

To what extent are lemmatisation and annotation relevant for deep learning assignments and textual motifs detection? The case-study of Peter Damian's letters (11th century)

Valérie Thon (UParis Cité, CERILAC; Uliège, LASLA)

Laurent Vanni (Université Côte d'Azur, BCL)

Dominique Longrée (Uliège, LASLA)

Introduction

- stylistic study : Peter Damian (1007-1072/73)

→ 20 letters (LASLA Encoding Initiative)

da - 048, 0012 - sentence n°: 634
sic sic videlicet cum petro claves aecclesiae merito facti participes optinebimus dum nosmet ipsos certam vivendi formam atque signaculum caeteris fidelibus exhibemus

N°	Form	Lemma	I.	Ana.	F.	U.	G.	SC	
1	SIC	SIC		60000					✓
2	SIC	SIC		60000					✓
3	VIDELICET	VIDELICET		60000					✓
4	CUM	CVM	2	70600					✓
5	PETRO	PETRVS	N	12F00					✓
6	CLAVES	CLAVIS		13L00					✓
7	AECCLISIAE	ECCLESIA		11D00					✓
8	MERITO	MERITO	2	60000					✓
9	FACTI	FACIO		5EJ44			4		✓
10	PARTICIPES	PARTICEPS	2	22J00			3		✓
11	OPTINEBIMUS	OBTINEO		52J13	&				✓
12	DUM	DVM	2	82011					✓
13	NOSMET	NOS		41L00					✓
14	IPSOS	IPSE		45L00			4		✓
15	CERTAM	CERTVS		21C00			2		✓
16	VIVENDI	VIVO		53D60					✓
17	FORMAM	FORMA		11C00					✓
18	ATQUE	ATQVE	1	81000					✓
19	SIGNACULUM	SIGNACVLVM		12C00					✓
20	CAETERIS	CETERVS		21N00			1		✓
21	FIDELIBUS	FIDELIS	1	13N00			1		✓
22	EXHIBEMUS	EXHIBEO		52J11	-			CD	✓

Tools

Add line after word:

Delete Line N°:

Go to Sentence N°:

[LASLA Codes Table](#)



Introduction

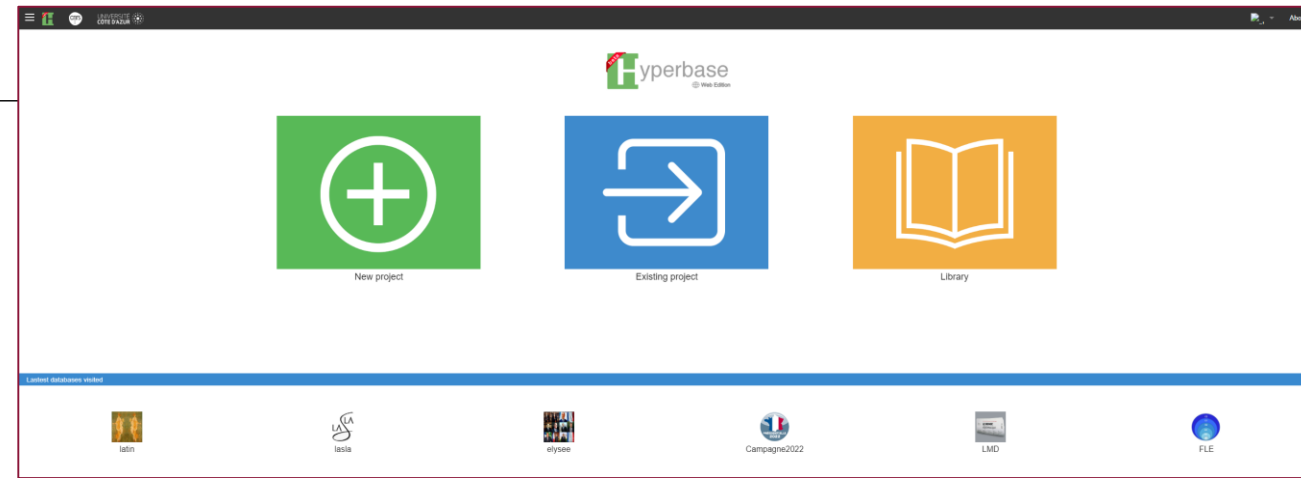
- stylistic study : Peter Damian (1007-1072/73)

→ textual motifs (Longrée et al., 2008)

*recurrent (phonological, syntactic,
metric,...) structure characteristic of an
author or even a genre*



Introduction



- *Hyperdeep*

MODEL 1 : 12 classical authors
(non lemmatised)

MODEL 2 : 12 classical authors
(lemmatised + morphosyntactically
annotated with LEI)

Peter Damian : 20 letters
(LASLA - LEI)

→ Extraction of features responsible for the
model's decision: **textual motifs** ?

Test corpus

AUTHOR	TEXT(S)
Plautus	Comedies
Cornelius Nepos	<i>Vitae</i>
Cicero	Dialogues + speeches
Cesar	<i>Bellum civile</i>
Sallust	<i>De catilinae coniuratione + Bellum Jugurthinum</i>
Livy	<i>Ab urbe condita</i>
Seneca	Letters + philosophical treatises
Quintus Curtius	<i>Historia Alexandri magni</i>
Petronius	<i>Satyricon</i>
Tacitus	<i>Historiae</i>
Pliny	<i>Epistulae</i>
Suetonius	<i>Vitae Caesarum</i>

Model 1

Example: Letter 162

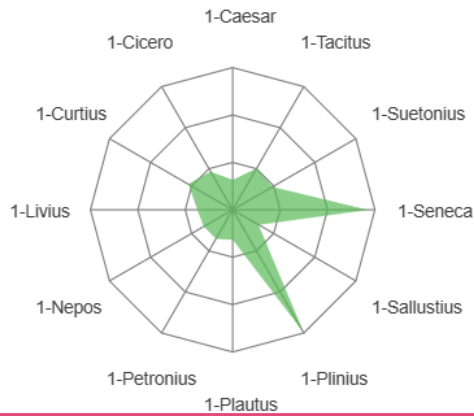
Predict Accuracy : 98.05 % Loss : 6.15

Hyperdeep À partir d'un copier/coller

Prédiction : 1-Plinius 38%

voir aussi	taux d'attribution
1-Seneca	35%
1-Curtius	7.1%
1-Tacitus	6.0%
1-Suetonius	5.4%
1-Cicero	4.8%
1-Livius	1.8%
1-Nepos	1.2%
1-Petronius	1.2%

Profil général



Passages-clés

[...] cor non ire de quibus utique scriptum est Populus hic labiis me honorat cor autem eorum longe est a me . **Quibus** etiam **apostolus** quasi personaliter exprobrans ait **Qui** ergo alium doces teipsum non **doces** . Qui praedicas non furandum furaris . Qui dicis non mechandum mecharis . **Qui** abhominaris [...]

[voir dans le texte](#)

◀ précédent | 1 | suivant ▶

[plus d'infos](#)

Go for searching!

Marqueurs linguistiques (TDS)

Forme



Motifs profonds

passage	TDS
ueritati .	77
iustum . Sepius [...] adulterinus [...] perimit	76
sed [...] pulvis	68
. Quibus [...] apostolus	67
Hic enim	67
ex [...] trahunt . Sicut	67
in saeculi fine persequitur	66
uidentur [...] Ad	66
tam expresse sicut clericis	61
Et tanquam	58
. De [...] utique	55
contingat .	55
lege constringitur etiam	53
accessu quanto	50

Texte intégral :

Domno Petro Lateranensis canonicae archipresbytero Petrus peccator monachus fraternae glutinum karitatis . Tantum ualet iaculum quanta uirtus est bellatoris . Nam quantalibet limetur et exacuat industria si brachium uibrantis elanguet insignis uictoriae manubias non reportat . At ubi bellator est strenuus per optunsum quoque gladium cesis hostibus aliquando fit triumphus . **Nam** et Goliath insigni munitus gladio non modo perdidit sed et percussus occubuit et Samson cum uili **mandibula asini** mille de Philistiim uiros occidit . Hinc est quod cum lether primogenitus Gedeonis gladium timuisset educere Zebee et Salmana reges Madian dixerunt Gedeoni Tu surge et irruere in nos quia iuxta aetatem robur est hominis . Itaque qui contra clericorum luxuriam **tamquam** tudites iaculum fabrico in nullius eum melius quam in tuis manibus pono . Tanto quippe zelo **uidi** te semper aduersus eorum perditos mores medullitus inardescere ut Finees uel Heliae iudicis incendio non egere adeo ut plus in sugillationis tuae uereantur obloquium quam ipsum quoque timeant papale decretum . **Nam uelut** egregius canis aulae regiae custos nocturnos fures claris uaubatibus impetis eos que ne libidinis suae facibus palacium regale comburant mordicus apprehendis . **Non** enim te latet quia presbyteris diaconis et subdiaconis tam districe tam rigide regula castitatis indicitur ut omni catholicorum patrum concurrente iudicio a contrahendis nuptiis funditus abscondantur . Et certe satis est ratione plenissimum ut post episcopos in his quoque tribus gradibus niueus pudicitiae candor effloreat qui nimirum sacris mysteriis familiaris appropinquant . Nam quia dominicum corpus in uirginalis uteri templo cohaluit nunc etiam a ministris suis continentis pudicitiae mundiciam quaerit . Plane sicut concipienti deifica uirgini non uirile semen influxit sed uirtus sancti Spiritus obumbravit ita nunc in altari positum sacramentum eadem sancti Spiritus uirtus est quae uiuificat et necesse est ut hoc pudica manus et inpolluta **contingat** . Sed quoniam super hoc themate prolixius iam alibi disputauimus et primo quidem pia memoriae Nikolao papae deinde Taurinensi episcopo postremo etiam comitissae Adalaidi uirginitatem **scribentes** diuersa scripturarum exempla congressimus ad illa te frater uenerande transmittito illis te uelut armis militaribus induo atque ut alter Iosue urbem lericho subuertere nitaris impello . Superest ergo nunc **ut** eosdem clericos ipse paulisper aggrediar quatinus quod in camino pectoris feruet ita **labiis** quasi rimis patentibus euaporet . Conuenio uos o prolectarii uxorii ac mulierum dominantium ditionibus inseruiti ut quid canonicae sanctionis iura confunditis et a sanctis patribus institutam puritatis aecclesiasticae regulam **uiolatis** . Pro pudor libidinis uestrae legibus colla summittitis et promulgata per ora doctorum sancti Spiritus edicta calcatis . Ecce per uos apostolorum labor euertitur et aedificium quod doctores aecclesiae fundauerunt liquet quod uobis impugnantibus obruatur . **Et** ita fit ut lex

Model 1 : Seneca

Example: Letter 162

Predict Accuracy : 98.05 % Loss : 6.15

Hyperdeep

Prédiction : 1-Seneca 35%

voir aussi	taux d'attribution
1-Plinius	38%
1-Curtius	7.1%
1-Tacitus	6.0%
1-Suetonius	5.4%
1-Cicero	4.8%
1-Livius	1.8%
1-Nepos	1.2%
1-Petronius	1.2%

Passages-clés

[...] Ecce absorbebit fluium et non **mirabitur habet enim** fiduciam quod influat Iordanis in os eius . Quod autem dicitur fuisse positum in foramine petrae potest non inconuenienter intellegi intra septa aecclesiae . Quasi enim in foramine petrae clerici recluduntur dum intra aecclesiae limina suis excubare ministeriis sedula frequentatione **iubentur** . [...]

[voir dans le texte](#)

◀ précédent | 1 | suivant ▶

[plus d'infos](#)

possunt ista congruere

Texte intégral :

Domno Petro Lateranensis canonicae archipresbytero Petrus peccator monachus fraternae glutinum karitatis . Tantum ualet iaculum quanta uirtus est bellatoris . Nam quantalibet limetur et exacuatur industria si brachium uibrantis elanguet insignis uictoriae manubias non reportat . At ubi bellator est strenuus per optunsum quoque gladium cesis hostibus aliquando **fit** triumphus . Nam et Goliath insigni munitus gladio **non modo** perdidit **sed** et percussus occubuit et Samson cum uili mandibula asini mille de Philistiim uiros occidit . Hinc est quod cum lether primogenitus Gedeonis gladium timuisset educere Zebee et Salmana reges Madian dixerunt Gedeoni Tu surge et irru in **nos** quia iuxta aetatem robur **est** hominis . Itaque qui contra clericorum **luxuriam tamquam** tudites iaculum fabrico in nullius eum melius quam in tuis manibus pono . Tanto quippe **zelo** uidi te semper aduersus eorum perditos mores medullitus inardescere ut Finees uel Heliae iudiceres incendio non egere adeo ut plus in sugillationis tuae uereantur obloquium quam ipsum quoque **timeant** papale decretum . Nam uelut **egregius** canis aulae **regiae** custos nocturnos fures claris uaubatibus impetis eos que ne libidinis suae facibus palacium regale comburant mordicus apprehendis . Non enim **te latet** quia presbyteris diaconis et subdiaconis tam **districte** tam **rigide** regula castitatis indicitur ut omni catholicorum patrum concurrente iudicio a **contrahendis** nuptiis funditus abscondantur . Et certe satis est ratione plenissimum ut post episcopos in his quoque tribus gradibus niueus pudicitiae candor effloreat qui nimirum sacris mysteriis familiarius appropinquant . Nam quia dominicum corpus in uirginalis uteri templo cohaluit nunc etiam a ministris suis continentis pudicitiae mundiciam quaerit . Plane sicut concipienti deificae uirgini non uirile semen influxit sed uirtus sancti Spiritus obumbravit ita nunc in altari positum sacramentum eadem sancti Spiritus uirtus est quae uiuificat et necesse est **ut** hoc pudica manus et inpolluta contingat . Sed quoniam super hoc themate prolixius iam alibi disputauimus et primo quidem pia memoriae Nikolao papae deinde Taurinensi episcopo postremo etiam comitissae Adalaidi uiritim scribentes diuersa scripturarum **exempla conguessimus ad** illa te frater uenerande transmitto

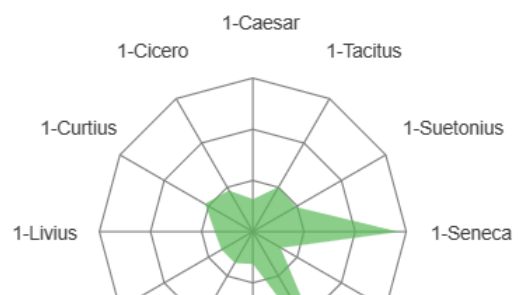
Marqueurs linguistiques (TDS)

Forme

Motifs profonds





passage	TDS
et iste quomodo uiuet [...] non	178
trepidat quem	141
longe decurrit abscondere .	131
exprobrans ait Qui	129
te latet quia	128
iubentur .	118
sequitur Tu [...] uigila	115
sponsus ait in	115
testatur inuenit cum	115
regem sordidis	111
possunt ista congruere	110
mirabitur habet enim	109
illa terribilia	109
suam in	109
qui nolunt audire	105
uiros occidit	103
esset [...] Et ait Dominus	101

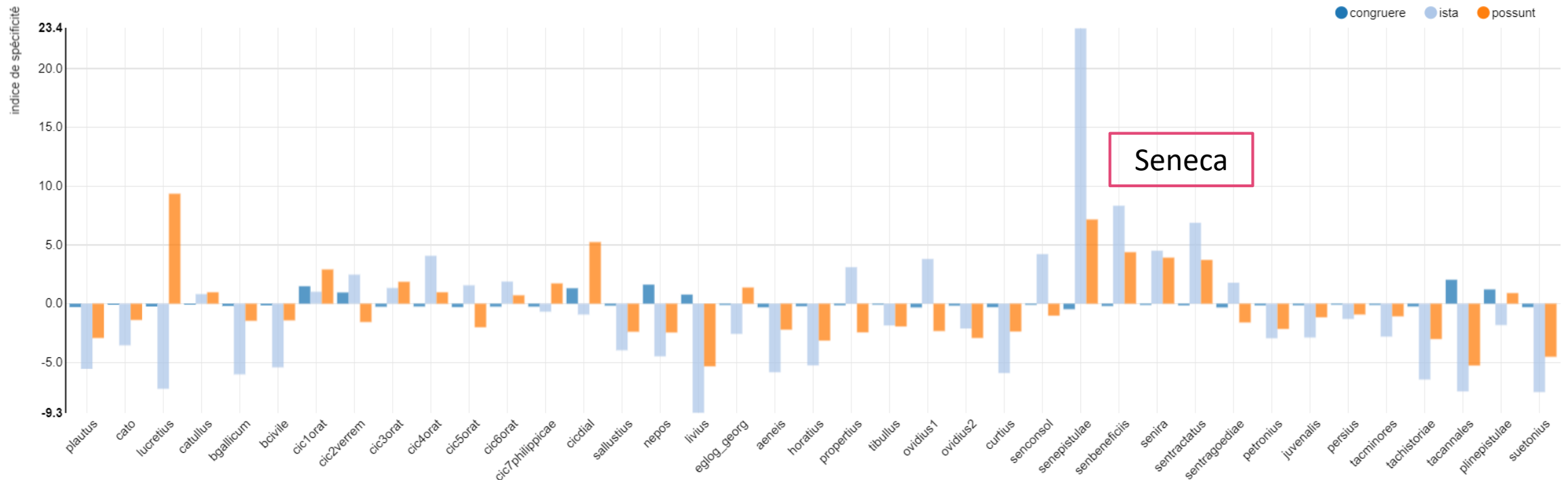
Profil général



Senecan motif?

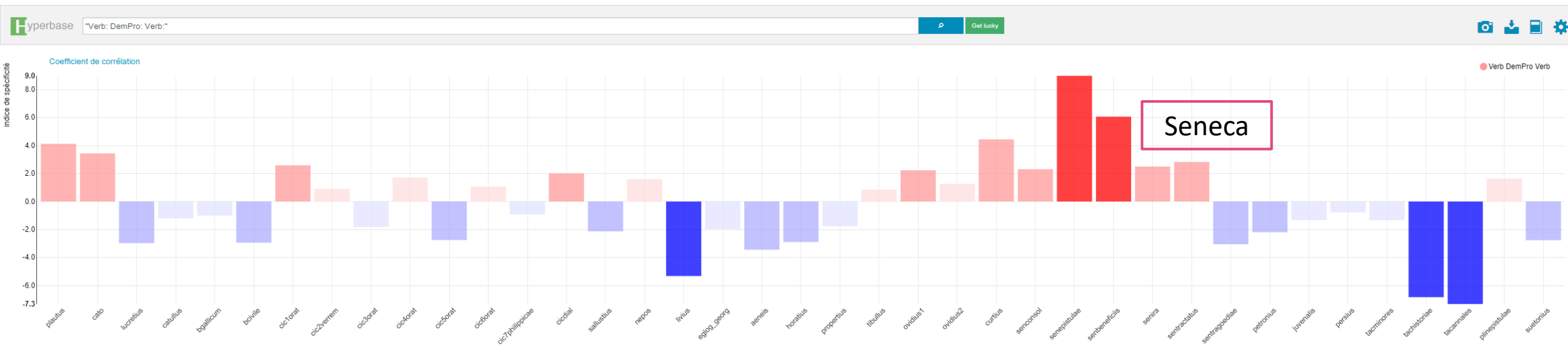
- 'Latin' database (*Hyperbase*): 'Verb: DemPro Verb:' 'Verb: RelPro Verb:'

Hyperbase    



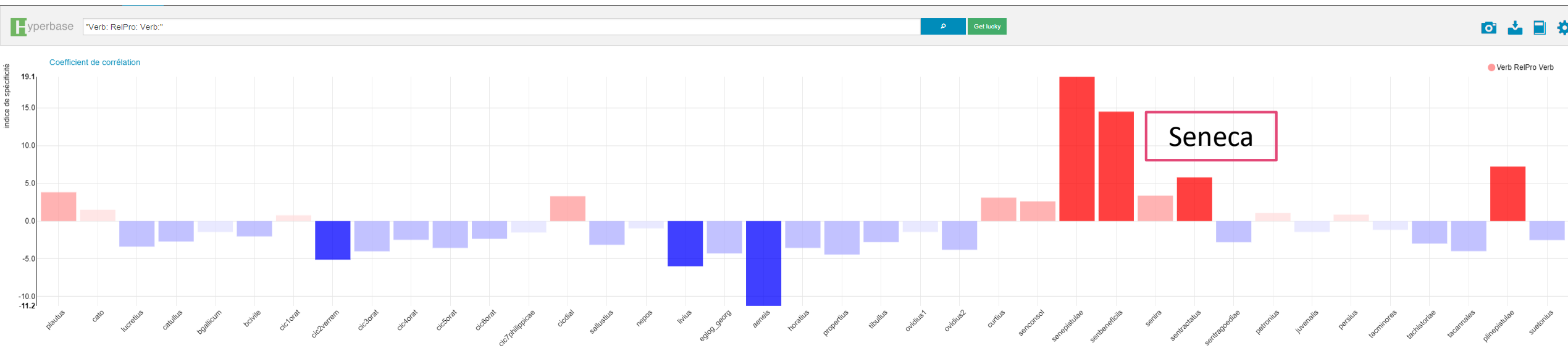
Senecan motif?

- 'Latin' database (*Hyperbase*): 'Verb: DemPro Verb:'



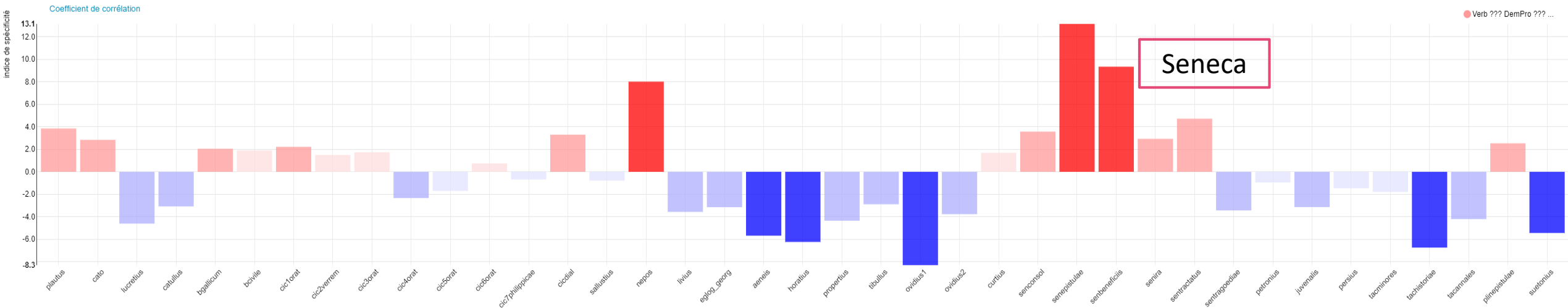
Senecan motif?

- 'Latin' database (*Hyperbase*): 'Verb: DemPro Verb:' 'Verb: RelPro Verb:'



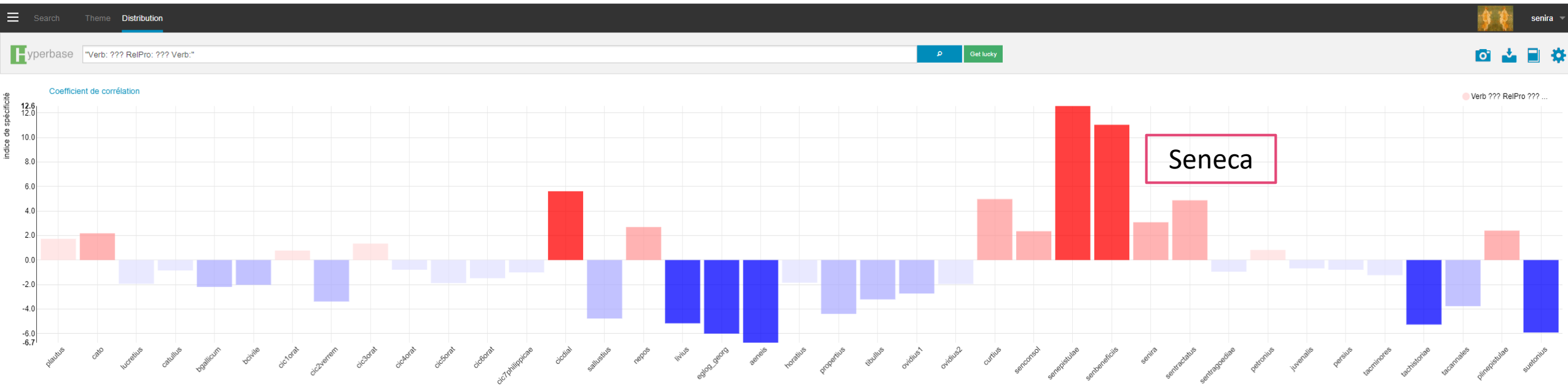
Senecan motif?

- ‘Latin’ database (*Hyperbase*): ‘Verb: ??? DemPro: ??? Verb:’



Senecan motif?

- 'Latin' database (*Hyperbase*): 'Verb: ??? DemPro: ??? Verb:' 'Verb: ??? RelPro: ??? Verb:'



Model 2

- Development of the motif: 'Verb ??? Pronoun ??? Verb'

Letter 162 (~ Seneca):

Passages-clés

Verb

[...] Eufraten **ET_2** absconde illud ibi in foramine **PETRA** . quod cum **FACIO** et in Euftrate sicut iussus fuerat abscondisset **POST_2** plurimos **dies** **Verb:AnomConj:Sing:Ind:Pres:Act:3Pers:CodeSub00**: dominus ad **DemPro:Acc:Sing:Masculin: Verb:3Conj:Sing:Imper:Pres:Act:2Pers:CodeSub00**: uade **AD_2** Eufraten **et** tolle inde lumbar . et abii inquit **Prep**: Eufraten et tuli lumbar de loco ubi **ABSCONDO ILLE** et ecce [...]

[voir dans le texte](#)

Verb

[◀ précédent](#) | 40 | [suivant ▶](#)

[plus d'infos](#)

Pronoun

→ *ait dominus ad eum: surge, vade [...]*

Model 2

- Other examples?

Letter 162 → *ait dominus ad eum: surge, vade [...]*

Letter 162 → *ait: omnis homo qui audit verba mea [...]*

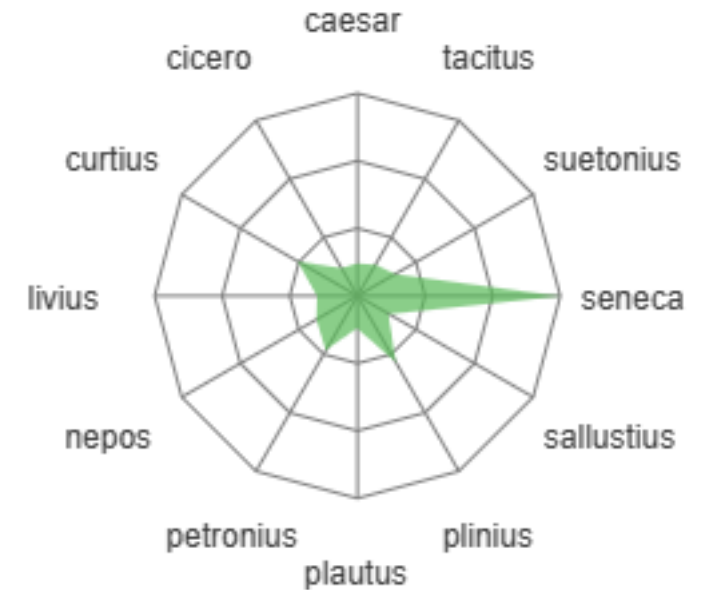
Letter 61 → *apostolo testante qui ait: qui adheret meretrici unum corpus efficitur [...]*

Letter 61 → *Facti siquidem culpam habet qui, quod potest, negligit emendare [...]*

} Bible?

Conclusion

- *Hyperdeep*
 - model 1: forms
 - model 2: forms, lemmas, part-of-speech
- Dialogue?
- Seneca: Peter Damian or biblical language?



author classification (letter 162)

Bibliography

- ↓ [1] Brunet, Étienne and Vanni, Laurent. «Deep learning et authentification des textes». *Texto ! Textes et cultures* 24.1 (2019): 1-34.
- ↓ [2] Eder, Maciej, Rybicki, Jan and Kestemont, Mike. «Stylometry with R: A Package for Computational Text Analysis». *The R Journal* 8.1 (2016): 107-122.
- ↓ [3] Fürst, Alfons. «Epistulae Senecae ad Paulum et Pauli ad Senecam». In *Brill's Companion to Seneca. Philosopher and Dramatist*, edited by Andreas Heil and Gregor Damschen, 213-214. Leiden-Boston: Brill, 2014.
- ↓ [4] Longrée, Dominique and Poudat, Céline. «New Ways of Lemmatizing and Tagging Classical and post-Classical Latin: the LATLEM project of the LASLA». In *Proceedings of the 15th International Colloquium on Latin Linguistics*, edited by Peter Anreiter and Manfred Kienpointner, 683-694. Innsbruck, 2010.
- ↓ [5] Longrée, Dominique, Xuan, Luong and Mellet, Sylvie. «Les motifs : un outil pour la caractérisation topologique des textes». In *Actes des JADT 2008, 9èmes Journées internationales d'Analyse statistique des Données Textuelles*, edited by Serge Heiden and Bénédicte Pincemin, 733-744. Lyon: Presses Universitaires de Lyon, 2008.
- ↓ [6] Martini, Paola Supino. «L'inventario del secolo XII della biblioteca di Santa Croce di Fonte Avellana». In *Studi sulle società e le culture del Medioevo per Girolamo Arnaldi*, edited by Ludovico Gatto and Paola Supino Martini, 629-642. Firenze: All'Insegna del Giglio, 2002.

Bibliography

- ↓ [7] Mayer, Ronald. «Seneca Redivivus: Seneca in the Medieval and Renaissance World». In *The Cambridge Companion to Seneca*, edited by Shadi Bartsch and Alessandro Schiesaro, 277-288. Cambridge: Cambridge University Press, 2015.
- ↓ [8] Reindel, Kurt (éd). *Die Briefe des Petrus Damiani. Teil 1 - 4. Die Briefe der deutschen Kaiserzeit (MGH)*. München: Monumenta Germaniae Historica, 1983-1993.
- ↓ [9] Thon, Valérie, Vanni, Laurent and Longrée, Dominique. «Le deep learning auxiliaire de l'ADT dans le choix de textes à étiqueter en vue d'un corpus de comparaison : à propos de l'étude stylistique des lettres de Pierre Damien». In *Proceedings of the 16th International Conference on Statistical Analysis of Textual Data (2 vols.)*, edited by Michelangelo Misuraca, Germana Scepi and Maria Spano, 834-841. Naples-Cosenza: VADISTAT Press/Edizioni Erranti, 2022.
- ↓ [10] Vanni, Laurent, Corneli, Marco, Longrée, Dominique, Mayaffre, Damon and Precioso, Frédéric. «Hyperdeep : deep learning descriptif pour l'analyse de données textuelles». *Lexicometrica* (2020): 1-12.
- ↓ [11] Vanni, Laurent, Corneli, Marco, Mayaffre, Damon and Precioso, Frédéric. «From text saliency to linguistic objects: learning linguistic interpretable markers with a multi-channels convolutional architecture». *Corpus 24* (2023).
- ↓ <https://doi.org/10.4000/corpus.7667>.
- ↓ [12] Vanni, Laurent, Ducoffre, MMélanie, Mayaffre, Damon, Precioso, Frédéric, Longrée, Dominique, Elango, Veeresh, Buitrago, Nazly Santos, Gonzalez, Juan, Galdo, Luis and Aguilar, Carlos. «Text Deconvolution Saliency (TDS): a deep tool box for linguistic analysis». In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 548-557. Melbourne, 2018.
- ↓ [13] Vanni, Laurent, Mayaffre, Damon and Longrée, Dominique. «ADT et deep learning, regards croisés. Phrases-clefs, motifs et nouveaux observables». In *Proceedings of the 14th International Conference on Statistical Analysis of Textual Data*, edited by Domenica Fioredistella Iezzi, Livia Celardo and Michelangelo Misuraca, 459-466. Rome: UniversItalia, 2018.
- ↓ [14] Verkerk, Philippe, Ouvrard, Yves, Fantoli, Margherita and Longrée, Dominique. «L.A.S.L.A. and Collatinus: a convergence in lexis». In *Studi e saggi linguistici*, edited by Laura Tesconi, 1-26. Pisa: Edizioni ETS, 2020