

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

Aleksandra Miletic  
CLLE-ERSS, Université de Toulouse - Jean Jaurès

17 avril 2018

# Table des matières

<b>1</b>	<b>Remarques introductives</b>	<b>5</b>
1.1	Comment lire ce guide . . . . .	5
1.2	Quelques principes d’annotation morphosyntaxique de base . . . . .	6
1.2.1	Correspondance tag-token 1 :1 . . . . .	6
1.2.2	Résolution des cas ambigus basée sur le contexte . . . . .	7
<b>2</b>	<b>Présentation du traitement par classe de mots</b>	<b>8</b>
2.1	Les noms . . . . .	8
2.1.1	Les sous-catégories de noms . . . . .	8
2.1.2	Noms du genre grammatical féminin désignant des êtres du sexe masculin . . . . .	10
2.1.3	Noms composés contenant un trait d’union . . . . .	10
2.2	Les adjectifs . . . . .	11
2.2.1	Les sous-catégories d’adjectifs . . . . .	11
2.2.2	Degré de comparaison de l’adjectif . . . . .	13
2.2.3	Traits <i>genre</i> et <i>nombre</i> pour la sous-catégorie <i>prisvojni</i> . . . . .	13
2.2.4	Adjectifs (pro)nominalisés . . . . .	13
2.2.5	Adjectifs déverbaux dérivés des formes des participes passif et actif ( <i>glagolski pridev trpni</i> et <i>glagolski pridev radni</i> ) . . . . .	14
2.2.6	Adjectifs composés . . . . .	14
2.3	Les verbes . . . . .	15
2.3.1	Trait <i>forme verbale</i> . . . . .	15
2.3.2	Verbes auxiliaires dans les formes surcomposées . . . . .	17
2.3.3	Trait <i>négation</i> . . . . .	17
2.4	Les pronoms . . . . .	19
2.4.1	Les sous-catégories des pronoms . . . . .	20
2.4.2	Annotation du genre et du nombre . . . . .	20
2.4.3	Pronoms indéfinis discontinus . . . . .	20
2.4.4	Distinction entre le pronom personnel fléchi et l’auxiliaire . . . . .	22

2.4.5	Le pronom <i>što</i> . . . . .	22
2.5	Les numéraux . . . . .	24
2.5.1	Le genre et le nombre des numéraux . . . . .	24
2.5.2	Paucal . . . . .	25
2.5.3	Formes en <i>-ak</i> . . . . .	25
2.6	Les adverbes . . . . .	26
2.6.1	Sous-catégories des adverbes . . . . .	26
2.6.2	Degré de comparaison des adverbes . . . . .	26
2.6.3	Forme <i>kad</i> ‘quand’ . . . . .	27
2.7	Conjonctions . . . . .	27
2.8	Les prépositions, les interjections, les particules, les mots étrangers et la ponctuation . . . . .	28
<b>3</b>	<b>Gestion des cas de figure spécifiques</b>	<b>29</b>
3.1	Prépositions . . . . .	29
3.2	Listes des particules . . . . .	30
3.3	Autres cas de figure . . . . .	32
	<b>Bibliographie</b>	<b>35</b>

## Liste des tableaux

1.1	Exemple d'annotation morphosyntaxique . . . . .	5
2.1	Nom : traits utilisés et leurs valeurs possibles . . . . .	8
2.2	Nom : distribution des traits en fonction de la sous-catégorie . . . . .	9
2.3	Nom propre polylexical . . . . .	9
2.4	Nom collectif . . . . .	10
2.5	Nom de famille . . . . .	10
2.6	Adjectifs : traits utilisés et leurs valeurs possibles . . . . .	11
2.7	Adjectifs : distribution des traits en fonction de la sous-catégorie . . . . .	11
2.8	Adjectif : exemples des sous-catégories différentes . . . . .	12
2.9	Adjectif : degré de comparaison . . . . .	13
2.10	Verbe : traits utilisés et leurs valeurs possibles . . . . .	15
2.11	Verbe : distribution des traits en fonction de la forme verbale . . . . .	16
2.12	Verbe auxiliaire : formes surcomposées . . . . .	17
2.13	Pronom : traits utilisés et leurs valeurs possibles . . . . .	19
2.14	Pronom : distribution des traits en fonction de la sous-catégorie . . . . .	19
2.15	Pronom : genre et nombre . . . . .	21
2.16	Pronom : discontinuité dans le GP . . . . .	21
2.17	Pronom <i>vs</i> auxiliaire . . . . .	22
2.18	Pronom <i>što</i> : exemple 1 . . . . .	23
2.19	Pronom <i>što</i> : exemple 2 . . . . .	23
2.20	Pronom <i>što</i> : exemple 3 . . . . .	23
2.21	Numéral : traits utilisés et leurs valeurs possibles . . . . .	24
2.22	Numéral : distribution des traits en fonction de la sous-catégorie . . . . .	24
2.23	Numéral <i>jedan</i> . . . . .	24
2.24	Numéral <i>dva</i> : fléchi <i>vs</i> invariable . . . . .	25
2.25	Adverbe : traits utilisés et leurs valeurs possibles . . . . .	26
2.26	Adverbe : distribution des traits en fonction de la sous-catégorie . . . . .	26
2.27	Conjonctions : traits utilisés et leurs valeurs possibles . . . . .	27

2.28	Conjonctions : exemples des sous-catégories . . . . .	27
2.29	Autres classes : étiquettes POS . . . . .	28

# 1. Remarques introductives

Ce document s’articule comme suit : la première partie présente quelques principes généraux de l’annotation morphosyntaxique telle que définie dans le cadre du projet ParCoLab et propose une grille de lecture de ce document. La deuxième partie est la plus conséquente : elle contient la présentation du traitement adopté pour chaque classe de mots. Chaque sous-section donne les traits morphosyntaxiques traités pour la classe de mots donnée, les valeurs possibles des traits et les règles d’annotation pour les cas de figure nécessitant un traitement spécifique. La dernière partie du document contient l’indication des traitements adoptés pour certaines constructions problématiques. Dans cette partie aussi, les règles d’annotation sont organisées selon la classe de mots.

## 1.1 Comment lire ce guide

L’annotation morphosyntaxique de ParCoLab se fait en utilisant un tableur (un fichier Excel). Les fichiers à traiter se présentent dans un format verticalisé : une ligne correspond à un token du texte. Elle contient le token lui-même dans la première colonne, suivi de son annotation morphosyntaxique dans les colonnes suivantes.

Le traitement morphosyntaxique défini dans le cadre du projet ParCoLab met en place une annotation riche en informations : au-delà de l’indication de la classe et de la sous-catégorie distributionnelle (cf. nom commun, pronom démonstratif), les étiquettes indiquent également les traits morphosyntaxiques pertinents pour chaque classe de mots. À titre d’illustration, les informations accordées aux noms comprennent également l’indication du genre, du nombre et du cas, alors que le traitement des verbes fait appel à la personne, le nombre, le genre, la forme verbale et la présence ou l’absence de la négation. Dans le cadre de l’annotation manuelle, on utilise des indications explicites des valeurs de ces traits : elles sont épelées en toutes lettres (cf. tableau 1.1).

<b>vidim</b>	V	glavni	prezent	prvo lice	jednina	---	---
<b>devojk</b>	N	zajednicka	akuzativ	jednina	zenski rod		

TABLE 1.1 – Exemple d’annotation morphosyntaxique

Notons que ce ne sont pas les mêmes traits morphosyntaxiques qui sont indiqués pour les deux classes de mots. Par conséquent, les mêmes colonnes ne correspondent pas aux mêmes traits dans les deux cas. Or, pour que l’annotation manuelle soit exploitable par la suite, il est **essentiel** que les traits soient indiqués toujours dans le même ordre pour une classe de mots donnée. Dans la suite du document, nous présentons donc systématiquement les traits **dans l’ordre** dans lequel ils doivent apparaître dans l’annotation manuelle.

Pour faciliter ce travail, une macro VisualBasic a été implémentée dans les fichiers d’annotation manuelle. Cette macro permet la modification des en-têtes des colonnes en fonction de la valeur de la colonne contenant l’indication de la classe de mots. La modification s’active en passant d’une ligne du tableur à l’autre avec les flèches ↑ et ↓ sur le clavier. Si cette fonctionnalité ne démarre pas à l’ouverture du fichier, il est possible de la lancer en utilisant le raccourci **CTRL + q**, ou bien en allant dans l’onglet **Affichage/Macros/Affichage des macros** et en cliquant sur le bouton **Exécuter**.

Dans la partie 2 de ce document, nous définissons le traitement de chaque classe de mots. Nous présentons tous les traits morphosyntaxiques encodés, ainsi que leurs valeurs possibles. Dans le cadre de l’annotation manuelle, il est essentiel d’utiliser exactement les mêmes valeurs : attention à l’absence des diacritiques serbes.

Pour chaque classe de mots, dans un premier tableau nous présentons tous les traits utilisés, ainsi que toutes leurs valeurs possibles. Ici, les traits sont présentés dans le même ordre dans lequel ils doivent être utilisés en annotation manuelle. Ensuite, dans un deuxième tableau, nous définissons la distribution des traits en fonction de la sous-catégorie morphosyntaxique : souvent, certaines sous-catégories présentent des restrictions quant à certains traits.

Si une case du tableau est renseignée comme ‘[tous]’, ceci signifie que toutes les valeurs du trait donné sont valides pour la sous-catégorie en question. Si la case contient ‘---’, ceci veut dire que le trait donné n’est pas marqué dans la sous-catégorie donnée ; la même indication doit être reproduite dans l’annotation manuelle. Enfin, si la case contient une ou plusieurs valeurs concrètes (*mnozina* ‘pluriel’, *drugo lice* ‘2e personne’), ces valeurs seules sont valides pour la sous-catégorie donnée.

## 1.2 Quelques principes d’annotation morphosyntaxique de base

### 1.2.1 Correspondance tag-token 1 :1

L’identification des unités à annoter au niveau morphosyntaxique est basée sur le principe suivant : une forme orthographique ne peut porter qu’une seule étiquette, et

une étiquette ne peut être attachée qu'à une seule forme à la fois. Ce principe permet la réalisation d'un étiquetage du premier niveau (chaque forme orthographique se voit attribuer un tag), qui, du point de vue linguistique, n'est pas toujours correct : on dissocie ainsi les unités polylexicales et on leur impose une interprétation analytique. Pourtant, un étiquetage qui prendrait en compte les unités polylexicales serait problématique à ce stade, premièrement pour la question de la définition des unités polylexicales, mais aussi à cause de l'existence des unités discontinues en serbe (notamment les formes verbales complexes). L'approche adoptée permet d'éviter ces problèmes et facilite ainsi un maintien de cohérence de l'annotation. Cette première annotation pourra donner lieu ensuite à un traitement plus pertinent du point de vue linguistique.

### 1.2.2 Résolution des cas ambigus basée sur le contexte

Une partie de l'ambiguïté relevée dans le traitement d'un corpus peut provenir des formes capables d'assumer les rôles prototypiques de plusieurs classes de mots. Si un token, qui appartient typiquement à une classe de mots, prend dans le contexte le comportement d'une autre catégorie grammaticale, il est annoté selon sa fonction syntaxique. L'exemple typique est celui des adjectifs qui se trouvent nominalisés dans un contexte spécifique : le mot *mrtav* 'mort' est un adjectif ; cependant, dans l'exemple *razmišljati o mrtvima* 'penser aux morts', il est employé indépendamment d'un nom et fonctionne lui-même comme un groupe nominal. Par conséquent, il sera annoté comme un nom commun. D'une manière comparable, la forme *iza* 'derrière', qui est typiquement une préposition, n'introduit pas un groupe prépositionnel dans l'exemple *Parkirao sam se iza* 'Je me suis garé derrière'. Ici cette forme connaît un emploi adverbial, et peut être remplacée par un adverbe locatif typique comme *daleko* 'loin'. Par conséquent, dans ce contexte, *iza* sera annoté comme adverbe.

Les cas de figure de ce type qui sont relativement systématiques sont rappelés et illustrés dans les sections dédiées aux classes de mots.



## 2. Présentation du traitement par classe de mots

Cette partie du guide est dédiée à la présentation du traitement de chaque classe de mots. Dans chaque section, nous indiquons d’abord les traits morphosyntaxiques pertinents et l’ordre dans lequel ils se présentent dans le cadre de l’annotation manuelle, ainsi que les valeurs valides pour chaque trait. Nous indiquons ensuite les spécificités du traitement si celui-ci ne coïncide pas avec la tradition grammaticale serbe. Si nécessaire, certains cas de figure particuliers sont également traités.

### 2.1 Les noms

Le traitement des noms fait appel aux traits morphosyntaxiques suivants : la classe de mots, la sous-catégorie, le cas, le nombre et le genre. L’ordre des traits et leurs valeurs possibles sont donnés dans le tableau 2.1, alors que la distribution des traits en fonction de la sous-catégorie est présentée dans le tableau 2.2.

POS	Sous-cat.	Cas	Nombre	Genre
N	zajednicka vlastita zbirna	nominativ genitiv dativ akuzativ vokativ instrumental lokativ	jednina mnozina	muski rod zenski rod srednji rod

TABLE 2.1 – Nom : traits utilisés et leurs valeurs possibles

#### 2.1.1 Les sous-catégories de noms

Stanojčić & Popović (2012) distinguent 6 sous-catégories de noms : noms propres (*vlastite*), communs (*zajedničke*), massifs (*gradivne*), abstraits (*apstraktne*) et déverbaux (*glagolske*) (équivalentes des mêmes sous-catégories en français), ainsi que les noms collectifs (*zbirne*), qui en serbe, à la différence du français et de l’anglais, comprennent les

Exemples	POS	Sous-cat.	Cas	Nombre	Genre
noms communs (cf. <i>konj</i> ‘cheval’, <i>stolica</i> ‘chaise’)	N	zajednicka	[tous]	[tous]	[tous]
noms propres (cf. <i>Beograd</i> ‘Belgrade’)	N	vlastita	[tous]	[tous]	[tous]
noms collectifs (cf. <i>granje</i> ‘branches’, <i>pilad</i> ‘poussins’)	N	zbirna	[tous]	jednina	[tous]

TABLE 2.2 – Nom : distribution des traits en fonction de la sous-catégorie

noms ayant la forme et le comportement d’un nom au singulier, mais le sémantisme du pluriel (cf. *pilad* ‘poussins’, qui désigne un ensemble de poussins, mais se décline comme le singulier du nom *noć* ‘nuit’).

Les noms massifs (*voda* ‘eau’, *brašno* ‘farine’), abstraits (*ljubav* ‘amour’, *mržnja* ‘haine’) et déverbaux (*crtanje* ‘dessin’, *pevanje* ‘chant’) ne se distinguent pas par leur comportement morphologique des autres noms communs : ils suivent les mêmes modèles de déclinaison. Par conséquent, il a été décidé d’annoter les 4 sous-classes comme noms communs (valeur du trait sous-catégorie **zajednicka**).

La distinction des noms propres a été gardée : dans la suite de l’enrichissement du corpus, cette propriété peut être exploitée dans la reconnaissance des entités nommées. La valeur du trait *sous-catégorie* pour ces noms est donc **vlastita**. Cette étiquette s’applique aux noms des personnes et des animaux (cf. *Marko*, *Jovanović*, *Pera*, *Žuća* etc.), mais aussi aux noms des entités géographiques (cf. *Beograd*, *Italija*) ou des institutions (cf. *RTS*, *Kolarac*, etc.). Dans le cas d’un nom propre polylexical (cf. *Novi Sad*, *Fruška gora*), c’est la tête nominale qui est annotée comme nom propre, alors que le modifieur (dans ce cas, l’adjectif qualificatif) est traité en fonction de sa classe grammaticale. Un exemple est donné dans le tableau 2.3.

<b>od</b>	Prep					
<b>Fruške</b>	A	opsti	genitiv	jednina	zenski rod	pozitiv
<b>gore</b>	N	vlastita	genitiv	jednina	zenski rod	

TABLE 2.3 – Nom propre polylexical

Les noms collectifs sont distingués pour leur comportement spécifique dans l’accord. Ils sont par ailleurs identifiables par leurs terminaisons (cf. *-nje* dans *granje* ‘ensemble de branches’, *-ad* dans *pilad* ‘ensemble de poussins’, etc.) La valeur du trait sous-catégorie pour ces noms est donc **zbirna** (cf. tableau 2.4).

En ce qui concerne les autres traits :

<b>po</b>	Prep				
<b>lišću</b>	N	zbirna	lokativ	jednina	srednji rod

TABLE 2.4 – Nom collectif

- toutes les valeurs de tous les traits sont possibles pour la sous-catégorie **zajednicka** ;
- les noms collectifs (sous-catégorie **zbirna**) sont systématiquement annotés comme singulier vu le patron de déclinaison qu’ils suivent ;
- pour les noms propres (sous-catégorie **vlastita**), toutes les valeurs de tous les traits s’appliquent, à l’exception des noms de famille des femmes, qui sont invariables et dont il est impossible de déterminer le genre à partir des critères de surface (cf. *gospoda Petrović*, *gospode Petrović* etc.). De manière arbitraire mais systématique, ces formes-là sont traitées comme un nominatif singulier, sans indication de genre.

<b>gospode</b>	N	zajednicka	genitiv	jednina	zenski rod
<b>Petrović</b>	N	vlastita	nominativ	jednina	---

TABLE 2.5 – Nom de famille

### 2.1.2 Noms du genre grammatical féminin désignant des êtres du sexe masculin

Le genre des noms tels *deda* ‘grand-père’, *tata* ‘papa’, *vladika* ‘évêque’ etc. est annoté selon leur genre naturel, vu que c’est souvent celui-ci qui impose l’accord aux dépendants du nom.

Dans le cas des noms tels *pijanica* ‘ivrogne’, *sudija* ‘juge’, etc., qui sont du genre grammatical féminin et qui peuvent désigner des êtres des deux sexes, le genre est annoté en fonction du contexte (par exemple, le genre imposé à un adjectif qui s’accorde avec le nom). Si aucun indice de ce type n’est disponible, le nom est annoté comme masculin.

### 2.1.3 Noms composés contenant un trait d’union

Dans la phase de la tokénisation du corpus, les noms composés contenant un trait d’union (cf. *žena-leptir* ‘femme-papillon’) ont été séparés. Dans la phase de l’annotation morphosyntaxique, il faut corriger ce fait en fusionnant les deux lignes de sorte à rétablir le token initial relié par un trait (celui de la touche 6 sur le clavier français). Les traits morpho-syntaxiques de ces formes sont ensuite annotés en considérant la forme complexe.

## 2.2 Les adjectifs

Le traitement des adjectifs fait appel aux traits suivants : la classe de mots, la sous-catégorie, le cas, le nombre, le genre et le degré de comparaison. L'ordre des traits et leurs valeurs possibles sont donnés dans le tableau 2.6, alors que la distribution des traits en fonction de la sous-catégorie est présentée dans le tableau 2.7.

POS	Sous-cat.	Cas	Nombre	Genre	Degré de comp.
A	opisni pokazni prisvojni neodredjeni upitni odnosni	nominativ genitiv dativ akuzativ vokativ instrumental lokativ	jednina mnozina	muski rod zenski rod srednji rod	pozitiv komparativ superlativ

TABLE 2.6 – Adjectifs : traits utilisés et leurs valeurs possibles

Exemples	POS	Sous-cat.	Cas	Nombre	Genre	Degré de comp.
qualificatifs (cf. <i>lep</i> ‘beau’, <i>letnji</i> ‘estival’)	A	opisni	[tous]	[tous]	[tous]	[tous]
autres ( <i>taj</i> ‘ce’, <i>neka</i> ‘(une) certaine’, <i>kakvo</i> ‘quel’)	A	pokazni prisvojni neodredjeni upitni odnosni	[tous]	[tous]	[tous]	---

TABLE 2.7 – Adjectifs : distribution des traits en fonction de la sous-catégorie

### 2.2.1 Les sous-catégories d’adjectifs

Le traitement des adjectifs dans le cadre de ce projet diffère de celui proposé par les grammaires traditionnelles serbes. Par exemple, Stanojčić & Popović (2012) distinguent les adjectifs qualificatifs (*lep* ‘beau’, *hrabar* ‘courageux’), possessifs (*Markov* ‘qui appartient à Marko’, *školski* ‘qui appartient à/relatif à l’école’), massifs (*zlatan* ‘en or’, *drven* ‘en bois’), temporels (*današnji* ‘d’aujourd’hui’, *godišnji* ‘annuel’) et spatiaux (*desni* ‘de droite’, *gornji* ‘supérieur’, ‘d’en haut’) (2012 : 93). Or, cette classification repose sur des critères sémantiques. Par ailleurs, tous les types d’adjectifs énumérés partagent le même

comportement morphosyntaxique. Il a donc été décidé de ne pas introduire ces distinctions sémantiques dans notre jeu d'étiquettes : tous ces sous-types d'adjectifs sont traités comme des adjectifs qualificatifs et leur trait *sous-catégorie* a la valeur ***opisni***.

Sont également traitées comme adjectifs les formes traditionnellement identifiées comme *pridevske zamenice* 'pronoms adjectivaux', quand elles figurent au sein d'un GN. Cette décision est motivée par le fait que ces formes ont le comportement adjectival typique : elles suivent les mêmes patrons de déclinaison que les adjectifs, et à l'intérieur d'un groupe nominal, elles occupent la position canonique d'un adjectif – à gauche du nom. Les valeurs possibles du trait sous-catégorie comprennent donc aussi les valeurs ***pokazni*** 'démonstratif', ***prisvojni*** 'possessif', ***neodredjeni*** 'indéfini', ***upitni*** 'interrogatif' et ***odnosni*** 'relatif'.

La valeur ***pokazni*** regroupe les formes des séries *taj, ova, ona* ; ***prisvojni*** s'applique aux paradigmes de *moj, tvoj, njegov, njen, svoj*, etc., la valeur ***neodredjeni*** 'indéfini' regroupe les formes traditionnellement qualifiées comme pronoms généraux (*opšte*), négatifs (*odrične*) et indéfinis (*neodredene*) et s'applique donc aux paradigmes de *svaki, svakakav, svačiji, neki, nekakav, nečiji, nikoji, nikakav, ničiji*, etc. ***Upitni*** correspond aux formes fléchies de *koji, kakav, koliki, čiji* etc. employées dans un contexte interrogatif, et ***odnosni*** concerne les mêmes formes, mais dans un contexte relatif. Le tableau 2.8 donne quelques illustrations.

<b>pokazni</b>	<b>taj</b> pas, <b>ova</b> sveska, <b>ona</b> pravila
<b>prisvojni</b>	<b>moj</b> pas, <b>tvoja</b> sveska, <b>njegova</b> pravila
<b>neodredjeni</b>	<b>svaki</b> pas, <b>nijedna</b> sveska, <b>neka</b> pravila
<b>upitni</b>	<b>Kojeg</b> psa si video? <b>Kakvu</b> svesku si kupio? <b>Čija</b> pravila slediš?
<b>odnosni</b>	ljudi <b>čija</b> pravila sledim

TABLE 2.8 – Adjectif : exemples des sous-catégories différentes

#### NB1

Les formes des sous-catégories ***pokazni***, ***prisvojni*** et ***neodredjeni*** peuvent également connaître de véritables emplois pronominaux : *Hoću ovu, a ne njegovu* 'Je veux celle-ci et pas la sienne'. Dans ce cas, ces formes sont traitées comme des pronoms. Pour les traits qui s'appliquent dans ce cas, voir la section 2.4 Pronoms.

**NB 2**

Il est également à noter que le trait **prisvojni** ne s'applique pas aux adjectifs dénominatifs comme *Markov* 'qui appartient à Marko', *Marijin* 'qui appartient à Marija', *vojnički* 'militaire' ou *školski* 'scolaire'. Ces adjectifs sont traités comme **opisni**, et la sous-catégorie **prisvojni** est limitée aux seules formes possessives considérées traditionnellement comme pronoms (cf. *moj*, *tvoj*, *njen* etc.).

### 2.2.2 Degré de comparaison de l'adjectif

Pour la sous-catégorie **opisni** 'qualificatif', le degré de comparaison par défaut est le positif, même pour les adjectifs dont le sémantisme se prête mal à la comparaison (cf. *sutrašnji* 'de demain', *letnji* 'estival', *školski* 'scolaire' etc.). En revanche, le degré de comparaison n'est pas marqué pour les autres sous-catégories et ce fait est indiqué en utilisant la valeur '---' dans la colonne correspondant au degré de comparaison.

<b>moj</b>	A	prisvojni	nominativ	jednina	muski rod	---
<b>školski</b>	A	opisni	nominativ	jednina	muski rod	pozitiv
<b>pribor</b>	N	zajednicka	nominativ	jednina	muski rod	

TABLE 2.9 – Adjectif : degré de comparaison

### 2.2.3 Traits *genre* et *nombre* pour la sous-catégorie **prisvojni**

Pour les adjectifs possessifs, le genre et le nombre indiqués sont celui du possédé. Le genre et le nombre du possesseur ne sont pas marqués. Il en est de même pour la personne du possesseur. Par conséquent, les possessifs dans les exemples *njihov stan* 'leur appartement', *njen stan* 'son appartement à elle', *moj stan* 'mon appartement' sont tous annotés comme singulier masculin.

### 2.2.4 Adjectifs (pro)nominalisés

En accord avec nos principes d'annotation de base, les adjectifs nominalisés sont traités en fonction de leur rôle syntaxique. De manière générale, les adjectifs qualificatifs nominalisés sont annotés comme noms, alors que les sous-catégories **neodredjeni**, **pokazni** et **prisvojni** dans les emplois indépendants d'un nom sont considérés comme des pronoms (cf. NB1 ci-dessus). Par conséquent, la forme *mrtvih* dans l'exemple *enciklopedija mrtvih* 'encyclopédie des morts' est traitée comme un nom. Les formes *ovu* et *njegovu* dans

l'exemple *Hoću ovu, a ne njegovu* ‘Je veux celle-ci et pas la sienne’ sont traitées comme des pronoms.

### 2.2.5 Adjectifs déverbaux dérivés des formes des participes passif et actif (*glagolski pridev trpni* et *glagolski pridev radni*)

Il existe en serbe une sous-classe d’adjectifs qui coïncident formellement avec la forme du participe passif (cf. *otvoren* ‘ouvert’, *kupljen* ‘acheté’, *začaran* ‘enchanté’, etc) ou actif (cf. *zalutao* ‘égaré’, *procvetao* ‘fleuri’, etc.). Le traitement de ces formes est décrit en détail dans la section 2.3 Verbes.

### 2.2.6 Adjectifs composés

Dans la phase de la tokénisation du corpus, les adjectifs composés contenant un trait d’union (cf. *kabalističko-astrološki*, *jugo-zapadni*) ont été séparés. Dans la phase de l’annotation morphosyntaxique, il faut corriger ce fait en fusionnant les deux lignes de sorte à rétablir le token initial relié par un trait (celui de la touche 6 sur le clavier français). L’annotation des traits est faite en considérant l’unité dans sa totalité.

## 2.3 Les verbes

Le traitement des verbes comprend les traits morphosyntaxiques suivants : la classe de mots, la sous-catégorie, la forme verbale, la personne, le nombre, le genre et la négation. Précisons qu'ici la sous-catégorie correspond en effet au statut principal ou auxiliaire du verbe. L'ordre des traits et leurs valeurs possibles sont donnés dans le tableau 2.10. Ici, la distribution des traits est définie en fonction de la forme verbale, et non pas selon la sous-catégorie, étant donné que cette première est plus pertinente pour la présence ou l'absence de certains traits (cf. tableau 2.11).

POS	Sous-cat.	Forme verbale	Personne	Nombre	Genre	Négation
V	glavni pomocni	aorist	prvo lice	jednina	muski rod	negiran
		futur	drugo lice	mnozina	zenski rod	---
		imperativ	trece lice		srednji rod	
		imperfekat				
		infinitiv				
		particip_pro				
		particip_sad				
		particip_radni				
		particip_trpni				
		present				

TABLE 2.10 – Verbe : traits utilisés et leurs valeurs possibles

### 2.3.1 Trait *forme verbale*

L'utilisation des valeurs *present*, *futur*, *aorist*, *imperfekat*, *imperativ* et *infinitiv* correspondent à la définition de ces formes telles que présentées dans (Stanojčić & Popović, 2012). En revanche, les étiquettes utilisées pour les formes de différents participes ont été simplifiées : *particip\_radni* correspond à *glagolski pridev radni*, *particip\_trpni* à *glagolski pridev trpni*, *particip\_sad* à *glagolski prilog sadašnji*, et *particip\_pro* à *glagolski prilog prošli*.

Les formes dites du participe passif, telles que *otvoren* 'ouvert', *donet* 'apporté', et celles du participe actif comme *zalutao* 'égaré', *procvetao* 'fleuri', ont en serbe deux fonctionnements distincts : elles peuvent faire partie des formes verbales composées et être réellement des participes, ou avoir le rôle d'épithète ou d'attribut et être en effet des adjectifs, cf. *Ana je zalutala* 'Ana s'est égarée' vs *zalutala mačka* 'un chat égaré'.

La distinction entre ces deux fonctionnements devient problématique quand ces formes sont accompagnées de celles du verbe *biti* 'être' : cette combinaison des formes peut être interprétée soit comme une forme verbale composée, où la forme ambiguë correspondrait à



Exemples	POS	Sous-cat.	Forme	Personne	Nombre	Genre	Nég.
les temps ( <i>radim</i> ‘je travaille’, <i>pričaće</i> ‘elle parlera’)	V	[tous]	prezent futur aorist imperfekat	[tous]	[tous]	---	[tous]
impératif ( <i>radi!</i> ‘travaille!’)	V	glavni	imperativ	drugo lice	[tous]	---	[tous]
				prvo lice	mnozina	---	[tous]
infinitif ( <i>raditi</i> ‘travailler’)	V	glavni	infinitiv	---	---	---	[tous]
participe actif ( <i>radio</i> ‘travaillé’)	V	[tous]	particip_radni	---	[tous]	[tous]	[tous]
participe passif ( <i>rađen</i> ‘travaillé’)	V	glavni	particip_trpni	---	[tous]	[tous]	---
participe présent ( <i>ra- deći</i> ‘tra- vaillant’)	V	glavni	particip_sad	---	---	---	[tous]
participe passé ( <i>ra- divši</i> ‘ayant travaillé’)	V	glavni	particip_pro	---	---	---	[tous]

TABLE 2.11 – Verbe : distribution des traits en fonction de la forme verbale

un participe (et serait donc traitée comme un verbe), soit comme un emploi attributif du verbe *biti* ‘être’, où la forme ambiguë serait plutôt un adjectif utilisé comme attribut de sujet (ou *imenski predikativ* dans la terminologie serbe). C’est le cas des exemples comme *Hotel je otvoren* ‘L’hôtel est ouvert’, *Ključ je izgubljen* ‘La clé est perdue’, etc. Même si des critères sémantiques pourrait permettre à un annotateur humain de faire la distinction, elle est impossible à opérer sur base des critères de surface accessibles à un étiqueteur statistique. Par conséquent, le critère de distinction qui a été adopté est le suivant : **toute occurrence** d’une forme correspondant à un participe accompagnée du verbe *biti* ‘être’ est annotée comme participe, et la forme du verbe *biti* ‘être’ est considérée comme verbe auxiliaire. C’est donc le cas des formes soulignées dans les exemples *Hotel je otvoren* ‘L’hôtel est ouvert’ ou *Ključ je izgubljen* ‘La clé est perdue’. En revanche, les occurrences de ces formes dans la position de modifieur d’un nom (*atribut* en serbe), comme *otvoren hotel* ‘l’hôtel ouvert’, *izgubljen ključ* ‘la clé perdue’, etc., sont annotées comme adjectifs.

### 2.3.2 Verbes auxiliaires dans les formes surcomposées

Le serbe dispose de deux formes verbales surcomposées : le potentiel passé (équivalent dans un certain degré du conditionnel passé français) et le plus-que-parfait. Dans l'exemple *On je bio došao* 'Il était venu', le plus-que-parfait *je bio došao* est constitué de *je bio*, parfait du verbe *jesam* 'être', et de *došao*, participe actif du verbe *doći* 'venir'. Le parfait du verbe *jesam* lui-même consiste en *je*, présent du verbe *jesam*, qui est la forme de l'auxiliaire, et en *bio*, participe passé du verbe *biti*, le verbe principal.

Ici, le parfait du verbe *jesam* devrait être annoté en tant que verbe auxiliaire faisant partie du plus-que-parfait. Or, ceci serait contraire au principe de la correspondance 1 :1 entre les tokens et les étiquettes. On traite donc l'auxiliaire composé de manière suivante : on attribue l'étiquette du verbe auxiliaire à chacune des formes qui le constituent, alors que la forme du participe passé porte celle du verbe principal. Un exemple est donné dans le tableau 2.12.

<b>je</b>	V	pomocni	prezent	trece lice	jednina	---	---
<b>bio</b>	V	pomocni	particip_radni	---	jednina	muski rod	---
<b>došao</b>	V	glavni	particip_radni	---	jednina	muski rod	---

TABLE 2.12 – Verbe auxiliaire : formes surcomposées

#### NB3

Les verbes modaux ou aspectuels complémentés par un infinitif ne sont pas considérés comme temps composés : dans *trebati raditi* 'falloir travailler' ou *prestati raditi* lit. 'arrêter travailler', les deux verbes ont le statut d'un verbe principal.

### 2.3.3 Trait *négation*

Ce trait s'applique **seulement** aux formes niées **synthétiques** des verbes *biti* 'être', *imati* 'avoir' et *hteti* 'vouloir'. Autrement dit, il est marqué pour les formes comme *nisam* 'je ne suis pas', *nisi* 'tu n'es pas', *nije* 'il n'est pas', etc., *nemam* 'je n'ai pas', *nemaš* 'tu n'as pas', *nema* 'il n'a pas' etc., *neću* 'je ne veux pas', *nećeš* 'tu ne veux pas', *neće* 'il ne veut pas', etc. Il est également pertinent pour les formes du verbe défectif *nemoj* 'ne fais pas' (*nemoj* 'ne fais pas', *nemojte* 'ne faites pas', *nemojmo* 'ne faisons pas'). Dans le cas des formes niées **analytiques**, comme *ne radi* 'il ne travaille pas', la négation est marquée par la particule, et non pas par la forme verbale. Dans ce cas, la valeur du trait *négation*

pour le verbe est '---'.

## 2.4 Les pronoms

Le traitement des pronoms fait appel aux traits morphosyntaxiques suivants : la classe de mots, la sous-catégorie, le cas, le nombre, le genre et la personne. L'ordre des traits et leurs valeurs possibles sont donnés dans le tableau 2.13, et la distribution des traits en fonction de la sous-catégorie est présentée dans le tableau 2.14.

POS	Sous-cat.	Cas	Nombre	Genre	Personne
P	licna pokazna neodredjena odnosna povratna upitna brojna	nominativ genitiv dativ akuzativ vokativ instrumental lokativ	jednina mnozina	muski rod zenski rod srednji rod	prvo lice drugo lice trece lice

TABLE 2.13 – Pronom : traits utilisés et leurs valeurs possibles

Exemples	POS	Sous-cat.	Cas	Nombre	Genre	Personne
personnels ( <i>ja</i> 'je', <i>mi</i> 'nous')	P	licna	[tous]	[tous]	---	prvo lice drugo lice
		licna	[tous]	[tous]	[tous]	trece lice
démonstratifs, indéfinis, numéraux ( <i>onaj</i> 'celui-là', <i>jedan</i> 'un (certain)')	P	pokazna neodredjena brojna prisvojna	[tous]	[tous]	[tous]	---
réflexif ( <i>se</i> , <i>sebe</i> 'se, soi')	P	povratna	genitiv dativ akuzativ instrumental lokativ	---	---	---
interrogatifs et relatifs ( <i>ko</i> 'qui', <i>koji</i> 'qui')	P	upitna odnosna	[tous]	[tous]	[tous]	---

TABLE 2.14 – Pronom : distribution des traits en fonction de la sous-catégorie

### 2.4.1 Les sous-catégories des pronoms

Les formes traditionnellement appelées *pridevske zamenice* ‘pronoms adjectivaux’ ont un double traitement en fonction de leur contexte. Si elles figurent indépendamment d’un nom, elles sont traitées comme pronoms. Les valeurs ***pokazna*** ‘démonstratif’, ***neodredjena*** ‘indéfini’ et ***prisvojna*** ‘possessif’ s’appliquent donc aux mêmes formes que dans la section 2.2, mais dans les cas où elles sont utilisées en dehors d’un GN.

La valeur ***brojna*** ‘numéral’ s’applique aux emplois pronominaux des formes *jedan* ‘un’, *prvi* ‘premier’ et *drugi* ‘deuxième, autre’, comme dans l’exemple *Jedan je stigao, ali drugi kasni* ‘L’un est arrivé, mais l’autre est en retard’.

### 2.4.2 Annotation du genre et du nombre

Le genre n’est pas annoté sur les pronoms personnels à la 1<sup>ère</sup> et à la 2<sup>e</sup> personne. Pour les pronoms personnels à la 3<sup>e</sup> personne, le genre est annoté, même si au niveau formel il n’est pas distingué à tous les cas (cf. la forme *ga* peut correspondre au masculin ou au neutre, alors que *ih* peut correspondre à chacun des trois genres). L’annotateur identifiera la bonne valeur à partir du contexte.

Pour les indéfinis et les relatifs, seuls les traits morphosyntaxiques formellement marqués seront encodés. Pour les pronoms tels que *neko* ‘quelqu’un’, *nešto* ‘quelque chose’, *svako* ‘chacun’, *svašta* ‘toute chose’, *niko* ‘personne’, *ništa* ‘rien’, *iko* ‘personne/quelqu’un’, *išta* ‘rien/quelque chose’, nous indiquons seul le type du pronom (indéfini) et le cas. Il en est de même pour les pronoms relatifs et interrogatifs *ko* ‘qui’, *šta/što* ‘que’. En revanche, pour les relatifs *koji* ‘qui’, *kakav* ‘quel’, *koliki* ‘de quelle taille’, le genre et le nombre sont annotés aussi, et c’est également le cas des démonstratifs *taj*, *ovaj* et *onaj* (cf. tableau 2.17).

Quant au traitement du pronom réflexif, sa forme clitique *se* est systématiquement annotée comme étant à l’accusatif. En revanche, pour la forme pleine, tous les cas obliques sont éligibles.

### 2.4.3 Pronoms indéfinis discontinus

Il existe une série des pronoms indéfinis en serbe dérivés des pronoms interrogatifs *ko* ‘qui’ et *šta* ‘quoi’ par préfixation en *ni-* et en *i-* (cf. *niko* ‘personne’, *ništa* ‘rien’, *iko* ‘qui que ce soit’, *išta* ‘quoi que ce soit’). Si ces formes se trouvent dans un syntagme prépositionnel, elles deviennent discontinues : le préfixe se détache de la base et la préposition vient s’insérer entre les deux : *ni za koga* ‘pour personne’, *ni o čemu* ‘de rien’, *i prema kome* ‘envers qui que ce soit’, *i sa čim* ‘avec quoi que ce soit’. Pour ces cas de figure, la solution suivante a été adoptée : le préfixe est annoté comme particule, la base du pronom porte

Pronom	Sous-cat.	Cas	Nombre	Genre	Pers.
brojne ( <i>jedan, drugi, prvi</i> )	brojna	[tous]	[tous]	[tous]	---
<i>niko</i>	neodredjena	[tous]	---	---	---
<i>ništa</i>	neodredjena	[tous]	---	---	---
<i>neko</i>	neodredjena	[tous]	---	---	---
<i>nešto</i>	neodredjena	[tous]	---	---	---
<i>iko</i>	neodredjena	[tous]	---	---	---
<i>išta</i>	neodredjena	[tous]	---	---	---
<i>ko</i> (odnosna)	odnosna	[tous]	---	---	---
<i>što</i> (odnosna)	odnosna	[tous]	---	---	---
<i>se</i>	povratna	akuzativ	---	---	---
<i>sebe</i>	povratna	genitiv dativ akuzativ instrumental lokativ	---	---	---
<i>svi</i>	neodredjena	[tous]	---	---	---
<i>koji</i>	odnosna	[tous]	---	---	---
pokazne ( <i>taj, ta, to ; onaj, ona, ono ; ovaj, ova, ovo</i> )	pokazna	[tous]	---	---	---

TABLE 2.15 – Pronom : genre et nombre

l'étiquette du pronom indéfini, alors que la préposition est étiquetée de manière habituelle. Voir les exemples dans le tableau 2.16.

<b>ni</b>	Part					
<b>za</b>	Prep					
<b>koga</b>	P	neodredjena	akuzativ	---	---	---
<b>ni</b>	Part					
<b>iz</b>	Prep					
<b>jednog</b>	P	neodredjena	genitiv	jednina	muski rod	---

TABLE 2.16 – Pronom : discontinuité dans le GP

#### 2.4.4 Distinction entre le pronom personnel fléchi et l’auxiliaire

Dans les exemples comme *Unosio je hladnoću u pregrštima, ili je gurao pred sobom*, la deuxième forme *je* est ambiguë : il n’est pas clair si elle correspond au pronom personnel *ona* ‘elle’ à l’accusatif singulier (qui serait alors une reprise de *hladnoća* ‘le froid’) ou bien à l’auxiliaire *jesam* ‘être’ à la troisième personne du singulier, qui ferait partie du parfait avec le participe *gurao* ‘poussé’. Un test simple permet cependant de confirmer qu’il s’agit du pronom : il suffit de mettre la forme verbale à la première personne. Dans l’exemple *Unosio sam hladnoću u pregrštima, ili je gurao pred sobom*, la forme *je* ne peut représenter que le pronom, vu que l’auxiliaire à la première personne du singulier a la forme *sam*, comme dans la première proposition. Ceci montre que dans la deuxième proposition l’auxiliaire est éliminé, et le pronom maintenu. Par conséquent, dans ce type d’exemples, la forme *je* doit être annotée comme pronom personnel.

<b>Unosio</b>	V	glavni	particip_radni	---	jednina	muski rod	---
<b>je</b>	V	pomocni	prezent	trece lice	jednina	---	---
<b>hladnoću</b>	N	zajednicka	akuzativ	jednina	zenski rod		
<b>u</b>	Prep						
<b>pregrštima</b>	N	zajednicka	instrumental	mnozina	zenski rod		
<b>,</b>	Z						
<b>ili</b>	C	koordinirani					
<b>je</b>	P	licna	akuzativ	jednina	zenski rod	trece lice	
<b>gurao</b>	V	glavni	particip_radni	---	jednina	muski rod	---
<b>pred</b>	Prep						
<b>sobom</b>	P	povratna	instrumental	---	---	---	

TABLE 2.17 – Pronom *vs* auxiliaire

#### 2.4.5 Le pronom *što*

La forme *što* est ambiguë. Elle peut représenter un pronom relatif, comme dans les exemples *Pas što laje* (= *Pas koji laje*) ‘le chien qui aboie’, *Knjiga što čitam* (= *Knjiga koju čitam*) ‘le livre que je lis’. Il se comporte ici comme un relatif typique : il a une double fonction syntaxique, celle de l’introducteur de la relative, mais aussi celle d’un élément syntaxique dans la relative elle-même (sujet dans le premier exemple, objet direct dans le deuxième). En revanche, sa nature dans l’exemple suivant est moins claire : *kuća što sam je kupio* lit. ‘la maison que je l’ai achetée’. Ici, le rôle de l’objet direct dans la relative est assuré par le pronom personnel *je* ‘la’, et la forme *što* ‘que’ est en effet réduite à la fonction de subordination. Par conséquent, dans les exemples des relatives utilisant un pronom personnel au lieu d’un relatif à double emploi, la forme *što* sera traitée comme un subordonnant.

<b>kuća</b>	N	zajednicka	nominativ	jednina	zenski rod		
<b>što</b>	C	subordinirani					
<b>sam</b>	V	pomocni	prezent	prvo lice	jednina	---	---
<b>je</b>	P	licna	akuzativ	jednina	zenski rod	trece lice	
<b>kupio</b>	V	glavni	particip_radni	---	jednina	muski rod	---

TABLE 2.18 – Pronom *što* : exemple 1

Il faut souligner que ceci n'arrive jamais avec le sujet de la relative : les exemples comme *\*Pas što on laje* 'le chien qui il aboie' sont agrammaticaux. Dans l'exemple *pas što laje*, la forme *što* est considérée comme un pronom relatif au nominatif (cf. tableau 2.19).

<b>pas</b>	N	zajednicka	nominativ	jednina	muski rod		
<b>što</b>	P	odnosna	nominativ	---	---	---	
<b>laje</b>	V	glavni	prezent	trece lice	jednina	---	---

TABLE 2.19 – Pronom *što* : exemple 2

Il faut également faire la distinction entre les emplois relatifs de cette forme et de ses emplois en tant que complétif. Dans l'exemple *Raduje me to što mi je rekao* (au sens 'Ce qu'il m'a dit me réjouit'), il s'agit d'un relatif, dont l'antécédent est *to* et qui a la fonction du sujet dans la relative. En revanche, dans la phrase *Raduje me to što dolazi* (au sens 'Le fait qu'il vient me réjouit'), la forme *što* est un subordonnant introduisant une complétive qui exprime le contenu résumé par le pronom *to*. Il est utile de noter que *to* est remplaçable par *ono* dans le premier cas, mais pas dans le deuxième (sans changer le sens). Les deux traitements sont illustrés dans le tableau 2.20.

<b>Raduje</b>	V	glavni	prezent	trece lice	jednina	---	---
<b>me</b>	P	licna	akuzativ	jednina	---	prvo lice	
<b>to</b>	P	pokazna	nominativ	jednina	srednji rod	---	
<b>što</b>	P	odnosna	nominativ	---	---	---	
<b>mi</b>	P	licna	dativ	jednina	---	prvo lice	
<b>je</b>	V	pomocni	prezent	trece lice	jednina	---	---
<b>rekao</b>	V	glavni	particip_radni	---	jednina	muski rod	---
<b>Raduje</b>	V	glavni	prezent	trece lice	jednina	---	---
<b>me</b>	P	licna	akuzativ	jednina	---	prvo lice	
<b>to</b>	P	pokazna	nominativ	jednina	srednji rod	---	
<b>što</b>	C	subordinirani					
<b>dolazi</b>	V	glavni	prezent	trece lice	jednina	---	---

TABLE 2.20 – Pronom *što* : exemple 3



## 2.5 Les numéraux

Le traitement des numéraux fait appel aux traits morphosyntaxiques suivants : la classe de mots, la sous-catégorie, le cas, le nombre et le genre. L'ordre des traits et leurs valeurs possibles sont donnés dans le tableau 2.21, alors que la distribution des traits en fonction de la sous-catégorie est présentée dans le tableau 2.22.

POS	Sous-cat.	Cas	Nombre	Genre
Num	opsti redni zbirni	nominativ genitiv dativ akuzativ vokativ instrumental lokativ	jednina mnozina	muski rod zenski rod srednji rod

TABLE 2.21 – Numéral : traits utilisés et leurs valeurs possibles

Exemples	POS	Sous-cat.	Cas	Nombre	Genre
<i>jedan</i> ‘un’, <i>dva</i> ‘deux’	Num	opsti	[tous]	[tous]	[tous]
<i>tri</i> ‘trois’, <i>četiri</i> ‘quatre’	Num	opsti	[tous]	---	---
à partir de <i>pet</i> ‘cinq’	Num	opsti	---	---	---
ordinaux ( <i>prvi</i> ‘premier’)	Num	redni	[tous]	[tous]	[tous]
collectifs ( <i>dvoje</i> ‘deux’, <i>troje</i> ‘trois’)	Num	zbirni	[tous]	mnozina	[tous]

TABLE 2.22 – Numéral : distribution des traits en fonction de la sous-catégorie

### 2.5.1 Le genre et le nombre des numéraux

Pour les formes du numéral *jedan* ‘un’, nous annotons le cas, le nombre et le genre (cf. tableau 2.23).

<b>jedne</b>	Num	opsti	genitiv	jednina	zenski rod		
<b>večeri</b>	N	zajednicka	genitiv	jednina	zenski rod		

TABLE 2.23 – Numéral *jedan*

En ce qui concerne le numéral *dva* ‘deux’, deux cas de figure existent. Si le numéral est décliné en accord avec la fonction du groupe nominal dans la phrase, on annote le cas, le nombre et le genre (cf. *Priča sa dvama učenicima* lit. ‘Il parle avec deux.INS élèves.INS).

Dans le cas contraire, où le numéral reste invariable et impose au nom la forme du paucal (cf. *Priča sa dva učenika* lit. ‘Il parle avec deux élèves.PAUC), *dva* ‘deux’ est traité comme un numéral invariable (cf. le traitement de *pet* ‘cinq’) et par conséquent on n’annote que la sous-catégorie. Voir les exemples dans le tableau 2.24.

Un traitement parallèle est mis en place pour les numéraux *tri* ‘trois’ et *četiri* ‘quatre’ : s’ils sont déclinés au même cas que le nom, nous annotons le cas ; sinon, nous considérons qu’il s’agit d’un numéral invariable.

<b>Priča</b>	V	glavni	prezent	trece lice	jednina	---	---
<b>sa</b>	Prep						
<b>dvama</b>	Num	osnovni	instrumental	mnozina	muski rod		
<b>učenicima</b>	N	zajednicka	instrumental	mnozina	muski rod		
<b>Priča</b>	V	glavni	prezent	trece lice	jednina	---	---
<b>sa</b>	Prep						
<b>dva</b>	Num	osnovni	---	---	---		
<b>učenika</b>	N	zajednicka	genitiv	jednina	muski rod		

TABLE 2.24 – Numéral *dva* : fléchi *vs* invariable

## 2.5.2 Paucal

Les formes spéciales dites de *paukal* ‘paucal’, résidu du dual de l’ancien slave, imposées aux noms par les numéraux *dva* ‘deux’, *tri* ‘trois’ et *četiri* ‘quatre’, sont traitées comme **génitif singulier**. Ceci est fait dans un souci de limiter le nombre de valeurs possibles pour la catégorie du cas.

## 2.5.3 Formes en *-ak*

Les formes à sémantisme approximatif comme *dvadesetak* ‘à peu près vingt’ sont considérées comme des numéraux collectifs, mais leur genre et nombre ne sont pas annotés.

## 2.6 Les adverbes

Le traitement des adverbes fait appel aux traits morphosyntaxiques suivants : la classe de mots, la sous-catégorie et le degré de comparaison. L'ordre des traits et leurs valeurs possibles sont donnés dans le tableau 2.25, alors que la distribution des traits en fonction de la sous-catégorie est indiquée dans le tableau 2.26.

POS	Sous-cat.	Degré de compar.
Adv	opsti odnosni upitni neodredjeni	pozitiv komparativ superlativ

TABLE 2.25 – Adverbe : traits utilisés et leurs valeurs possibles

Exemples	POS	Sous-cat.	Degré de compar.
généraux ( <i>sutra</i> ‘de-main’, <i>mirno</i> ‘tranquille-ment’)	Adv	opsti	[tous]
autres ( <i>kada</i> ‘quand’, <i>nikada</i> ‘jamais’)	Adv	odnosni upitni neodredjeni	---

TABLE 2.26 – Adverbe : distribution des traits en fonction de la sous-catégorie

### 2.6.1 Sous-catégories des adverbes

Nous abandonnons la classification sémantique proposée par Stanojčić & Popović (2012). Nous distinguons les adverbes relatifs (*odnosni*, cf. *Otišao je tamo gde sam mu rekla* ‘Il est allé là où je lui ai dit’), interrogatifs (*upitni*, cf. *Gde je otišao?* ‘Où est-il allé?’) et indéfinis (*neodredjeni*, cf. *nikako* ‘aucunement’, *nikada* ‘jamais’, *nekako* ‘d’une certaine façon’, *nekada* ‘autrefois’ etc.). Tous les autres adverbes sont considérés comme *opsti*.

### 2.6.2 Degré de comparaison des adverbes

Ce trait n’est pas marqué sur les sous-catégories *odnosni*, *upitni* et *neodredjeni*. Il a la valeur par défaut de *pozitiv* dans la sous-catégorie *opsti*, même pour les adverbes qui ne se comparent a priori pas, comme *letos* ‘l’été dernier’ ou *levo* ‘à gauche’.

### 2.6.3 Forme *kad* ‘quand’

Cette forme est traitée comme une conjonction quand elle introduit une subordonnée temporelle, cf. *Reći ću mu kad dođe* ‘Je le lui dirai quand il viendra’. En revanche, elle est traitée comme adverbe interrogatif (*upitni*) dans les interrogatives directes et indirectes, cf. *Kad Marko dolazi?* ‘Quand est-ce que Marko vient?’ et *Pita kad Marko dolazi* ‘Il demande quand vient Marko’, et comme adverbe relatif (*odnosni*) dans les relatives, cf. *dan kad je stigao* ‘le jour où il est arrivé’. Il est utile de noter que les emplois relatifs de la forme *kad* peuvent toujours être remplacés par un relatif non ambigu : *dan u koji je stigao* lit. ‘le jour dans lequel il est venu’.

## 2.7 Conjonctions

Le traitement des conjonctions fait appel à la classe de mots et à la sous-catégorie. L’ordre des traits et leurs valeurs possibles sont donnés dans les tableaux 2.27 et 2.28.

POS	Sous-catégorie
C	subordinirani koordinirani

TABLE 2.27 – Conjonctions : traits utilisés et leurs valeurs possibles

Exemples	POS	Sous-catégorie
<i>i</i> ‘et’, <i>ili</i> ‘ou’, <i>ali</i> ‘mais’	C	koordinirani
<i>jer</i> ‘parce que’, <i>da</i> ‘que’, <i>kad</i> ‘quand’	C	subordinirani

TABLE 2.28 – Conjonctions : exemples des sous-catégories

Certaines formes qui fonctionnent comme conjonctions de coordination peuvent également être employées comme des particules. Ceci est notamment le cas des formes *i* et *pa*.

*I* est traité comme conjonction dans les emplois où il établit réellement une coordination entre deux éléments syntaxiques, cf. *Petar i Marko* ‘Petar et Marko’, *Sedim i jedem* ‘Je suis assis et je mange’, *lep pas i divlja mačka* ‘un beau chien et un chat sauvage’. Il est considéré comme une particule quand il est employé au sens de ‘aussi’ : *I moja majka je tu* lit. ‘Et ma mère est là’, ‘Ma mère est là aussi’.

*Pa* est considéré comme conjonction de coordination dans les exemples comme *Oprao je zube pa je legao da spava* ‘Il s’est brossé les dents puis il s’est couché’. En revanche, il est traité comme particule dans les emplois comme *Pa rekla sam ti da to ne radiš* ‘Mais

je t’avais dit de ne pas le faire’. Pour plus d’informations, voir la section 3.2, dédiée au traitement des particules.

## 2.8 Les prépositions, les interjections, les particules, les mots étrangers et la ponctuation

Pour ces classes de mots nous n’indiquons que la classe de mots. Les étiquettes sont données dans le tableau 2.29. Des listes des prépositions et des particules ont été compilées à partir de (Mrazović, 2009) et sont disponibles dans la suite de ce document. Le traitement des particules y est présenté en détail.

Classe	POS
Prépositions	Prep
Particules	Part
Mots étrangers	X
Ponctuation	Z

TABLE 2.29 – Autres classes : étiquettes POS

## 3. Gestion des cas de figure spécifiques

### 3.1 Prépositions

Afin de faciliter leur identification dans le corpus, nous reprenons ici une liste de prépositions proposée par Mrazović (2009).

bez	mimo	pod	protiv	umesto
blagodareći	na	podno	put	unutar
blizu	nad	pokraj	putem	uoči
čelo	nadno	polovinom	radi	uprkos
dno	nadohvat	pomoću	sa	usled
do	nadomak	poput	saglasno	usred
dovrh	nakon	pored	saobrazno	usuprot
duž	nakraj	posle	sem	ususret
ispod	namesto	posred	shodno	uvrh
ispred	naspram	posredovanjem	silom	uz
iz	nasred	posredstvom	skraj	uzduž
iza	nasuprot	potkraj	sledstveno	van
između	navrh	poviše	slično	više
iznad	niz	povodom	sred	vrh
izuzev	niže	povrh	sredinom	za
izvan	o	pozadi	suprotno	zarad
ka	od	pre	tokom	zaslugom
kod	oko	pred	u	zbog
kraj	okolo	preko	udno	zahvaljujući
krajem	osim	prema	uinat	
kroz	po	pri	uključujući	
među	početkom	prilikom	ukraj	

## NB4

Certaines de ces formes sont ambiguës : elles peuvent fonctionner comme prépositions, mais aussi comme d'autres classes de mots. Il s'agit typiquement d'adverbes, mais aussi de verbes et de noms. Par conséquent, il faut prendre en considération leur contexte syntaxique. Pour être considérées comme des prépositions, ces formes doivent introduire un complément nominal ou pronominal, dans un cas imposé par la forme, cf. *Izašao je zahvaljujući Ani* 'Il est sorti grâce à Ana', où *zahvaljujući* introduit le nom *Ani*, qui est au datif, imposé par la préposition. En revanche, dans *Izašao je zahvaljujući se Ani* 'Il est sorti en remerciant Ana', la forme *zahvaljujući* est en effet un participe. Cf. également les oppositions *Ostavio ga je nadohvat ruke* 'Il l'a laissé à portée de main' (préposition) vs *Pobeda je bila nadohvat* 'La victoire était proche' (adverbe) ; *Kupio je sve, uključujući televizor* 'Il a tout acheté, y compris la télé' (préposition) vs *Začutao je uključujući televizor* 'Il s'est tu en allumant la télé' (participe) ; *Obratio se novinarima prilikom konferencije* 'Il s'est adressé aux journalistes lors de la conférence de presse' (préposition) vs *Obraća se novinarima svakom prilikom* 'Il s'adresse aux journalistes à toutes occasion' (nom).

### 3.2 Listes des particules

Notre traitement des particules est basé sur celui de Mrazović (2009, p. 466-512). L'auteure introduit 4 types de particules, basés largement sur des distinctions sémantiques ; nous ne les reprenons pas ici. L'auteure souligne également que certaines particules coïncident formellement avec des adverbes et indique que la distinction entre ces deux classes de mots se base sur le fait que les adverbes sont des dépendants verbaux qui servent à préciser ou situer le contenu verbal, alors que les particules réfèrent à la phrase entière et servent à exprimer l'attitude de l'énonciateur. Cependant, cette différenciation est délicate à opérer en contexte. Pour faciliter le travail des annotateurs, nous donnons d'abord la liste des particules censées être non ambiguës (par rapport aux adverbes). Ensuite, nous reprenons celles marquées comme ambiguës par Mrazović (2009) et donnons des indices pour opérer la désambiguïsation.

## Liste des particules non ambiguës par rapport aux adverbes

a	doduše	međutim	obavezno	uglavnom
ako	dosta	možda	očito	umalo
ala	e	naime	odveć	uopšte
ama	eno	najzad	odviše	uostalom
bar	eto	naprotiv	oko	upravo
barem	evo	naravno	ono	vala
baš	gle	naročito	otprilike	valjda
besumnje	hajde	navodno	pa	veoma
bezmalomalo	i	nažalost	samo	verovatno
bogme	inače	ne	srećom	vrlo
čak	ionako	neka	štaviše	zaista
čas	ipak	nemoguće	što	zamalo
časkom	izgleda	neosporno	suviše	zapravo
da	jedino	nepotrebno	sve	zar
dabogda	jedva	nešto	ta	zbilja
dabogme	još	nesumnjivo	takođe	
dakako	kao	ni	taman	
dakle	li	nipošto	tek	
de	ma	no	to	

Liste des particules **ambiguës** par rapport aux adverbes

badava	konačno	onako	sigurno	teško
daleko	lično	onda	skoro	već
dobro	malo	potpuno	slobodno	više
dosta	mnogo	praktično	slučajno	
gotovo	najmanje	prilično	svakako	
jednostavno	nikako	prosto	tačno	
još	obično	sad	tako	

En ce qui concerne les formes ambiguës, les critères suivants peuvent être utilisés pour identifier de quelle classe de mots il s'agit :

1. construction **Particule + da** : des particules en serbe ont la capacité d'introduire cette construction spécifique, cf. *Skoro/gotovo/prosto da ne poveruješ, Teško da će stići, Nikako da stigne*. Si une forme ambiguë se trouve dans cette construction, elle sera traitée comme particule.



2. reformulation du type **Particule + je + da/što** : si une forme ambiguë permet ce type de reformulation, cela montre qu'elle porte sur toute la phrase, et non seulement sur le processus verbal (cf. *Sigurno si umorna* 'Tu es certainement fatiguée' => *Sigurno je da si umorna* 'Il est certain que tu es fatiguée'). Par conséquent, elle sera traitée comme particule .
3. présence des virgules : étant donné qu'elles portent sur toute la phrase, les particules peuvent se trouver en dehors de sa structure syntaxique proprement dite. Dans ce cas, elles sont séparées par des virgules du reste de la phrase, cf. *Jednostavno, treba se odlučiti*. Ce type d'utilisation des formes ambiguës sera traité comme particule.
4. questions **kako ?** 'comment ?', **kada ?** 'quand ?', **gde ?** 'où ?', **koliko ?** 'combien ?' : si la forme ambiguë répond à l'une de ces questions dans la phrase, elle sera traitée comme adverbe (cf. *Govori jednostavno* 'Il parle simplement' => **kako govori ?** 'comment parle-t-il ?' *jednostavno* 'simplement').

Des exemples des oppositions adverbe - particule :

1. *Sigurno si umorna* (particule [critère 2]) vs *Kako je govorio ? - Sasvim sigurno* (adverbe [critère 4])
2. *Teško da će doći* (particule [critère 1]) vs *Kako si uspeo da ga ubediš ? - Teško*. (adverbe [critère 4])
3. *Badava se trudiš* (particule [critère 2]) vs *Jabuke su badava* (adverbe [critère 4])
4. *Jednostavno, treba se odlučiti* (particule [critère 3]) vs *Oblači se jednostavno* (adverbe [critère 4])

### 3.3 Autres cas de figure

1. Adjectifs
  - (a) **ma koliki** : *ma* = particule, *koliki* = adjectif indéfini. Voir le traitement des pronoms discontinus (cf. section 2.4.3).
2. Pronoms
  - (a) **sve to** je doneo : *sve* = adjectif indéfini, *to* = pronom démonstratif
  - (b) **svi mi** mu verujemo : *svi* = adjectif indéfini, *mi* = pronom personnel
  - (c) **onaj novi** je došao : *onaj* = adjectif démonstratif, *novi* = nom commun
  - (d) **onaj koji** je došao : *onaj* = pronom démonstratif, *koji* = pronom relatif
3. Numéraux

- (a) **drugi** : traité comme numéral ordinal seulement dans le sens de ‘deuxième’ cf. *Ovo mi je druga knjiga ove nedelje* ‘C’est le deuxième livre que je lis cette semaine’. Dans le sens de ‘autre, différent’, il est annoté comme adjectif qualificatif, cf. *Nađi neki drugi način* ‘Trouve un moyen différent’. S’il est utilisé indépendamment d’un nom, il est traité comme pronom numéral, cf. *Stigao je i drugi* ‘Le deuxième est arrivé aussi’.

#### 4. Noms

- (a) L’instrumental de certains noms s’est grammaticalisé et fonctionne également comme un mot invariable. Des exemples :
1. **većinom** : *Većinom su zadovoljni* ‘Ils sont majoritairement contents’ (adverbe) *vs* *Zadovoljan je većinom odgovora* ‘Il est content de la majorité des réponses’ (nom)
  2. **početkom** : *Došao je početkom popodneva* ‘Il est venu en début de l’après-midi’ (préposition) *vs* *Razočaran je početkom filma* ‘Il est déçu du début du film’ (nom)
  3. **tokom** : *Upoznala sam ga tokom leta* ‘Je l’ai rencontré durant l’été’ (préposition) *vs* *Nezadovoljan je tokom stvari* ‘Il n’est pas content du déroulement des choses’ (nom)
  4. **mahom** : *Nalazili su se mahom u gradu* ‘Ils se retrouvaient majoritairement en ville’ (adverbe)

#### 5. Mots invariables

- (a) **kao**
- conjonction si utilisé au sens comparatif : *Voli ga kao brata* ‘Elle l’aime comme un frère’
  - particule quand il est possible de le remplacer par *tobože* : *On je kao bolestan* ‘Il est soi-disant malade’
  - adverbe quand il est utilisé au sens de ‘en tant que’ : *Radi kao novinar* ‘Il travaille comme journaliste’
- (b) **stoga, zato** : des adverbes, même quand ils se trouvent en tête de proposition, cf. *Stoga/Zato je došao da proveriti* ‘Ainsi il est venu s’assurer’. Bien qu’ils semblent introduire ici une subordonnée (utilisée indépendamment d’une proposition principale), ces formes sont en effet mobiles : *Došao je stoga/zato da proveriti*. Il ne s’agit donc pas de véritables subordonnants. La situation est différente s’ils font partie des unités polylexicales *stoga što* et *zato što* : ces unités sont effectivement des subordonnants et elles introduisent des subordonnées.

Néanmoins, même à l'intérieur de ces expressions, les formes *stoga* et *zato* sont annotées comme adverbes, et c'est la forme *što* qui porte l'étiquette de conjonction subordonnée.

- (c) **tako da** : *tako* = adverbe, *da* = conjonction de subordination
- (d) **kao da** : *kao* = conjonction de subordination, *da* = conjonction de subordination
- (e) **kao i** : *kao* = conjonction de subordination, *i* = particule

## Bibliographie

Pavica Mrazović. *Gramatika srpskog jezika za strance*. Izdavačka knjižarnica Zorana Stojanovića, 2009.

Živojin Stanojčić and Ljubomir Popović. *Gramatika srpskog jezika*. Zavod za udžbenike, 2012.