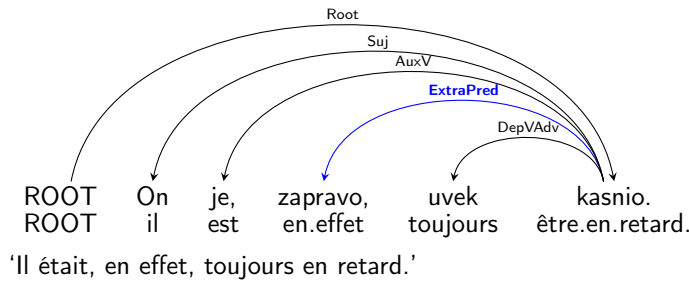
Le point problématique qui se pose ici est que l'extra-prédicatif paraît modifier le verbe au même titre qu'un ajout adverbial classique, plutôt que de modifier la phrase. Pour remédier à ce problème, nous introduisons une étiquette spécialisée **ExtraPred** pour les éléments extra-prédicatifs, ce qui permet d'identifier facilement ce type de dépendant en corpus (cf. section 2.7.6).



(7)



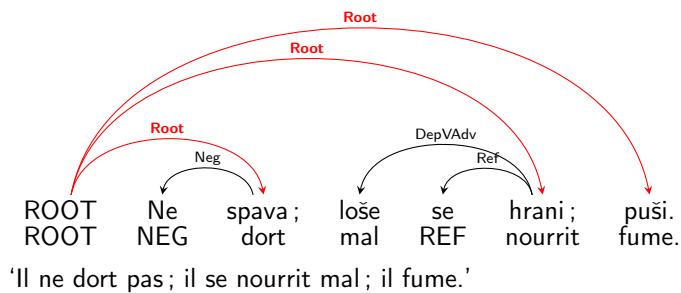
(8)



### Juxtaposition de plusieurs propositions

Un autre cas de figure où l'on peut considérer le rattachement de plusieurs éléments à la racine externe est celui où plusieurs propositions indépendantes sont enchaînées en une phrase sans être en coordination. Dans ce cas, il peut paraître intuitif de rattacher chacun des verbes principaux directement à la racine externe (cf. exemple 9).

(9)



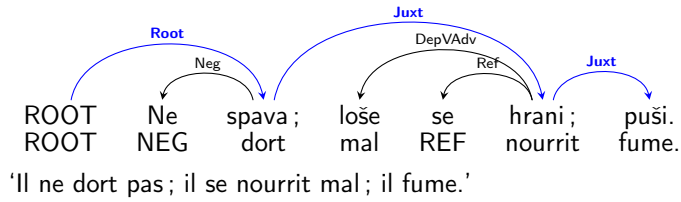
Cependant, pour maintenir la cohérence du traitement, nous considérons que le seul descendant de la racine externe est la tête de la première proposition et que les autres ont une relation de juxtaposition par rapport à elle. Le traitement adopté est donc celui montré dans l'exemple 10.

La juxtaposition est décrite en détail dans la section 2.11.

## 2.3 Dépendants du verbe

Cette section est destinée à la description du traitement des fonctions régies par le verbe.

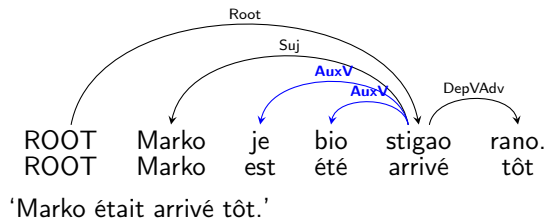
(10)



### 2.3.1 Verbe auxiliaire

La relation **AuxV** a pour gouverneur le verbe principal, et pour dépendant le verbe auxiliaire d'un temps composé. Dans le cas des temps surcomposés, tous les verbes auxiliaires sont reliés directement au verbe principal par cette relation, indépendamment les uns des autres.

(11)

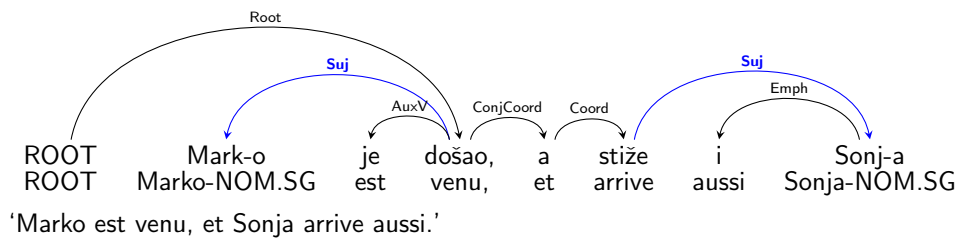


Ce traitement est parallèle à celui adopté au niveau morphosyntaxique, où tous les éléments du verbe auxiliaire d'une forme verbale surcomposée sont traités comme des auxiliaires.

### 2.3.2 Sujet

La relation **Suj** correspond à la notion de *sujet grammatical* dans la tradition syntaxique serbe. Cet élément se réalise typiquement sous forme d'un **élément nominal au nominatif** qui répond à la question *Ko ?* 'qui-sujet-humain' ou *Šta ?* 'quoi-sujet-humain'. Le gouverneur de cette relation est donc le verbe principal de la proposition, alors que le dépendant peut avoir la forme d'un nom, pronom ou numéral au nominatif.

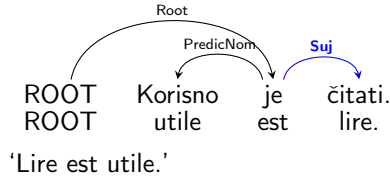
(12)



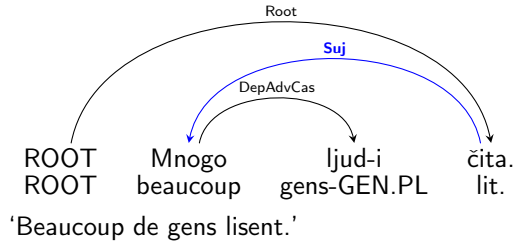
Il est également possible d'avoir un **infinitif** dans cette fonction (exemple 13).

Le sujet peut également être représenté par un syntagme partitif, dont la tête est l'**adverbe de quantité** (exemple 14).

(13)



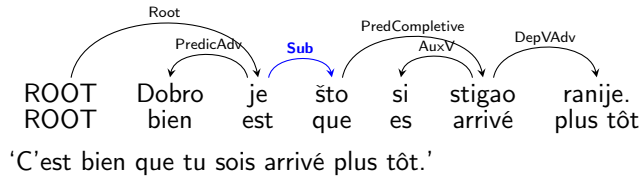
(14)



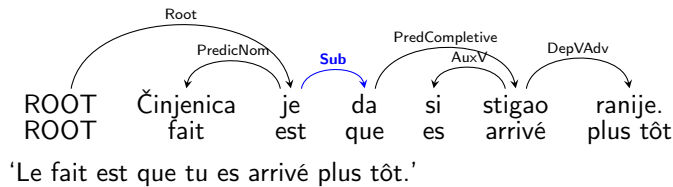
### Complétives en *da* et *što*

À la différence de la tradition grammaticale serbe, qui considère que les propositions en *da* et *što* dans une phrase de type *Dobro je da si tu* 'C'est bien que tu sois là' ont la fonction du sujet, nous considérons en effet qu'il s'agit des phrases impersonnelles, sans sujet, et que les propositions sont en effet des subordonnées complétives qui dépendent du verbe. Pour ces cas de figure, nous proposons donc les traitements illustrés dans les exemples 15 et 16.

(15)



(16)



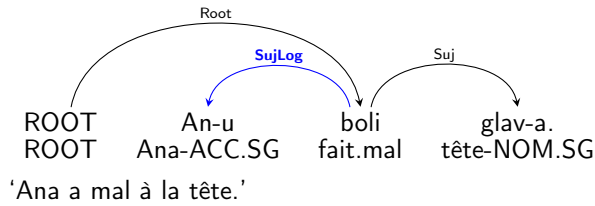
Le traitement des complétives est expliqué en détail dans la section 2.8.2.

### 2.3.3 Sujet logique

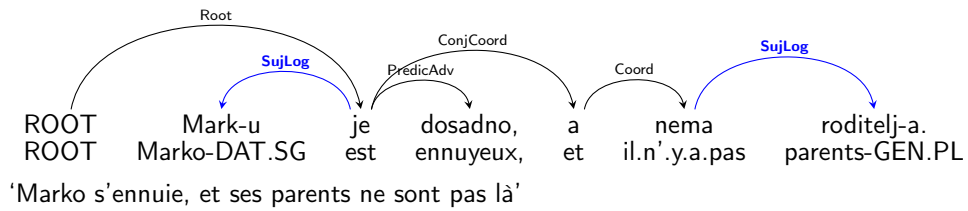
L'étiquette **SujLog** correspond à la fonction du sujet logique de la syntaxe serbe (cf. Stanojčić and Popović, 2012; Ivić, 2005). Cette fonction représente de manière générale l'**expérienteur d'un processus verbal** ou le **sujet d'un énoncé existentiel**. Il peut se réaliser **au datif, au génitif ou à l'accusatif**. Le sujet logique fait partie de la structure argumentale du verbe dont il dépend, et le remplacement par un sujet grammatical n'est pas possible. En revanche, certains verbes ouvrent deux places pour les deux types de sujet (cf.

exemple 17). Ces faits confirment que le sujet grammatical et le sujet logique représentent deux fonctions syntaxiques distinctes. D'autres constructions qui contiennent le sujet logique sont listées dans les exemples 18 à 23.

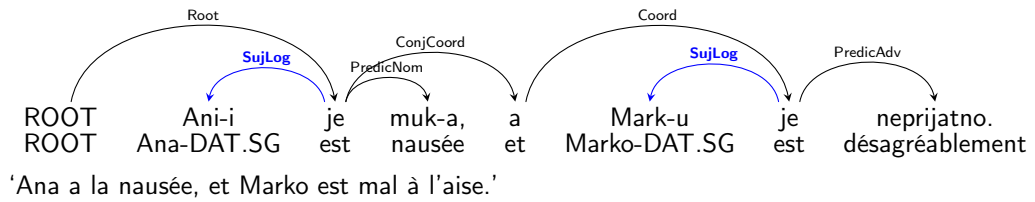
(17)



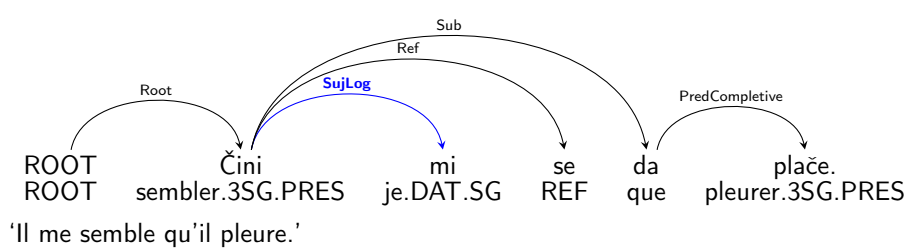
(18)



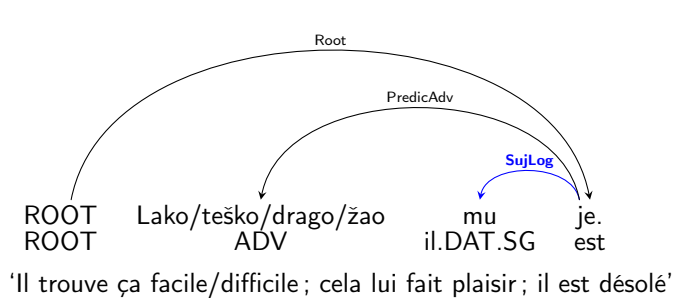
(19)



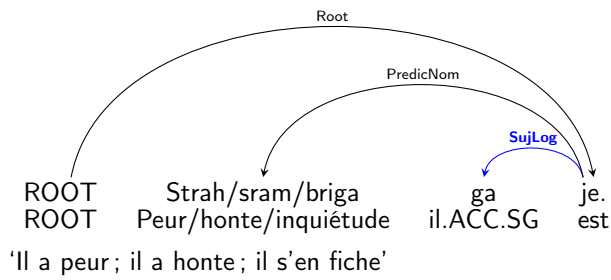
(20)



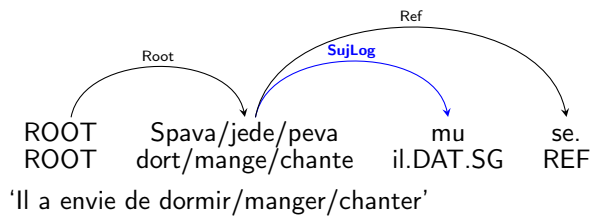
(21)



(22)



(23)

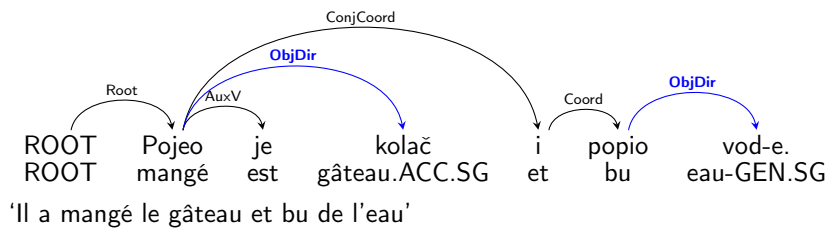


À la différence de (Stanojčić and Popović, 2012), nous ne considérons pas le datif dit possessif (cf. *Popeo se ocu na kolena*, lit. ‘Il est monté père.DAT sur les genoux’ = ‘Il est monté sur les genoux de son père’) comme un sujet logique. Ce datif, tout comme le datif éthique, est traité comme un objet indirect (cf. section 2.3.5).

### 2.3.4 Objet direct

Cette fonction est représentée par l’étiquette **ObjDir**. L’objet direct en serbe a typiquement la forme d’un **élément nominal** (nom, pronom ou numéral) **à l’accusatif**. Il existe également l’objet direct **au génitif**, le génitif en question étant un génitif partitif, ou bien un génitif dit ‘slave’ (utilisé après la négation). Les deux réalisations (celle à l’accusatif et celle au génitif) sont annotées à l’aide de l’étiquette **ObjDir** (cf. exemple 24).

(24)

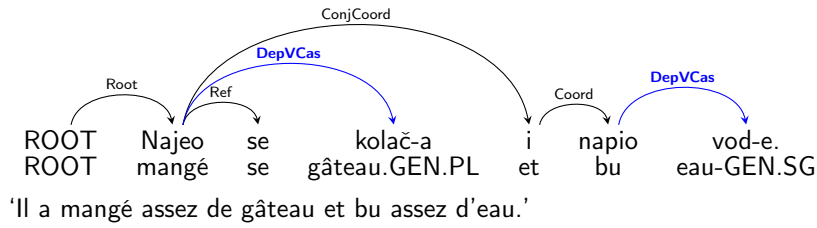


Pour décider si un génitif est un objet direct, nous utilisons le **test** suivant : si le génitif peut être remplacé par l’accusatif, il s’agit effectivement d’un objet direct (cf. *Popio je vode* => *Popio je vodu*). En revanche, le dépendant au génitif qui ne peut pas être remplacé par un accusatif n’est pas considéré comme objet direct. Il est annoté comme dépendant verbal sous forme d’un nom fléchi (étiquette **DepVCas**, cf. exemple 25<sup>6</sup>). Ici, il n’est pas possible de remplacer le génitif par l’accusatif : *Najeo se kolača* => \**Najeo se kolači*, *Napio se vode* => \**Napio se vodu*.

L’utilisation de l’étiquette **DepVCas** est présentée en détail dans la section 2.3.11.

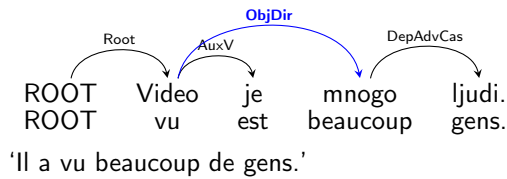
6. Les verbes *najesti se* et *napiti se* ont une interprétation aspectuelle spécifique : ils ont une lecture perfective associée à l’idée d’une action effectuée à satiété.

(25)



Tout comme dans le cas du sujet, l'objet direct peut également être représentée par un adverbe de quantité introduisant un syntagme partitif (cf. exemple 26).

(26)



### 2.3.5 Objet indirect

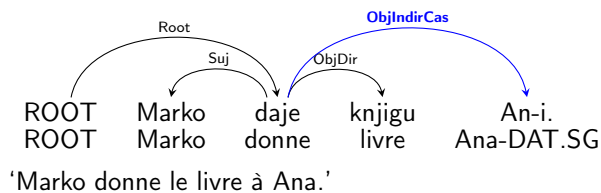
Cette étiquette est consacrée à un sous-ensemble des relations regroupées sous le nom de l'objet indirect dans (Stanojčić and Popović, 2012). En effet, ces auteurs considèrent comme objet indirect tout argument verbal qui n'est pas un objet direct et qui n'a pas un sémantisme adverbial net. Les caractéristiques de surface de ces éléments sont très disparates : il peut s'agir aussi bien des groupes nominaux à des cas différents que des groupes prépositionnels introduits par un nombre élevé de prépositions différentes (les auteurs en citent 9). Pour éviter de rassembler cette grande diversité de dépendants sous une étiquette dénotant une fonction bien précise, nous avons préféré réserver la fonction de l'objet indirect à deux cas de figure certes spécifiques mais prototypiques et de traiter les autres avec des étiquettes sous-spécifiées en **DepV** (cf. section 2.3).

Quant à l'objet indirect, nous introduisons deux étiquettes : **ObjIndirCas**, qui correspond à l'objet indirect casuel, exprimé sous forme d'un élément nominal fléchi au datif, et **ObjIndirPrep**, objet indirect prépositionnel, qui correspond à l'objet indirect d'un verbe de parole ou de processus mental introduit par la préposition *o* 'de', complétée par un nom au locatif.

#### Objet indirect casuel

**ObjIndirCas** correspond à la réalisation prototypique de l'objet indirect : il s'agit d'un **élément nominal au datif** exprimant typiquement le bénéficiaire ou le destinataire d'un processus verbal.

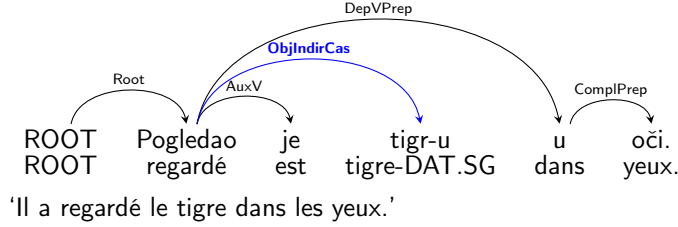
(27)



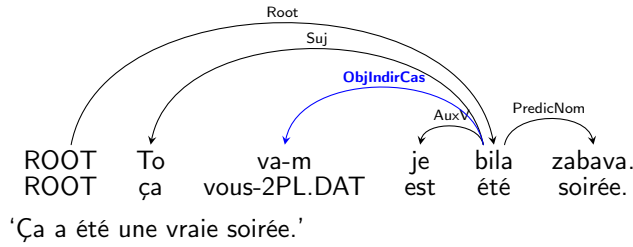
Cette étiquette est également accordée à d'autres types des dépendants verbaux au datif, comme le **datif possessif** (exemple 28) et le **datif éthique** (exemple 29). Même s'il ne s'agit

pas de véritables objets indirects dans ces deux cas, ces datifs ont le même comportement syntaxique que l'objet indirect, et le seul paramètre de distinction est le sémantisme du verbe. Ceci dépasse donc le cadre d'une annotation syntaxique de surface et sera traité dans une étape ultérieure du développement du corpus.

(28)



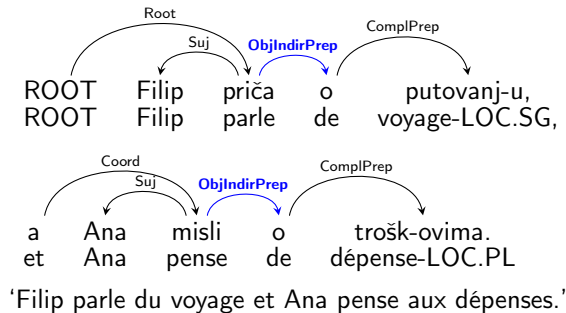
(29)



## Objet indirect prépositionnel

Nous considérons comme objet indirect prépositionnel (**ObjIndirPrep**) les dépendants des verbes de parole et de processus mentaux ayant la forme d'un **groupe prépositionnel introduit par la préposition o 'de' complétée par un nom au locatif**. Tout autre dépendant prépositionnel d'un verbe (qui n'est pas un prédicatif) est traité comme **DepVPrep** (cf. section 2.3.12).

(30)

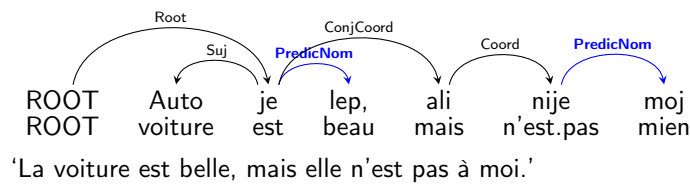


### 2.3.6 Prédicatif nominal

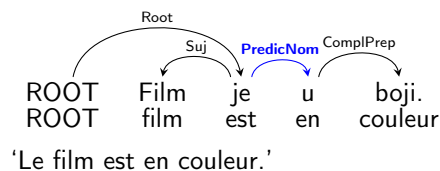
Dans la tradition syntaxique serbe (et, plus généralement, slave), le prédicatif désigne un dépendant nominal non-objet dans l'expression d'un prédicat. La notion du prédicatif nominal (**PredicNom**) correspond à un élément à nature nominale introduit par le verbe *biti* 'être', qui exprime une caractéristique du sujet du même verbe. Cette fonction est donc équivalente de l'attribut du sujet dans la tradition grammaticale française, mais limitée à un

seul gouverneur (le verbe *être*). Le dépendant de cette relation peut avoir la forme d'un **nom**, **pronom**, **adjectif ou numéral au nominatif**, ou bien d'une **préposition** introduisant un groupe prépositionnel.

(31)



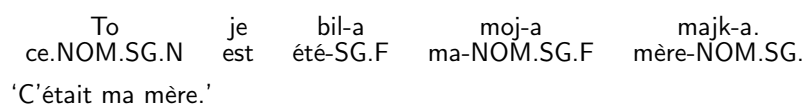
(32)



### Les phrases comme *To je bila moja majka* 'C'était ma mère'

En serbe, le sujet impose au verbe l'accord en nombre, genre et personne. Si l'on considère la phrase dans l'exemple 33, nous constatons que le verbe porte les marques d'accord avec le nom, qui est du genre féminin, mais pas avec le pronom démonstratif, qui est du genre neutre.

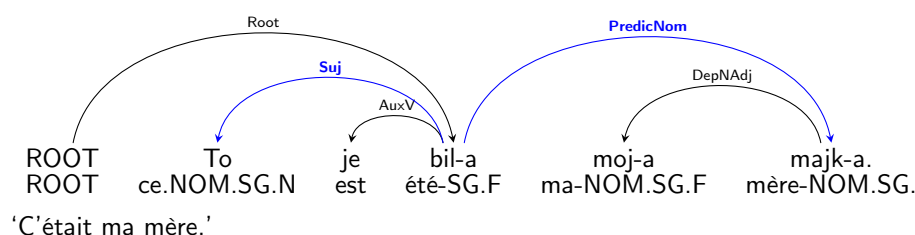
(33)



Cet indice morphosyntaxique pourrait nous amener à considérer que c'est le nom qui a le rôle du sujet, et que le pronom est en effet le prédicatif nominal. Cependant, le pronom démonstratif est fonctionnellement équivalent ici d'un groupe nominal ou d'un pronom personnel, réalisations type du sujet en serbe : *Profesorka matematike/ona/to je bila moja majka* 'La professeure de maths/elle/c'était ma mère'. Nous adoptons donc l'analyse suivante : le pronom démonstratif est traité comme sujet, alors que le nom (ou l'élément à nature nominale) a la fonction du prédicatif nominal.



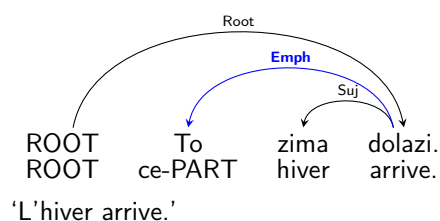
(34)



**NB**

Ce cas de figure n'est pas à confondre avec les phrases dans lesquelles la forme *to* 'ce' n'est pas un pronom, mais une particule, et dans lesquelles elle n'a pas de rôle syntaxique. Dans la phrase *To zima dolazi* 'C'est l'hiver qui arrive', le verbe *dolaziti* 'arriver, venir' ouvre une seule place dans sa structure argumentale : celle du sujet, qui est dans cet exemple occupée par le nom *zima* 'hiver'. La particule *to* 'ce' n'a donc pas de place dans la structure syntaxique de la phrase. Son rôle est de l'ordre emphatique : elle met en valeur le contenu propositionnel. Par conséquent, ce type d'occurrences de la forme *to* est annoté avec la relation **Emph** (cf. exemple 35). L'utilisation de cette étiquette est décrite en détail dans la section 2.7.7.

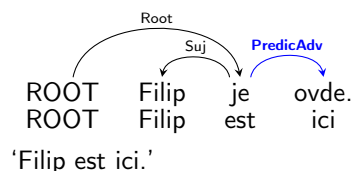
(35)



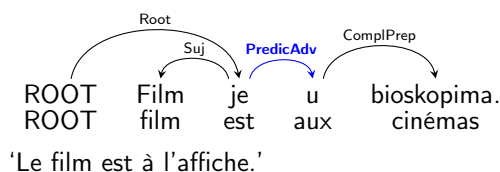
### 2.3.7 Prédicatif adverbial

Le prédicatif adverbial (**PredicAdv**) est également un dépendant du verbe *biti* 'être', mais à sens adverbial. Il peut se réaliser comme un **adverbe** ou comme une **préposition** introduisant un groupe prépositionnel à sens adverbial.

(36)



(37)



**NB**

Comme mentionné dans la section 2.3.6, un groupe prépositionnel gouverné par le verbe *biti* 'être' peut également avoir le rôle d'un prédicatif nominal. Il faut donc veiller à faire la distinction entre l'exemple 37, où l'on trouve un prédicatif adverbial, et l'exemple 32, qui illustre un prédicatif nominal sous forme d'un groupe prépositionnel. Pour identifier la bonne étiquette, le test suivant peut être utilisé : si le groupe prépositionnel peut être remplacé par un adjectif, il s'agit d'un prédicatif nominal, et s'il peut être remplacé par un adverbe, il s'agit d'un prédicatif adverbial.

### 2.3.8 Prédicatif complémentaire

La fonction appelée *dopunski predikativ* 'prédicatif complémentaire' en serbe correspond dans la syntaxe française à l'attribut du sujet ou de l'objet direct avec les verbes obligatoirement attributifs autres que le verbe *être*. Il s'agit donc des dépendants obligatoires des verbes comme *zvati (se)* '(s)'appeler', *proglasiti (se)* '(se) proclamer', *smatrati (se)* '(se) considérer', etc. En syntaxe serbe (cf. Stanojčić and Popović, 2012), le prédicatif complémentaire regroupe aussi bien les prédicatifs qui définissent le sujet que ceux qui sont liées à l'objet direct. Dans le cadre de ce projet, nous utilisons deux étiquettes distinctes selon ce critère, à savoir **PredicComplSuj** pour le prédicatif complémentaire lié au sujet et **PredicComplObj** pour le prédicatif complémentaire lié à l'objet direct.

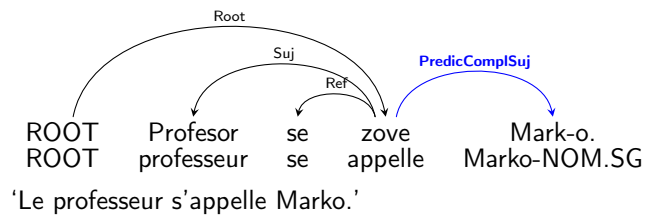
#### Prédicatif complémentaire lié au sujet

S'il détermine le sujet, le prédicatif complémentaire peut se réaliser sous les formes suivantes :

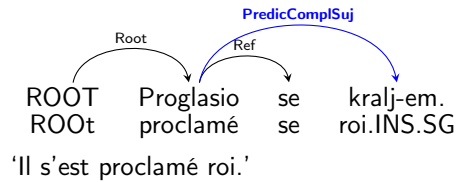
- un *nom*, *pronom*, *numéral* ou *adjectif* au **nominatif** ou à l'**instrumental**, ou
- la **préposition za** 'pour' complétée par un **accusatif**.

Dans le corpus, il est annoté en utilisant l'étiquette **PredicComplSuj**.

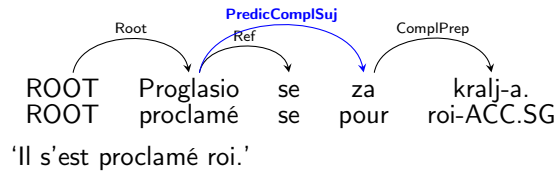
(38)



(39)



(40)



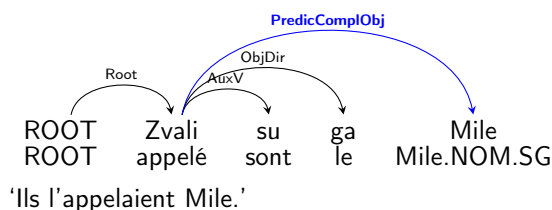
### Prédicatif complémentaire lié à l'objet direct

Le prédicatif lié à l'objet direct peut avoir les mêmes formes que celui lié au sujet :

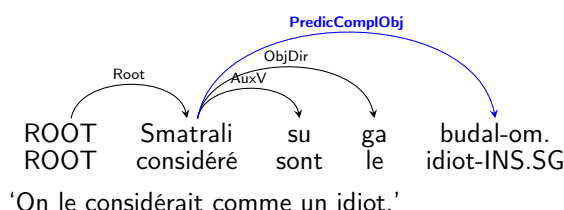
- *nom*, *pronom*, *numéral* ou *adjectif* au **nominatif** ou à l'**instrumental**, ou
- un *groupe prépositionnel* introduit par la **préposition** *za* '**pour**' complétée par un **accusatif**.

Ce qui distingue cette construction de la précédente est la présence de l'objet direct dans la phrase. Dans le corpus, il est annoté avec l'étiquette **PredicComplObj**.

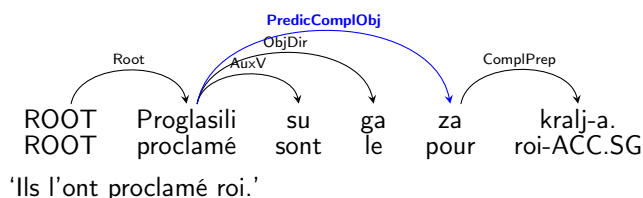
(41)



(42)



(43)

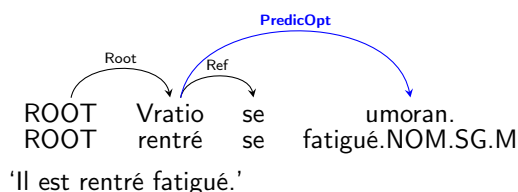


### 2.3.9 Prédicatif optionnel

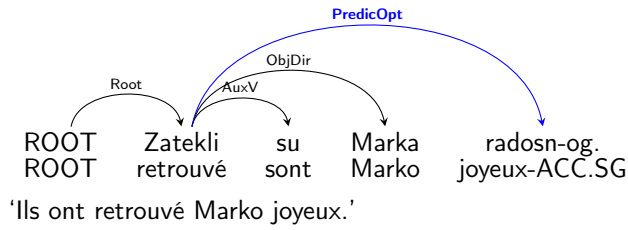
Le prédicatif optionnel (**PredicOpt**), désigné typiquement par le terme de *aktuelni kvalifikativ* dans la tradition grammaticale serbe, correspond à un dépendant verbal qui qualifie le sujet ou l'objet direct dans le cadre du processus verbal. En français, il correspond à la fonction de l'attribut du sujet ou de l'objet direct avec les verbes occasionnellement attributifs ou à l'épithète détachée. C'est un dépendant optionnel qui peut s'associer à différents verbes, comme *zateći* 'retrouver', *stići* 'arriver', *krenuti* 'partir', etc. Il a trois réalisations principales :

- sous forme d'un *adjectif*, il peut être au **nominatif** (s'il modifie le sujet) ou à l'**accusatif** (s'il modifie l'objet direct) ;
- sous forme d'un *nom*, *pronom* ou *numéral*, il est au **génitif** (peu importe l'élément qu'il modifie) ;
- sous forme d'un **groupe prépositionnel**, indépendamment de la fonction à laquelle il est lié.

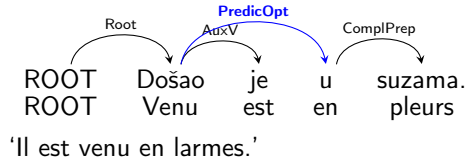
(44)



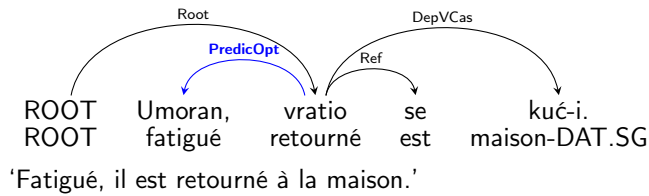
(45)



(46)

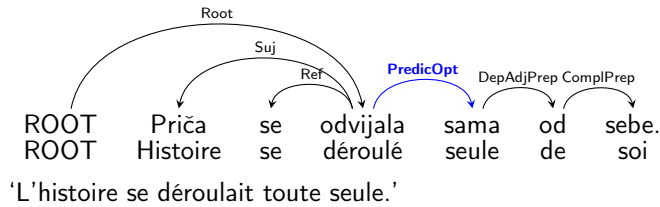


(47)



Le prédicatif optionnel peut également se trouver en tête de la phrase (cf. exemple 47).  
La construction *sam od sebe* 'tout seul, de son gré' relève souvent du prédicatif optionnel :

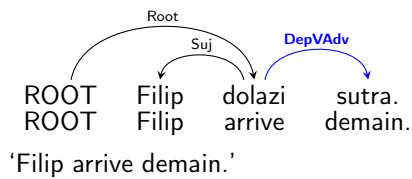
(48)



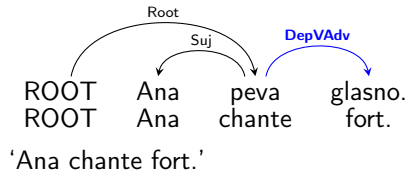
### 2.3.10 Dépendant sous forme d'un adverbe

Comme il a déjà été indiqué, dans le cadre du projet *ParCoLab*, nous ne faisons pas la distinction entre les ajouts et les arguments verbaux. Par conséquent, tout **dépendant adverbial d'un verbe** qui n'est pas un prédicatif adverbial est annoté comme **DepVAdv**.

(49)



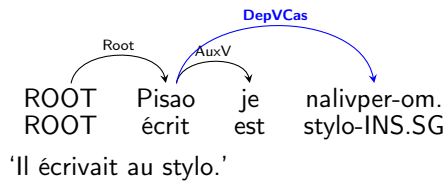
(50)



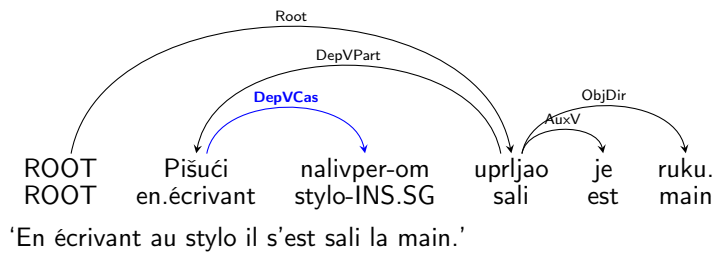
### 2.3.11 Dépendant sous forme d'un nom fléchi

L'étiquette **DepVCas** regroupe tous les **dépendants casuels d'un verbe** qui ne représentent pas un sujet, sujet logique, objet direct, objet indirect ou prédicatif. Il s'agit notamment des groupes nominaux exprimant des valeurs comme celle de l'instrument, l'accompagnement, le temps, etc. Le gouverneur de cette relation peut être le verbe principal d'une proposition (cf. exemple 51) ou bien un participe présent ou passé (cf. exemple 52).

(51)



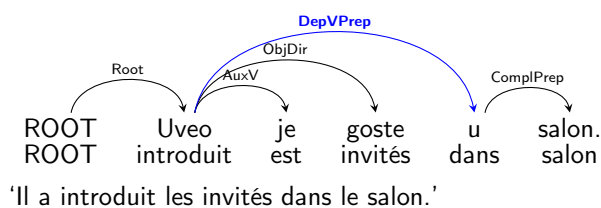
(52)



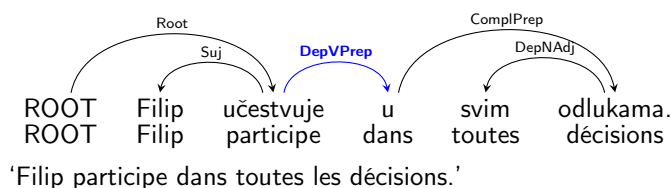
### 2.3.12 Dépendant sous forme d'une préposition

L'étiquette **DepVPrep** regroupe tous les **dépendants prépositionnels d'un verbe** qui ne représentent pas un objet indirect prépositionnel ou un prédicatif. Elle s'applique donc aux cas de figure considérés par Stanojčić and Popović (2012) comme des objets indirects (*strahovati od nečega* 'avoir peur de quelque chose'), aux ajouts à sens adverbial (*sedeti u sobi* 'être assis dans la pièce'), ou aux compléments adverbiaux (*ići u školu* 'aller à l'école').

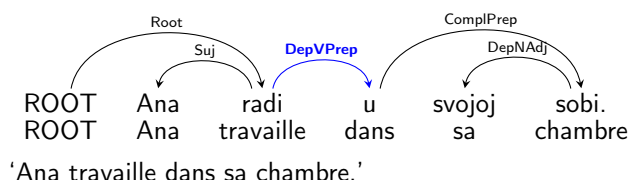
(53)



(54)



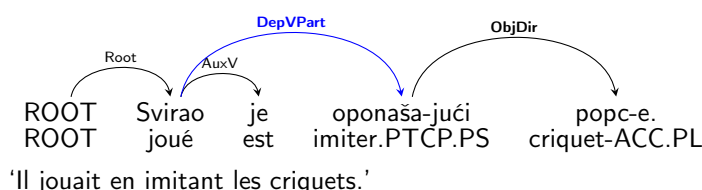
(55)



### 2.3.13 Propositions participiales

Outre les deux participes utilisés dans la construction des formes verbales complexes<sup>7</sup>, le serbe dispose de deux autres formes appelées des participes : le participe présent (*glagolski prilog sadašnji* 'adverbe déverbal présent', par ex. *radeći* 'en travaillant') et le participe passé (*glagolski prilog prošli* 'adverbe déverbal passé', par ex. *uradivši* 'ayant fait'). Malgré leur sémantisme qui pourrait être qualifié d'adverbial, ces participes ont un comportement fondamentalement différent de celui des adverbes. Plus particulièrement, ils gardent la structure argumentale du verbe dont ils sont la réalisation et peuvent avoir des dépendants verbaux typiques, notamment des objets directs et indirects. Vu cette capacité, on pourrait les rapprocher des subordonnées, mais dans la tradition grammaticale serbe on considère comme propositions seulement les constructions ayant une forme verbale finie pour noyau (cf. Stanojčić and Popović, 2012; Mrazović, 2009; Ivić, 2005). Par conséquent, nous introduisons une étiquette spéciale pour le traitement de ces formes : **DepVPart**. Le gouverneur de la relation est le verbe principal, et le dépendant est obligatoirement un autre verbe sous forme d'un **participe présent ou passé**. Les dépendants éventuels des participes sont, quant à eux, traités comme les dépendants d'un verbe principal.

(56)

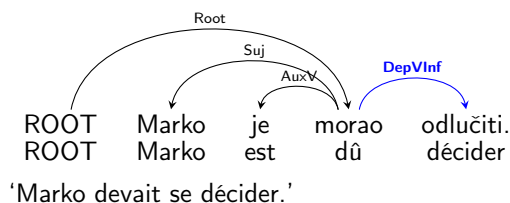


### 2.3.14 Prédicat complexe : verbe modal ou aspectuel introduisant un infinitif

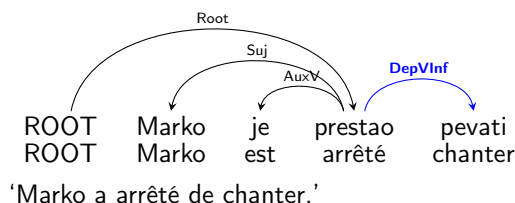
En serbe, les verbes modaux et aspectuels peuvent introduire deux types de constituants : un infinitif (cf. *moгу raditi* lit. 'je.peux travailler') ou une complétive en *da* 'que' introduisant un verbe au présent (cf. *moгу da radim* lit. 'je.peux que je.travaille'). Le premier type de construction est traité avec l'étiquette **DepVInf**, cf. les exemples 57 et 58.

7. Dans le cadre du projet *ParCoLab*, nous dénotons comme *participe actif* la forme connue comme *glagolski pridev radni* 'adjectif verbal actif' dans la tradition grammaticale serbe, alors que le terme *glagolski pridev trpni* 'adjectif verbal passif' est traduit comme *participe passif*.

(57)



(58)



Le cas de figure où ces verbes introduisent une complétive est présenté dans la section dédiée à ce type de subordonnée (cf. section 2.8.2).

### 2.3.15 Traitement des enchaînements des dépendants

Il est possible qu'une phrase présente plusieurs éléments qui semblent avoir le même rôle syntaxique par rapport au verbe : *Ana, Filip i Alan su došli* 'Ana, Filip et Alain sont venus', ou *Filip je putovao u Francusku, u Italiju, u Liban*. 'Filip est allé en France, en Italie, au Liban'. Dans ce cas, deux traitements sont possibles, en fonction du type de dépendant en question.

Pour les dépendants suivants :

- Suj,
- SujLog,
- ObjDir,
- ObjIndirCas,
- ObjIndirPrep,
- PredicNom,
- PredicAdv,
- PredicComplSuj,
- PredicComplObj et
- ComplPrep,

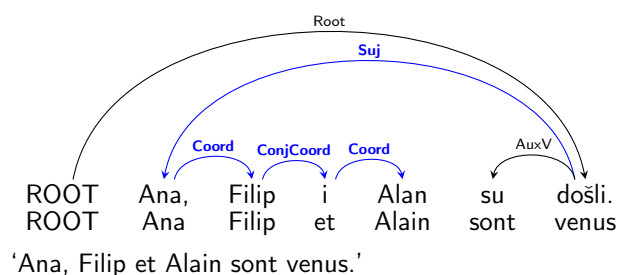
nous considérons qu'il s'agit d'un seul dépendant complexe, composé de plusieurs dépendants en coordination entre eux. Autrement dit, dans la phrase *Ana, Filip i Alan su došli* 'Ana, Filip et Alain sont venus', il y aura une seule relation *Suj*, qui reliera le premier élément de l'enchaînement au verbe, et les suivants seront considérés comme des dépendants coordonnés. Ce traitement se justifie par la structure argumentale des verbes : le verbe n'ouvre qu'une position pour un sujet ou un objet<sup>8</sup>, et par conséquent, le fait d'annoter plusieurs dépendants avec l'une de ces fonctions fausserait la représentation de la structure argumentale des verbes en corpus. L'annotation correcte du premier exemple est montrée dans l'exemple 59.

Il faut noter que s'il n'y avait pas de conjonction de coordination dans la phrase, on considérerait tout de même qu'il s'agit d'une coordination. (v. section 2.10).

8. Avec quelques exceptions possibles en serbe, notamment *pitati* 'demander', qui exige un double accusatif (*pitati nekoga nešto* lit. 'demander quelqu'un quelque chose').

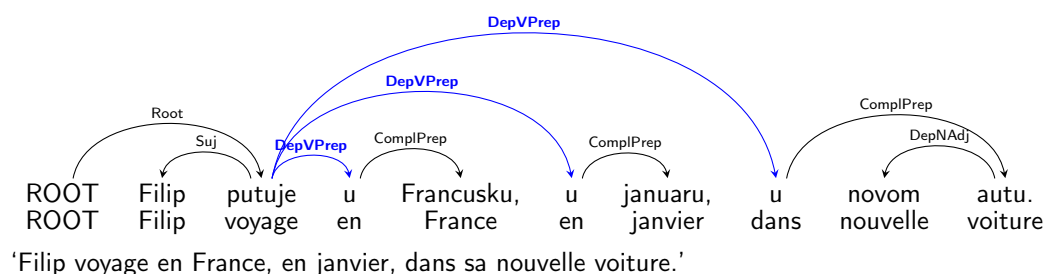


(59)



En revanche, pour tout autre dépendant verbal, nous adoptons le rattachement direct de chaque élément au verbe gouverneur.

(60)



Ce traitement reflète le fait que les verbes peuvent admettre plusieurs réalisations des dépendants qui n'appartiennent pas à leur structure argumentale.

Nous avons ainsi passé en revue l'ensemble des dépendants du verbe. Dans la section suivante, nous présentons les dépendants du nom.

## 2.4 Dépendants du nom

Cette section regroupe les relations dont le gouverneur est un nom (ou une forme à comportement nominal, comme un pronom ou un numéral). Il s'agit principalement des dépendants adjectivaux, casuels et prépositionnels. Précisons que nous abandonnons la classification traditionnelle de (Stanojčić and Popović, 2012) en *kongruentni atribut*, *padežni atribut* et *atributiv* : les critères de distinction de ces fonctions étaient trop disparates pour permettre une identification fiable en corpus. À leur place, nous introduisons des étiquettes sous-spécifiées à l'instar de celles utilisées pour les dépendants verbaux. Leur utilisation est expliquée dans la suite.

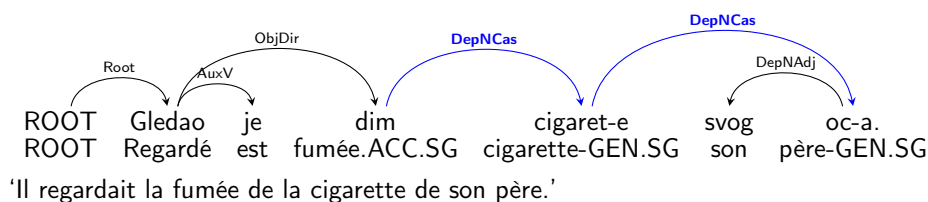
### 2.4.1 Dépendant sous forme d'un nom fléchi

La relation **DepNCas** s'applique aux dépendants du nom sous forme d'un **élément nominal à un cas oblique** (génitif, datif, accusatif ou instrumental) (cf. exemple 61).

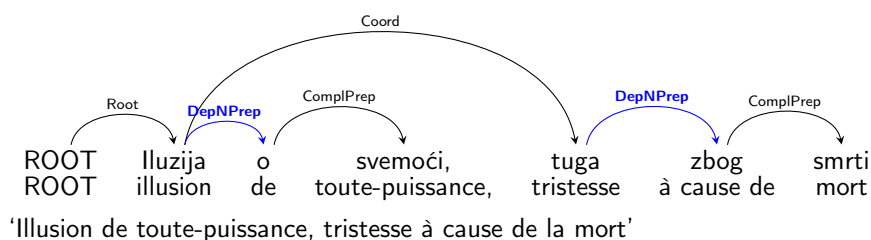
### 2.4.2 Dépendant sous forme de préposition

L'étiquette **DepNPrep** regroupe les dépendants prépositionnels d'un nom, peu importe leur sémantisme (cf. exemple 62).

(61)



(62)

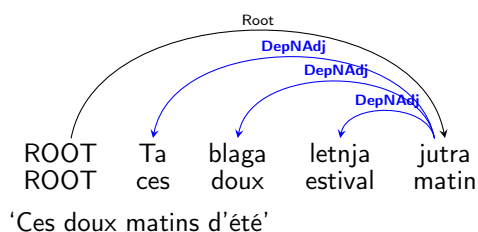


### 2.4.3 Dépendant sous forme d'adjectif

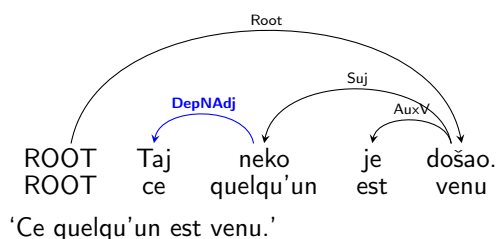
L'étiquette **DepNAdj** regroupe tous les dépendants adjectivaux du nom. Il s'agit ici aussi bien des adjectifs qualificatifs (*lep auto* 'belle voiture') que des autres sous-catégories (cf. *naš auto* 'notre voiture'). Elle s'applique également aux **adjectifs dérivés des participes**.

Le gouverneur de la relation est un nom ou un pronom, alors que le dépendant prend la forme d'un **adjectif ou d'un numéral qui s'accorde avec le nom**.

(63)

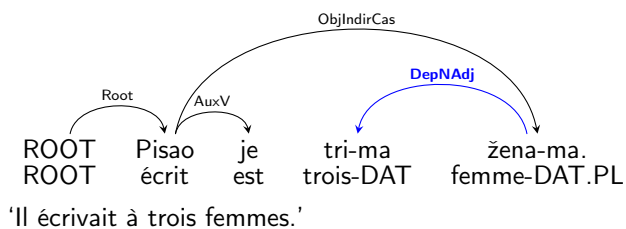


(64)

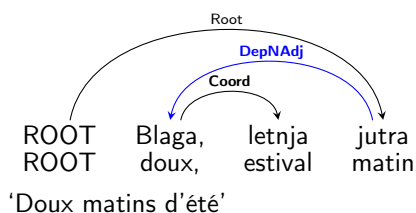


Il est à noter que lorsque plusieurs adjectifs sont placés au même niveau (sans virgule intervenante, cf. exemple 63), ils sont rattachés au nom gouverneur indépendamment les uns des autres. En revanche, si les adjectifs sont séparés par une virgule, on considère qu'il s'agit d'un seul dépendant complexe, contenant des adjectifs coordonnés (cf. exemple 66).

(65)



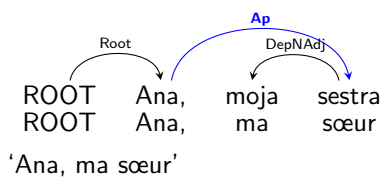
(66)



#### 2.4.4 Apposition

L'apposition est marquée par l'étiquette **Ap**. Cette étiquette s'applique, bien évidemment, aux cas prototypiques de l'apposition. Le gouverneur et le dépendant de la relation sont des noms, qui doivent être au même cas. Le dépendant apporte une précision par rapport au gouverneur, et il est le plus souvent séparé du contexte par des virgules (cf. exemple 67).

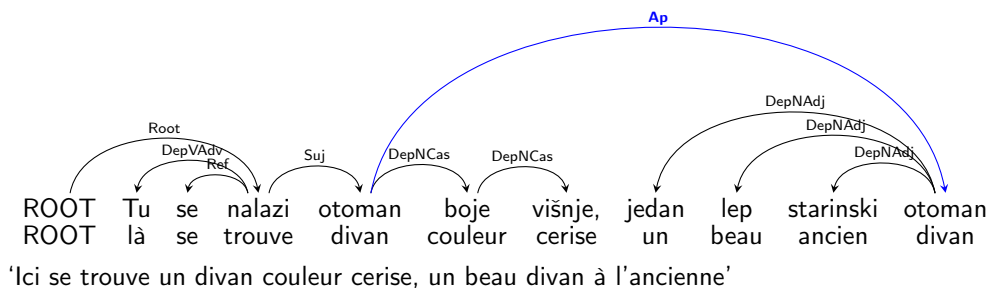
(67)



Cependant, l'étiquette **Ap** est également utilisée pour plusieurs cas de figure qui ne relèvent pas traditionnellement de l'apposition. Il s'agit notamment de la **répétition** du même élément dans la phrase, des **noms des personnes**, des **honorifiques** antéposés aux noms et des **enchaînements des constituants suivis d'un pronom de reprise**. Tous les cas de figure sont détaillés dans la suite.

L'exemple 68 illustre la **répétition** du même élément dans la phrase.

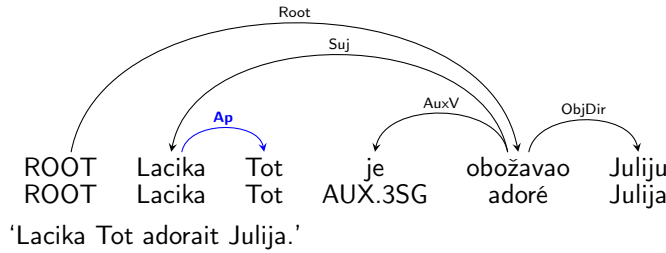
(68)



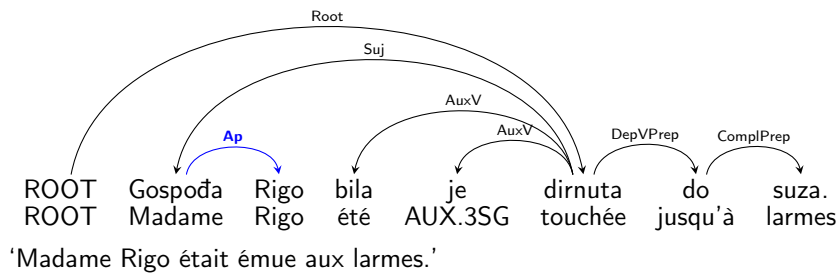
Dans le cas des **noms de personne** composés du prénom et du nom de famille, le premier élément (typiquement le prénom) porte l'étiquette de la fonction exercée par le groupe

dans la phrase, et il gouverne à son tour le deuxième élément, qui porte l'étiquette **Ap**. Le même principe s'applique dans le cas des honorifiques devant un nom de famille, et d'autres cas relevant de la relation syntaxique connue sous le nom de *atributiv* dans la tradition grammaticale serbe. Dans ce type d'application, nous considérons systématiquement que la tête du syntagme est le nom à l'extrême gauche du groupe, et l'apposition s'établit de gauche à droite, en cascade si plusieurs éléments entretenant le même type de rapport s'enchaînent dans la phrase (cf. exemples 69, 70, 71).

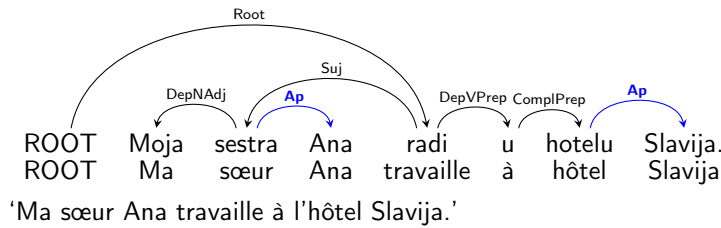
(69)



(70)



(71)



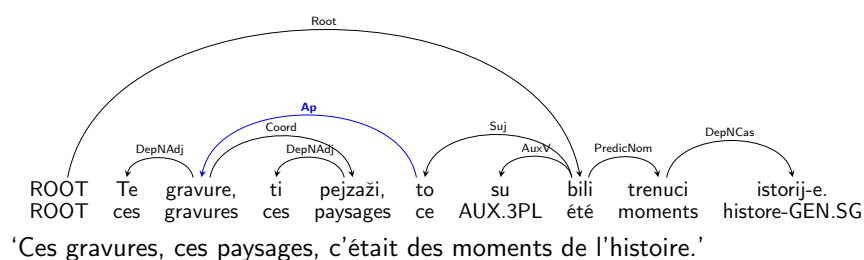
L'étiquette **Ap** est utilisée aussi dans le traitement d'un phénomène syntaxique plus complexe. Il s'agit des cas où un élément de la structure syntaxique est représenté par une sorte d'énumération, reprise ensuite par le pronom démonstratif *to*. Dans les cas rencontrés jusqu'à présent en corpus, il s'agit typiquement du sujet.

Comme les éléments de l'énumération ont le plus souvent des traits morphosyntaxiques disparates, alors que le prédicat s'accorde (le plus souvent) avec le pronom *to*, nous annotons le pronom comme sujet, et considérons qu'il gouverne le premier élément de l'énumération *via* la relation **Ap**. Les éléments de l'énumération sont reliés entre eux par la relation **Coord**, en cascade (cf. l'exemple 72).

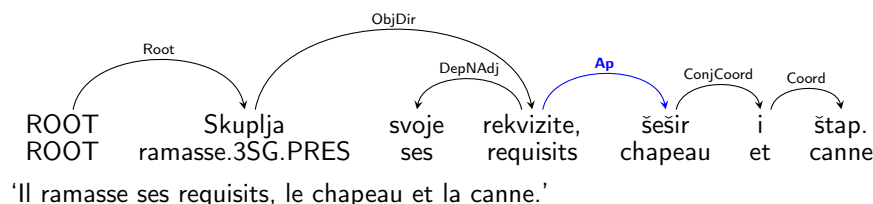
Il peut parfois être délicat de déterminer s'il s'agit d'une apposition ou d'une coordination, cf. exemple 73.

Ici, le seul critère disponible est de l'ordre référentiel : si les éléments enchaînés ont le même référent (autrement dit, s'ils désignent la même entité), il s'agit d'une apposition ;

(72)

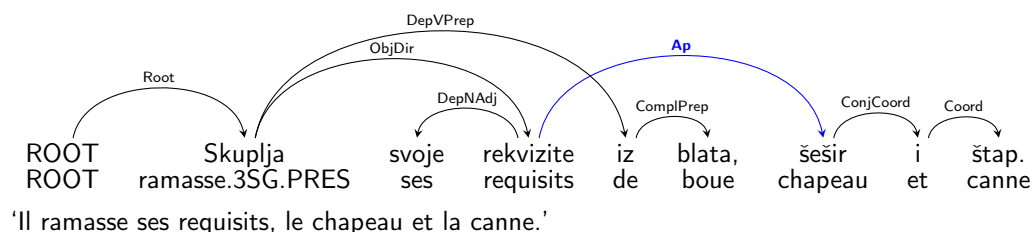


(73)



sinon, il s'agit d'une coordination. Un autre critère qui peut être utile est l'insertion d'un autre élément entre le premier nom et le reste du groupe nominal (cf. exemple 74).

(74)



Le positionnement du groupe prépositionnel indique qu'il s'agit plutôt d'une apposition, vu que ce type de clivage d'une coordination est moins probable.

Cet aperçu des dépendants du nom montre que leur inventaire est beaucoup plus restreint que celui des dépendants du verbe. Dans la section qui suit, nous présentons les dépendants de l'adjectif.

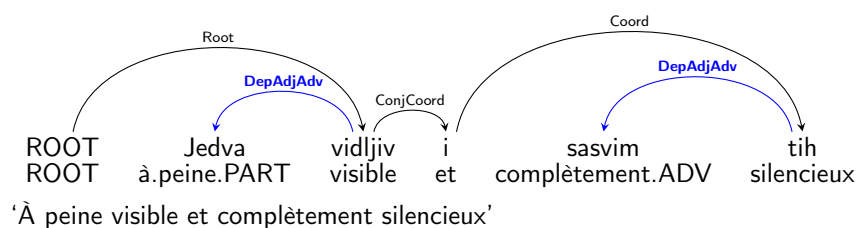
## 2.5 Dépendants de l'adjectif

Cette section regroupe les relations gouvernées par l'adjectif. En effet, en serbe l'adjectif peut avoir des dépendants adverbiaux (typiquement des intensifieurs comme *vrlo* 'très'), mais aussi des dépendants casuels et prépositionnels, et il existe également un cas de figure dans lequel un adjectif en gouverne un autre. Tous ces types de dépendants sont présentés ci-dessous.

### 2.5.1 Dépendant sous forme d'un adverbe

L'étiquette **DepAdjAdv** s'applique aux **dépendants adverbiaux** d'un adjectif. Cette relation peut également être portée par une particule à sens adverbial gouvernée par un adjectif (cf. exemple 75).

(75)

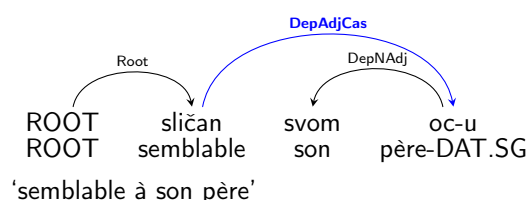


## 2.5.2 Dépendant sous forme d'un nom fléchi

L'étiquette **DepAdjCas** est appliquée aux **dépendants casuels** d'un adjectif. Il s'agit des éléments nominaux à un cas oblique (génitif, datif, accusatif ou instrumental).

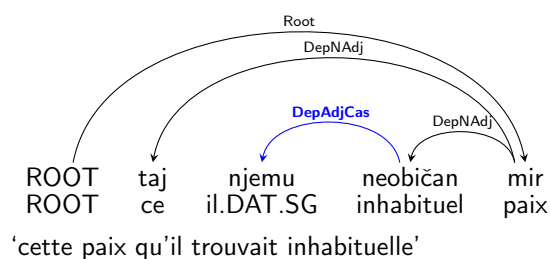
Il peut s'agir d'un dépendant exigé par l'adjectif, comme dans l'exemple 76.

(76)



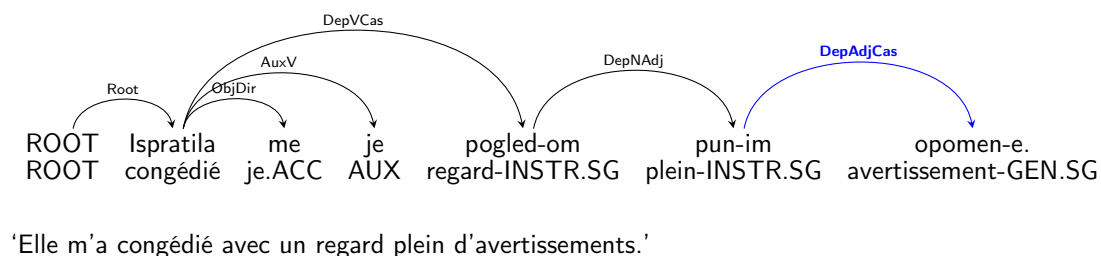
Il peut également s'agir des compléments qui ne sont pas exigés par l'adjectif (cf. exemple 77).

(77)



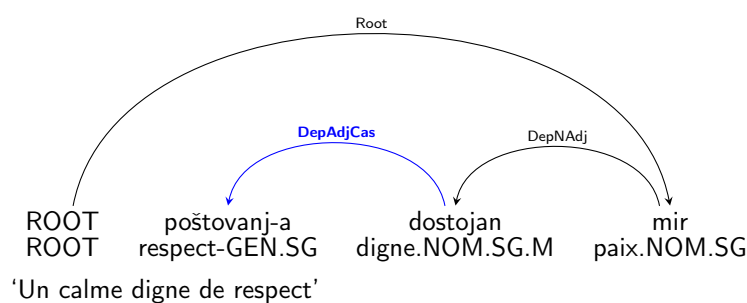
Il faut remarquer le fait que les adjectifs ayant un dépendant sous forme d'un nom fléchi sont souvent post-posés au nom (cf. exemple 78).

(78)



En revanche, si un tel adjectif est antéposé au nom, son dépendant sous forme de nom fléchi se trouve antéposé à lui (cf. exemple 79).

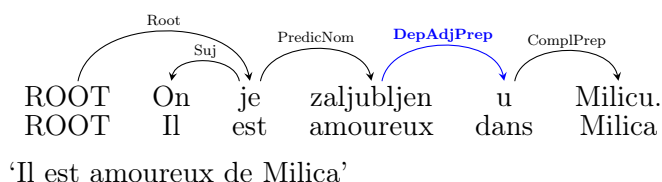
(79)



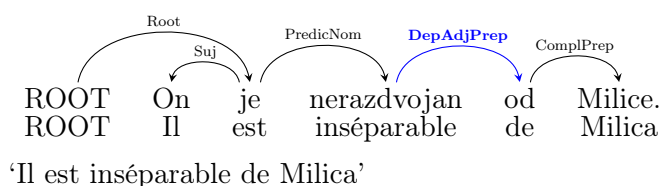
### 2.5.3 Dépendant prépositionnel

L'étiquette **DepAdjPrep** est dédiée aux **dépendants prépositionnels** d'un adjectif. Tout comme dans le cas des dépendants casuels, il peut s'agir d'un dépendant exigé par l'adjectif (cf. exemples 80 et 81).

(80)

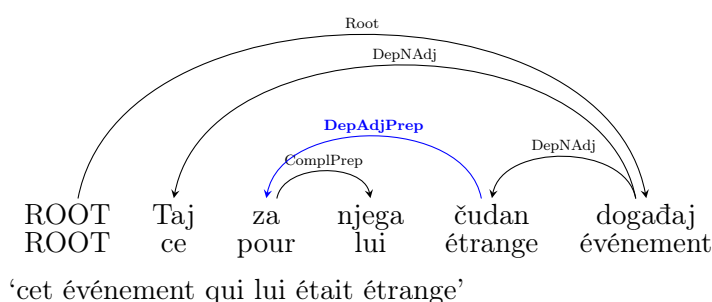


(81)



Cependant, il peut également s'agir des dépendants qui ne sont pas exigés par l'adjectif (cf. exemple 82).

(82)

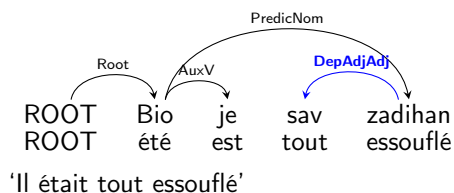


Nous remarquons le même patron de linéarisation que pour les adjectifs dotés d'un dépendant casuel : typiquement, un adjectif doté d'un dépendant prépositionnel se trouve à droite du nom. S'il est tout de même antéposé au nom, son dépendant prépositionnel se trouve à sa gauche (cf. exemple 82).

### 2.5.4 Construction *sav* 'tout' + Adjectif

L'étiquette **DepAdjAdj** est utilisée pour la construction dans laquelle un adjectif (typiquement qualificatif) gouverne l'adjectif indéfini *sav* 'tout' (cf. exemple 83).

(83)



À la différence du français, où cette fonction est exercée par un adverbe, en serbe il s'agit d'un adjectif, ce qui est confirmé par sa capacité d'être décliné, ainsi que par la variation en nombre et en genre : *Zatekli su ga sveg zadihanog* 'Ils l'ont retrouvé tout.ACC essoufflé'. Nous déterminons que l'adjectif qualificatif gouverne l'indéfini grâce au fait que ce dernier peut être omis de la phrase sans compromettre sa grammaticalité (*Bio je zadihan*, *Zatekli su ga zadihanog*), alors que le qualificatif non (\**Bio je sav*, *Zatekli su ga sveg*).

Nous avons ainsi parcouru les dépendants possibles d'un adjectif. Dans la suite, nous présentons les dépendants de l'adverbe.

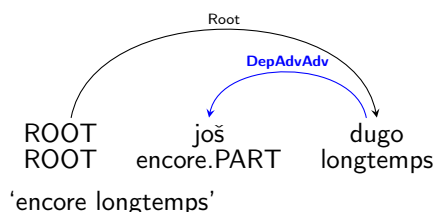
## 2.6 Dépendants de l'adverbe

Cette section est consacrée aux relations destinées au traitement des dépendants de l'adverbe. En serbe, un adverbe peut avoir des dépendants adverbiaux, prépositionnels ou casuels. Les trois types sont présentés dans la suite.

### 2.6.1 Dépendant sous forme d'adverbe

La relation **DepAdvAdv** est consacrée à l'annotation des **dépendants adverbiaux** d'un adverbe sous forme d'un autre adverbe ou de particule (cf. exemple 84).

(84)



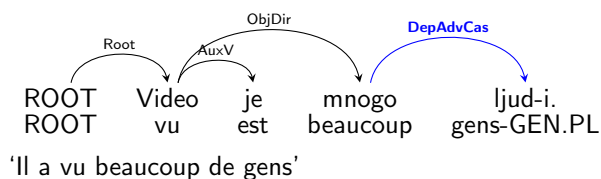
### 2.6.2 Dépendant sous forme de nom fléchi

La relation **DepAdvCas** est destinée à l'annotation des dépendants casuels d'un adverbe. Il s'agit en fait des syntagmes dits *partitifs* dans la tradition grammaticale du serbe : la tête du syntagme est un adverbe de quantité comme *malo* 'peu' ou *mnogo* 'beaucoup', qui exige un complément au génitif. Dans ce cas, nous considérons que la tête de la relation est l'adverbe, et le complément au génitif est annoté comme **DepAdvCas**.<sup>9</sup> Notons que l'adverbe lui-même porte la fonction de l'ensemble du syntagme (cf. exemple 85).

9. Les syntagmes partitifs peuvent également avoir une tête nominale, notamment sous forme d'un nom exprimant la quantité comme *čša* 'verre' ou *gomila* 'tas', qui exigent le même type de complément au génitif. Ce cas de figure est traité avec l'étiquette **DepNCas** (cf. la section 2.4.1).



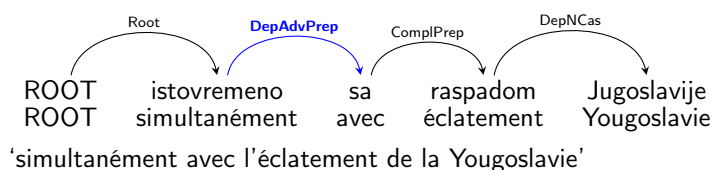
(85)



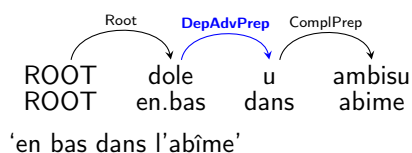
### 2.6.3 Dépendant sous forme de préposition

L'étiquette **DepAdvPrep** sert à annoter les **dépendants prépositionnels** d'un adverbe (cf. les exemples 86 à 89).

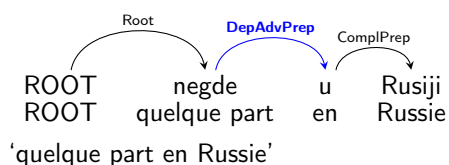
(86)



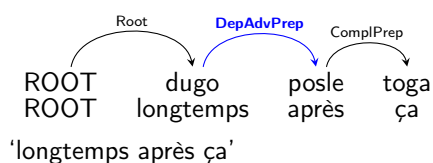
(87)



(88)



(89)

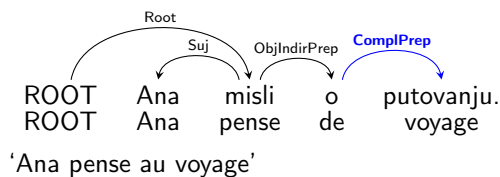


## 2.7 Relations diverses

### 2.7.1 Complément de préposition

L'étiquette **ComplPrep** est utilisée pour relier la préposition avec son dépendant. Pour rappel, la préposition elle-même est annotée avec l'étiquette de la fonction exercée par le groupe prépositionnel dans la phrase.

(90)

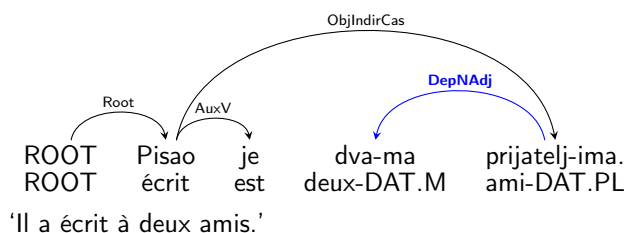


### 2.7.2 Complément de numéral

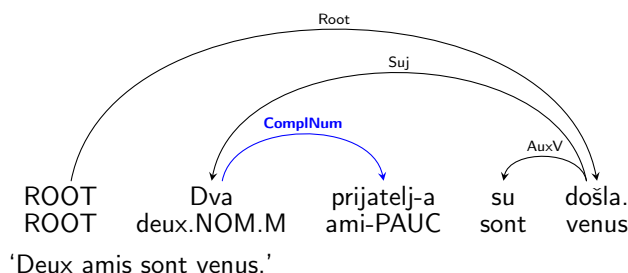
Les numéraux cardinaux *dva* 'deux', *tri* 'trois' et *četiri* 'quatre' en serbe se déclinent. Ils prennent donc par défaut le même cas que le nom (cf. exemple 91). Dans ce cas de figure, nous considérons que c'est le nom qui impose la forme au numéral, et que c'est par conséquent le nom qui est le gouverneur de la relation. Comme le numéral se comporte dans ce contexte comme tout adjectif antéposé au nom, on lui accorde l'étiquette **DepNAdj** (cf. section 2.4.3).

Cependant, dans certains cas, le numéral peut ne pas se décliner lui-même et imposer par ailleurs une forme spécifique aux noms masculins et neutres (cf. exemple 92). La forme en question s'appelle *paucal* et elle est un résidu du dual de l'ancien slave, dont l'apparition est limitée à ce contexte spécifique. Dans cette configuration, c'est le numéral qui impose la réalisation morphosyntaxique du nom, ce qui implique que c'est le numéral qui gouverne la relation. Le critère d'omissibilité corrobore également cette hypothèse : *Dva su došla* 'Deux sont venus' reste grammatical, ce qui n'est pas le cas de *\*Priatelj-a su došla* 'Amis sont venus'. Par conséquent, nous annotons le numéral avec la fonction exercée par le groupe nominal dans la phrase, et le nom est rattaché au numéral via l'étiquette **ComplNum** (cf. exemple 92).

(91)



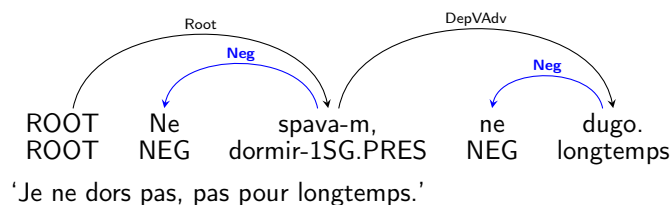
(92)



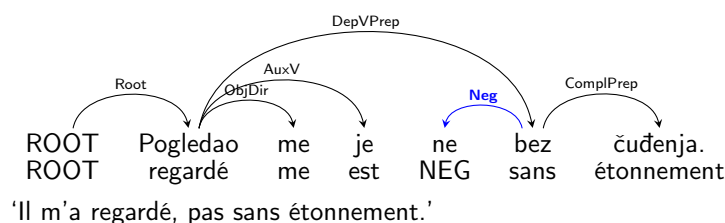
### 2.7.3 Négation

L'étiquette **Neg** est accordée à la particule de négation *ne* 'non, pas' quel que soit son gouverneur. En effet, elle est porteuse de la négation verbale, mais peut également être associée à d'autres parties du discours, par exemple aux noms ou aux adverbes (cf. exemples 93 et 94).

(93)



(94)

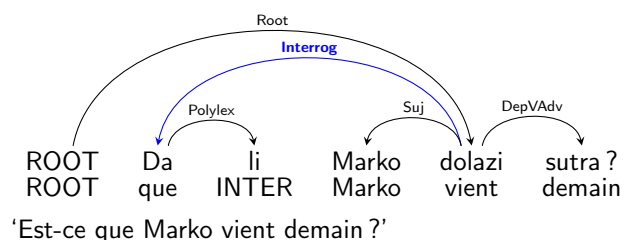


## 2.7.4 Interrogation

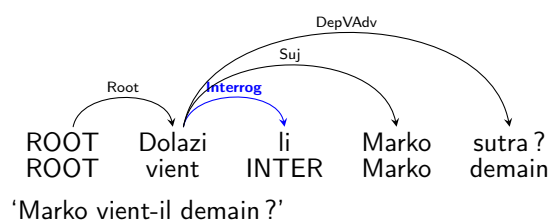
La relation **Interrog** est utilisée pour annoter les **marqueurs de la modalité interrogative** comme *da li* (cf. exemple 95) et *li* (cf. exemple 96). Le gouverneur de la relation est le verbe principal de la proposition interrogative.

Le marqueur complexe *da li* est considéré comme une forme polylexicale ; par conséquent, ses éléments sont reliés entre eux par l'étiquette **Polylex** (cf. section 2.7.8).

(95)



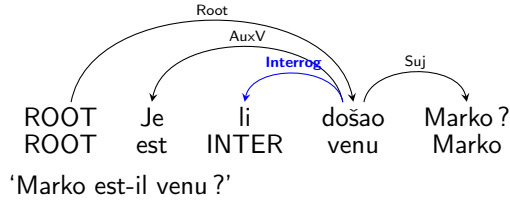
(96)



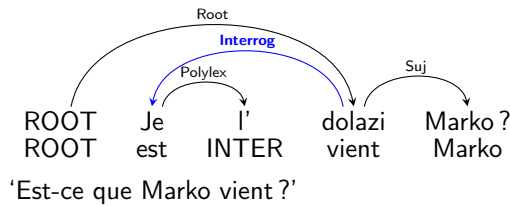
Il faut faire attention à distinguer les cas dans lesquels la forme *je li* est un marqueur d'interrogation de ceux où il s'agit d'une combinaison de la forme fléchie du verbe *jesam* 'être' et du marqueur simple *li* : dans l'exemple *Je li došao Marko ?* 'Marko est-il venu ?', la forme *je* correspond effectivement à une forme de l'auxiliaire *jesam* 'être' faisant partie de la forme complexe du parfait *je došao* 'est venu'. Ceci peut être démontré en substituant un sujet au pluriel, qui entraîne également la modification de l'auxiliaire à cause des règles d'accord : *Jesu li došli prijatelji ?* 'Les amis sont-ils venus ?', et non *\*Je li došli prijatelji ?*. Dans ce cas, *je* doit être traité en accord avec sa nature et sa fonction dans la proposition (cf.

exemple 97). En revanche, dans le langage parlé, *je li* (et sa forme abrégée *je l'*) peut avoir la fonction d'un véritable marqueur d'interrogation : *Je l' dolazi Marko ?*. Ici, la forme *je* ne peut pas être interprétée comme une forme fléchie du verbe *jesam* 'être', et *je li* bénéficie du même traitement que *da li* (cf. exemple 98).

(97)



(98)

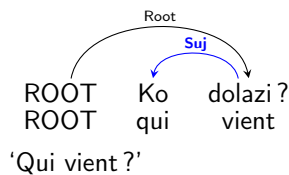


## Traitement des phrases interrogatives

Le traitement de l'interrogation totale indiquée par un marqueur d'interrogation a été montré ci-dessus. L'interrogation partielle bénéficie, en revanche, d'un traitement différent.

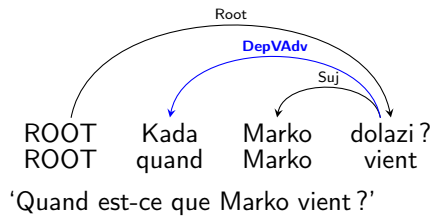
Dans le cas des questions comme *Ko dolazi ?* 'Qui vient ?' ou *Kada počinje film ?* 'Quand est-ce que commence le film ?', les formes interrogatives en gras ne sont pas de purs marqueurs de modalité. Il s'agit en effet d'un pronom interrogatif dans le cas de *ko* 'qui', et d'un adverbe interrogatif dans celui de *kada*. Les deux formes exercent des fonctions par rapport au verbe : dans le premier cas, le pronom a le rôle du sujet, et dans le deuxième, l'adverbe a le rôle d'un dépendant adverbial du verbe. Pour montrer l'équivalence de *ko* et d'un sujet typique, il suffit de comparer les phrases *Marko dolazi* 'Marko vient' et *Ko dolazi ?* 'Qui vient ?' : *Marko* et *ko* sont tous les deux au nominatif, et expriment l'agent du processus verbal. Le pronom interrogatif porte donc l'étiquette du sujet (cf. exemple 99). Un procédé comparable peut être utilisé pour l'analyse du deuxième exemple : *Kada dolazi Marko ?* 'Quand est-ce que Marko vient ?' peut être comparé à *Sutra dolazi Marko*, lit. 'Demain vient Marko'. Dans cette phrase canonique, il est facile d'identifier le rôle de l'adverbe *sutra* : il s'agit d'un dépendant du verbe *dolaziti* sous forme d'un adverbe. Il porterait donc l'étiquette **DepVAdv**. L'adverbe interrogatif *kada* dans la phrase de départ exerce exactement le même rôle par rapport à son verbe ; par conséquent, il porte la même étiquette (cf. exemple 100).

(99)



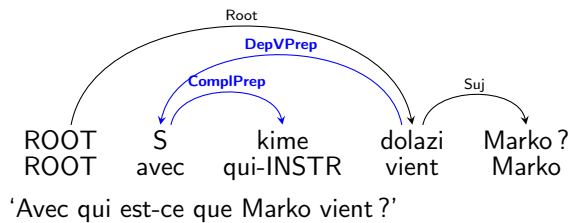
Dans le traitement des phrases interrogatives, il est donc utile de rétablir l'ordre canonique de mots et de remplacer la forme interrogative par un mot équivalent (nom, pronom ou

(100)



adverbe, en fonction du cas traité). Par exemple, la phrase *S kime Marko dolazi?* 'Avec qui vient Marko?' peut se transformer en *Marko dolazi s kime* 'Marko vient avec qui', puis en *Marko dolazi s Anom* 'Marko vient avec Ana'. En analysant cette phrase à ordre de mots canonique, nous voyons que la préposition *s* 'avec' a la fonction **DepVPrep** par rapport au verbe *dolaziti* 'venir', et que la forme à l'instrumental *Anom* est le complément de cette préposition. Le même traitement peut être transposé sur la phrase de départ : la proposition *s* 'avec' est un **DepVPrep** du verbe *dolaziti* 'venir', alors que l'instrumental *kime* est le **ComplPrep** de la préposition (cf. exemple 101).

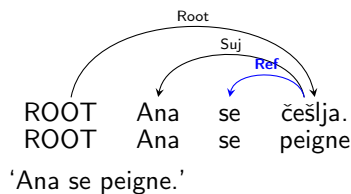
(101)



### 2.7.5 Réflexif

L'étiquette **Ref** est utilisé pour annoter le pronom réflexif sous forme clitique *se* 'se'.<sup>10</sup> Le gouverneur de la relation est le verbe pronominal.

(102)



Bien qu'on puisse distinguer plusieurs types de pronoms réfléchifs, ils seront tous annotés avec l'étiquette unique **Ref**. Ce choix est dû au fait qu'il n'y a pas de critères syntaxiques suffisamment fiables qui permettent de faire cette distinction. Dans la phrase 103, il semble évident que le pronom réflexif a le rôle de l'objet direct par rapport au verbe *osvežiti* 'rafraîchir'; il peut par ailleurs être remplacé par un autre type d'objet (cf. *Krenuo sam da ga osvežim* 'Je partis le rafraîchir'). Mais le statut du réflexif dans la phrase 104 est moins clair. À la base, *prekrstiti* 'signer' est un verbe transitif et le pronom réflexif devrait donc avoir le rôle de l'objet direct. Cependant, dans la langue contemporaine, ce verbe semble réduit à sa forme pronominale, et un remplacement du pronom réflexif par un autre objet n'est pas

10. Remarque : en serbe, il n'existe qu'une forme du pronom réflexif pour tous les nombres et toutes les personnes.

possible : \**Majka me prekrsti* ‘Ma mère me signa’. Il est donc difficile d’établir les critères de différenciation même dans le cadre d’une annotation manuelle, alors que pour un parser les différences seront presque certainement inaccessibles et très difficiles à généraliser. Par conséquent, nous retenons une seule étiquette pour tous les types du pronom réflexif.

(103)

Krenuh da se osveži-m.  
partis que REF rafraîchir-1SG.PRES.  
‘Je partis me rafraîchir.’

(104)

Moja majka se prekrsti.  
ma mère REF signa.  
‘Ma mère se signa.’

### 2.7.6 Éléments extra-prédicatifs

Les **éléments extra-prédicatifs** sont traités à l’aide de l’étiquette **ExtraPred**. Comme il a été mentionné dans la section 2.2.3, on peut considérer que ces éléments, étant des modificateurs phrastiques, doivent se trouver au même niveau de la structure syntaxique que le contenu propositionnel. En revanche, dans le cadre de ce projet, ils seront annotés comme étant gouvernés par la tête de la proposition qu’ils modifient. Cette approche permet de minimiser la création des arcs non-projectifs, étant donné que les éléments extra-prédicatifs peuvent apparaître à l’intérieur de la proposition aussi.

### 2.7.7 Emphase

L’étiquette **Emph** est destinée aux éléments n’ayant pas de rôle syntaxique net dans la phrase et dont la fonction sémantique est de souligner une partie du contenu phrastique. Il s’agit typiquement de particules et d’adverbes, mais aussi d’interjections. Le gouverneur doit être déterminé au cas par cas, mais il s’agira idéalement de la racine de la proposition dans laquelle figure l’emphase.

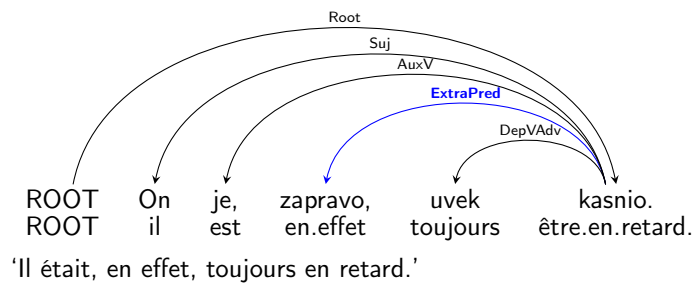
En ce qui concerne les interjections, étant donné qu’elles portent le plus souvent sur la totalité de la proposition, elles sont rattachées à la racine de la proposition dans laquelle elles figurent, qu’il s’agisse d’une racine verbale ou non.

Un emploi particulier concerne la particule *to* soulignant la totalité du contenu d’une proposition (exemple 110). À ne pas confondre avec le sujet sous forme du pronom démonstratif *to* : dans le cas où il s’agit d’une particule, la position du sujet est déjà occupée par un autre élément de la phrase, et la forme *to* ne correspond en effet à aucune fonction syntaxique auprès du verbe.

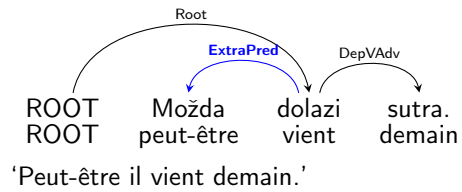
Les formes qui ont typiquement le rôle des conjonctions de coordination, tel que *i* ‘et’ et *ni* ‘ni’, peuvent se trouver dans le rôle d’emphase (cf. exemples 111, 112). Le critère pour distinguer ces cas de figure de la coordination se trouve dans le fait qu’avec l’emphase il n’est pas possible d’identifier l’autre élément de la coordination. L’emphase peut également porter sur d’autres types d’éléments, par exemple sur des noms ou des groupes prépositionnels (cf. exemples 115 et 116).

Il est à noter que cette étiquette ne doit être utilisée que si aucune autre ne s’applique. Autrement dit, si une particule dépend d’un verbe, elle sera préférablement traitée comme **DepVAdv**, et non pas comme **Emph**.

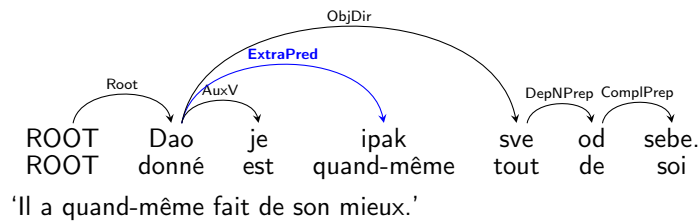
(105)



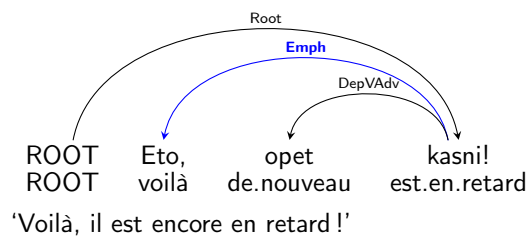
(106)



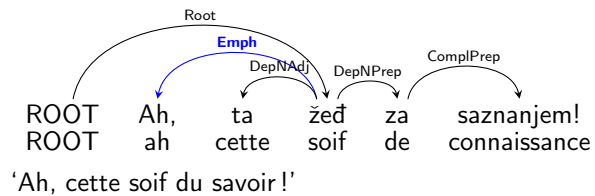
(107)



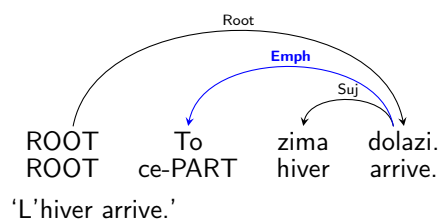
(108)



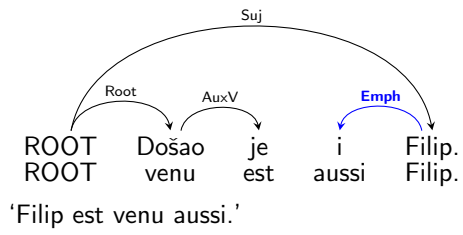
(109)



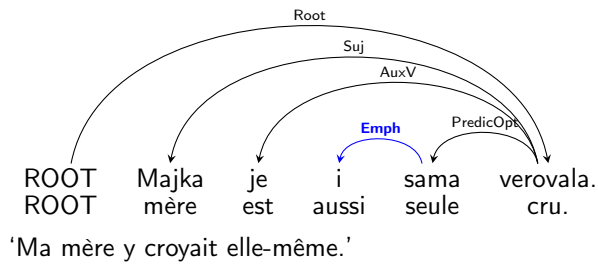
(110)



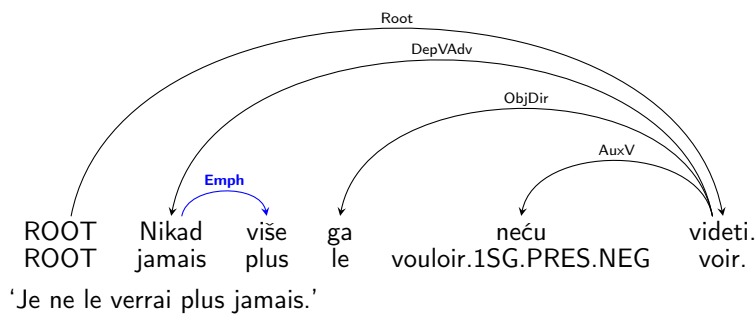
(111)



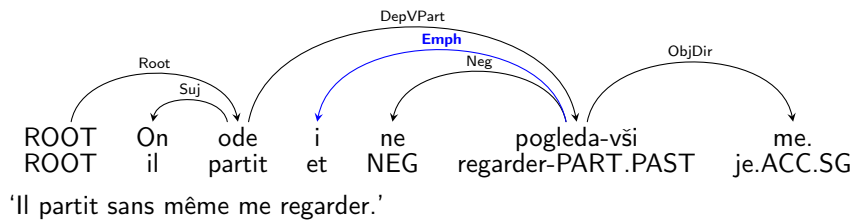
(112)



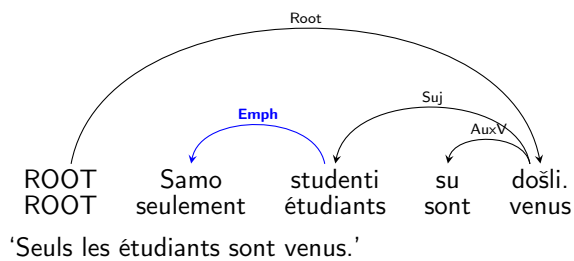
(113)



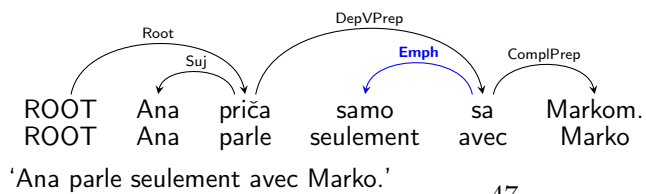
(114)



(115)



(116)





## 2.7.8 Éléments polylexicaux

L'étiquette **Polylex** est une abréviation de *polylexical* et on l'utilise pour relier les éléments d'une unité polylexicale entre eux. Elle est utilisée seulement pour les véritables figements grammaticaux : elle concerne en premier lieu les locutions conjonctives et adverbiales, et n'est appliquée ni aux phraséologismes, ni aux collocations. En revanche, elle est utilisée pour l'annotation des numéraux cardinaux et ordinaux complexes, ainsi que pour le traitement des pronoms indéfinis polylexicaux (*ma ko* 'toute personne', *bilo ko* 'n'importe qui'), ainsi que lors du clivage d'un pronom indéfini clivé par une préposition (cf. *ni za šta* 'pour rien'). Elle est également utilisée pour relier les formes des mots étrangers entre elles quand ces formes-là constituent des suites de tokens. Des exemples de ces différents cas de figure sont donnés dans la suite.

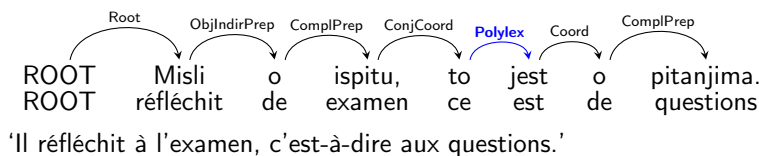
Dans le traitement de ces formes, nous retenons deux principes de base : le premier élément de l'unité polylexicale porte l'étiquette de la fonction exercée par l'unité entière, et les éléments de l'unité polylexicale sont reliés entre eux par la relation **Polylex** de gauche à droite, en cascade.

**NB**

Les exemples donnés dans la suite recensent les structures traitées comme polylexicales dans le corpus à présent. Ce recensement n'est sans doute pas exhaustif (de nouvelles structures polylexicales peuvent être rencontrées en corpus), mais les constructions qui y figurent doivent être systématiquement traitées en accord avec le Guide.

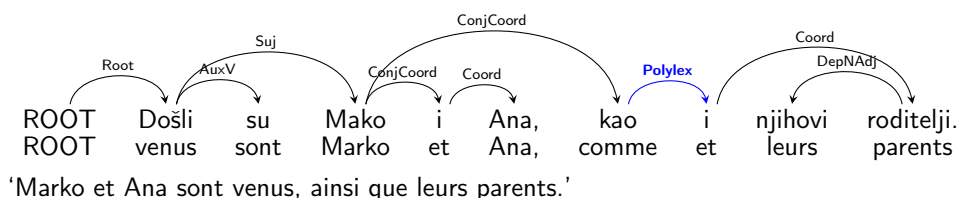
Il existe aussi bien des conjonctions de coordination que des conjonctions de subordination polylexicales.<sup>11</sup>

(117)



Strictement parlant, il ne s'agit pas ici d'une coordination, mais d'une reprise d'un élément du discours avec le but d'apporter une précision. Ce rapport relève cependant des relations de discours, et non pas des relations syntaxiques. Il est par ailleurs évident que les deux éléments sont mis en parallèle. Pour ces raisons, nous décidons de traiter ce type de structures comme coordination.

(118)



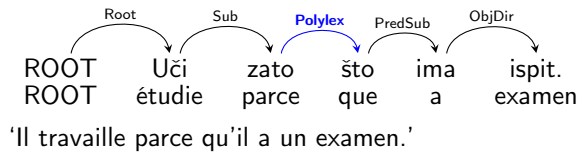
La situation est comparable dans l'exemple 118 : il s'agit, au niveau discursif, d'une précision plutôt que d'une coordination. Or, ce dernier élément est mis en parallèle avec

11. Pour le traitement de la coordination de base, voir section 2.10

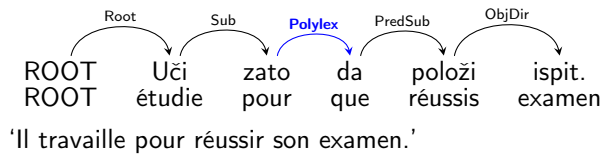
la première partie de la coordination, et la forme *kao* semble avoir le rôle de souligner la conjonction de coordination *i*. Par conséquent, nous choisissons de considérer qu’il s’agit ici d’une forme de coordination.

Dans la suite, nous listons d’autres locutions conjonctives polylexicales et indiquons leur traitement.

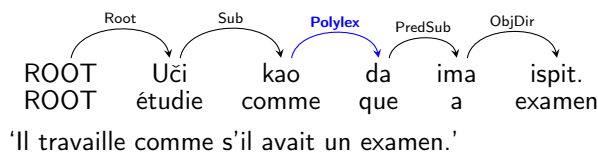
(119)



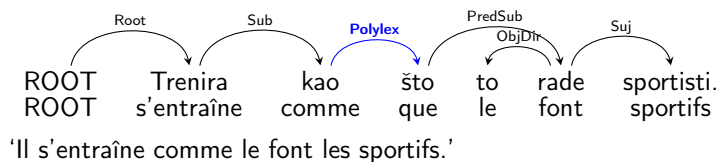
(120)



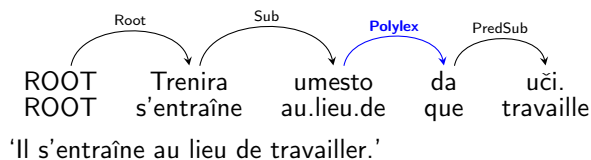
(121)



(122)



(123)

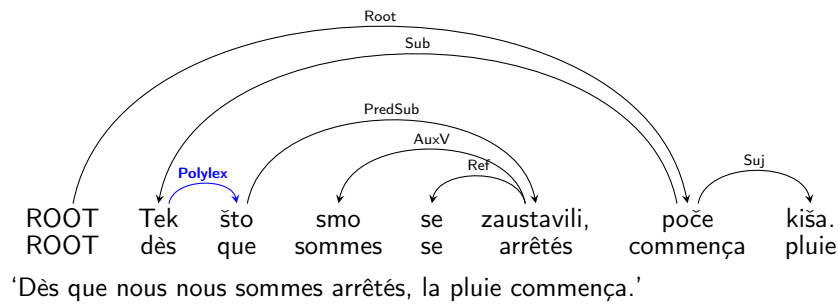


Les propositions comparatives exprimant une comparaison d'inégalité introduite par *nego što* relèvent des structures corrélatives (cf. exemple 127). Pour plus de détails sur ce sujet, voir la section 2.8.5.

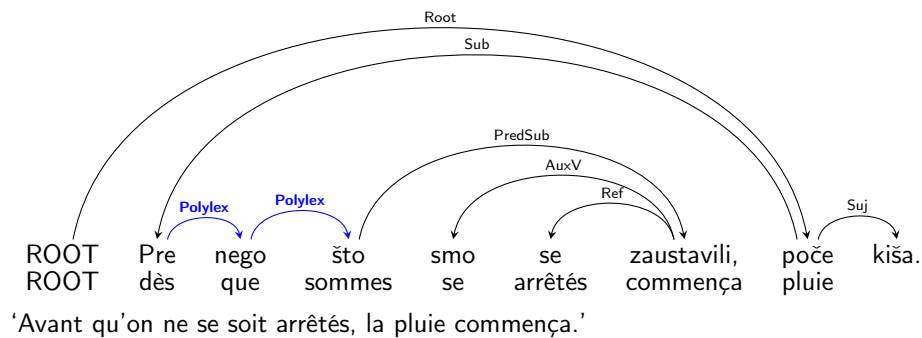
Une attention spéciale doit être accordée à la structure *tako da* 'de sorte que', qui peut avoir deux rôles, et par conséquent, deux analyses syntaxiques différentes.

Dans le premier cas possible, *tako* 'ainsi' est un véritable adverbe de manière qui introduit une **structure corrélatrice** (cf. exemple 129).

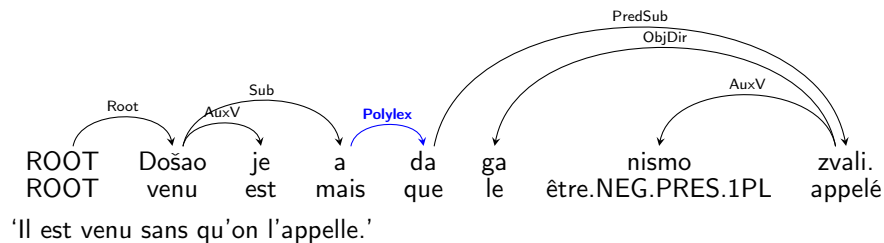
(124)



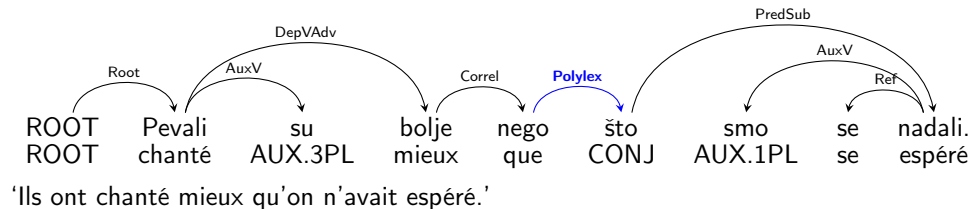
(125)



(126)



(127)

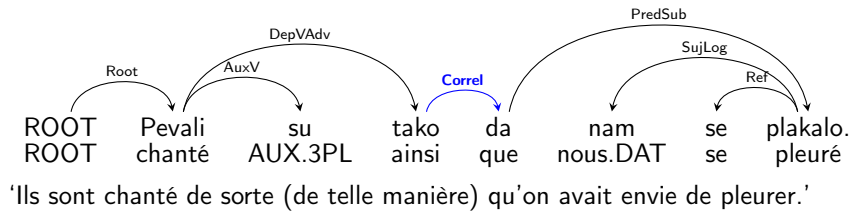


Dans le deuxième cas de figure, il s'agit d'une **conjonction polylexicale**, et la forme *tako* n'est plus un dépendant verbal sous forme d'adverbe (cf. exemple 129).

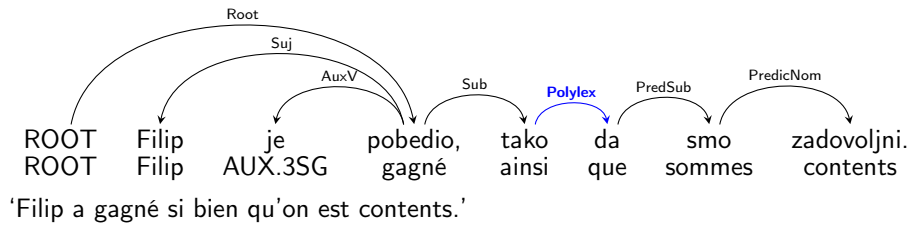
Les **numéraux cardinaux et ordinaux complexes** (*dvadest dva* 'vingt-deux', *dvadeset drugi* 'vingt-deuxième') comme des éléments polylexicaux, qu'ils contiennent la conjonction *i* ou non.

Cet emploi concerne également l'expression de l'heure :

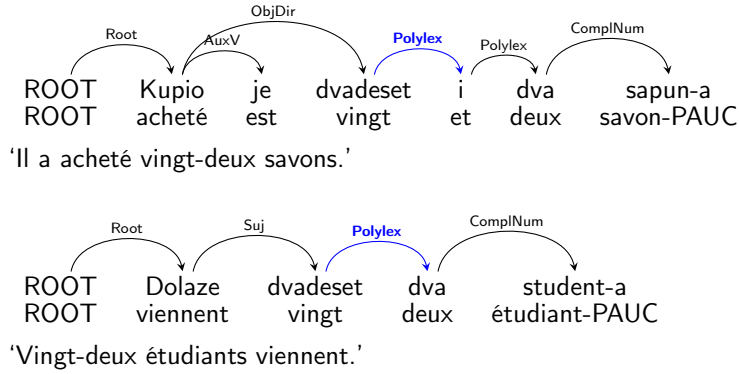
(128)



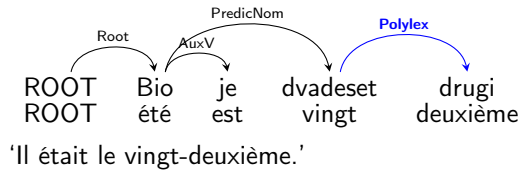
(129)



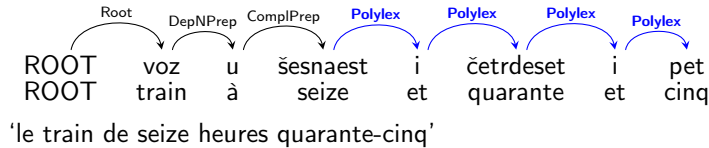
(130)



(131)

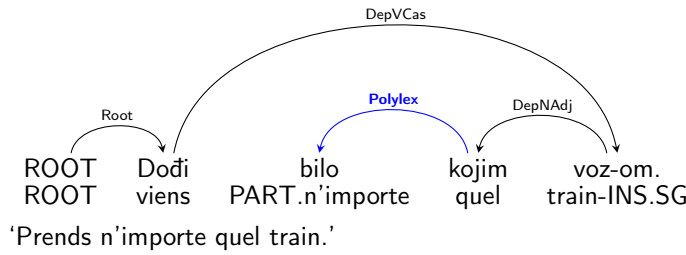


(132)

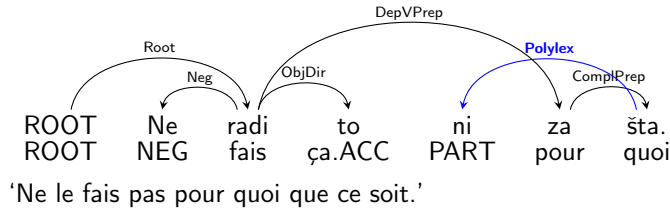


Cette relation est également utilisée pour relier les parties des **pronoms et adjectifs indéfinis polylexicaux**, comme *bilo koji* 'n'importe lequel', *ma ko* 'n'importe qui', mais aussi dans les cas où un **pronom indéfini simple** comme *ništa* 'rien' ou *iko* 'd'aucuns' a le rôle du complément d'une préposition et se voit **clivé** par cette préposition (cf. *ni za šta* 'pour rien', *i za koga* 'pour qui que ce soit').

(133)

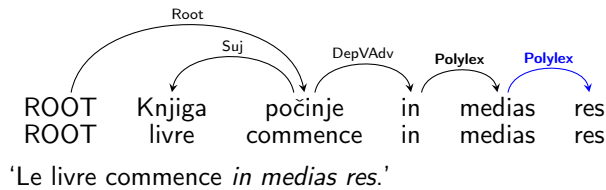


(134)

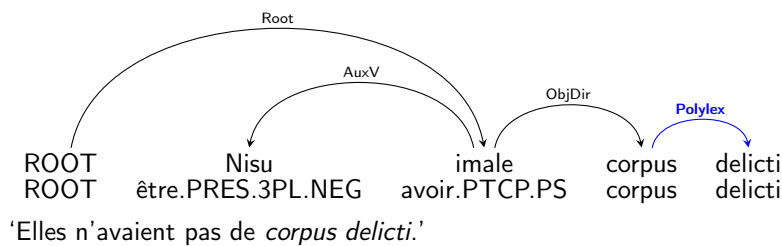


En ce qui concerne les **suites de plusieurs tokens provenant d'une langue étrangère**, le premier est annoté avec la fonction que le groupe exerce dans la phrase, alors que les autres sont reliés en cascade par l'étiquette **Polylex**.

(135)



(136)

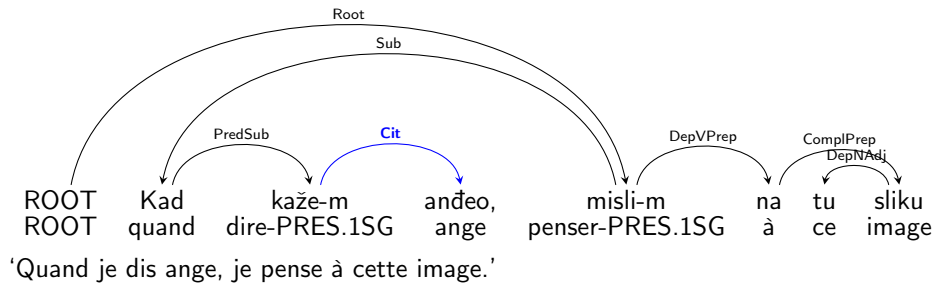


## 2.7.9 Citations et emplois métalinguistiques

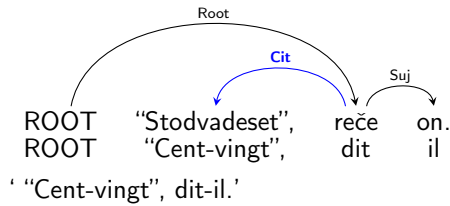
L'étiquette **Cit** est utilisée pour marquer les emplois métalinguistiques de mots individuels ainsi que pour les segments du discours rapporté averbaux.

Ces éléments pourraient être considérés comme des objets directs des verbes de parole qui les introduisent. En revanche, pour le premier exemple, cette approche signifierait qu'une forme au nominatif peut exercer la fonction de l'objet direct, ce qui semble peu justifié. Pour ce qui est du deuxième cas de figure, le verbe introducteur peut facilement ne pas être un verbe transitif direct : *Sto dvadeset, uvredi se on* 'Cent-vingt, s'offusqua-t-il.' Pour simplifier le traitement, nous choisissons de considérer que ces formes ont un lien moins direct avec le reste de l'arbre syntaxique et les annotons simplement comme morceaux de discours intervenus dans la phrase.

(137)



(138)



### 2.7.10 Ponctuation

Toutes les ponctuations sont traitées par l'étiquette **Ponct**. L'approche adoptée est celle signalée comme la plus performante par A. Urieli : chaque signe de ponctuation est rattaché au premier token précédent qui n'en est pas une. Cette approche permet d'avoir un traitement systématique des ponctuations, facile à effectuer de manière automatique. Par conséquent, il n'est pas nécessaire d'annoter la ponctuation dans le cadre du traitement manuel.

## 2.8 Subordination

Dans le cadre du projet *ParCoLab*, nous ne reprenons pas la typologie traditionnelle des subordonnées en serbe. Nous introduisons un traitement plus global, qui distingue 5 types de subordonnées basé sur leurs propriétés syntaxiques : celui des subordonnées adverbiales à subordonnant simple (section 2.8.1), celui des complétives (section 2.8.2), celui des interrogatives indirectes (section 2.8.3), celui des relatives (section 2.8.4) et celui des subordonnées corrélatives (les comparatives et les consécutives - section 2.8.5).

Le traitement de base de la subordination est celui des subordonnées adverbiales : ces propositions sont introduites par un subordonnant dont la seule fonction syntaxique est d'assurer l'inclusion de la subordonnée dans la proposition principale. Les quatre traitements restants représentent des variations de ce traitement de base. Nous distinguons les complétives du fait de leur statut spécifique par rapport au verbe qui les introduit (elles font souvent partie de la structure argumentale du verbe). Les relatives et les interrogatives indirectes ont la spécificité d'être introduites par un subordonnant à double fonction syntaxique, qui, outre son rôle de subordination, effectue aussi une fonction à l'intérieur de la subordonnée. Enfin, les propositions corrélatives ont une structure syntaxique spécifique qui mérite un traitement à part. Le traitement de chacun de ces cas de figure est présenté dans la suite.

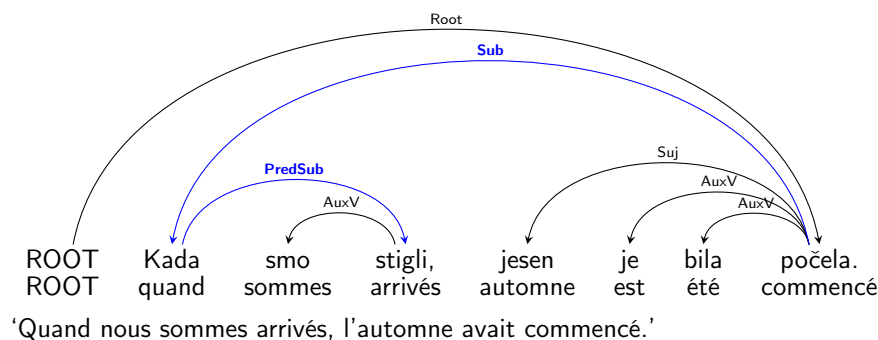
### 2.8.1 Subordonnées adverbiales à subordonnant mono-fonctionnel

Nous ne faisons pas la distinction entre les différents types sémantiques de subordonnées adverbiales. Le seul critère pour appliquer le traitement décrit dans cette section à une proposition subordonnée est qu'elle soit dotée d'un subordonnant à fonction syntaxique unique, servant seulement à établir le lien avec la proposition principale. À titre d'exemple, ceci est

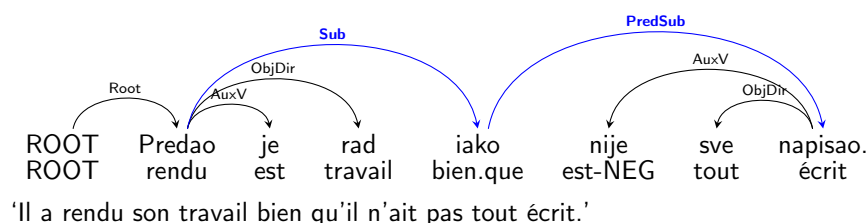
le cas des conjonctions comme *kada* ‘quand’, *pošto* ‘après que, puisque’, *jer* ‘parce que’, *iako* ‘bien que’, *ako* ‘si’, etc.

Deux étiquettes sont liées à ce traitement : **Sub** est l’étiquette qui relie le verbe de la principale et le **subordonnant**, et **PredSub** est utilisée pour établir le lien entre le subordonnant et le **verbe principal de la subordonnée**.

(139)



(140)



S’il s’agit d’un subordonnant polylexical, comme *nakon što* ou *zato što*, il faut faire appel à l’étiquette **Polylex**, dont l’utilisation est présentée dans la section 2.7.8.

**NB**

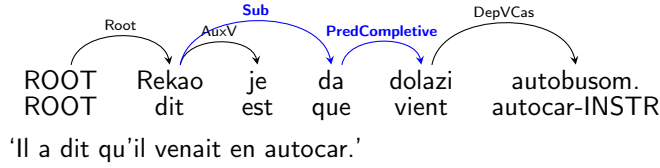
Ce traitement *ne concerne pas* les subordonnées en *da* ‘que’ complétant les verbes aspectuels ou modaux, ni les propositions déclaratives. Voir la section 2.8.2.

## 2.8.2 Subordonnées complétives

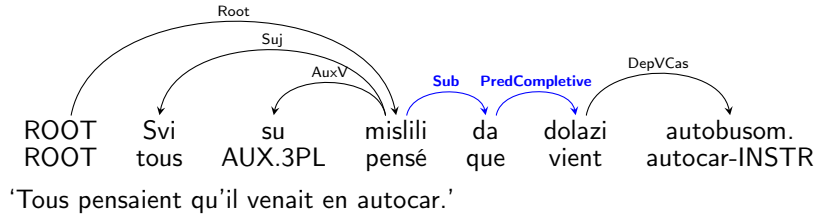
Nous traitons comme complétives les **propositions** introduites par des **verbes de parole** (*govoriti* ‘parler’, *reći* ‘dire’, *pričati* ‘raconter’, etc.) ou **de processus mentaux** (*misliti* ‘penser’, *smatrati* ‘considérer’, *verovati* ‘croire’, etc.), introduites par la **conjonction *da* ‘que’** ou ***kako* ‘comment’**, ainsi que les propositions introduites par des **verbes de perception** (cf. *gledati* ‘regarder’, *slušati* ‘écouter’) avec la conjonction ***kako* ‘comment’**. La relation entre le verbe de la principale et le **subordonnant** est la même que pour les propositions adverbiales : **Sub**. En revanche, pour indiquer la nature spécifique des complétives, la relation allant du subordonnant vers le **verbe principal de la subordonnée** est **PredComplative** (prédicat de la complétive).

Les complétives dans les exemples 153, 142 et 143 représentent l’objet direct du verbe qui les introduit. Nous avons cependant écarté la possibilité de les traiter comme tel pour le fait que toutes les complétives en serbe ne sont pas des objets directs. Par exemple, les verbes *verovati* ‘croire’, *spremati se* ‘se préparer’ et *uspeti* ‘réussir’ admettent tous des compléments

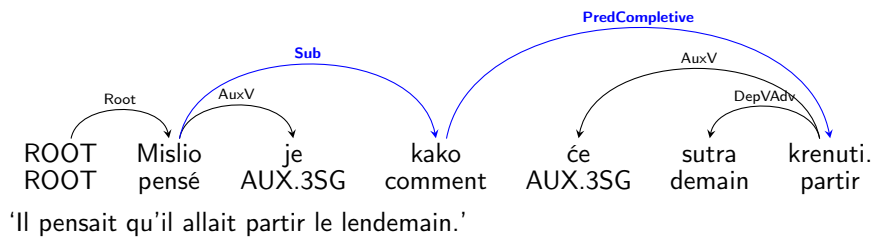
(141)



(142)



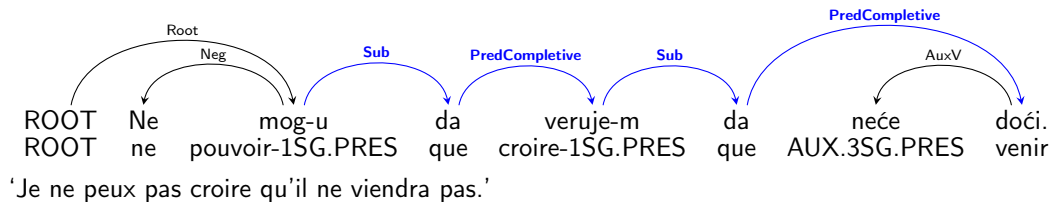
(143)



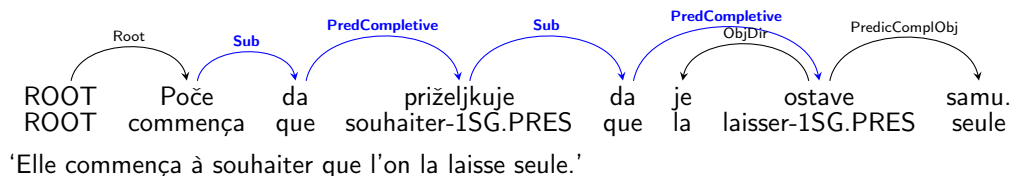
sous forme d'une proposition en *da*. Cependant, tous les trois ont des constructions de base différentes : *verovati u nešto* 'croire en quelque chose', *spremati se na nešto* lit. 'se préparer à quelque chose', *uspeti u nečemu* lit. 'réussir en quelque chose'. On peut donc difficilement affirmer que les complétives en *da* 'que' introduites par ces verbes soient des objets directs, vu que les verbes eux-mêmes n'ouvrent pas cette position dans leur structure argumentale. Ainsi, toutes ces propositions sont simplement traitées comme des complétives, alors que leur statut par rapport au verbe gouverneur reste sous-spécifié.

Nous considérons également comme complétives les constructions traitées dans la tradition grammaticale serbe comme **prédicats complexes**. Autrement dit, les compléments des verbes modaux (*morati* 'devoir', *moći* 'pouvoir', etc.) et aspectuels (*početi* 'commencer', *prestati* 'arrêter', etc.) sous forme de la construction *da* 'que' +  $V_{present}$  sont annotés comme complétives.

(144)



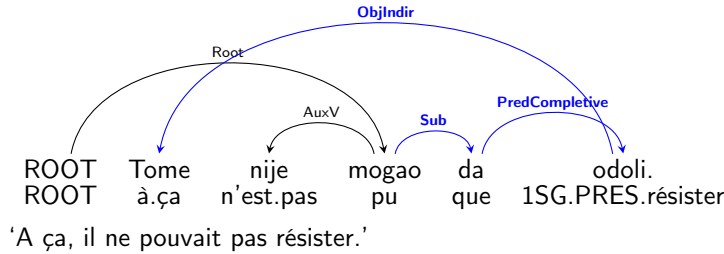
(145)





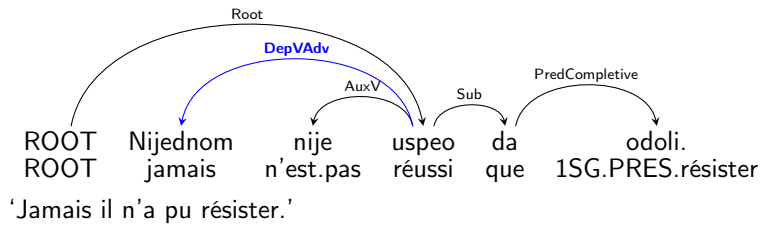
Dans ces constructions, il faut veiller à identifier correctement le verbe qui gouverne les dépendants nominaux : un objet direct ou indirect du verbe de la complétive peut se trouver en dehors de sa proposition. Il doit tout de même être rattaché au verbe qui l'introduit, et non pas au verbe introducteur de la relative (cf. exemple 146).

(146)



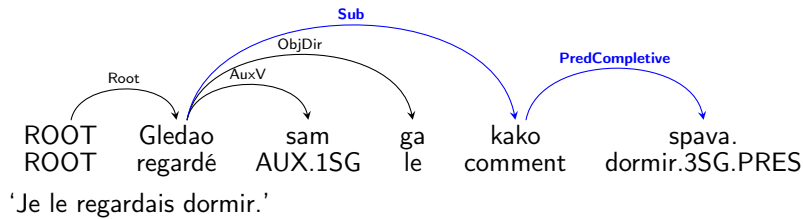
Or, dans le cas des dépendants adverbiaux, il n'est pas possible d'identifier lequel des deux verbes est la tête en utilisant des critères de surface. Par conséquent, ce type de dépendant sera rattaché au verbe introducteur de la complétive.

(147)



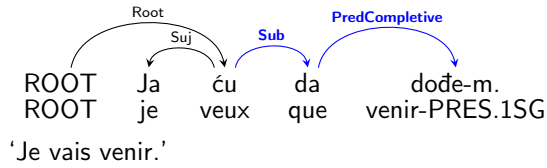
Une autre sous-classe des complétives concerne les verbes de perception introduisant une proposition en *kako* 'comment'. Ces propositions sont équivalentes de la constructions infinitivales en français du type *Je le regardais dormir*, d'autant plus que la proposition principale en serbe contient également un objet direct à l'accusatif.

(148)



Nous annotons de la même manière la **forme irrégulière du futur** comme dans *ja ću da dodem* 'je vais venir', construite de l'auxiliaire *hteti* 'vouloir' qui n'est pas suivi d'un infinitif (comme c'est le cas pour les formes régulières du futur), mais d'une proposition en *da* 'que' contenant le verbe principal au présent. Dans ce cas de figure, étant donné que la forme du verbe *hteti* 'vouloir' n'est pas suivie d'un participe, nous considérons qu'il s'agit d'un verbe principal, et la construction *da* 'que' +  $V_{present}$  est traitée comme une complétive.

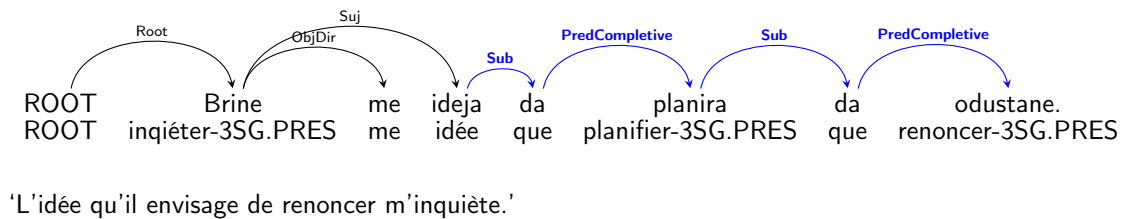
(149)



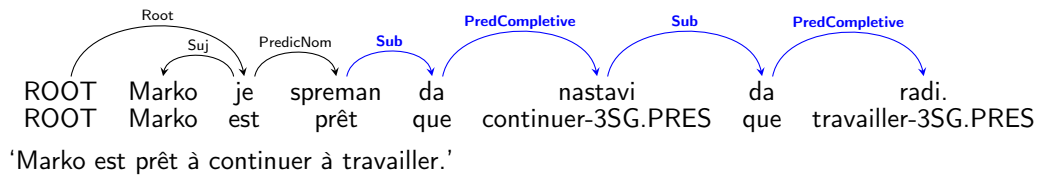
### Complétives introduites par des parties du discours autres qu'un verbe

Les complétives peuvent également être introduites par certains noms et adjectifs, voire par des particules. Elles expriment alors le contenu encapsulé par le mot introducteur : *ideja da će Marko doći, rešen da ne odustane, jedva da diše*.

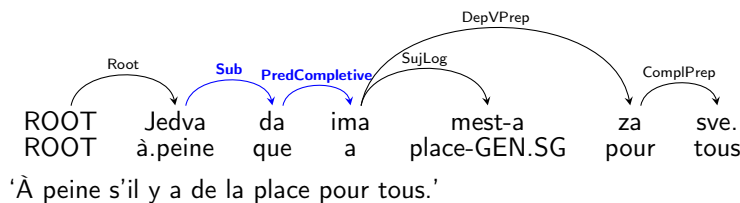
(150)



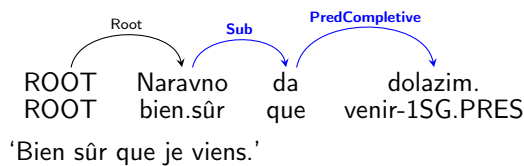
(151)



(152)



(153)



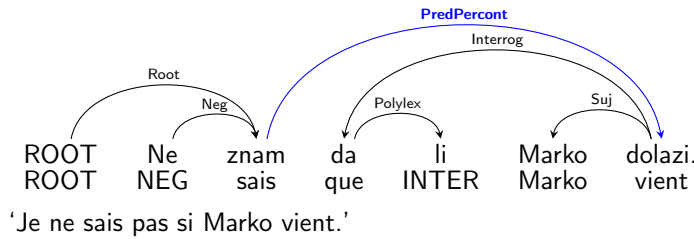
### 2.8.3 Subordonnées interrogatives indirectes

Les subordonnées interrogatives indirectes diffèrent des subordonnées adverbiales et des complétives par le fait que leur subordonnant a une double fonction syntaxique : il établit le lien entre la principale et la subordonnée, mais il a également une fonction à l'intérieur de la subordonnée.<sup>12</sup> Le comportement syntaxique des formes interrogatives a déjà été expliqué dans la section 2.7.4 et il a été constaté qu'un interrogatif peut avoir la fonction d'un sujet,

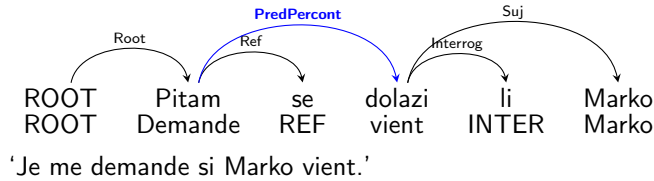
12. Ceci est également le cas des relatives : voir la section 2.8.4.

objet direct ou indirect, dépendant verbal ou dépendant nominal au sein de la proposition dans laquelle il figure. Il en est de même des interrogatifs dans les interrogatives indirectes subordonnées. En revanche, comme il a déjà été mentionné, dans ce cas de figure les interrogatifs ont également une deuxième fonction : celle de la subordination. On souhaiterait donc marquer les deux dépendances de ces formes : celle par rapport à la proposition subordonnée elle-même, mais aussi celle par rapport au verbe de la principale. Or, ceci est impossible : comme il a été expliqué dans les règles générales de l'analyse en dépendances, un élément de l'arbre ne peut avoir qu'un seul gouverneur. Il faut donc choisir entre les deux dépendances. Comme l'interrogatif occupe souvent une position syntaxique qui ne peut pas rester vide dans la proposition interrogative, nous préférons encoder celle-ci. Les interrogatives indirectes sont donc traitées comme suit : c'est le verbe de la principale qui est relié directement au verbe de la subordonnée par la relation **PredPercont** (= prédicat de percontative), et le contenu de la subordonnée est traité en accord avec les principes décrits dans la section 2.7.4.

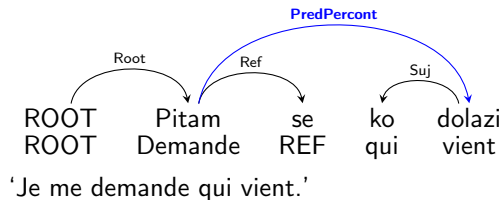
(154)



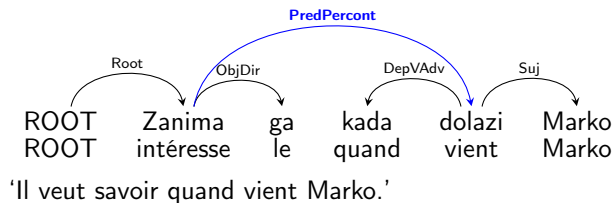
(155)



(156)

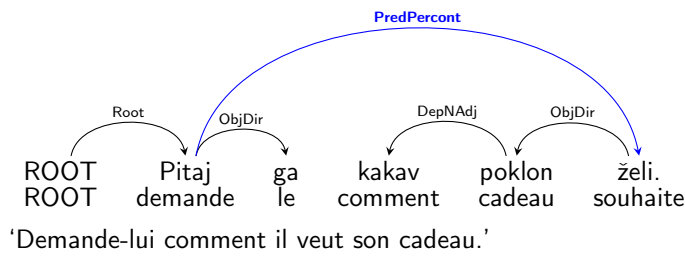


(157)



Pour une méthode qui facilite l'identification de la fonction syntaxique de l'interrogatif au sein de la subordonnée, voir la section 2.7.4.

(158)



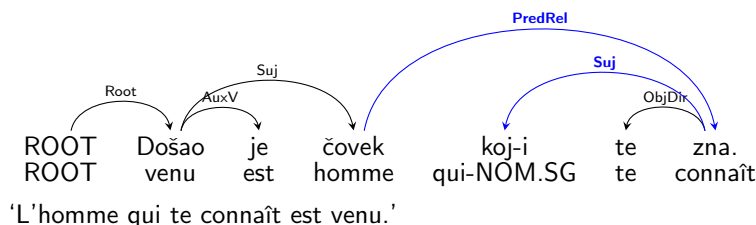
### Rappel

Dans le cas des interrogatives exprimant une interrogation totale, le marqueur d'interrogation n'a pas de rôle spécifique par rapport au prédicat de la subordonnée. Par conséquent, il est relié au verbe de la subordonnée interrogative par la relation **Interrog**.

## 2.8.4 Subordonnées relatives

Les relatifs exhibent un comportement syntaxique proche de celui des interrogatifs : au-delà de leur fonction de subordination, ils exercent également une fonction syntaxique à l'intérieur de la proposition subordonnée qu'ils introduisent. La même question se pose donc ici : il faut décider si le relatif doit être gouverné par son antécédent en tant que subordonnant ou bien par l'élément dont il dépend à l'intérieur de la relative. Ici aussi, nous optons pour la fonction du relatif exercée à l'intérieure de la subordonnée. Ceci est notamment fait pour maintenir la représentation de la structure argumentale des verbes dans les relatives. Par conséquent, le lien entre la proposition principale et la relative s'établit sans passer par la forme relative. À la différence du traitement des interrogatives indirectes, le lien ne s'établit pas entre les prédicats de la principale et de la subordonnée : la relative est plutôt introduite par l'antécédent du relatif, ce qui permet d'indiquer le rôle de cet élément dans la structure relative. L'étiquette utilisée pour relier l'**antécédent au prédicat de la relative** est **PredRel** (= prédicat de relative). Le relatif est annoté avec l'étiquette qui exprime le mieux sa fonction à l'intérieur de la relative (tout comme les interrogatifs) (cf. exemples ci-dessous).

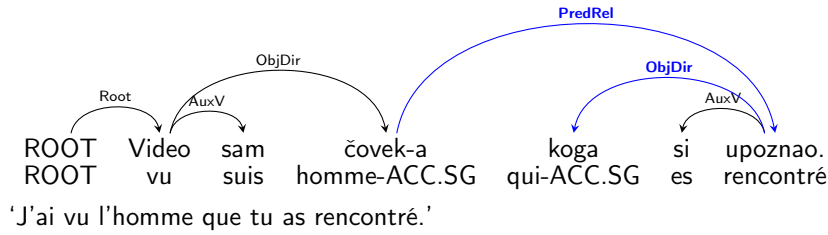
(159)



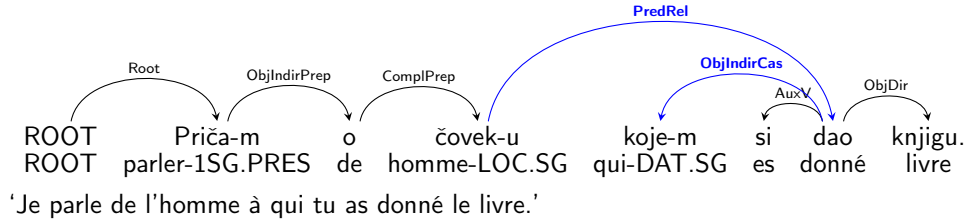
Il faut souligner que l'antécédent peut avoir des fonctions différentes dans la proposition principale, et il en est de même pour le relatif au sein de la relative.

Le relatif peut également ne pas dépendre directement du verbe de la relative, mais faire partie d'une construction prépositionnelle introduite par le verbe (cf. exemple 162). Dans l'identification de la fonction du relatif, il peut être utile de remplacer le pronom relatif par un pronom personnel ou démonstratif et de rétablir l'ordre des mots canonique dans la relative : *o kojem sam ti pričao* peut se transformer ainsi en *o njemu sam ti pričao*, et ensuite

(160)

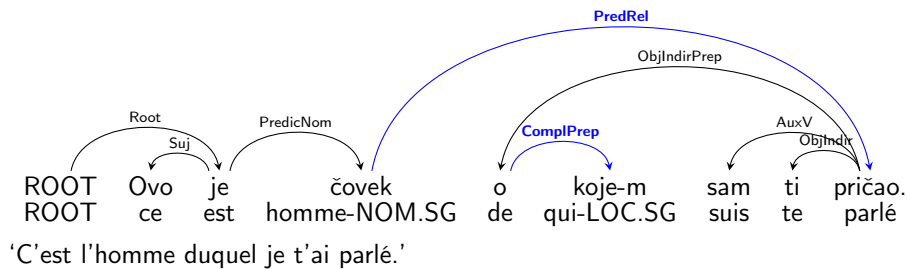


(161)



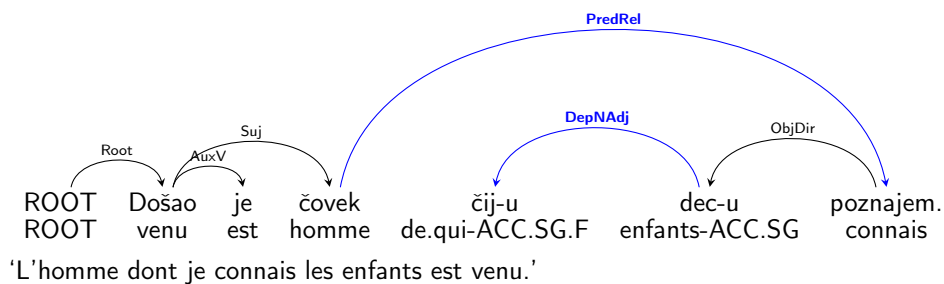
en *pričao sam ti o njemu*. Cette manipulation peut aider à voir plus nettement que le verbe *pričao* introduit la préposition *o*, qui, à son tour, introduit une forme au locatif (*njemu* dans la proposition transformée, ou bien *kojem* dans la proposition de départ). La proposition correspond donc à la relation **ObjIndirPrep**, et le pronom (dans les deux cas) a le rôle du complément de la préposition.

(162)



L'adjectif relatif *čiji* 'de qui' fonctionne comme tout autre adjectif : il s'accorde pleinement avec le nom dont il dépend, et il lui est antéposé. Par conséquent, il est traité en utilisant la relation **DepNAdj**.

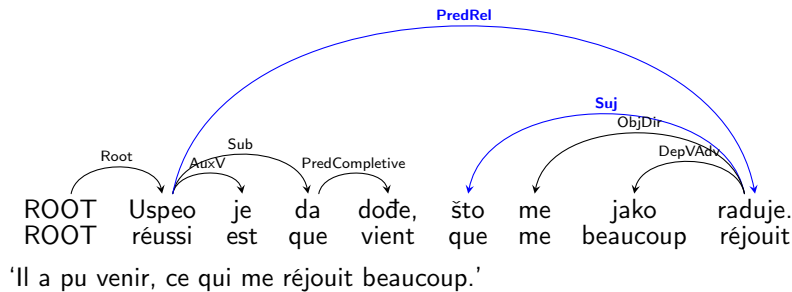
(163)



Il existe également des **relatives qui ont pour antécédent toute la proposition principale**, comme dans *Uspeo je da dode, što me jako raduje* 'Il a pu venir, ce qui me

réjouit beaucoup’. Comme le prédicat de la relative est introduit toujours par la tête de l’antécédent, dans ce cas de figure, c’est le prédicat de la principale qui introduit le prédicat de la relative (cf. exemple 164).

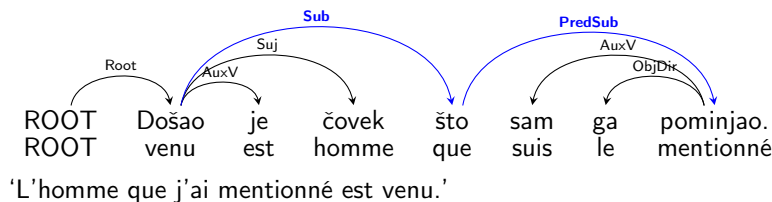
(164)



**NB**

Les relatives introduites par *sto* 'que' dans lesquelles la fonction censée être exercée par le relatif est reprise par le pronom personnel ne sont pas traitées de la même manière. En effet, leur traitement est celui des subordonnées à subordonnant simple (cf. exemple 165). Ceci se justifie par le fait que le relatif n'a plus de rôle syntaxique par rapport au verbe de la relative, celui-ci étant déjà accompagné d'un objet direct (le pronom personnel *ga* 'le').

(165)

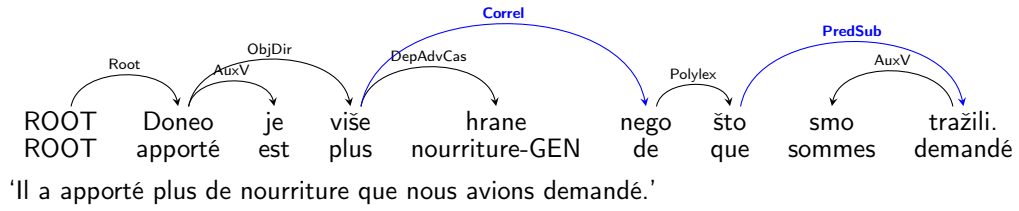


### 2.8.5 Subordonnées corrélatives

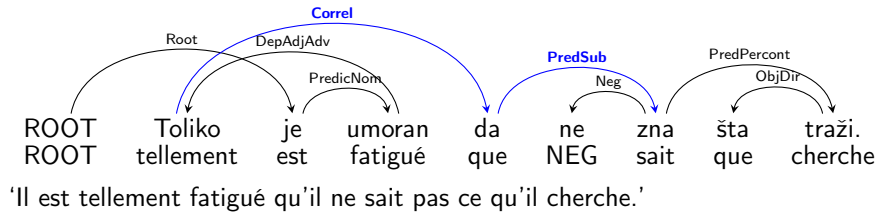
Cette section est consacrée aux subordonnées introduites par un corrélatif : il s'agit d'un élément qui a une fonction syntaxique dans la principale, et qui appelle un autre élément dans la subordonnée : la présence du premier terme étant subordonnée à la présence du second et réciproquement, on parle de corrélation. Il s'agit notamment de propositions consécutives et comparatives.

La relation de subordination s'établit donc entre le corrélatif et le subordonnant via la relation **Correl**. Le subordonnant gouverne à son tour le verbe de la subordonnée à travers la relation **PredSub**.

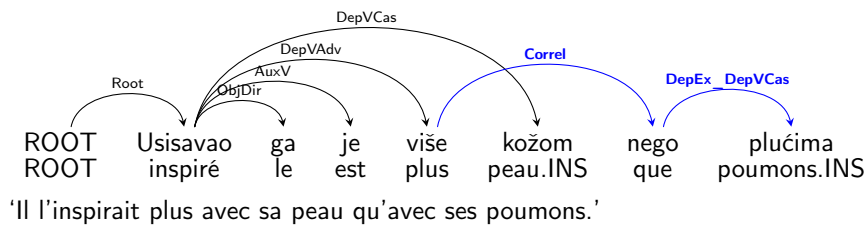
(166)



(167)



(168)



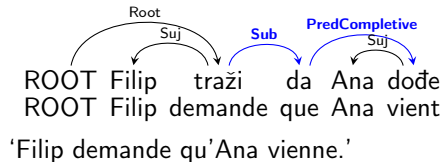
Dans l'exemple 168, nous trouvons également une proposition averbale, dont le verbe a été éliminé. Le traitement de l'ellipse est présenté en détail dans la section 2.12.

## 2.8.6 Ambiguïté des subordonnées en *da*

Une attention particulière doit être accordée au traitement des subordonnées introduites par la conjonction *da*. En effet, cette conjonction est polysémique et peut introduire différents types de subordonnées.

Premièrement, la conjonction *da* est la conjonction prototypique pour les **complétives** (cf. section 2.8.2). Notons que dans ce cas la subordonnée est introduite par un verbe de parole ou de processus mental et correspond souvent à un argument du verbe. Par conséquent, elle est difficilement omissible (cf. exemple 169).

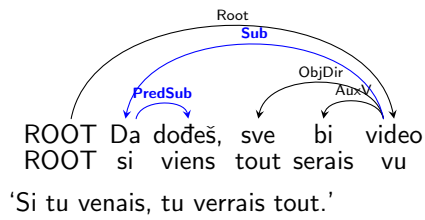
(169)



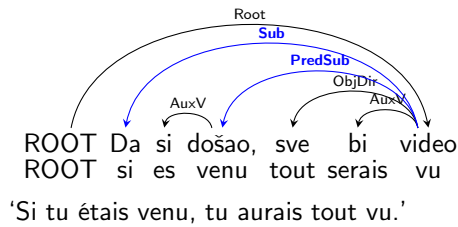
Deuxièmement, la conjonction *da* peut introduire une proposition hypothétique irréaliste (cf. exemples 170 et 171). Notons que dans ce cas la conjonction introduit le présent d'un verbe imparfaitif (cf. exemple 170) ou le parfait (cf. exemple 171).

Troisièmement, la conjonction *da* peut également introduire une proposition finale (cf. exemple 172). Dans ce cas, le verbe dans la subordonnée est au présent ou au potentiel, et ces

(170)

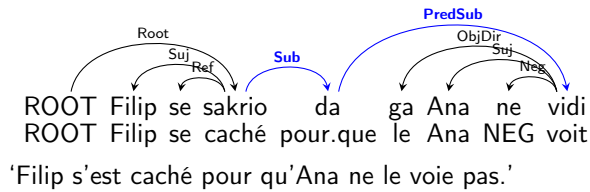


(171)



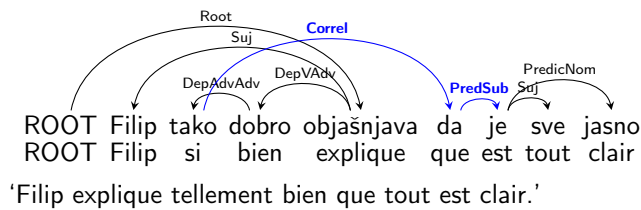
deux formes verbales sont interchangeable (cf. *Filip se sakrio da ga Ana ne vidi => Filip se sakrio da ga Ana ne bi videla*).

(172)



Enfin, elle peut également introduire une consécutive, mais dans ce cas, elle se combine avec un corrélatif dans la principale (cf. exemple 173). Le traitement des structures corrélatives est décrit en détail dans la section 2.8.5.

(173)

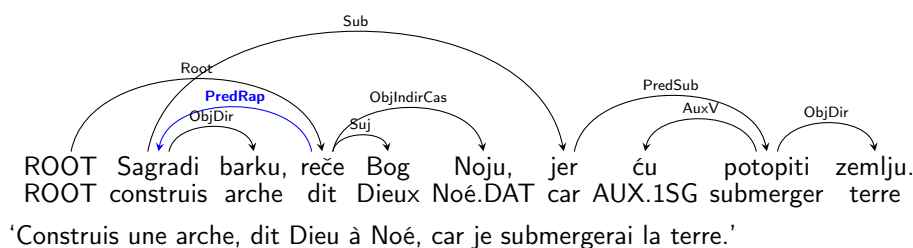


## 2.9 Discours indirect

Si un élément de **discours rapporté** contient un verbe principal, autrement dit, s'il a la structure d'une **proposition indépendante**, il est traité par la étiquette **PredRap**. Cette relation est gouvernée par le verbe introducteur du discours rapporté. Si le verbe introducteur se trouve au milieu de la phrase rapportée, les éléments séparés sont reliés par les mêmes relations qui s'appliqueraient dans une phrase typique (cf. exemple 174). En revanche, si le discours rapporté ne contient pas de verbe, il relève plutôt de la relation **Cit** (cf. section 2.7.9).



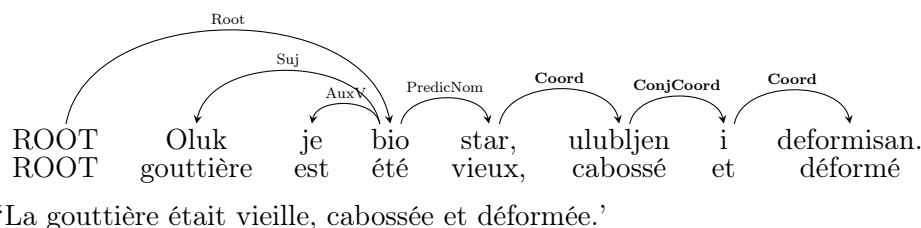
(174)



## 2.10 Coordination

Pour les structures coordonnées, nous adoptons un traitement en cascade. Le premier coordonné porte l'étiquette qui exprime la fonction syntaxique de la structure coordonnée dans la phrase. Tous les coordonnés entre le premier et la conjonction de coordination sont annotés en cascade comme **Coord**. La conjonction est gouvernée par le coordonné immédiatement précédent à l'aide de l'étiquette **ConjCoord**. Le dernier coordonné dépend de la conjonction à l'aide de l'étiquette **Coord** (cf. exemple 175).

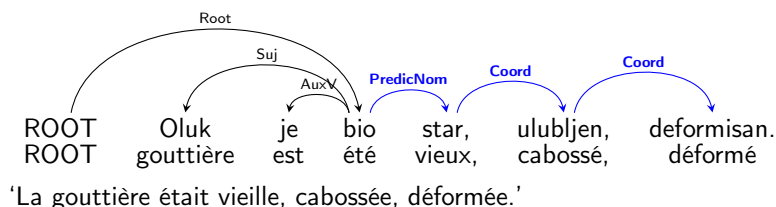
(175)



**NB**

Nous considérons également comme coordination les enchaînements de plusieurs éléments séparés par des virgules, sans conjonction de coordination entre les deux derniers éléments (cf. exemple 176). Ce phénomène est considéré comme juxtaposition dans la littérature linguistique, mais il est traité comme coordination dans plusieurs autres corpus (FTBDep, PDT). Comme ce traitement permet de simplifier le traitement de ce phénomène, nous l'adoptons ici.

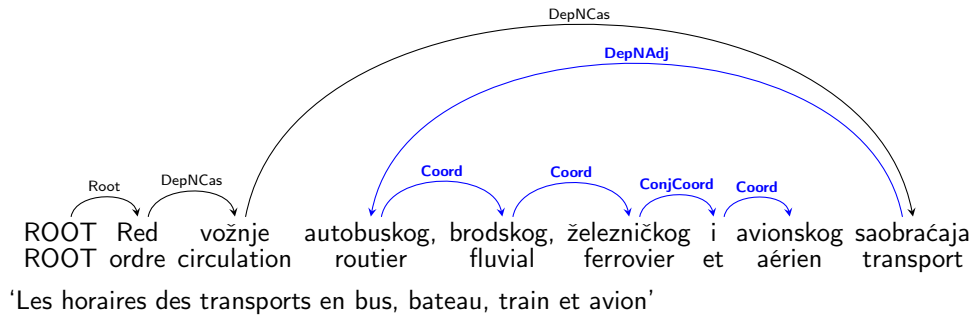
(176)



Les adjectifs antéposés au nom peuvent également être coordonnés (cf. exemple 177).

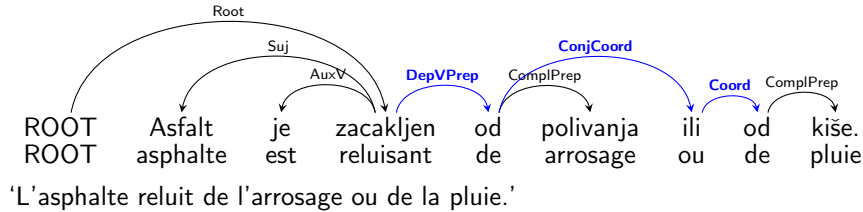
La relation de coordination s'établit entre les **têtes** des éléments coordonnés. Par conséquent, il faut veiller à bien identifier les structures mises en parallèle par la coordination et à sélectionner les bons gouverneurs pour la relation **Coord**. Comparez l'exemple 178, où ce sont

(177)

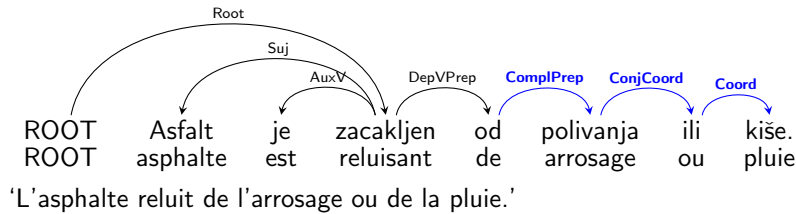


des groupes prépositionnels qui sont coordonnés, avec l'exemple 179, où ce sont les compléments de la préposition. Dans l'exemple 180, la coordination porte sur les deux compléments du nom *tama* 'obscurité', alors que dans l'exemple 181, elle met en parallèle les deux objets directs *tamu* 'obscurité' et *daljinu* 'distance'.

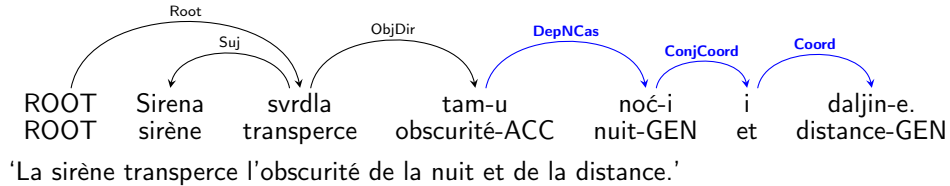
(178)



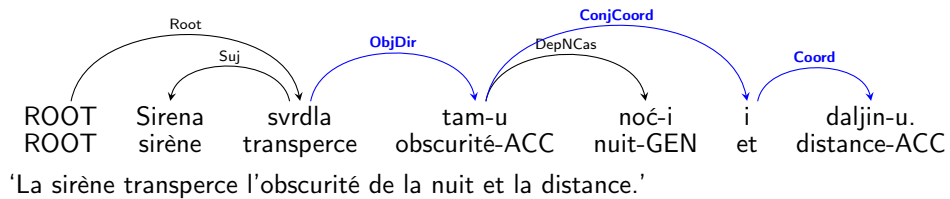
(179)



(180)

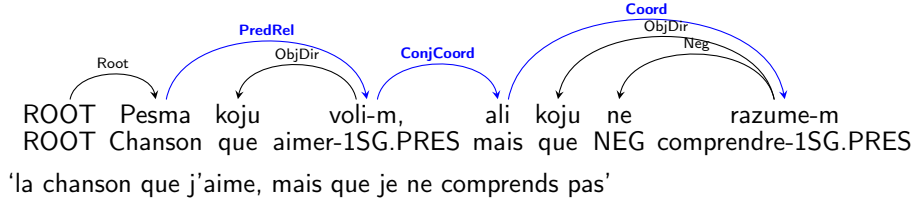


(181)



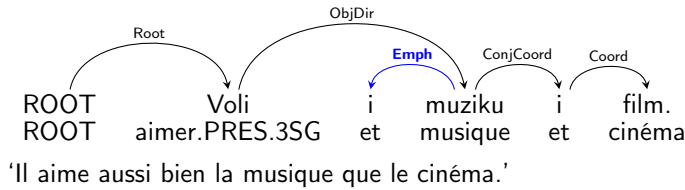
Dans le cas de la **coordination des relatives**, c'est entre les **verbes principaux des relatives** que la coordination s'établit. En cas d'absence de conjonction de coordination, la relation **Coord** relie directement les deux verbes, et sinon, la relation **ConjCoord** part du verbe de la première relative vers la conjonction de coordination, et la relation **Coord** s'établit entre la conjonction et le verbe de la deuxième relative (cf. exemple 182).

(182)

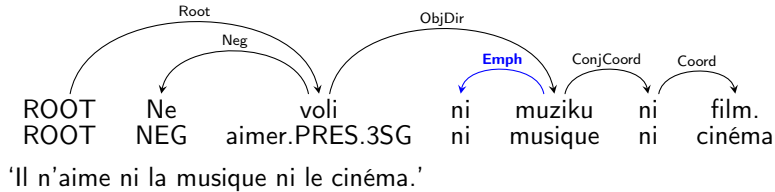


Les structures en *i - i* comme *Voli i muziku i film* 'Il aime et la musique et le cinéma' et en *ni - ni* comme *Ne voli ni muziku, ni film* 'Il n'aime ni la musique, ni le film' bénéficient d'un traitement spécifique. En effet, étant donné que la première des deux formes répétées est optionnelle, (cf. *Voli muziku i film*, *Ne voli muziku ni film*), on considère que ce n'est que la deuxième qui a le rôle d'un coordonnant. La première est interprétée comme un élément d'emphase. Voir les exemples ci-dessous.

(183)



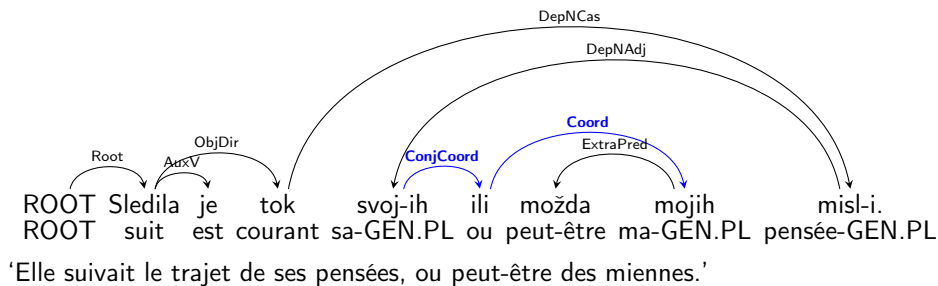
(184)



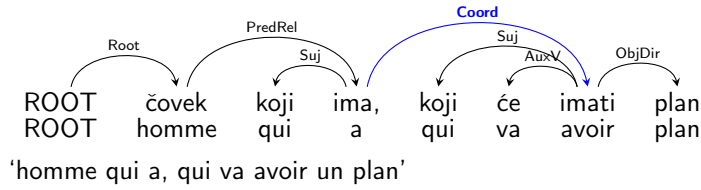
## Constituants incomplets

Dans le cas des constituants incomplets, si pertinent, la coordination peut être appliquée.

(185)

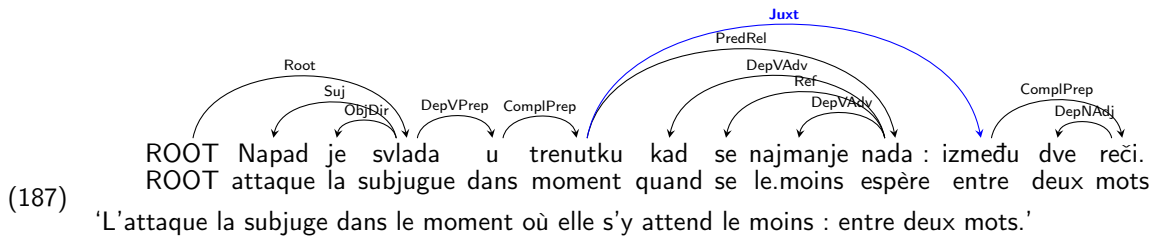


(186)

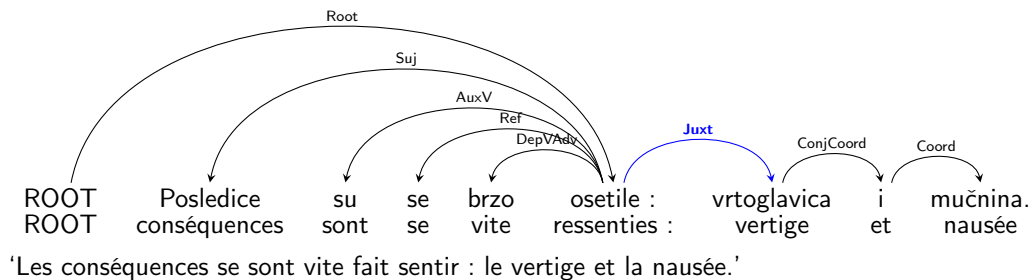


## 2.11 Juxtaposition

Cette relation est réservée aux éléments de haut niveau (deux propositions, une proposition et un syntagme) qui sont liés au niveau discursif, mais qui paraissent indépendants l'un de l'autre au niveau syntaxique et entre lesquels aucune autre étiquette syntaxique ne s'applique. Ces éléments sont typiquement séparés par deux points ou par des points-virgules. Le dépendant, qui est l'élément qui se trouve à la périphérie de la structure de la phrase, est gouverné par l'élément avec lequel il est mis en parallèle (cf. exemple 187). S'il n'est pas possible d'identifier cet élément, on utilise comme gouverneur la tête du segment à gauche (cf. exemple 188).

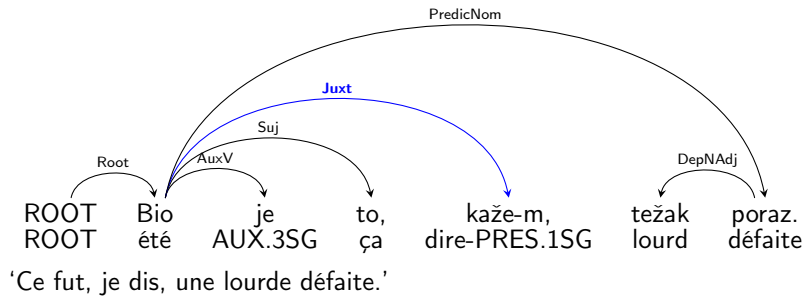


(188)



Cette relation est également utilisée pour annoter les verbes de la parole (typiquement des verbes pouvant servir d'introducteur du discours indirect) insérés dans la phrase sans qu'il s'agisse du discours rapporté. Ils sont gouvernés par le verbe principal de la proposition dans laquelle ils se trouvent (cf. exemple 189).

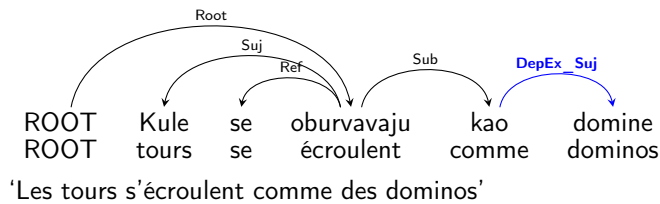
(189)



## 2.12 Ellipse

Le traitement de l'ellipse reste une question ouverte aussi bien en linguistique théorique qu'en TAL. Pour cette première annotation de notre corpus, nous retenons l'approche mise en place dans le treebank tchèque PDT : une forme dépendante d'un élément phrastique élidé est gouvernée par le gouverneur de l'élément élidé, en utilisant l'étiquette de la relation que la forme concernée a par rapport à son gouverneur élidé préfixé de **DepEx** (pour 'dépendance externe'). À priori, n'importe quelle étiquette du jeu peut être modifiée de cette manière pour permettre le traitement de l'ellipse. De nombreux cas de figure sont possibles ; quelques-uns des cas les plus fréquents sont illustrés dans la suite.

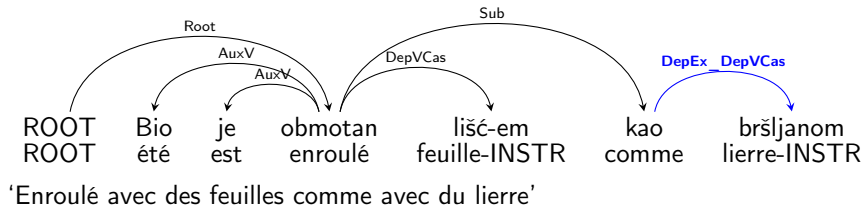
(190)



Pour identifier la fonction syntaxique que la forme porterait par rapport à son gouverneur élidé, il convient de reconstituer la proposition élidée. Dans l'exemple 190, la proposition reconstituée correspond à *Kule se oburvavaju kao što se oburvavaju domine* 'Les tours s'écroulent comme s'écroulent des dominos'. À l'intérieur de cette proposition, la forme *domine* a la fonction du sujet. Par conséquent, on lui accorde l'étiquette **DepEx\_Suj** dans la phrase de départ. Quant à son gouverneur, c'est le subordonnant : c'est lui qui serait le gouverneur du verbe de la subordonnée si celui-ci était présent dans la phrase.

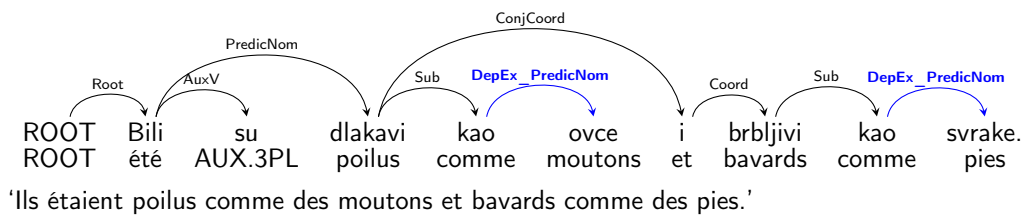
Une approche comparable peut être transposée à l'exemple 191. Ici, la proposition complète serait *kao da je obmotan bršljanom* 'comme s'il était enroulé du lierre', mais le verbe, qui gouvernerait normalement la forme à l'instrumental, est élidé. Par conséquent, c'est la conjonction qui est le gouverneur de la forme à l'instrumental, et comme cette forme aurait la fonction **DepVCas** par rapport au verbe élidé, c'est l'étiquette **DepEx\_DepVCas** qui unit la conjonction à la forme nominale.

(191)



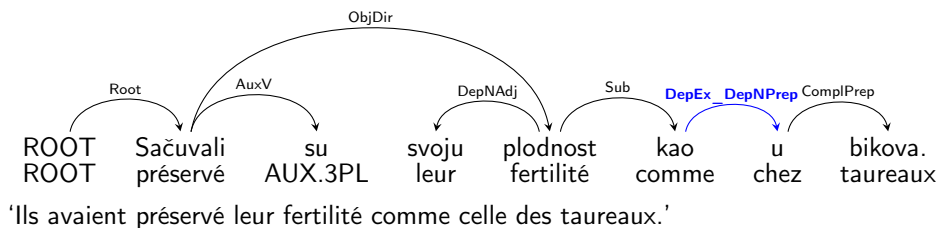
Dans l'exemple 192, les propositions rétablies sont les suivantes : *Bili su dlakavi kao što su ovce dlakave i brbljivi kao što su svrake brbljive* 'Ils étaient poilus comme sont poilus les moutons et bavards comme sont bavardes les pies'. Par conséquent, les adjectifs introduits par *kao* 'comme' ont le rôle du prédicatif nominal par rapport à un verbe *être* élidé.

(192)



Ce type de construction peut également être gouverné par d'autres parties du discours, si la comparaison porte sur un dépendant nominal (cf. exemple 193).

(193)



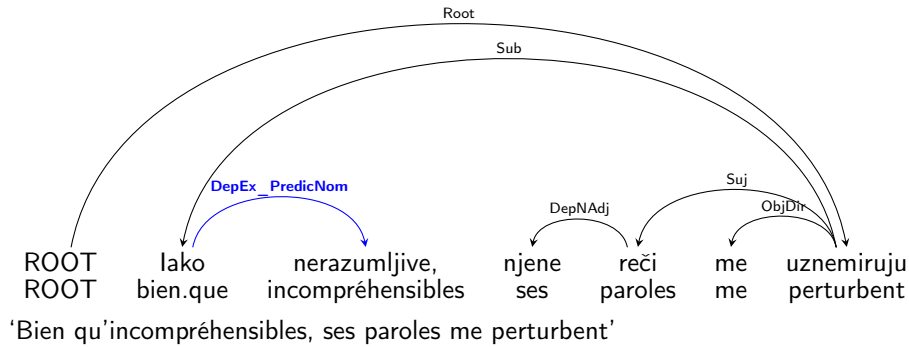
Ici, il s'agit en effet d'une ellipse à plusieurs niveaux : le groupe prépositionnel dépend d'un nom ellidé, qui dépendrait, lui, du verbe, qui est ellidé aussi (cf. *sačuvali su svoju plodnost koja je bila kao plodnost u bikova* 'ils avaient préservé leur fertilité qui était comme la fertilité des taureaux'). Nous encodons ici un seul niveau d'ellipse et marquons la dépendance de ce groupe prépositionnel par rapport à un nom absent de la phrase.

La même approche est appliquée aux **conjonctions de subordination introduisant des syntagmes** : dans ce cas, nous considérons qu'il s'agit d'une proposition subordonnée avec le verbe élidé et nous appliquons le traitement illustré dans l'exemple 194, qui correspond à un prédicatif nominal.

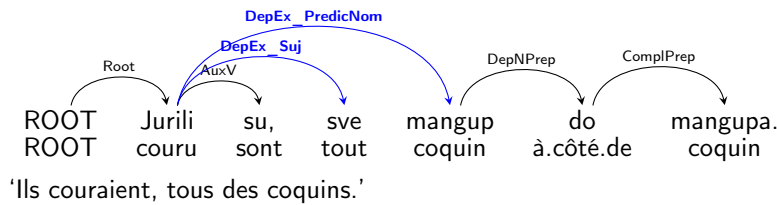
Un cas spécifique de l'ellipse concerne la construction en *sve* 'tout' comme *sve mangup do mangupa* lit. 'tous coquin à côté de coquin', 'que des coquins'. Elle est traitée comme dans l'exemple 195.

Il faut préciser aussi qu'il peut y avoir des situations dans lesquelles le **gouverneur de l'ellipse** est **ambigu**. Par exemple, dans la phrase *Kroz kosu joj proviruje pupak kao kiklopsko oko* 'Son nombril transperce ses cheveux tel un œil de Cyclope', la conjonction *kao* 'comme' peut être introduite par le verbe (et dans ce cas, la phrase s'interpréterait comme *Pupak*

(194)

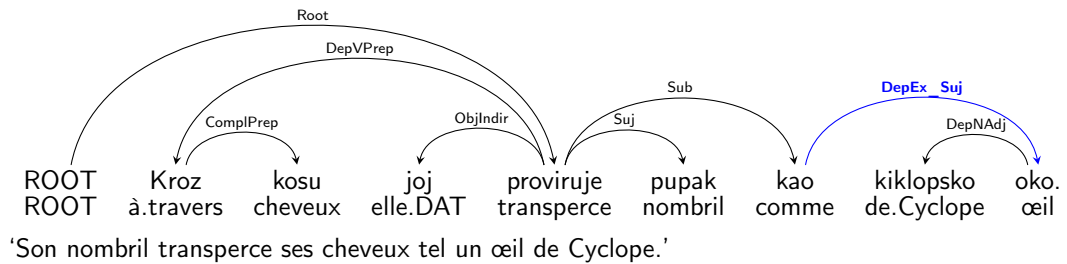


(195)

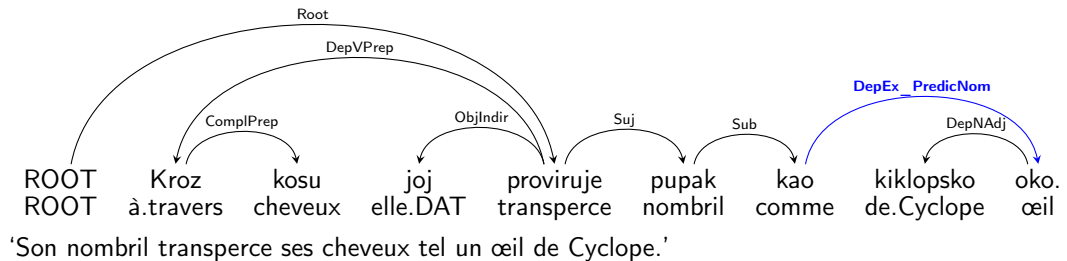


*joj proviruje kroz kosu kao što proviruje kiklopsko oko* 'Son nombril transperce ses cheveux comme le ferait un œil de Cyclope'), ou bien par le nom *pupak* 'nombril' (et alors le phrase s'interprète comme *Kroz kosu joj proviruje pupak, koji je kako kiklopsko oko* 'Ses cheveux sont transpercés par son nombril, qui est comme un œil de Cyclope'). Le premier cas correspond à l'analyse montrée dans l'exemple 196, alors que le deuxième est illustré dans l'exemple 197.

(196)



(197)



Dans ce type d'exemples, la décision du traitement à adopter revient à la discrétion de l'annotateur. Il est cependant invité à signaler tout exemple ambigu à l'annotateur expérimenté pour que celui-ci puisse l'analyser et l'inventorier si nécessaire.

# Bilan

Le guide recense les 48 relations de dépendance utilisées dans l’annotation syntaxique du corpus ParCoTrain-Synt, ainsi que le traitement de l’ellipse. Les étiquettes se répartissent comme suit :

- 17 dépendants du verbe,
- 4 dépendants du nom,
- 4 dépendants de l’adjectif,
- 3 dépendants de l’adverbe,
- 7 étiquettes pour le traitement de la subordination,
- 2 étiquettes pour le traitement de la coordination,
- 10 étiquettes pour d’autres types de dépendants.

Dans la suite, nous proposons un index des exemples par étiquette syntaxique.



# Index des exemples par étiquette

Ap : 67, 68, 69, 70, 71, 72, 73, 74.  
AuxV : 2, 7, 8, 12, 15, 16, 24, 26, 28, 34, 41, 42, 43, 45, 46, 51, 52, 53, 56, 57, 58, 59, 61, 64, 65, 69, 70, 72, 78, 83, 85, 91, 92, 94, 97, 105, 107, 111, 112, 113, 115, 118, 124, 125, 126, 129, 131, 136, 140, 153, 142, 143, 146, 147, 149, 159, 160, 161, 163, 164, 165, 167, 168, 170, 171, 174, 175, 176, 178, 179, 185, 186, 188, 189, 191, 192, 193, 195.  
Cit : 138.  
ComplNum : 92.  
ComplPrep : 28, 30, 32, 37, 40, 43, 46, 48, 53, 54, 55, 60, 62, 70, 71, 74, 80, 81, 82, 86, 87, 88, ??, 90, 94, 101, 107, 109, 116, 117, 132, 134, 138, 153, 161, 178, 179, 187, 193, 195, 196, 197.  
ConjCoord : 12, 18, 19, 24, 25, 31, 59, 73, 74, 75, 117, 118, 175, 177, 178, 179, 180, 181, 182, 183, 184, 185, 188, 192.  
Coord : 4, 12, 18, 19, 24, 25, 30, 31, 59, 62, 66, 72, 73, 74, 75, 117, 118, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 188, 192.  
Correl : 127, 129, 167, 168, 173.  
DepAdjAdj : 83.  
DepAdjAdv : 75, 167.  
DepAdjCas : 76, 77, 78, 79.  
DepAdjPrep : 48, 80, 81, 82.  
DepAdvAdv : 84, 173.  
DepAdvCas : 13, 26, 85, 167.  
DepAdvPrep : 86, 87, 88, ??.  
DepEx\_DepNPrep : 193.  
DepEx\_DepVCas : 168, 191.  
DepEx\_PredicNom : 192, 194, 195, 197.  
DepEx\_Suj : 190, 195, 196.  
DepNAdj : 3, 34, 54, 55, 61, 63, 64, 65, 66, 67, 68, 71, 72, 73, 74, 76, 77, 78, 79, 82, 91, 109, 118, 133, 138, 158, 163, 177, 185, 187, 189, 193, 194, 196, 197.  
DepNCas : 61, 68, 72, 86, 177, 180, 181, 185.  
DepNPrep : 62, 107, 109, 132, 195.  
DepVAdv : 7, 8, 9, 10, 15, 16, 49, 50, 68, 93, 95, 96, 100, 105, 106, 108, 113, 127, 129, 135, 143, 147, 157, 164, 168, 173, 187, 188.  
DepVCas : 25, 47, 51, 52, 78, 133, 153, 142, 168, 191.  
DepVInf : 57, 58.  
DepVPart : 52, 56, 114.  
DepVPrep : 28, 53, 54, 55, 60, 70, 71, 74, 94, 101, 116, 134, 138, 153, 178, 179, 187, 196, 197.  
Emph : 12, 35, 108, 109, 110, 112, 113, 114, 115, 116, 183, 184.  
ExtraPred : 7, 8, 105, 106, 107, 185.  
Interrog : 95, 96, 97, 98, 154, 155.  
Juxt : 10, 187, 188, 189.  
Neg : 9, 10, 93, 94, 114, 134, 153, 154, 167, 169, 182, 184.

ObjDir : 1, 2, 4, 24, 26, 27, 41, 42, 43, 45, 52, 53, 56, 61, 69, 73, 74, 78, 85, 94, 107, 113, 114, 119, 120, 121, 122, 126, 134, 136, 140, 153, 149, 157, 158, 159, 160, 161, 163, 164, 165, 167, 168, 170, 171, 169, 174, 180, 181, 182, 183, 184, 185, 186, 187, 193, 194.  
 ObjIndir : 146, 162, 196, 197.  
 ObjIndirCas : 27, 28, 65, 91, 161, 174.  
 ObjIndirPrep : 30, 90, 117, 161.  
 Polylex : 95, 98, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 129, 131, 132, 133, 134, 135, 136, 154, 167.  
 PredCompletive : 15, 16, 20, 153, 142, 143, 146, 147, 149, 164, 169.  
 PredPercent : 154, 155, 156, 157, 158, 167.  
 PredRap : 174.  
 PredRel : 159, 160, 161, 163, 164, 182, 186, 187.  
 PredSub : 5, 6, 119, 120, 121, 122, 123, 124, 125, 126, 129, 138, 140, 165, 167, 170, 171, 169, 173, 174.  
 PredicAdv : 15, 18, 19, 21, 36, 37.  
 PredicComplObj : 41, 42, 43, 153.  
 PredicComplSuj : 38, 39, 40.  
 PredicNom : 6, 13, 16, 19, 22, 28, 31, 32, 34, 72, 80, 81, 83, 129, 131, 153, 162, 167, 173, 175, 176, 189, 192.  
 PredicOpt : 44, 45, 46, 47, 48, 112.  
 Ref : 9, 10, 20, 23, 25, 38, 39, 40, 44, 47, 48, 68, 102, 124, 125, 129, 155, 156, 169, 187, 188, 190.  
 Root : 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 30, 31, 32, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, ??, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 105, 106, 107, 108, 109, 110, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 129, 131, 132, 133, 134, 135, 136, 138, 140, 153, 142, 143, 146, 147, 149, 154, 155, 156, 157, 158, 159, 160, 161, 163, 164, 165, 167, 168, 169, 170, 171, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197.  
 Sub : 6, 15, 16, 20, 119, 120, 121, 122, 123, 124, 125, 126, 129, 138, 140, 153, 142, 143, 146, 147, 149, 164, 165, 169, 170, 171, 174, 190, 191, 192, 193, 194, 196, 197.  
 Suj : 1, 2, 4, 5, 6, 7, 8, 12, 13, 17, 27, 28, 30, 31, 32, 34, 35, 36, 37, 38, 48, 49, 50, 54, 55, 57, 58, 59, 60, 64, 68, 69, 70, 71, 72, 80, 81, 90, 92, 95, 96, 97, 98, 99, 100, 101, 102, 105, 110, 111, 112, 114, 115, 116, 118, 122, 124, 125, 129, 135, 138, 140, 142, 149, 153, 154, 155, 156, 157, 159, 162, 163, 164, 165, 169, 173, 174, 175, 176, 178, 179, 180, 181, 186, 187, 188, 189, 190, 194, 196, 197.  
 SujLog : 17, 18, 19, 20, 21, 22, 23, 129, 153.

# Bibliographie

- Noam Chomsky. *Some concepts and consequences of the theory of government and binding*. MIT press, 1982.
- Noam Chomsky. *Lectures on government and binding : The Pisa lectures*. Walter de Gruyter, 1993.
- Gerald Gazdar, Klein Ewald, Geoffrey Pullum, and Ivan Sag. *Generalized phrase structure grammar*. Harvard University Press, 1985.
- Jan Hajič, Jarmila Panevová, Eva Buráňová, Zdeňka Urešová, and Alla Bémová. Annotations at analytical level. Instructions for annotators. *UK MFF ÚFAL, Praha, Czech Republic*. URL <http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/a-layer/pdf/a-man-en.pdf> (2012-03-18), 1999.
- Peter Hellwig. Dependency unification grammar. In *Proceedings of the 11th Conference on Computational Linguistics, COLING '86*, pages 195–198, Stroudsburg, PA, USA, 1986. Association for Computational Linguistics. doi: 10.3115/991365.991423. URL <http://dx.doi.org/10.3115/991365.991423>.
- Richard A Hudson. *Word grammar*. Blackwell Oxford, 1984.
- Milka Ivić, editor. *Sintaksa savremenog srpskog jezika*. Institut za srpski jezik SANU, Beograd, 2005.
- Igor Mel'čuk. *Dependency syntax : Theory and practice*. State University Press of New York, 1988.
- Pavica Mrazović. *Gramatika srpskog jezika za strance*. Izdavačka knjižarnica Zorana Stojanovića, 2009.
- Carl Pollard and Ivan A Sag. *Head-driven phrase structure grammar*. University of Chicago Press, 1994.
- Petr Sgall, Eva Hajicová, and Jarmila Panevová. *The meaning of the sentence in its semantic and pragmatic aspects*. Springer Science & Business Media, 1986.
- Živojin Stanojčić and Ljubomir Popović. *Gramatika srpskog jezika*. Zavod za udžbenike, 2012.
- Pasi Tapanainen and Timo Järvinen. A non-projective dependency parser. In *Proceedings of the fifth conference on Applied natural language processing*, pages 64–71. Association for Computational Linguistics, 1997.