

Richard Renault & Stéphane Ferrari

CRISCO, EA 4255

Université de Caen Normandie

Langues & Langage à la croisée des Disciplines

workshop : *WS1 Métrique, musication*

Paris, septembre 2024

**ANALYSE AUTOMATIQUE
D'UN CORPUS DE TEXTES VERSIFIÉS
(métrique et syntaxe)**

Richard Renault & Stéphane Ferrari

CRISCO, EA 4255

Université de Caen Normandie

Langues & Langage à la croisée des Disciplines

workshop : *WS1 Métrique, musication*

Paris, septembre 2024

ANALYSE AUTOMATIQUE D'UN CORPUS DE TEXTES VERSIFIÉS (métrique et syntaxe)

1. Présentation du corpus
2. Problématique des enjambements
3. *Être* et *avoir* en fin de vers
4. Évaluation des outils d'analyse syntaxique

1. Présentation du corpus

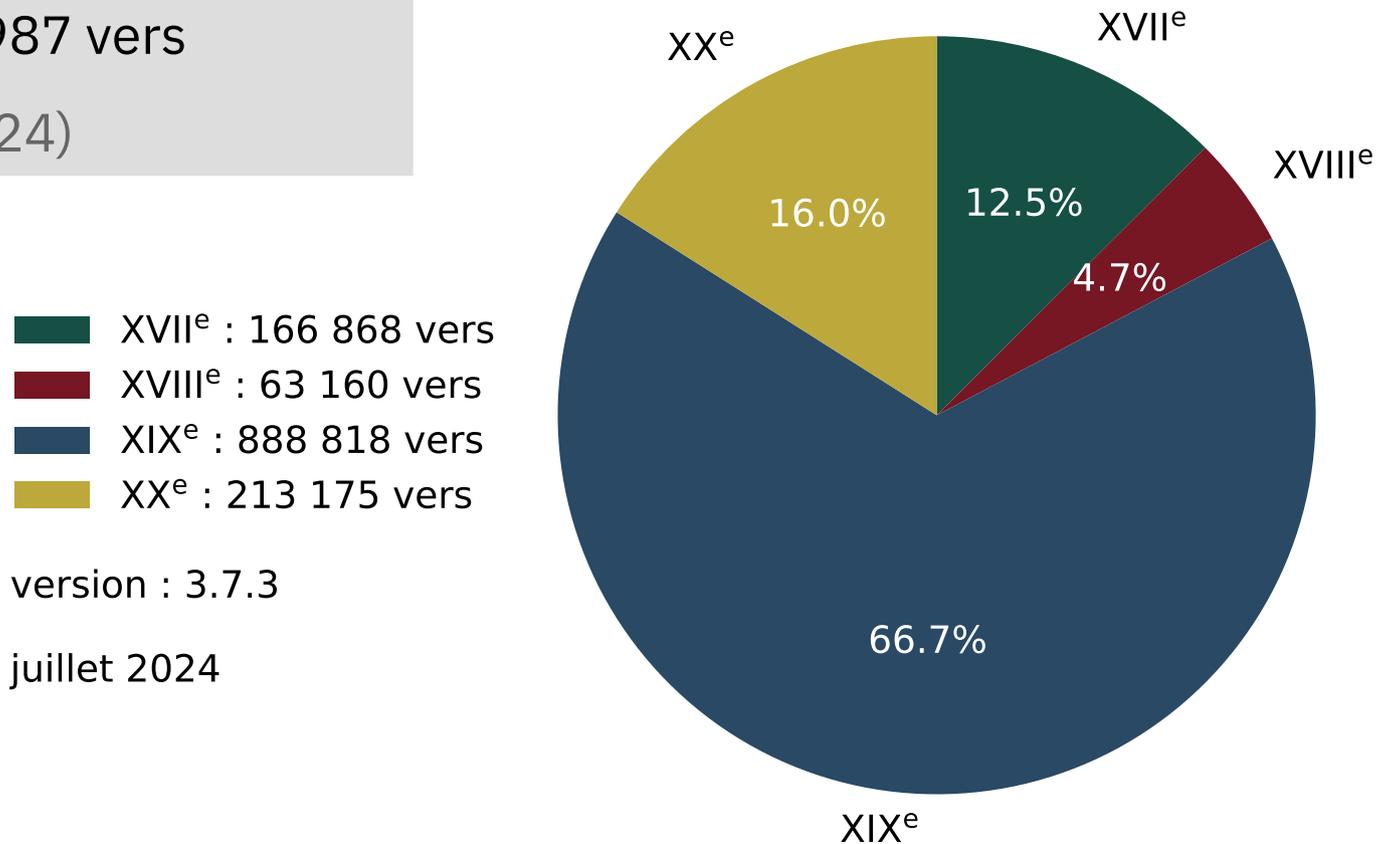
Corpus Malherbæ

- 302 auteurs
- 638 recueils de poésies
- 26 179 poèmes
- 141 pièces de théâtre
- 1 321 987 vers

(juillet 2024)

RÉPARTITION DU CORPUS SELON LES SIÈCLES (date d'édition du recueil)

pour un total de 1 332 021 vers



- XVII^e : 166 868 vers
- XVIII^e : 63 160 vers
- XIX^e : 888 818 vers
- XX^e : 213 175 vers

version : 3.7.3

juillet 2024

1. Présentation du corpus

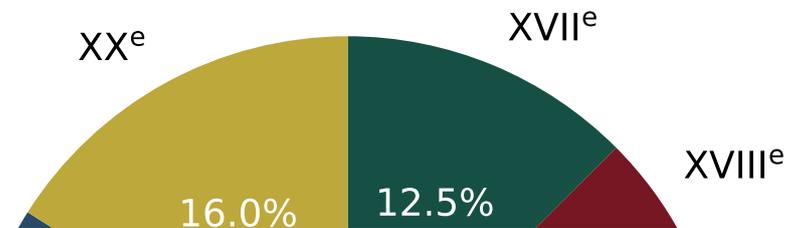
Corpus Malherbæ

- 302 auteurs
- 638 recueils de poésies
- 26 179 poèmes
- 141 pièces de théâtre
- 1 321 987 vers

(juillet 2024)

RÉPARTITION DU CORPUS SELON LES SIÈCLES (date d'édition du recueil)

pour un total de 1 332 021 vers



sous-corpus :

■ corpus *Malherbe* (CRISCO)

■ corpus *Le Rire des vers*

(Anne-Marie Bories, Université de Bâle)

<https://slw-comicverse.dslw.unibas.ch/prima.php?lang=fr>

■ corpus *Pamela_Puntel*

(Pamela Puntel, Università Degli Studi Di Udine et Université Lumière Lyon 2)

<https://ihrim.ens-lyon.fr/auteur/puntel-pamela>

■ corpus *Romain Benini (Filles du peuple ?)*

<https://books.openedition.org/enseditions/17162>

1. Présentation du corpus

Les composantes du traitement automatique :

- La nature des noyaux syllabiques
- Le mètre des vers et le profil métrique des poèmes
- Les rimes, les strophes et la forme globale des poèmes
- La concordance syntaxe/mètre (en cours)

2. Problématique des enjambements

Identification des enjambements et évaluation de la concordance

Principe de meilleure concordance (Principe de concordance optimale)

Toutes choses étant égales par ailleurs, on tend spontanément à comprendre un texte en sorte que la structure métrique et le sens, ou du moins la structure grammaticale convergent le mieux possible.

B. de Cornulier, *L'Art Poétique*, 1995

Implémentation de l'approche de Dell et Benini (2020)

Français Dell et Romain Benini,
La concordance chez Racine,
Rapport entre structure grammaticale et forme métrique dans le théâtre de Racine
Classique Garnier. 2020

2. Problématique des enjambements

Identification des enjambements et évaluation de la concordance

3 cas de figures à prendre en compte :

Concordance

Concordance différée/enjambante

Mazaleytrat, 1990, (concordance différée)

Roubaud, 1978 (concordance enjambante)

Quicherat, 1882

Tobler, 1885

Mourgues, 1724 (enjambement vicieux)

...

Discordance

■ concordance de fin d'hémistiche ou de sous-vers

[[Ils sortirent tous deux.] [— La ville était déserte.]]

Théophile Gautier, ALBERTUS OU L'ÂME ET LE PÉCHÉ, LXXXV, 1833

■ concordance de fin de vers

[[Je suis las de la vie] [et ne veux pas mourir ;]]

[[Mes pieds ne peuvent plus] [ni marcher ni courir ;]]

Théophile Gautier, LA COMÉDIE DE LA MORT, *Thébaïde*, 1838

2. Problématique des enjambements

Identification des enjambements et évaluation de la concordance

Distinction entre discordance et concordance enjambante

■ enjambements en fin de vers

[Bienheureux, j'allongeai][les jambes sous la table]

discordance

[Verte : je contemplai][les sujets très naïfs]

concordance enjambante

[De la tapisserie.][— Et ce fut adorable,]

[Quand la fille aux tétons][énormes, aux yeux vifs,]

Arthur Rimbaud, *Au Cabaret-Vert, cinq heures du soir*, POÉSIES I, 1869-1970

■ enjambements en fin d'hémistiche

[Bienheureux, j'allongeai][les jambes sous la table]

[Verte : je contemplai][les sujets très naïfs]

concordance enjambante

[De la tapisserie.][— Et ce fut adorable,]

[Quand la fille aux tétons][énormes, aux yeux vifs,]

discordance

Arthur Rimbaud, *Au Cabaret-Vert, cinq heures du soir*, POÉSIES I, 1869-1970

2. Problématique des enjambements

Identification des enjambements et évaluation de la concordance

Typologie des enjambements :		typographie	liste	catégorie	dépendance syntaxique
■ syllabe		■			
■ morphème	- affixe	■			
	- base	■			
■ clitique	- clitique	■	■	■	■
	- base	■	■	■	■
■ composant (mot composé)		■			
■ mot	- préposition		■	■	■
	- conjonction		■	■	■
	- verbe			■	■
■ constituant syntaxique				■	■

2. Problématique des enjambements

Identification des enjambements et évaluation de la concordance

Typologie des enjambements :		typographie	liste	catégorie	dépendance syntaxique
■ syllabe		■			
■ morphème	- affixe	■			
	- base	■			
■ clitique	- clitique	■	■	■	■
	- base	■	■	■	■
■ composant (mot composé)		■			
■ mot	- préposition		■	■	■
	- conjonction		■	■	■
	- verbe			■	■
■ constituant syntaxique				■	■

J'en vis un affamé plus que l'est un moineau,
 – Voilà qui tient du diantre ! –
 Qui me parut manger tout simplement une o-
 -Melette avec son ventre ;
 Raoul Ponchon, *Vieux messieurs*, LA MUSE FRONDEUSE, 1971

2. Problématique des enjambements

Identification des enjambements et évaluation de la concordance

Typologie des enjambements :		typographie	liste	catégorie	dépendance syntaxique
■ syllabe		■			
■ morphème	- affixe	■			
	- base	■			
■ clitique	- clitique	■	■	■	■
	- base	■	■	■	■
■ composant (mot composé)		■			
■ mot	- préposition		■	■	■
	- conjonction		■	■	■
	- verbe			■	■
■ constituant syntaxique				■	■

Nous, que notre jeu soit **super-**
Fin, tout autant que le Dimanche
 Stéphane Mallarmé, *Triolets*, VERS DE CIRCONSTANCE, 1920

Paraît-il, de ma mine **affreuse-**
Ment peuple et sans nul galbe exquis
 Paul Verlaine, *Ecce iterum crispinus*, DÉDICACES, 1890

Prête-la moi, je te le **rend-**
Rai gaillard et digne d'envie.
 Paul Verlaine, *Triolets à une vertu pour s'excuser du peu*, FEMMES, 1890

2. Problématique des enjambements

Identification des enjambements et évaluation de la concordance

Typologie des enjambements :		typographie	liste	catégorie	dépendance syntaxique
■ syllabe		■			
■ morphème	- affixe	■			
	- base	■			
■ clitique	- clitique	■	■	■	■
	- base	■	■	■	■
■ composant (mot composé)		■			
■ mot	- préposition		■	■	■
	- conjonction		■	■	■
	- verbe			■	■
■ constituant syntaxique				■	■

Pour aimer et chercher le qu'**en-**
Dira-t-on, et : zut pour ce zeste !
Paul Verlaine, XVI, DANS LES LIMBES, 189

Pour moi le fol Amour, et viens, au clair de **la**
Lune. Allons vers Cythère ou bien vers Pampelune,
Théodore de Banville, *Variations*, DANS LA FOURNAISE, 1892

Monsieur le... Tiens, au fait, qu'**avez-**
Vous été sur terre ? — Poète.
Edmond Rostand, *Au ciel*, LES MUSARDISES, 1911

2. Problématique des enjambements

Identification des enjambements et évaluation de la concordance

Typologie des enjambements :		typographie	liste	catégorie	dépendance syntaxique
■ syllabe		■			
■ morphème	- affixe	■			
	- base	■			
■ clitique	- clitique	■	■	■	■
	- base	■	■	■	■
■ composant (mot composé)		■			
■ mot	- préposition		■	■	■
	- conjonction		■	■	■
	- verbe			■	■
■ constituant syntaxique				■	■

De *Malacca, Manon Lescaut* avec deux **eaux-Fortes** (l'une piquée et l'autre disparue)...
 Tristan Derème, *CXLIX, LA VERDURE DORÉE*, 1908

2. Problématique des enjambements

Identification des enjambements et évaluation de la concordance

Typologie des enjambements :		typographie	liste	catégorie	dépendance syntaxique
■ syllabe		■			
■ morphème	- affixe	■			
	- base	■			
■ clitique	- clitique	■	■	■	■
	- base	■	■	■	■
■ composant (mot composé)		■			
■ mot	- préposition		■	■	■
	- conjonction		■	■	■
	- verbe			■	■
■ constituant syntaxique				■	■

Je me croyais en Décembre,
 Ça doit tenir **au**
Thermomètre de ma chambre
 Qui marque zéro.
 Raoul Ponchon, *To be or not to be*, LA MUSE VAGABONDE, 1947

Et j'étais parti pour ne plus la revoir ; **mais**
Comme j'avais gardé la clef, je revenais,
 Georges de Porto-Riche, *Elle dormait...*, PRIMA VERBA, 1872

Dans le tourment de sa pensée il **regardait**
L'épanouissement de ce rêve nocturne ;
 Raymond de la TAILHÈDE, *Ombres*, POÉSIES, 1938

2. Problématique des enjambements

Identification des enjambements et évaluation de la concordance

Typologie des enjambements :		typographie	liste	catégorie	dépendance syntaxique
■ syllabe		■			
■ morphème	- affixe	■			
	- base	■			
■ clitique	- clitique	■	■	■	■
	- base	■	■	■	■
■ composant (mot composé)		■			
■ mot	- préposition		■	■	■
	- conjonction		■	■	■
	- verbe			■	■
■ constituant syntaxique				■	■

Nos mains se touchaient, le fluide
S'échappait du bout de nos doigts

Raoul Ponchon, *Spiritisme*, LA MUSE FRONDEUSE, 1971

2. Problématique des enjambements

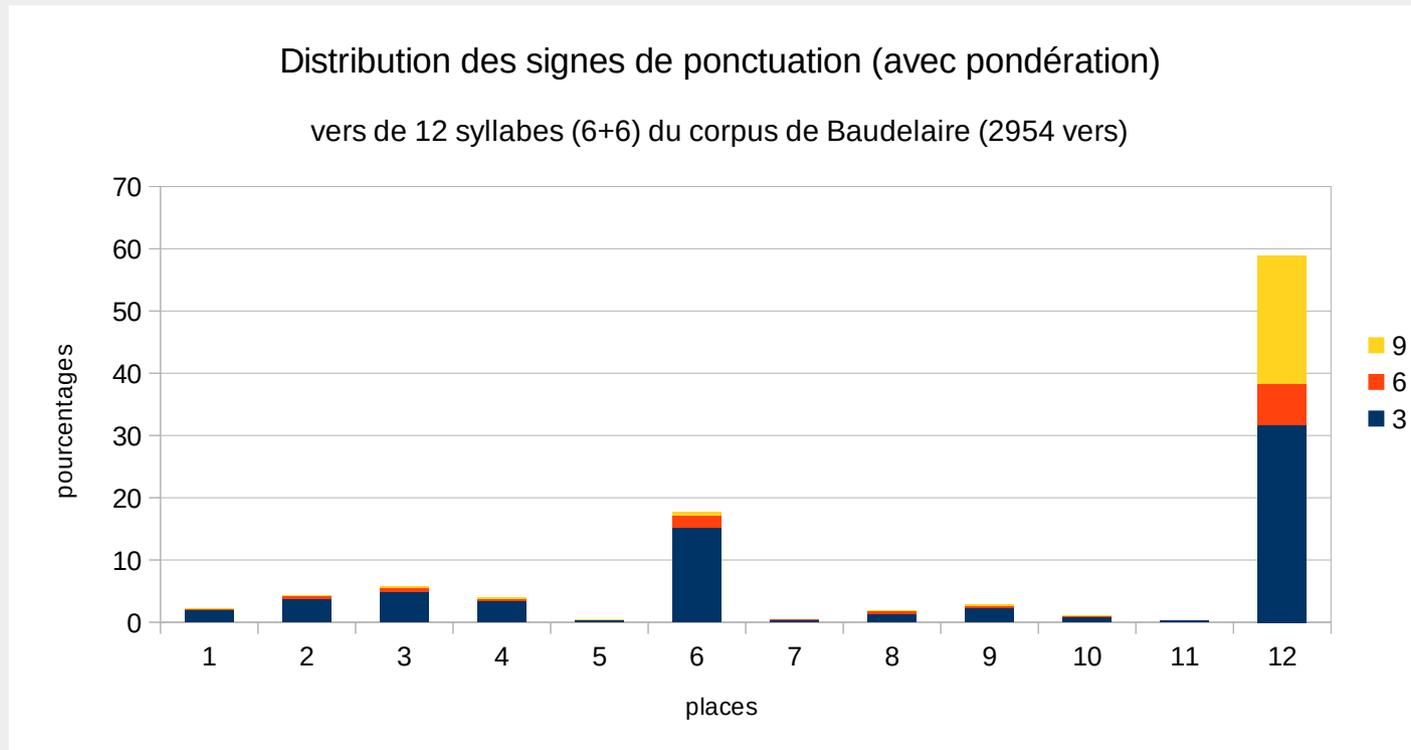
Identification des enjambements et évaluation de la concordance

Typologie des enjambements :		typographie	liste	catégorie	dépendance syntaxique
■ syllabe		■			
■ morphème	- affixe	■			
	- base	■			
■ clitique	- clitique	■	■	■	■
	- base	■	■	■	■
■ composant (mot composé)		■			
■ mot	- préposition		■	■	■
	- conjonction		■	■	■
	- verbe			■	■
■ constituant syntaxique				■	■

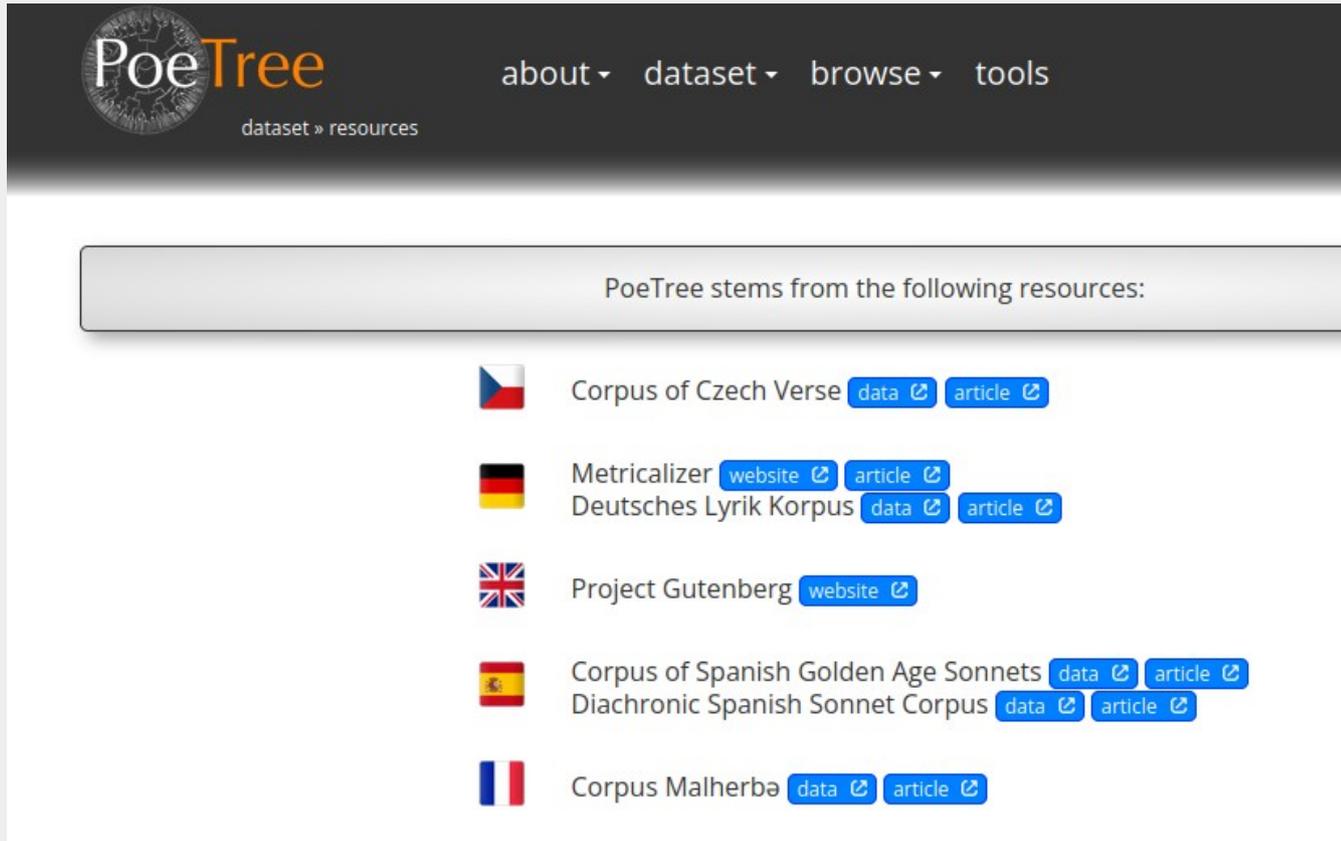
syntaxe

2. Problématique des enjambements

Identification des enjambements et évaluation de la concordance



2. Problématique des enjambements



The screenshot shows the PoeTree website interface. At the top left is the logo 'PoeTree' with a circular tree icon and the text 'dataset » resources'. To the right are navigation links: 'about', 'dataset', 'browse', and 'tools'. Below this is a grey box containing the text 'PoeTree stems from the following resources:'. Underneath, there is a list of five resources, each with a flag icon, a name, and one or more blue buttons with icons for 'data' and 'article'.

Resource	Flags	Buttons
Corpus of Czech Verse	Czech	data, article
Metricalizer	Germany	website, article
Deutsches Lyrik Korpus	Germany	data, article
Project Gutenberg	United Kingdom	website
Corpus of Spanish Golden Age Sonnets	Spain	data, article
Diachronic Spanish Sonnet Corpus	Spain	data, article
Corpus Malherbæ	France	data, article

Petr Plecháč,

Institute of Czech Literature, Czech Academy of Sciences: Prague, CZ

<https://lindat.mff.cuni.cz/services/udpipe/>

<https://zenodo.org/records/10907309>

2. Problématique des enjambements



The screenshot shows the PoeTree website interface. At the top left is the logo "PoeTree" with a circular emblem and the text "dataset » resources". To the right are navigation links: "about", "dataset", "browse", and "tools". Below the logo is a search bar. Underneath the search bar are five language selection icons: Czech, German, English, Spanish, and French. Below the icons, the name "Petr Plecháč" and affiliation "Institute of Czech Literature, Czech Academy of Sciences" are listed. Two URLs are provided: <https://lindat.mff.cuni.cz/search?query=La+chair+est+triste,+hélas!+et+j'ai+lu+tous+les+livres.> and <https://zenodo.org/records/1234567>. On the right side of the interface, a JSON data structure is displayed, representing the parsing of the sentence "La chair est triste, hélas ! et j'ai lu tous les livres.".

```
{
  "id": 0,
  "stanza_id": 0,
  "text": "La chair est triste, hélas ! et j'ai lu tous les livres.",
  "part": false,
  "words": [
    {
      "id": 1,
      "form": "La",
      "lemma": "le",
      "upos": "DET",
      "xpos": "_",
      "feats": "Definite=Def|Gender=Fem|Number=Sing|PronType=Art",
      "head": 2,
      "deprel": "det",
      "deps": "_",
      "sentence": 0
    },
    {
      "id": 2,
      "form": "chair",
      "lemma": "chair",
      "upos": "NOUN",
      "xpos": "_",
      "feats": "Gender=Fem|Number=Sing",
      "head": 4,
      "deprel": "nsubj",
      "deps": "_",
      "sentence": 0
    }
  ]
}
```

2. Problématique des enjambements

Identification des enjambements et évaluation de la concordance

Paul VALÉRY
ALBUM DE VERS ANCIENS
1920

César

	César, calme César, le pied sur toute ch	ose,	6+6	a	
	Les poings durs dans la barbe, et l'œil sombre peu	plé	6+6	b	
	D'aigles et des combats du couchant contem	plé,	6+6	b	
	Ton cœur s'enfle, et se	sent toute-puissante C	ause.	6+6	a
5	Le lac en vain palpite et lèche son lit r	ose ;	6+6	a	
	En vain d'or précieux brille le jeune	blé ;	6+6	b	
	Tu durcis dans les nœuds de ton corps rassem	blé	6+6	b	
	L'ordre, qui doit enfin fendre ta bouche cl	ose.	6+6	a	
	L'ample monde, au delà de l'immense hor	iz on,	6+6	c	
10	L'Empire attend l'éclair, le décret, le t	is on	6+6	c	
	Qui changeront le soir en furieuse aur	lore.	6+6	d	
	Heureux là-bas sur l'onde, et bercé du ha	sard,	6+6	e	
	Un pêcheur indolent qui flotte et chante, ign	ore	6+6	d	
	Quelle foudre s'amasse au centre de Cé	sar.	6+6	e	

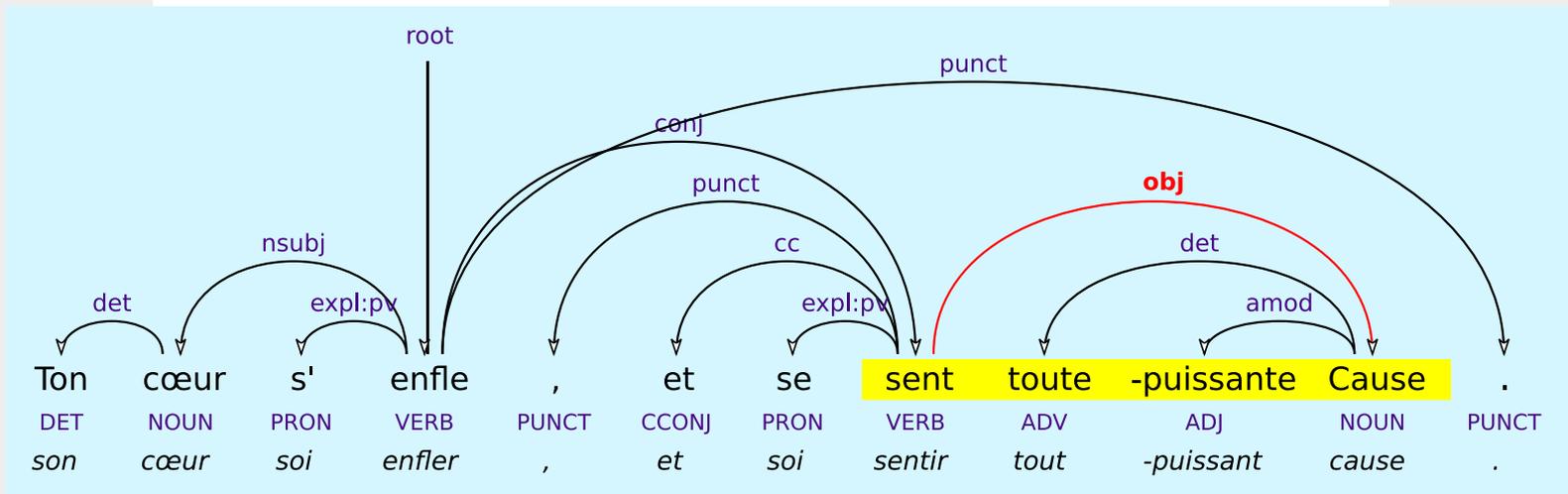
2. Problématique des enjambements

Identification des enjambements et évaluation de la concordance

Paul VALÉRY
ALBUM DE VERS ANCIENS
 1920

César

César, calme César, | le pied sur toute ch^{ose}, 6+6 a
 Les poings durs dans la barbe, | et l'œil sombre peu^{plé} 6+6 b
 D'aigles et des combats | du couchant contem^{plé}, 6+6 b
 Ton cœur s'enfle, et se ^{sent} | toute-puissante C^{ause}. 6+6 a



Un pêcheur indolent | qui flotte et chante, ign^{ore} 6+6 d
 Quelle foudre s'amasse | au centre de Cé^{sar}. 6+6 e

3. Être et avoir en fin de vers

Effectifs et pourcentages

Corpus Malherbè (1 321 987 vers)

- Extraction des couples de vers dont le premier se termine par une forme du verbe **être** :
(non suivie d'une ponctuation)

- **850** occurrences brutes

après filtrage :

- **565** occurrences

- Extraction des couples de vers dont le premier se termine par une forme du verbe **avoir** :
(non suivie d'une ponctuation)

- **289** occurrences brutes

après filtrage :

- **259** occurrences

3. Être et avoir en fin de vers

Effectifs et pourcentages

Corpus Malherbè (1 321 987 vers)

	effectif	pourcentage
être		
auxiliaire	50	9 %
copule/verbe	515	91 %
total	565	

	effectif	pourcentage
avoir		
auxiliaire	47	18 %
verbe	217	82 %
total	264	

3. Être et avoir en fin de vers

Effectifs et pourcentages

Corpus Malherbæ (1 321 987 vers)

	effectif	pourcentage
être		
auxiliaire	50	9 %
copule/verbe	515	91 %
total	565	

	effectif	pourcentage
avoir		
auxiliaire	47	18 %
verbe	217	82 %
total	264	

	effectif	pourcentage
pronoms		
clitiques	49	4 %
pro. forts	1147	96 %
total	1196	

3. Être et avoir en fin de vers

Exemples

Pronoms

[[Vous ne saurez jamais,][quand je bois, combien **je**]]

[Me fiche du Danube !]

Raoul Ponchon, LA MUSE AU CABARET, *Les Cabarets d'hier et d'aujourd'hui*, 1920

[Mais, crois-le bien, ce n'est pas **moi**]

[Qui lui dictai ses Ordonnances !]

Théodore de Banville, OCCIDENTALES, *La Criminelle*, 1875

3. Être et avoir en fin de vers

Exemples

être

auxiliaire

discordance

[[Pour boire ce breuvage ? ↵]] En meurt-on ? ↵ Tu **seras**]]

[[**Tout d'abord comme pris**]] **de vin.** — Tu sentiras]]

Alfred de Musset, PREMIÈRES POÉSIES, *Don Paez*, 1829

concordance enjambante

[[Et prononce déjà]][des mots qui ne **seront**]]

[[**Dits par le genre humain**]][que dans trois mille années.]]

Victor Hugo, LA FIN DE SATAN, *Le Gibet I*, 1886

copule

discordance

[[Quand vous vous échauffez]][ainsi, comme vous **êtes**]]

[[**Jolie** ! ↵ Oh ! laissez-moi,]][monsieur, ou je me jette]]

Alfred de Musset, PREMIÈRES POÉSIES, *Les Marrons du feu*, 1829

concordance enjambante

[Étant durs pour vous, nous **sommes**]

[**Doux pour le peuple aux abois,**]

Victor Hugo, LA LÉGENDE DES SIÈCLES, *Le Romencero du Cid*, 1877

3. Être et avoir en fin de vers

Exemples

avoir

auxiliaire

discordance

[[A l'exemple de la][cigale nous **avons**]]

[[**Chanté...** ← Si nous allions][danser ? ← Nous vous suivons !]]

Paul Verlaine, JADIS ET NAGUÈRE, *Les Uns et les autres*, 1884

concordance enjambante

[[En finir de l'enfant.][Certes, ainsi nous **eussions**]]

[[**Dans leur source extirpé][les révolutions.]]**

Victor Hugo, LES QUATRE VENTS DE L'ESPRIT, *Les Deux trouvailles de gallus*, 1881

verbe

discordance

[[Que vous ne m'aimez plus. ←][Bon, comme vous **avez**]]

[[**L'esprit fait !** — Pensez-vous,][madame, que j'oublie]]

Alfred de Musset, PREMIÈRES POÉSIES, *Les Marrons du feu*, 1829

concordance enjambante

[[Nous irons conquérir][le monde, et nous **aurons**]]

[[**Des filles et du vin,**][et tu feras ripaille,]]

Victor Hugo, LA LÉGENDE DES SIÈCLES, *Welf*, 1877

4. Évaluation des outils d'analyse syntaxique

Objectifs et méthode

- enrichissement prévu du corpus
 - catégories grammaticales
 - annotation syntaxique
- comparaison de différents outils pour conserver le(s) plus efficace(s)
 - treetagger
 - catégorisation grammaticale
 - chunking → vers une analyse syntaxique de surface (partielle)
 - *arbres de décision, possibilité d'augmenter le lexique pour l'adapter aux périodes couvertes par le corpus*
 - arborator, hops & udpipe
 - annotation syntaxique en dépendance
 - inclut une catégorisation grammaticale
 - *LLM – boîte noire, possibilité de « bootstrapping » c-à-d proposer de nouvelles données d'apprentissage*

4. Évaluation des outils d'analyse syntaxique

Phase 1

réalisée en collaboration avec

Natalia Romanova,

Rayan Ziane

& Asmaa Abdmeziem

- choix d'une segmentation commune
 - en mots (tokens)
 - en phrases (propositions pas trop longues)
 - *éviter les problèmes d'alignement*
permettre une comparaison sur les mêmes unités
- utilisation des outils
 - conversion des tags TT en UD
 - *les jeux d'étiquettes des 4 outils ne sont pas les mêmes*
nous prenons le standard UD comme référence

TT	
KON	conjunction
...	
VER:cond	verb conditional
VER:futu	verb futur
VER:impe	verb imperative
VER:impf	verb imperfect
VER:infverb	infinitive
VER:pper	verb past participle
VER:ppre	verb present participle
VER:pres	verb present
VER:simp	verb simple past
VER:subi	verb subjunctive imperfect
VER:subp	verb subjunctive present

UD	
CCONJ	coordinating conjunction
SCONJ	subordinating conjunction
...	
AUX	auxiliary
VERB	verb

4. Évaluation des outils d'analyse syntaxique

- mise en place d'annotations de référence (« gold standard »)
 - pour la catégorisation grammaticale
 - pour l'analyse syntaxique
 - évaluer les outils
et sélectionner le(s) plus efficace(s)

Méthode retenue

- plusieurs annotateurs sur un même texte
- amorce : experts & consensus = phase 1
- poursuite : campagne d'annotation, filtrage & vote = phase 2

Amorce réalisée pour la partie catégorisation

1. chaque annotateur propose sa version
au regard des résultats des 4 outils testés
2. les annotateurs se réunissent et décident par consensus

4. Évaluation des outils d'analyse syntaxique

Eval_NN.ods — LibreOffice Calc

Fichier Édition Affichage Insertion Format Styles Feuille Données Outils Fenêtre Aide

Liberation Sans 10 pt G I S A

G4

	A	B	C	D	E	F	G	H
1	Id	Tok	hops	arborator	treetagger	udpipe	correcteur	commentaire
2	# sent	id = HER14_8_1						
3	1	C'	PRON	PRON	PRON	PRON		
4	2	est	AUX	AUX	VERB	AUX		
5	3	un	DET	DET	DET	DET		
6	4	vallon	NOUN	NOUN	NOUN	NOUN	ADJ	
7	5	sauvage	ADJ	ADJ	ADJ	ADJ	ADP	
8	6	abrité	VERB	VERB	VERB	VERB	ADV	
9	7	de	ADP	ADP	ADP	ADP	AUX	
10	8	l'	DET	DET	DET	DET	CCONJ	
11	9	Euxin	PROPN	PROPN	NOUN	PROPN	DET	
12	10	;	PUNCT	PUNCT	PUNCT	PUNCT	INTJ	
13	11	Au	ADP	ADP	ADP	ADP	NOUN	
14	12	-dessus	ADV	ADV	PROPN	ADV	NUM	
15	13	de	ADP	ADP	ADP	ADP	PART	
16	14	la	DET	DET	DET	DET	PRON	
17	15	source	NOUN	NOUN	NOUN	NOUN		
18	16	un	DET	DET	DET	DET		
19	17	noir	NOUN	ADJ	ADJ	ADJ		
20	18	laurier	NOUN	NOUN	NOUN	NOUN		
21	19	se	PRON	PRON	PRON	PRON		
22	20	penche	VERB	VERB	VERB	VERB		
23	21	,	PUNCT	PUNCT	PUNCT	PUNCT		
24	22	Et	CCONJ	CCONJ	SCONJ	CCONJ		
25	23	la	DET	DET	DET	DET		
26	24	Nymphe	PROPN	NOUN	NOUN	PROPN		

HER14 LOZ45 MOL38 MOR36 MUS137 NOU132 PON61 REG20 VAL9 VIA30 modèle

Feuille 1 sur 13 Par défaut Français (France) Moyenne: ; NBVAL: 0 100 %

4. Évaluation des outils d'analyse syntaxique

Eval_NN.ods — LibreOffice Calc

Fichier Édition Affichage Insertion Format Styles Feuille Données Outils Fenêtre Aide

Liberation Sans 10 pt G I S A

G11 $fx \sum =$

	A	B	C	D	E	F	G	H
1	Id	Tok	hops	arborator	treetagger	udpipe	correcteur	commentaire
2	# sent	id = HER14_8_1						
3	1	C'	PRON	PRON	PRON	PRON		
4	2	est	AUX	AUX	VERB	AUX	AUX	
5	3	un	DET	DET	DET	DET		
6	4	vallon	NOUN	NOUN	NOUN	NOUN		
7	5	sauvage	ADJ	ADJ	ADJ	ADJ		
8	6	abrité	VERB	VERB	VERB	VERB		
9	7	de	ADP	ADP	ADP	ADP		
10	8	l'	DET	DET	DET	DET		
11	9	Euxin	PROPN	PROPN	NOUN	PROPN		
12	10	;	PUNCT	PUNCT	PUNCT	PUNCT	DET	
13	11	Au	ADP	ADP	ADP	ADP	INTJ	
14	12	-dessus	ADV	ADV	PROPN	ADV	NOUN	
15	13	de	ADP	ADP	ADP	ADP	NUM	
16	14	la	DET	DET	DET	DET	PART	
17	15	source	NOUN	NOUN	NOUN	NOUN	PRON	
18	16	un	DET	DET	DET	DET	PROPN	
19	17	noir	NOUN	ADJ	ADJ	ADJ	PUNCT	
20	18	laurier	NOUN	NOUN	NOUN	NOUN	SCONJ	
21	19	se	PRON	PRON	PRON	PRON	SYM	
22	20	penche	VERB	VERB	VERB	VERB	VERB	
23	21	,	PUNCT	PUNCT	PUNCT	PUNCT	X	
24	22	Et	CCONJ	CCONJ	SCONJ	CCONJ		
25	23	la	DET	DET	DET	DET		
26	24	Nymphe	PROPN	NOUN	NOUN	PROPN		

HER14 LOZ45 MOL38 MOR36 MUS137 NOU132 PON61 REG20 VAL9 VIA30 modèle

Feuille 1 sur 13 Par défaut Français (France) Moyenne: ; NBVAL: 0 100 %

4. Évaluation des outils d'analyse syntaxique

Eval_All_Gold.ods — LibreOffice Calc

Fichier Édition Affichage Insertion Format Styles Feuille Données Outils Fenêtre Aide

Libération Sans 10 pt G I S A

K27 \sum = PUNCT

	A	B	G	H	I	J	K
1	Id	Tok	AA	NR	RR	SF	Gold
2	# sent	id = HER14_8					
3	1	C'	PRON	PRON	PRON	PRON	PRON
4	2	est	VERB	AUX	VERB	VERB	AUX
5	3	un	DET	DET	DET	DET	DET
6	4	vallon	NOUN	NOUN	NOUN	NOUN	NOUN
7	5	sauvage	ADJ	ADJ	ADJ	ADJ	ADJ
8	6	abrité	PART	VERB		VERB	VERB
9	7	de	ADP	ADP	ADP	ADP	ADP
10	8	l'	DET	DET	DET	DET	DET
11	9	Euxin	PROPN	PROPN	PROPN	PROPN	PROPN
12	10	;	PUNCT	PUNCT	PUNCT	PUNCT	PUNCT
13	11	Au	ADP	ADP	ADP	ADP	ADP
14	12	-dessus	ADP	NOUN	NOUN	ADV	NOUN
15	13	de	ADP	ADP	ADP	ADP	ADP
16	14	la	DET	DET	DET	DET	DET
17	15	source	NOUN	NOUN	NOUN	NOUN	NOUN
18	16	un	DET	DET	DET	DET	DET
19	17	noir	ADJ	ADJ	ADJ	ADJ	ADJ
20	18	laurier	NOUN	NOUN	NOUN	NOUN	NOUN
21	19	se	PRON	PRON	PRON	PRON	PRON
22	20	penche	VERB	VERB	VERB	VERB	VERB
23	21	,	PUNCT	PUNCT	PUNCT	PUNCT	PUNCT
24	22	Et	CCONJ	CCONJ	CCONJ	CCONJ	CCONJ
25	23	la	DET	DET	DET	DET	DET
26	24	Nymphe	NOUN	NOUN	NOUN	PROPN	NOUN

Conso HER14 LOZ45 MOL38 MOR36 MUS137 NOU132 PON61

Feuille 2 sur 14 Par défaut Français (France) 100 %

4. Évaluation des outils d'analyse syntaxique

Résultats

- précision, rappel et F-mesure harmonique

TreeTagger				Arborator			HOPS			UDPipe			
classe	precision	recall	f1-score	support									
ADJ	0.93	0.82	0.87	0.92	0.82	0.87	0.88	0.81	0.85	0.93	0.91	0.92	122
ADP	0.91	0.96	0.93	0.96	0.88	0.92	0.97	0.84	0.90	0.98	0.88	0.93	141
ADV	0.93	0.93	0.93	0.79	0.82	0.80	0.79	0.84	0.81	0.96	0.95	0.95	55
AUX	0.00	0.00	0.00	0.87	0.91	0.89	0.80	0.91	0.85	0.79	1.00	0.88	22
CCONJ	1.00	0.98	0.99	0.98	1.00	0.99	1.00	0.97	0.98	1.00	1.00	1.00	58
DET	1.00	0.89	0.94	0.91	0.99	0.95	0.90	0.98	0.94	0.93	1.00	0.96	204
INTJ	0.80	0.67	0.73	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.33	0.50	6
NOUN	0.96	0.98	0.97	0.93	0.95	0.94	0.94	0.93	0.94	0.98	0.98	0.98	283
NUM	0.56	1.00	0.71	0.62	1.00	0.77	1.00	1.00	1.00	1.00	1.00	1.00	5
PRON	0.87	0.95	0.91	0.94	0.99	0.97	0.98	0.98	0.98	0.99	1.00	1.00	117
PROPN	0.54	0.88	0.67	0.40	1.00	0.57	0.17	1.00	0.30	0.62	1.00	0.76	8
PUNCT	1.00	1.00	1.00	0.99	0.93	0.96	1.00	0.92	0.96	1.00	1.00	1.00	213
SCONJ	0.77	0.85	0.81	1.00	0.75	0.86	0.94	0.80	0.86	0.95	0.95	0.95	20
VERB	0.79	0.94	0.86	0.89	0.90	0.89	0.86	0.84	0.85	0.94	0.91	0.92	140
accuracy			0.92			0.92			0.90			0.96	1394
macro avg	0.79	0.84	0.81	0.80	0.85	0.81	0.75	0.79	0.75	0.93	0.92	0.91	1394
weighted avg	0.92	0.92	0.92	0.93	0.92	0.92	0.92	0.90	0.91	0.96	0.96	0.96	1394
TT sans AUX													
classe	precision	recall	f1-score	support									
VERB	0.92	0.95	0.93	162									
accuracy			0.94	1394									
macro avg	0.86	0.91	0.88	1394									
weighted avg	0.94	0.94	0.94	1394									

- UDPipe est retenu pour cette première phase

4. Évaluation d'outils

- accords et désaccords
 - la mise en commun des annotations permet aussi la comparaison

	A	B	C	D	E	F	G	H
11	tokens	1398	1398	1398	1398	1398	1398	0
22	accords	1328	1324	1328	1337	1352	1353	0
33	désaccords	70	74	70	61	46	45	0
35		AA/NR	AA/RR	AA/SF	NR/RR	NR/SF	RR/SF	
36	tx accord	94,99 %	94,71 %	94,99 %	95,64 %	96,71 %	96,78 %	
37	tx désaccord	5,01 %	5,29 %	5,01 %	4,36 %	3,29 %	3,22 %	

- à l'avenir, utilisable pour proposer des campagnes d'annotations
 - filtrer les annotateurs atypiques
 - procéder au vote sur les annotateurs restants

4. Évaluation d'outils

Poursuite de la phase 1 – en cours

- annotation conjointe des résultats de l'analyse syntaxique



Projects Documentation Discussion

FR

Malherbe_eval > HER14_8

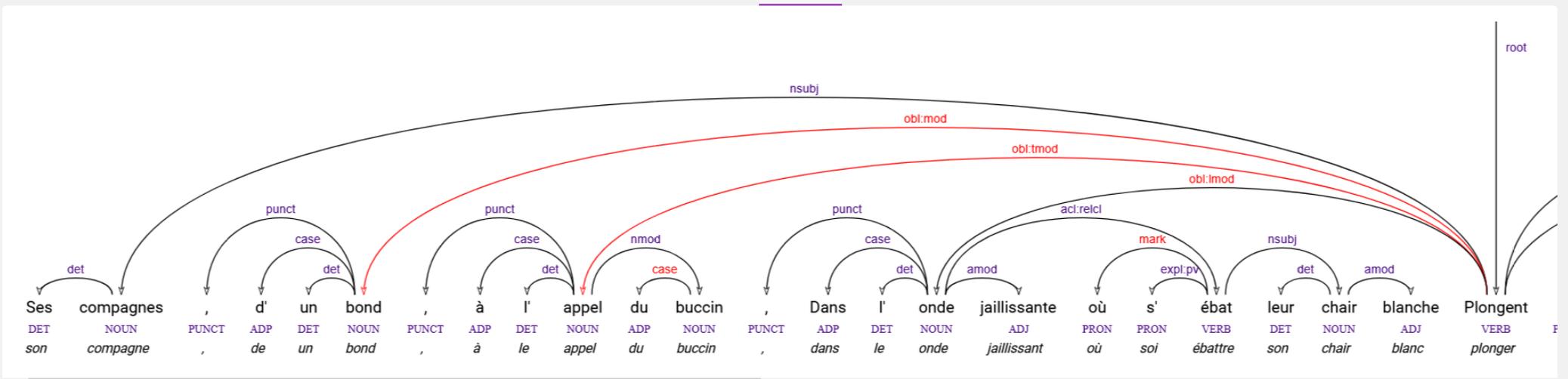
Filtre du texte Ta... FILTRER Order sentences by length initial SAVE PENDING TREES

1 HER14_8_1 C' est un vallon sauvage abrité de l' Euxin ; Au -dessus de la source un noir laurier se penche , Et la Nymphé , riant , suspendue à la branche , Frôle d' un pied craintif l' eau froide du bassin .

standro314 ziane.rayan4 richard.renault natroma UDpipe arborator HOPS

2 HER14_8_2 Ses compagnes , d' un bond , à l' appel du buccin , Dans l' onde jaillissante où s' ébat leur chair blanche Plongent , et de l' écume émergent une hanche , De clairs

natroma ziane.rayan4 richard.renault standro314 UDpipe arborator HOPS



Tags: text_en

3 HER14_8_3 Une gaîté divine emplit le grand bois sombre .

natroma ziane.rayan4 richard.renault standro314 UDpipe arborator HOPS

4 HER14_8_4 Mais deux yeux , brusquement , ont illuminé l' ombre .

natroma ziane.rayan4 richard.renault standro314 UDpipe arborator HOPS

5. Conclusion

- Ampleur de la tâche à venir pour l'évaluation de l'analyse syntaxique
- Mise en place d'une échelle de cohésion pour la typologie des enjambements à partir de la distribution des types sur l'ensemble du corpus
- Caractérisation d'un auteur ou d'une période à partir d'une évaluation qualitative et quantitative de la concordance
- Rapprochement de la concordance avec le mètre des vers pour une option de mètre alternatif