



# COLLATINUS : LEMMATISER ET ANALYSER DES TEXTES LATINS.

Yves Ouvrard  
Philippe Verkerk

# Qui sommes-nous ?

Yves Ouvrard  
Professeur de latin  
(à la retraite)

Philippe Verkerk  
Physicien  
Lab. PhLAM

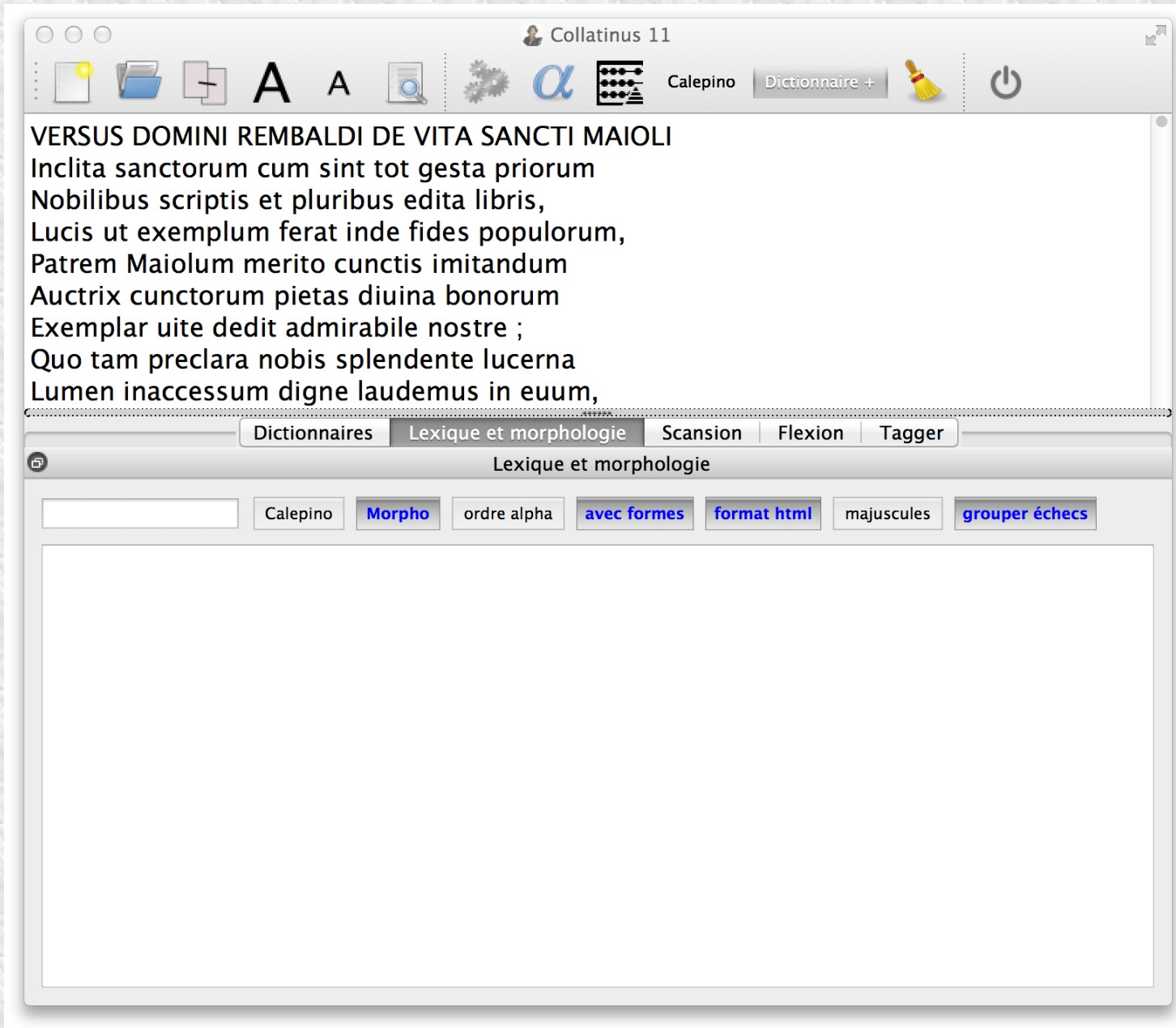
## Collatinus, ses objectifs ? (à l'origine)

- Préparer la liste du vocabulaire associé à un texte.
- Aide à la lecture pour les textes latins.

# Plan de l'exposé

- Visite guidée
  - Lemmatisation et analyse morpho.
  - Dictionnaires
  - Scansion
  - Flexion
  - Tagueur
- Choix techniques
- Principe de l'analyse
- Format des données
- Tagueur probabiliste
- Serveur interne
- Un pas vers le latin médiéval.

# L'interface



Barre d'outils

Fenêtre  
de texte

Onglets  
déplaçables

Fenêtre pour  
les résultats



# Lemmatisation et analyse

The screenshot shows the Collatinus 11 software interface. At the top, there's a menu bar with icons for file operations, text formatting, and a search icon. Below the menu bar, the main text area displays a Latin passage from the 'VITA SANCTI MAIOLI'. A yellow highlight is placed over the word 'tot' in the second line. A tooltip bubble appears over this word, displaying its definition: '• tot, adv. (825) : tant, si nombreux'. Below the text area, there's a tabbed interface with 'Dictionnaires', 'Lexique et morphologie', 'Scansion', 'Flexion', and 'Tagger'. The 'Lexique et morphologie' tab is active. Below the tabs, there's a search bar containing the word 'suis'. To the right of the search bar are several buttons: 'Calepino', 'Morpho', 'ordre alpha', 'avec formes', 'format html', 'majuscules', and 'grouper échecs'. The main results area below shows a list of lemmatized forms for 'suis':

- Inclita
  - *īnclītus, a, um* (30) : célèbre
- suis
  - *sūus, a, um, adj. et pron.* (7170) : adj. son pron. le sien
  - *sūum, i, n.* (154) : (et surtout pl. sua) son bien, ses biens, leurs biens
  - *sūs, suis, f.* (47) : cochon ; truie (f.)
  - *sūo, is, ere, sūi, sūtum* (2) : coudre
  - *sūi, orum, m.* : les siens

Une bulle d'aide s'ouvre avec la lemmatisation du mot sous le curseur

Un clic dans la fenêtre de texte lemmatise le mot

Les solutions sont rangées.

# TextiColor

- Les étudiants sont supposés connaître une liste de mots.
- Le texte est colorisé :
  - Mot connu
  - Mot inconnu
  - Forme médiévale

[Scripsit magister Johannes,  
Veenendaal, Batavia, Anno  
MMXVII.]

## Navigatio sancti Brendani abbatis

1.1 Sanctus Brendanus, filius Finlocha, nepotis Alti, de  
genere Eogeni Stagni Len, regionis Mumenensium ortus fuit.  
5 Erat vir magne abstinentie et in virtutibus clarus, trium  
miliū fere monachorum pater.

abstinentia, ae, f. = modestia  
(<modus), continentia, sobrietas  
miliū: gen. pl. < mille

1.2 Cum esset in suo certamine, in loco qui dicitur Saltus  
Virtutis Brendani, contigit ut quidam patrum ad eum,  
10 quadam vespera, venisset, nomine Barinthus, nepos Neil.

quidam patrum = pater aliquis

1.3 Cumque interrogatus esset multis sermonibus a predicto  
sancto patre, cepit lacrimari et prostrare se in terram et  
diucius permanere in oratione. At Sanctus Brendanus exivit  
15 illum de terra et osculatus est eum, dicens: "Pater, cur  
tristitiam habemus in adventu tuo? Nonne ad consolationem  
nostram venisti? Magis leticiam tu debes fratribus preparare.  
Indica nobis verbum Dei, atque refice animas nostras de  
diversis miraculis, quae vidisti in Oceano."

lacrimari: lacrimare | prostrare =  
prostrernere, prostravi, prostratum

20 1.4 Tunc Sanctus Barinthus, expletis his sermonibus sancti  
Brendani, cepit narrare de quadam insula, dicens: "Filiolus  
meus Mernoc, procurator pauperum Christi, confugit a facie  
mea et voluit se esse solitarium. Invenitque insulam iuxta  
25 Montem Lapidis, nomine Deliciosam. Post multum vero  
temporis, nunciatum est mihi quod plures monachos  
habuisset, et Deus multa mirabilia per illum ostendisset.  
Itaque perrexi, ut visitassem filiolum meum. Cumque  
appropinquarem, per trium dierum iter, in occursum mihi  
30 festinavit cum fratribus suis. Revelavit enim Dominus sibi  
adventum meum. Navigantibus nobis in predictam insulam,

ex-plere = h.l. finire

solitarius, i, m. = vir qui solus vivit

delici-osus = quod valde delectat

plures: multos

visita(vi)ssem: visitarem

per: post

sibi: ei/illi



# Consulter

VERSUS DOMINI REMBALDI DE VITA SANCTI MAIC  
Inclita sanctorum cum sint tot gesta priorum  
Nobilibus scriptis et pluribus edita libris,  
Lucis ut exemplum ferat inde fides populorum,  
Patrem Maiolum merito cunctis imitandum  
Auctrix cunctorum pietas diuina bonorum  
Exemplar uite dedit admirabile nostre ;  
Quo tam preclara nobis splendente lucerna  
Lumen inaccessum digne laudemus in eum,

Dictionnaires Lexique et morphologie

miles Gaffiot 2016

Gaffiot 2016 -- G. Gréco, M. De Wilde, B. Maréchal  
[gerardgreco.free.fr/spip.php?article47](http://gerardgreco.free.fr/spip.php?article47)  
[miles](#)

**mīlēs**,<sup>5</sup> *ītis*, m., soldat : CÆS., CIC., SALL., LIV., etc. ; v. *scribere, conscribere, deligere, conducere* | [sing. collectif] les soldats, l'armée : VIRG., LIV., TAC., etc. ; [en part.] infanterie [opposée à cavalerie] | [fig., en parl. d'une nymphe] *miles erat Phœbes* Ov. *M.* 2, 415, elle était de la suite de Diane | officier, fonctionnaire du palais impérial : DIG. 4, 6, 10 | pion [au jeu des latroncules] : Ov. *Tr.* 2, 1, 477. ➔ arch. *meiles, meilites* : INSCR. | *milex* GROM. 246, 19.

miles Lewis and Short 1879 mileon ↔ milesium <-sync

Lewis & Short, 1879 -- Perseus Digital Library (CC-BY-NC-SA)  
[www.perseus.tufts.edu/miles](http://www.perseus.tufts.edu/miles)

**mīlēs** (MEILES, Inscr. Mur. 582; late form, *milex*, Gromat. Vet. p. 246, 19), *ītis*, comm. [Sanscr root *mil-*, to unite, combine; cf.: *mille*, *milites*, quod trium millium primo legio fiebat, ac singulae tribus Titiensium, Ramnium, Lucerum milia singula militum mittebant, Varr. L. L. 5, § 89 Müll.], a *soldier*.

I. Lit.

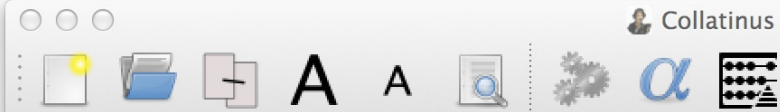
A. In gen.: *miles*, qui locum non tenuit, Cic. Clu. 46, 128: legere milites, to levy, raise, Pompei, ap. Cic. Att. 8, 12 a, 3: *scribere*, to enlist, enroll, Sall. J. 43, 3: *deligere*, Liv. 29, 1: *ordinare*, to form into companies, id. ib.: *mercede conducere*, to hire, take into one's pay, id. ib. 29, 5: *dimittere*, to dismiss, Cic. Fam. 3, 3, 2: *miles tremulus*, i. e. Priam, Juv. 10, 267: *miles cum die*, qui productus sit, aberat, neque excusatus erat, infrequens dabatur, Gell. 16, 4, 5.—

B. In partic., of foot-soldiers, infantry, in opp. to eques: *tripartito milites equitesque in expeditionem inisit*, Caes. B. G. 5, 10: v. eques.— Opp. to the general: *miles gregarius*, or miles alone, a common soldier, private: *strenui militis et boni imperatoris officia simul exsequebatur*, Sall. C. 60, 4: id. L. 62: Vell. 2, 18, 1: *volens militum*, Liv.

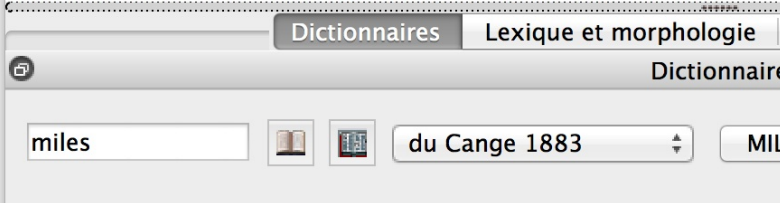
Simultanément  
deux dicos :  
ici, Gaffiot 2016  
et Lewis & Short



# Dicos de



VERSUS DOMINI REMBALDI DE VITA SANCTI MAIO  
Inclita sanctorum cum sint tot gesta priorum  
Nobilibus scriptis et pluribus edita libris,  
Lucis ut exemplum ferat inde fides populorum,  
Patrem Maiolum merito cunctis imitandum  
Auctrix cunctorum pietas diuina bonorum  
Exemplar uite dedit admirabile nostre ;  
Quo tam preclara nobis splendente lucerna  
Lumen inaccessum digne laudemus in eum,

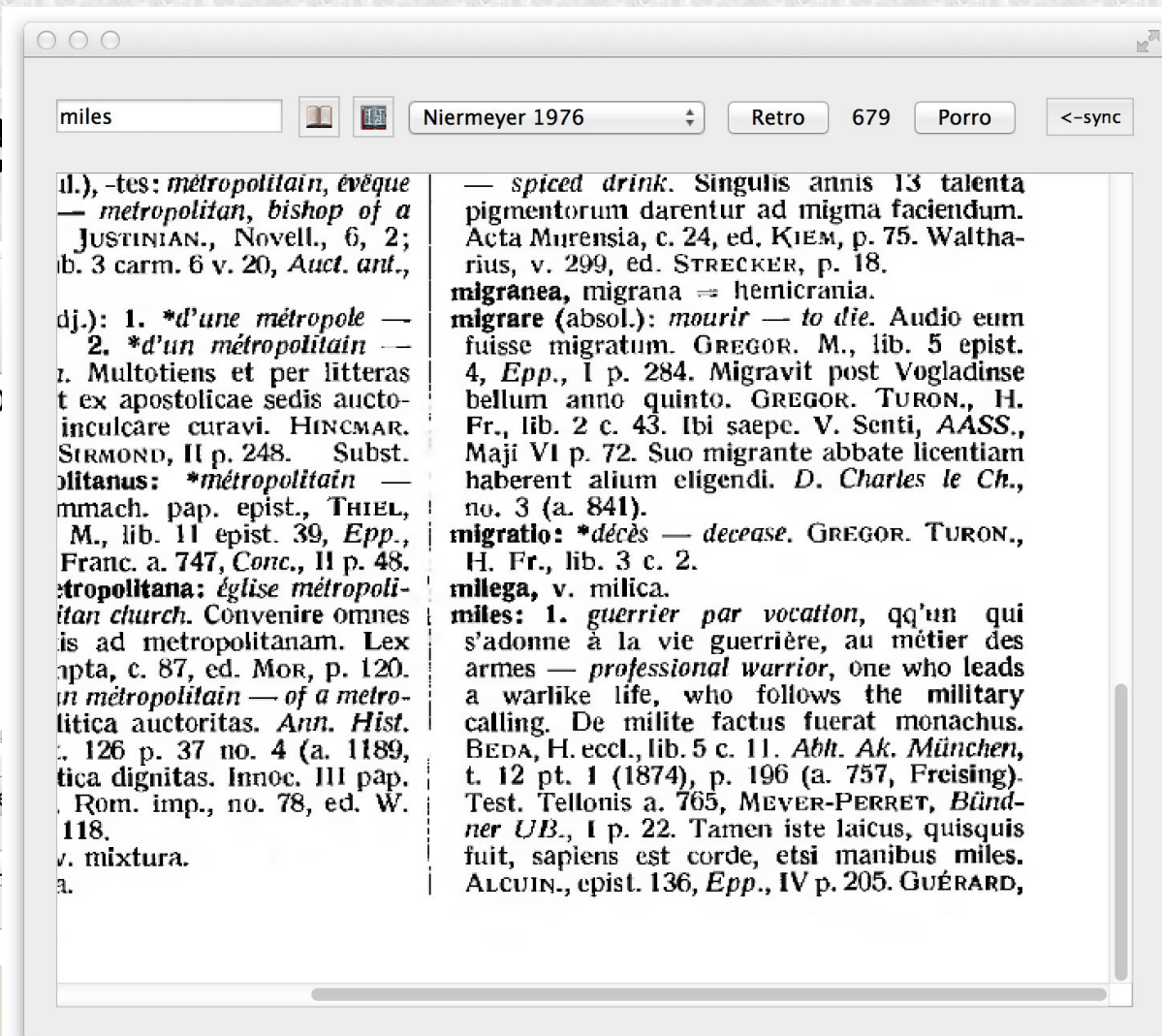


## MILES

« **Miles** » (par C. du Cange, 1678), dans du Cange, *latinitatis*, éd. augm., Niort : L. Favre, 1883–1887, t. 1, p. 100. [ducange.enc.sorbonne.fr/MILES](http://ducange.enc.sorbonne.fr/MILES)

**MILES**, apud Scriptores inferioris ætatis, is potissimum dicitur, qui Militari cingulo accinctus est, quem vulgo *Chevalier* appellamus. Gloss. Latino-Gall. Sangerm. : *Miles, Chevalier*. Fulcherius Carnotensis lib. 2. Hist. Hierosol. cap. 31. (ut cæteros omittam)

Milites nostri erant quingenti, exceptis illis qui Militari nomine non censebantur, tamen equitantes.



il.), -tes: *métropolitain*, évêque — *metropolitan*, bishop of a  
JUSTINIAN., Novell., 6, 2;  
ib. 3 carm. 6 v. 20, *Auct. ant.*,

dj.): 1. \*d'une métropole —  
2. \*d'un métropolitain —  
i. Multotiens et per litteras  
t ex apostolicae sedis aucto-  
inculcare curavi. HINCMAR.  
SIRMOND, II p. 248. Subst.  
olitanus: \**métropolitain* —  
mmach. pap. epist., THIEL,  
M., lib. 11 epist. 39, *Epp.*,  
Franc. a. 747, *Conc.*, II p. 48.  
tropolitanus: *église métropoli-  
tan church*. Convenire omnes  
is ad metropolitanam. Lex  
apta, c. 87, ed. MOR, p. 120.  
in métropolitain — of a metro-  
litica auctoritas. *Ann. Hist.*  
c. 126 p. 37 no. 4 (a. 1189,  
tica dignitas. Innoc. III pap.  
Rom. imp., no. 78, ed. W.  
118.  
v. mixtura.  
a.

— *spiced drink*. Singulis annis 13 talenta  
pigmentorum darentur ad migma faciendum.  
Acta Murensia, c. 24, ed. KIEM, p. 75. Waltha-  
rius, v. 299, ed. STRECKER, p. 18.

**migranea**, migrana = hemicrania.

**migrare** (absol.): *mourir* — *to die*. Audio eum  
fuisse migratum. GREGOR. M., lib. 5 epist.  
4, *Epp.*, I p. 284. Migravit post Vogladinse  
bellum anno quinto. GREGOR. TURON., H.  
Fr., lib. 2 c. 43. Ibi saepe. V. Senti, AASS.,  
Maji VI p. 72. Suo migrante abbate licentiam  
haberent alium eligendi. D. Charles le Ch.,  
no. 3 (a. 841).

**migratio**: \**décès* — *decease*. GREGOR. TURON.,  
H. Fr., lib. 3 c. 2.

**milega**, v. milica.

**miles**: 1. *guerrier par vocation*, qq'un qui  
s'adonne à la vie guerrière, au métier des  
armes — *professional warrior*, one who leads  
a warlike life, who follows the **military**  
calling. De milite factus fuerat monachus.  
BEDA, H. eccl., lib. 5 c. 11. *Abh. Ak. München*,  
t. 12 pt. 1 (1874), p. 196 (a. 757, Freising).  
Test. Tellonis a. 765, MEYER-PERRET, *Bünd-  
ner UB.*, I p. 22. Tamen iste laicus, quisquis  
fuit, sapiens est corde, etsi manibus miles.  
ALCUIN., epist. 136, *Epp.*, IV p. 205. GUÉRARD,



# Scansion

Ārmā (Ārmā) vīrūmquē cānō (cānō), Trōjāe quī prīmūs āb ōrīs (ōrīs)

-- ˘ u-u ˘˘ -- - -u u --

Ītālīām, fātō prōfūgūs, Lāvīniāquē (Lāvīniāquē) vēnīt (vēnīt)

-uu- -- uu- --u-˘ ˘u

lītōrā, mūlt[um] īll[e] ēt tērrīs jāctātūs (jāctātūs) ēt āltō

-uu -` -` - - -˘ u --

vī sūpērūm sāevāe mēmōrēm Jūnōnīs ōb īrām;

- uu- -- uu- --u u -u

mūltā (mūltā) quōqu[e] (quōqu[e]) ēt bēllō pāssūs, dūm cōndērēt ūrbēm,

-- ˘ ˘ - - - - -uu -u

īnfērrētquē dēōs Lātīō, gēnūs (gēnūs) ūndē Lātīnūm,

---u u- uu- u˘ -u u-u

Ālbānīquē pātrēs (pātrēs), ātqu[e] āltāe mōeniā Rōmāe.

---u ˘ -` - - -uu --

## Accentuation :

Spiritāli patri dōmino *Odiloni* abbāti, Syrus monachórum ínfimus, quidquid in christo felícus. Sepe mihi multúmque cunctánti beāti *Maioli* venerábiles actus stili prósequi offício, cum ad id óperis sancte recordatiónis *Warnerii* importúna perurgéret petítio, multa se obiciébant que ne id conárer non módica me formídine deterrébant.

# Flexion

sing.	M.	F.	N.
nom.	dōctŭs	dōctă	dōctŭm
uoc.	dōctě	dōctă	dōctŭm
acc.	dōctŭm	dōctăm	dōctŭm
gen.	dōctī	dōctăe	dōctī
dat.	dōctō	dōctăe	dōctō
abl.	dōctō	dōctă	dōctō

## Tagueur Probabiliste

- Choix de la meilleure lemmatisation en tenant compte du contexte.
- Statistiques sur les séquences de tags dans les textes (classiques) du LASLA.

# Plan de l'exposé

- Visite guidée
  - Lemmatisation et analyse morpho.
  - Dictionnaires
  - Scansion
  - Flexion
  - Tagueur
- Choix techniques
- Principe de l'analyse
- Format des données
- Tagueur probabiliste
- Serveur interne
- Un pas vers le latin médiéval.



# Collatinus 11

- Programme libre et ouvert (GNU GPL)
  - écrit en C++ (Qt 5)
  - les ressources sont des fichiers éditables
  - pas (encore) d'API
- Applications résidentes (Mac OS X, Windows, bientôt Linux)

<http://outils.biblissima.fr/fr/collatinus/>

- Version en ligne (Collatinus 10.2)

<http://outils.biblissima.fr/fr/collatinus-web/>

Arma virumque cano, Troiae qui primus ab oris  
 Italiam, fato profugus, Laviniaque venit  
 litora, multum ille et terris iactatus et alto  
 vi superum saevae memorem Iunonis ob iram;  
 multa quoque et bello passus, dum conderet urbem,  
 inferretque deos Latio, genus unde Latinum,  
 Albanique patres, atque altae moenia Romae.

Dictionnaires Lexique et morphologie Scansion Flexion Tagger

Lexique et morphologie

Calepino Morpho ordre alpha avec formes format html majuscules grouper échecs

- **Arma**
  - **ārma**, *orum, n. pl.* (1773) : armes
    - ārmă accusatif pluriel
    - ārmă nominatif pluriel
    - ārmă vocatif pluriel
  - **ārmo**, *as, are* (143) : armer, munir
    - ārmă 2ème singulier impératif présent actif
- **virumque**
  - **vīr**, *uri, m.* (2702) : homme ( par opp. à mulier, femme ), mari
    - vīrūm + quē accusatif singulier
- **cano**
  - **cāno**, *is, ere, cecini, cantum* (315) : chanter, résonner, retentir ; prédire, prophétiser ; jouer d'un instrument de musique
    - cānō 1ère singulier indicatif présent actif
  - **cānus**, *a, um* (78) : blanc (en parlant des cheveux ou de la barbe)
    - cānō ablatif neutre singulier
    - cānō ablatif masculin singulier
    - cānō datif neutre singulier
    - cānō datif masculin singulier
- **Troiae**
  - **Trōia** *ae, f* (210) : Troie

# Principe de l'analyse

- Pas de liste de formes fléchies  
→ Plus proche du mécanisme naturel
- Couper le mot en 2 (ttes les façons possibles)
- Chercher un radical et une désinence qui sont compatibles
- 80 000 lemmes et 140 paradigmes



# Enclitiques, assimilations et contractions

Si une forme n'est pas reconnue et qu'elle

- se termine par *-que*, *-ne* ou *-ve*, on coupe l'enclitique

- commence par un préfixe susceptible d'être assimilé (*ad-fero* → *affero*) ou par le résultat d'une assimilation, on procède à la substitution

- finit par une contraction (*amasse* → *amavisse*), on la développe

et on refait la recherche.

# Fichiers éditables

- Simple texte avec des séparateurs pour les champs (pas toujours les mêmes)
- Chaque ligne est une entité
- Une ligne commençant par un point d'exclamation “!” est un commentaire
- Le fichier commence, en général, avec une description du contenu et du format
- Tous les fichiers sont dans un même dossier

# Le fichier “modeles.la”

Il donne

- Les règles pour construire les radicaux
- Les désinences (les analyses morpho-syntaxiques étant numérotées)

Notion d'héritage pour ne pas répéter les désinences.

- \$lupus=ÿs;ě;ÿm;ĩ;ō;ō;ĩ;ĩ;ōs;ōrÿm;ĩs;ĩs
- modele:lupus  
R:1:2,0  
des:1-12:1:\$lupus
- modele:liberi  
pere:lupus  
R:1:1,0  
abs:1-6
- modele:templum  
pere:lupus  
R:1:2,0  
des:1-3:1:ÿm;ÿm;ÿm  
des:7-9:1:ă;ă;ă



# Les fichiers “lemmes.\*”

- Le fichier “lemmes.la” donne les informations essentielles.

`lēgītīmus|doctus|||a, um|65`

`lēgo2|amo|||as, are|34`

`lēgo|lego|lēg|lēct|is, ere, legi, lectum|619`

`lēgŭlus|lupus|||i, m.|4`

`lēgŭmĕn|corpus|lēgŭmĭn||īnĭs, n.|6`

- Les autres fichiers “lemmes.fr”, “lemmes.en” etc... donnent les traductions succinctes.

`lego2`:laisser par testament, léguer, envoyer en ambassade

`lego`:cueillir, choisir, lire

`legulus`:cueilleur d'olives, cueilleur de raisin

# Les autres fichiers

- Les irréguliers

dīvŭm:deus:10

dūc\*:duco:181

ēssĕ\*:edo:187

- Formes irrégulières exclusives ou pas.

- Les morphos

1:nominatif singulier

2:vocatif singulier

3:accusatif singulier

...

...

414:positif

415:infinitif futur actif

416:inv.

- Disponibles en plusieurs langues.

# Les chiffres

- 24 100 lemmes (Latin classique)
- 58 000 lemmes dans l'extension (mots plus rares)
- 140 “paradigmes”

- Lewis & Short ≈ 60 000 entrées
- Gaffiot 2016 ≈ 60 000 entrées
- K.E. Georges ≈ 50 000 entrées
- Gérard Jeanneau ≈ 50 000 entrées

Un programme a généré le lexique au format Collatinus

7 500 entrées attendent encore...

- Assimilation du préfixe  
*obfero* ↔ *offero*

- Contractions *fili* → *fili*  
*amasse* ↔ *amavisse*



# Plan de l'exposé

- Visite guidée
  - Lemmatisation et analyse morpho.
  - Dictionnaires
  - Scansion
  - Flexion
  - Tagueur
- Choix techniques
- Principe de l'analyse
- Format des données
- Tagueur probabiliste
- Serveur interne
- Un pas vers le latin médiéval.

# Tagueur probabiliste

- 2nd order hidden Markov model
- Pour chaque phrase, il “choisit” la séquence d'étiquettes (tags) la plus probable.
- Un tagueur probabiliste fonctionne-t-il avec le latin ? (ordre des mots assez “libre”)
- Pas d'évaluation du taux d'erreur (pas encore)

# Le jeu d'étiquettes (tagset)

- Trouver un compromis :
  - trop d'étiquettes = pas assez de statistiques
  - pas assez d'étiquettes = possibilités limitées

Si on ne prend que le POS, *oris* reste ambigu.
- Pour les formes déclinées : le POS, le cas et le nombre. Pour les verbes conjugués, le POS et le mode, avec une distinction pour le présent.
- 91 étiquettes : peut-être trop.



# Les statistiques

- Statistiques sur les textes lemmatisés du LASLA (Latin classique ; 2 000 000 mots).
- Décompte de toutes les séquences de trois étiquettes et de chaque lemme.
- Collatinus n'a pas de liste de formes, or un tagueur a besoin de la fréquence des formes.
  - Hypothèse forte : l'usage des cas et nombre est le même pour tous les mots.

# Le calcul

- Construire toutes les séquences possibles.
- Leur associer une probabilité par récurrence.
- Garder la plus probable.
- Limiter le nombre de séquences examinées : si deux séquences se terminent par les deux mêmes étiquettes, seule la plus probable a une chance de “survivre”.

Collatinus 11

VERSUS DOMINI REMBALDI DE VITA SANCTI MAIOLI  
 Inclita sanctorum cum sint tot gesta priorum  
 Nobilibus scriptis et pluribus edita libris,  
 Lucis ut exemplum ferat inde fides populorum,  
 Patrem Maiolum merito cunctis imitandum  
 Auctrix cunctorum pietas diuina bonorum  
 Exemplar uite dedit admirabile nostre :

Dictionnaires | Lexique et morphologie | Scansion | Flexion | Tagger

Tagger

Tagger probabiliste dérivé des [textes du LASLA](#) majuscules tout afficher

VERSUS DOMINI REMBALDI DE VITA SANCTI MAIOLI Inclita sanctorum cum sint tot gesta priorum Nobilibus scriptis et pluribus edita libris, Lucis ut exemplum ferat inde fides populorum, Patrem Maiolum merito cunctis imitandum Auctrix cunctorum pietas diuina bonorum Exemplar uite dedit admirabile nostre

snt n11 n41 r n61 a12 a12 a42 d v21 d w11 n42 a62 n62 c a62 w11 n62 n41 c n31 v21 d n11 n42 n31 d a62 w31 n11 a42 n11 a11 n42 n11 n41 v1 a11 p51 snt  
 avec la proba : 0.000112406 pour 3.33819e+18 branches.  
 Deuxième choix avec la proba : 0.000110851  
 snt d n41 r n61 a12 a12 a42 d v21 d w11 n42 a62 n62 c a62 w11 n62 n41 c n31 v21 d n11 n42 n31 d a62 w31 n11 a42 n11 a11 n42 n11 n41 v1 a11 p51 snt

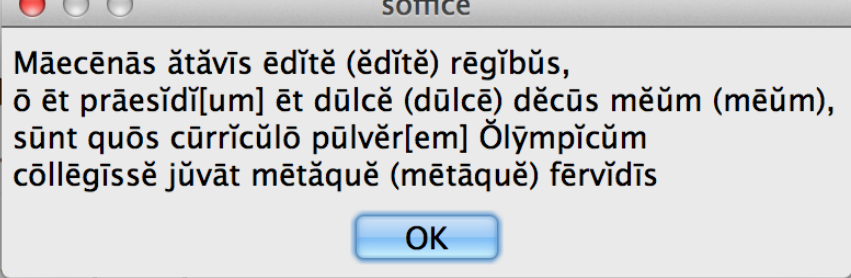
- **VERSUS** n11 (w11)  
 —> **vērsus**, *us, m.*, (230) : sillon, ligne, vers — vērsūs nominatif singulier
- **DOMINI** n41  
 —> **dōmīnus**, *i, m.* (675) : maître — dōmīnī génitif singulier
- **REMBALDI** : non trouvé
- **DE** r  
 —> **dē**, *prép. + abl.* (6164) : au sujet de, du haut de, de — dē
- **VITA** n61  
 —> **vīta**, *ae, f.* (2080) : vie — vītā ablatif singulier
- **SANCTI** a12 (a41)  
 —> **sānctus**, *a, um* (320) : sacré, saint — sānctī nominatif masculin pluriel
- **MAIOLI** : non trouvé
- **Inclita** a12 (a11)  
 —> **īnclītus**, *a, um* (30) : célèbre — īnclītā nominatif neutre pluriel
- **sanctorum** a42  
 —> **sānctus**, *a, um* (320) : sacré, saint — sānctōrūm génitif masculin pluriel
- **cum** d (c)  
 —> **cūm**, **quōm**, **quūm**, *prép. + abl. et adv.* (5910) : prép. avec; adv. (cum ... tum) tantôt [... tantôt] — cūm
- **sint** v21  
 —> **sūm**, *es, esse, fui* (40916) : être (en tête de phrase : il y a) — sīnt 3ème pluriel subjonctif présent actif
- **tot** d  
 —> **tōt**, *adv.* (825) : tant, si nombreux — tōt
- **gesta** w11

# Serveur interne

- Collatinus peut répondre à un autre programme. Par exemple, LibreOffice.

[1,01,1] Maecenas atavis edite regibus,  
o et praesidium et dulce decus meum,  
sunt quos curriculo pulverem Olympicum  
collegisse iuvat metaque feruidis

[1,01,5] terrarum  
hunc, s  
certat  
illum,



Māecēnās ātāvīs ēdītē (ēdītē) rēgībūs,  
ō ēt prāesīdī[um] ēt dūlcē (dūlcē) dēcūs mēūm (mēūm),  
sūnt quōs cūrrīcūlō pūlvēr[em] Ōlŷmpīcūm  
cōllēgīssē jūvāt mētāquē (mētāquē) fērvīdīs

OK

[1,01,10] quicquid de Libycis uerritur arc  
Gaudentem patrios findere sarculo  
agros Attaliciis condicionibus  
numquam demouneas, ut trabe Cypria  
Myrtoum pauidus nauta secet mare.



# Utiliser le serveur interne

- Écrire un programme qui appelle Collatinus :  
par exemple, un tagueur “standard” pour le LASLA
  - Liste de formes, avec lemme et analyse
  - Nombre exact d'occurrences
- Problème des formes qui n'ont pas encore été rencontrées : demander à Collatinus !
  - Analyse au format du LASLA.

# LASLA-Tagger

- Lemmatisation supervisée.
- Édition/correction du résultat du tagueur.
  - Objectif : zéro erreur !
- Nombre exact d'occurrences. Pour “patres” :  
253 voc. 110 nom. et 75 acc.  
Collatinus : 6 voc. 200 nom. et 320 acc.

# LASLA-Tagger : capture d'écran

LASLA_tagger				
6	partium	PARS	A342	Ref&0001/14,1,2
7	animis	ANIMVS	A262	Ref&0001/14,1,2
8	quas	QVI	1 J 32	Ref&0001/14,1,2
9	periculorum	PERICVLVM	A242	Ref&0001/14,1,2
10	uarietas	VARIETAS	A311	Ref&0001/14,1,2
11	fregerat	FRANGO	B3 1 1513	Ref&0001/14,1,2
12	et	ET	2 S	Ref&0001/14,1,3
13	laborum	LABOR	1 A342	Ref&0001/14,1,3
14	nondum	NONDVM	P	Ref&0001/14,1,3
15	tubarum	TVBA	A142	Ref&0001/14,1,3
16	cessante	CESSO	B161 411	Ref&0001/14,1,3
17	clangore	CLANGOR	A361	Ref&0001/14,1,3
18	uel	VEL	2 S	Ref&0001/14,1,3
19	milite	MILES	A361	Ref&0001/14,1,3
20	locato	LOCO	B161 442	Ref&0001/14,1,3
21	per	PER	R	Ref&0001/14,1,4
22	stationes	STATIO	A332	Ref&0001/14,1,4
23	hibernas	HIBERNVS	C1321	Ref&0001/14,1,4
24	Fortunae	FORTVNA	N A151	Ref&0001/14,1,4
25	saeuientes	SAEVIO	B432 411	Ref&0001/14,1,4
26	procellae	PROCELLA	A141	Ref&0001/14,1,4
27	tempestates	TEMPESTAS	A332	Ref&0001/14,1,4
28	alias	ALIAS	M 1	Ref&0001/14,1,5
29	rebus	RES	A562	Ref&0001/14,1,5
30	infudere	INFVNDIO	2 B3 2 1413	Ref&0001/14,1,5
31	communibus	COMMVNIS	C4621	Ref&0001/14,1,5
32	per	PER	R	Ref&0001/14,1,5
33	multa	MVLTVS	C1321	Ref&0001/14,1,5
34	illa	ILLE	I 32	Ref&0001/14,1,5
35	et	ET	2 S	Ref&0001/14,1,5/
36	dira	DIRVS	C1321	Ref&0001/14,1,5/

Fenêtre de sélection			
-9	Modifier la référence Ref&0001/14,1,5/		
-4	Supprimer ce mot...		
-2	Ajouter un mot...		
-1	Ajouter une analyse...		
0	<b>multa</b>	<b>MVLTVS</b>	<b>C1321</b>
1	multa	MVLTVS	C1121
2	multa	MVLTVS	C1611
3	multa	MVLTVS	C1111
4	multa	MVLTVM	1 A232
5	multa	MVLTA	1 A111
6	multa	MVLTA	2 M 1
7	multa	MVLTA	1 A161
8	multa	MVLTVM	1 A212
-1	Ajouter une analyse...		
-2	Ajouter un mot...		
-4	Supprimer ce mot...		
-9	Modifier la référence Ref&0001/14,1,5/		

# Plan de l'exposé

- Visite guidée
    - Lemmatisation et analyse morpho.
    - Dictionnaires
    - Scansion
    - Flexion
    - Tagueur
  - Choix techniques
  - Principe de l'analyse
  - Format des données
  - Tagueur probabiliste
  - Serveur interne
- Un pas vers le latin médiéval.



# Latin médiéval

- Option 1 : modifier les fichiers éditables pour les adapter à la langue.  
→ Nécessaire si on veut rendre l'évolution des sens.
- Option 2 : mettre un filtre pour s'accommoder des principales variations de graphie (et ajouter quelques mots nouveaux, ou désinences).

# Filtre médiéval

- Graphies médiévales :  
e pour ae ou oe, o pour au, f pour ph, etc...
- Difficile de remonter à la graphie classique
- “Médiévaliser” les radicaux et les désinences classiques.
- Appliquer le même filtre au texte  
→ hyper-urbanismes *aecclesia* pour *ecclesia*.

# Filtre Médiéval (2)

- Avec 7 règles de substitution, il n'y a plus que 112 mots non reconnus dans la vie de Maieul (sur 15 704). Contre 410, sans filtre.
- Beaucoup de noms propres.
- Mais aussi *explicit*, *archiepiscopeatus*, etc...
- Et des variantes à ajouter : *dam<sup>p</sup>natio*
- Sans oublier les typos : *per<sup>ue</sup>ueniemus*

# Faux positifs

- La simplification dans la graphie introduit de nouvelles ambiguïtés. Par exemple :  
vite → vitis (class.) vita (méd.)  
poete → poeta, paetus
- Mesurer une “distance” et choisir la solution la plus vraisemblable ?



# Perspectives

## ***Praelector* :**

- “Assistant grammatical ” pour aider à la lecture et à la compréhension de la phrase.  
→ “Traduction” de la phrase
- Coupler le tagueur et l'analyse grammaticale
- Statistiques sur le rythme (cursus)
- Recherche de vers dans la prose.

Merci de votre attention.

Yves.Ouvrard@collatinus.org

Philippe.Verkerk@univ-lille1.fr

<http://outils.biblissima.fr/fr/collatinus/>

<http://outils.biblissima.fr/fr/collatinus-web/>

Pour le Grec ancien : <http://outils.biblissima.fr/fr/eulexis/>