

# { antiquités numériques }

cours 3 : premiers exercices

# Créer de la donnée : comment faire ?

Transkribus demo odyssée > #1 11.1.2024, 10:48

ΟΔΥΣΣΕΙΑΣ Α΄ ΟΜΗ  
ΡΟΥ ΡΑΥΩΝΙΔΙΑΣ.

Αλφα θεῶμ ἀγορὴ Οδυσσεῖον Πατλάδη θέρος.

Νοῦσοι μοι γίνεται Μᾶστις τολύ-  
τροπομ, ὃς μάλα πολλὰ  
πλάγχθη, ἐπει Τροίης ἐρόμη πῆ-  
λιεθροῦ ἐπήρσε.  
Πολλῶν δ' αὐθεώπων ἕδην ἔ-  
σται, καὶ νόσοι ἐγνω.

Γολάδ' οὐ γὰρ τοντας τάθην ἄλγεα, ὃν κατέθυμόμη  
Ἄργυρόδηνος ἦμετε φυκῆμη, καὶ νόσοι ἐταίρωμ.  
Ἄλλ' ζελῶς ἐτάρρεταις ερρύνεταιο εἰμινός πορ,  
Αὐτῶν γη σφετέρησιν ἀπαδαλίκησιν ὅλεντο  
Νηπίοις, οἵ τε βόστεροιονος κατέλιοιο  
Ἡδιού, αὐτέρ δ, χωτησιν ἀφείλεται νόσιμοι πίκαροι.  
Τάδης ἀμόθην γε θεὰ θύγατερ διός, εἰπὲ καὶ νόμιμη.  
Ἐνθάδελοι μὲν ταῦτας ὅσιε φύγονται πολὺ ὀλεθροῦ  
Οἴκοι ἐσαν, τάλειμοι τε τεφερούστοι, οὐδὲ θάλασσαν,  
Τόν δ' οἴοι νόσοι κεχρημάτιον οὐδὲ γυανικά,  
Νύμφη τότην ερικε Καλυψώ σιασ θεάσιν  
Ἐμπάτεσι γλαφυροῖσι λιλαεομένη τόσιν  
Ἄλλ' ὅτε δηλετος οὐδετε πούειπλοι μάργην γίνεται,  
Τῶοι ἐπεκλώσαντο θεοί, οἴκους δε τένεδται  
Εἰς ιθάκην, γένθα τεφυγμόνθην ἀεθλωμ.

ΟΔΥΣΣΕΙΑΣ Α΄ ΟΜΗ  
ΡΟΥ ΡΑΥΩΝΙΔΙΑΣ.

Αλφα θεῶμ ἀγορὴ Οδυσσεῖον Πατλάδη θέρος.

Νοῦσοι μοι γίνεται Μᾶστις τολύ-  
τροπομ, ὃς μάλα πολλὰ  
πλάγχθη, ἐπει Τροίης ἐρόμη πῆ-  
λιεθροῦ ἐπήρσε.  
Πολλῶν δ' αὐθεώπων ἕδην ἔ-  
σται, καὶ νόσοι ἐγνω.

Γολάδ' οὐ γὰρ τοντας τάθην ἄλγεα, ὃν κατέθυμόμη  
Ἄργυρόδηνος ἦμετε φυκῆμη, καὶ νόσοι ἐταίρωμ.  
Ἄλλ' ζελῶς ἐτάρρεταις ερρύνεταιο εἰμινός πορ,  
Αὐτῶν γη σφετέρησιν ἀπαδαλίκησιν ὅλεντο

Aucune transcription pour cette page

Lancer la transcription automatique

Reconnaissance du texte Mise en page

Smart Search +50% Language Model Advanced Settings

Début de la reconnaissance Credits needed:  
Available (Personal | Collection) 282 | 0

Nom Search Mots Langue CER

Favorite Models 0

Public Models 3 Noscemus GM 6 667 127 LAT, GRE, GER 0.80%

Private Models 1 Noscemus GM 5 607 837 LAT, GRE, GER 0.60%

German Combined M1 8 090 801 GER 2.10%

Filtre

languages

Noscemus GM 6

ID: 52640

Created by NOSCEMUS project Jun 6, 2023

# Créer de la donnée : comment faire ?

Transkribus demo > odyssey #1 En cours 11.1.2024, 10:53

Region 1

νδρα μοι ρήνεπε Μούσα πολύ-  
τροπον, ός μάλα πολλά  
Πλαγχθη, ἐπει τραίσιερν πτο  
λιεθρον ἐπρεσε.  
Πολλῶν δ αύθρώπων ὅδην ἄ  
σεα, καὶ νόν ἔγνω.

Πολλα δ' ὅγ γυ πόντω πάθον ἀλγεα, ὃν καἀτθυμὸν  
Ἀρνύμθον κν τε ψυχὴν, και νόστον ἑταίρων.  
ἀλλ' οὐδ ὡς ἐτάρους ἐρρυτατιεμένος πορ,  
Αύτῶν γνς σφετόρησιν ἀτασθαλίσιν δλοντο  
Νύπτοι, οι καὰς Βοῦς ἐπερίονος κελιοι  
είσθιον, αὐτάρ ὁ, τοῖσιν ἀσφειλετα νόστιμον ἥμαρ.  
τὸν ἀμόθον γε θεὰ θύγατερ Λιός, εἰπὲ και ήμιν.  
ἐνθ' ἀλλοι μὲν πάντες ὅτοι φύγον ἀπὸν δλεθρον  
Οἰκοι ἔσαν, πόλεμὸν σε πεφουτότεν, κδὲ θάλασσαν,  
Τὸν δ' οἰον νόστου κεχρημένον ἥδε γυναικός,  
λύμφη πόντι ἔρυκε Καλυψώ δία θεάων  
ρυσπτιλαθροίσι Ιυλαιαθηνη πύσιεν  
Αλλ' ὅτε δήξεος κλθε ποριπλομόνων νιαυτῶν,  
Τῷ οι ἐπεκλώσαυτα θεοί, οικόν δε νέεσθαι

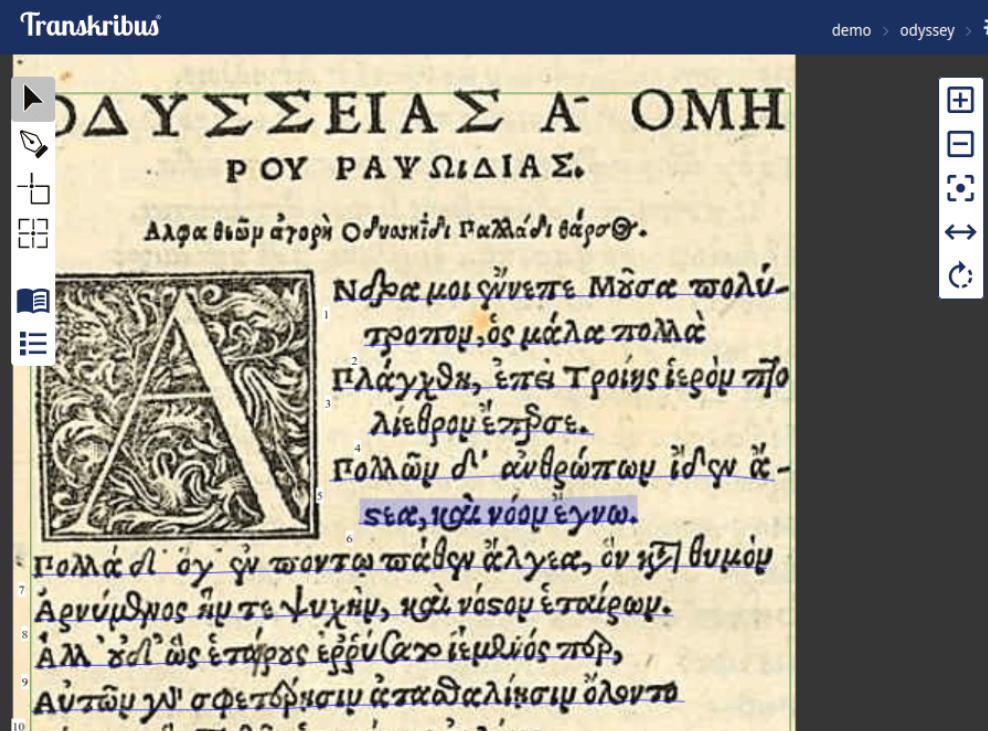
ΔΥΣΣΕΙΑΣ Α<sup>Γ</sup> ΟΜΗ  
ΡΟΥ ΡΑΨΩΙΔΙΑΣ.

Αλφα θεῶμ ἀτορὴ ο θυντήμι παλλάθι θάρσθ.

Nδρα μοι γῆνεπε Μάσα πολύ-  
τροπομ, ὃς μάλα πολλὰ  
Πλάγχθη, ἐπει τροίνις ερόμ πῆ-  
λιεθρομ ἐπρεσε.  
Πολλῶν δ' αὐθρώπων ἕδην ἄ-  
σεα, καὶ νόν ἔγνω.

Γολάδι οὐ γη τεστω ταύθην ἀλγεα, σι μηθι θυμόμ  
Ἄρνυμθος ἀμ τε ψυχὴν, καὶ νόσομ ἐταίρων.  
Ἄλλ' ζελῶς ἐτάρος ερρύνταχο εμβιός πορ,  
Αύτῶν γη σφετόρησιν ἀτασθαλίσιμ ὅλεντο

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20



# Créer de la donnée : comment faire ?

1—**ἀνδρα μοι ἔννεπε, μοῦσα, πολύτροπον,**  
—**δὲ μάλα πολλὰ πλάγχθη, ἐπεὶ Τροίης**  
—**ἱερὸν πτολίεθρον ἐπερσεν:** πολλῶν δ'  
—**ἀνθρώπων** ἕδεν ἄστεα καὶ νόον ἔγνω,  
—**πολλὰ δ' ὅ γ'** ἐν πόντῳ πάθεν ἄλγεα ὅν  
—**κατὰ θυμόν,** ἀρνύμενος ἦν τε ψυχὴν καὶ  
—**νόστον** ἔταιρων. ἀλλ' οὐδὲ ὡς ἔταρους  
—**έρρυσατο,** ιέμενός περ: αὐτῶν γὰρ  
—**σφετέρησιν** ἀτασθαλίησιν ὅλοντο,  
—**νήπιοι,** οἵ κατὰ βοῦς Ὑπερίονος Ἡελίοιο  
—**ἥσθιον:** αὐτὰρ ὁ τοῖσιν ἀφείλετο  
—**νόστιμον** ἥμαρ. τῶν ἀμόθεν γε, θεά,  
—**θύγατερ Διός,** εἰπὲ καὶ ἡμῖν. ἐνθ'  
—**ἄλλοι** μὲν πάντες, δσοι φύγον αἰπὺν

1+**νδρα μοι ρῆνεπε Μοῦσα πολύτροπον, ὃς**  
+**μάλα πολλὰ Πλαγχθη, ἐπεὶ τραίησίερν**  
+**πτολιεθρον ἐπρσε.** Πολλῶν δ ἀύθρώπων  
+**ἴδην ἄσεα,** καὶ νόον ἔγνω. Πολλα δ' ὅγ  
+**γυ πόντῳ πάθον** ἄλγεα, ὅν καὰτθυμὸν  
+**Ἀρνύμθνος κν τε ψυχὴν,** καὶ νόστον  
+**ἔταιρων.** ἀλλ' οὐδὲ ὡς ἔταρους  
+**έρρυτατιεμένός πορ,** Αύτῶν γνς  
+**σφετόρησιν** ἀτασθθαλίησιν ὅλοντο  
+**Νύπιοι,** οἵ καὰι βοῦς ἐπερίονος κελίοιο  
+**είσθιον,** αὐτὰρ ὁ, τοῖσιν ἀἀφειλετα  
+**νόστιμον** ἥμαρ. τὸν ἀμόθον γε θεὰ  
+**θύγατερ Λιὸς,** εἰπὲ καὶ ἡμῖν. ἐνθ'  
+**ἄλλοι μὲν πάντες** δτοι φύγον ἀιπὺν

De la genealogie & antiquite de  
Gargantua Chapitre.  
Et nous remets a la grande chomique Pantagruel  
lune reconnoistre la genealogie & antiquite dont nous  
est demeure Gargantua. En  
icelle nous entendez plus au long com  
ment les geans naissent en ce monde:  
et comment diceulz par signes directes  
yssit Gargantua pere de Pantagruel:  
et ne nous facherai, si pour le present  
je men deporte. Edien que la chose soit  
telle: que tant plus seroit remembree,  
tant plus elle plairoit a nos seignuries;  
comme nous ayons laudon de flacque qui  
dict estre aucuns propos, tels que ceux  
ce qui plus sont delectables, quand plus  
souuent sont redits. Pleust a dieu q'ym  
chacun sceust aussi certainement sa ge  
nealogie, depuis l'arche de Noe insques  
acec eage. Je pense que plusieurs sont  
amourd'huys empereurs, roys, ducs, prin  
ces, et papes, en la terre, lesquels sont des  
scendus de quelques porteurs de rogas  
sons & de coufretz. Comme au rebours  
plusieurs sont queux de l'hostiaire, sont  
fretens, & miserables: lesquels sont desces  
du sang & ligne de grands roys & em  
perateurs: atteindu l'admirable traistour des  
regnes & empries: des Assyriens, des He  
bes, des Medes, des Peres, des Pers  
es, des Macedones, des Macdonnes ca

De la genealogie & antiquite de  
Gargantua Chapitre.  
Et nous remets a la grande chomique Pantagruel  
lune reconnoistre la genealogie & antiquite dont nous  
est demeure Gargantua. En  
icelle nous entendez plus au long com  
ment les geans naissent en ce monde:  
et comment diceulz par signes directes  
yssit Gargantua pere de Pantagruel:  
et ne nous facherai, si pour le present  
je men deporte. Edien que la chose soit  
telle: que tant plus seroit remembree,  
tant plus elle plairoit a nos seignuries;  
comme nous ayons laudon de flacque qui  
dict estre aucuns propos, tels que ceux  
ce qui plus sont delectables, quand plus  
souuent sont redits. Pleust a dieu q'ym  
chacun sceust aussi certainement sa ge  
nealogie, depuis l'arche de Noe insques  
acec eage. Je pense que plusieurs sont  
amourd'huys empereurs, roys, ducs, prin  
ces, et papes, en la terre, lesquels sont des  
scendus de quelques porteurs de rogas  
sons & de coufretz. Comme au rebours  
plusieurs sont queux de l'hostiaire, sont  
fretens, & miserables: lesquels sont desces  
du sang & ligne de grands roys & em  
perateurs: atteindu l'admirable traistour des  
regnes & empries: des Assyriens, des He  
bes, des Medes, des Peres, des Pers  
es, des Macedones, des Macdonnes ca



# HTR

L'HTR

Handwritten Text Recognition

# / (reg)ex/

ν εωκε, τοι ουρανόδάων γαιήοχος ἀσκεκχόλωται, δὸν ὁφθαλμούλυφημον, δου κράτηώπεσσι· Θόωσα δέ με θυγάτηρ, ἀλὸς ἀτρυγγλαφυροῖσι **Ποσειδόν** θυμοῦσῃ **Ποσειδάων** ἐκτείνει, πλάζει δὲ ἡμεῖς οἶδε περιφριῶς ἔλθησι· **Ποσειδόν** οὐ μὲν γάρ τι δυνάκεκτηι θεῶν ἔριδαινείβετ' ἔπειτα θεὰ ν  
nd: \bΠοσειδ\p{L}+\b

[^]\*?@[^]\*?\.[^]\*

ένη ξείνω, Ταφίων ἡγήτορι, Μέντηδ' ἄρα μνηστῆρας ἀγήνορας· οἱ μὲν ἔπεισι προπάροιθε θυράων θυμὸν ἔτερον, ι ἐν δινοῖσι βοῶν, οὓς ἔκτανον αὐτοί· ες δ' αὐτοῖσι καὶ ὁτρηροὶ θεράποντες ν ἄρ' οἷνον ἔμισγον ἐνὶ κρητῆρσι καὶ αὗτε σπόργυοισι πολυτρήτοισι τραπέζας καὶ πρότιθεν, τοὶ δὲ κρέα πολλὰ δατεῖ πολὺ πρώτος ἦδε Τηλέμαχος θεοειδῆς· γὰρ ἐν μνηστῆρσι φίλον τετιημένος ἦτοι ενος πατέρ' ἐσθλὸν ἐνὶ φρεσίν, εἴ ποθιρῶν τῶν μὲν σκέδασιν κατὰ δώματα θεῖδ' αὐτὸς ἔχοι καὶ κτίμασιν οἴσιν ἀνά-

**Chercher un peu mieux dans les textes**

Les expressions régulières

# Un prérequis recommandé

Vous avez deux potentielles solutions pour utiliser les regex :

- le programme SublimeText
- l'interface en ligne RegexR

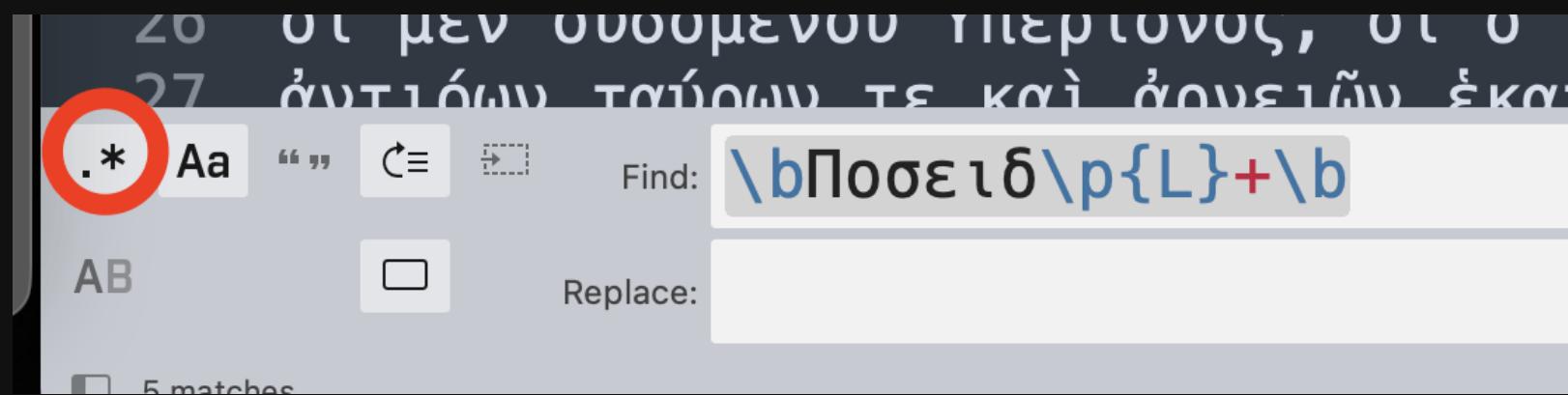
Mais les regex fonctionnent sur tous les outils d'écriture (libreoffice, word, etc.)

# Un prérequis recommandé

Attention à une petite chose :

lorsque vous recherchez une regex, vous devez enclencher l'option "chercher avec les regex", qui peut prendre différentes formes.

Sur sublimeText, cette option, c'est ".<sup>\*</sup>"



# Les expressions régulières ?

L'idée principale c'est :

- qu'on peut chercher des caractères littéraux,
- mais qu'on peut aussi chercher des chaînes plus complexes avec des caractères qui ont des propriétés particulières (méta-caractères)

# Les expressions régulières ?

Exemple :

- si dans l'onglet Recherche je n'active pas les regex, et que je cherche ".", je vais trouver les points, et c'est tout.
- si on active les regex, en cherchant ".", on trouve TOUS les caractères du texte.

# Les expressions régulières ?

DONC :

- il y a toute une liste de caractères qui veulent dire plus que ce qu'on voit,
- on peut donc chercher des choses beaucoup plus larges qu'avec la recherche simple.

# Les expressions régulières ?

Exemple :

- je veux chercher toutes les variantes de Ποσειδῶν dans un texte :

voilà la regex :

\bΠοσειδ\p{L}+\b

τοισιν εωκε, τοι ουρανον ει  
Ποσειδάων γαιήοχος ἀσκελὲς  
πος κεχόλωται, ὅν ὄφθαλμοῦ  
εον Πολύφημον, δου κράτος ἔσται  
Κυκλώπεσσι· Θόωσα δέ μιν τα  
νος θυγάτηρ, ἀλὸς ἀτρυγέτοις  
έεσι γλαφυροῖσι Ποσειδάωνι  
ĩ δὴ Ὄδυσῆα Ποσειδάων ἐνοσία  
κατακτείνει, πλάζει δ' ἀπὸ  
ἄγεθ' ἡμεῖς οἴδε περιφραζώμα  
ν, ὅπως ἔλθησι· Ποσειδάων δι  
λον· οὐ μὲν γάρ τι δυνήσεται  
των ἀέκητι θεῶν ἔριδαινέμεν  
  
· ἡμείβετ· ἔπειτα θεὰ νλαυκί

Find: \bΠοσειδ\p{L}+\b

# Les expressions régulières ?

Deux choses à retenir :

- les classes
- les quantificateurs

# Les expressions régulières ?

Les classes :

Avec une classe de regex, je peux donner un choix de jeux de caractères inconnus à mon ordinateur.

# Les expressions régulières ?

Exemple :

Je sais que Ποσειδῶν pourra avoir  
comme terminaison Ποσειδάωνι :  
donc je veux que mon ordinateur  
puisse me trouver à la fois :

- Ποσειδῶν
- Ποσειδάωνι

# Les expressions régulières ?

Donc voilà ce qu'il doit voir :

Je sais que Ποσειδῶν pourra avoir  
comme terminaison Ποσειδάωνι :  
donc je veux que mon ordinateur  
puisse me trouver à la fois :

- Ποσειδῶν
- Ποσειδάωνι

# Les expressions régulières ?

Comme en mathématiques : à une inconnue correspond un signe (ou une lettre). Donc, si on cherche :

- Ποσειδῶν
- Ποσειδάωντι

il faudra quelque chose comme :

Ποσειδ????

où certains des ? seront optionnels.

# Les expressions régulières ?

Or, il se trouve qu'on peut représenter  
ce ? de plein de manières.

Par exemple, ce ? peut être un tiret,  
une lettre, un ", etc.

Et pour chaque inconnue, on va avoir  
des classes de caractères qui vont  
marcher pour les attraper globalement.

# Les expressions régulières ?

Quand une inconnue peut-être remplacée par plusieurs lettres, on va laisser l'ordinateur le choix entre ces lettres.

Exemple :

- Ποσειδῶν
- Ποσειδάωντι

Ποσειδ????

- > ? n°1 : ô ou á
- > ? n°2 : v ou ω
- > ? n°3 : Ø ou ν
- > ? n°4 : Ø ou τ

# Les expressions régulières ?

Eh bien, ça, on peut le faire avec les regex.

Quand on laisse le choix pour des caractères, on crée ce qu'on appelle une **classe**.

Quand on veut qu'un caractère soit nécessaire présent ou absent, ou les deux, on utilise les **quantificateurs**.

# Les expressions régulières ?

Quand on sait qu'on veut un caractère, ou un autre, ou encore un autre, on va proposer, pour chaque caractère inconnu, une classe de choix.

Exemple :

- Ποσειδων
  - Ποσειδάωνι
- Ποσειδ?
- > Ποσειδ[ωά]

# Les expressions régulières ?

Attention, si on ne spécifie pas combien de fois cette classe doit être représentée (c'est-à-dire combien de caractères cachés elle concerne), la classe ne vaut que pour UN caractère inconnu

# Les expressions régulières ?

Donc, théoriquement, si je cherche :

- Ποσειδών
- Ποσειδάω

Je pourrais écrire quelque chose  
comme ça :

Ποσειδ[ῶά][νω]

Mais dès qu'on a des mots longs ou  
plein de choix, ça devient très  
compliqué.

C'est là qu'interviennent les  
**quantificateurs.**

# Les expressions régulières ?

Un quantificateur permet de dire combien de fois on veut que la classe soit représentée. En gros, combien de fois on laisse à l'ordinateur la possibilité, avec les choix à disposition, de remplir des inconnues.

Les quantificateurs s'écrivent entre { }.  
Ils se mettent directement après la classe (et ne s'appliquent qu'à la classe qui les précède).

# Les expressions régulières ?

Ils peuvent s'écrire de deux manières.

- soit on met le nombre exact de fois que la classe est autorisée :

exemple : elle est autorisée 3 fois

--> {3}

- soit on met un minimum et un maximum de fois qu'elle est autorisée (séparés par une , ):

exemple : autorisée entre 1 et 6 fois

--> {1,6}

# Les expressions régulières ?

Donc, avant pour trouver

- Ποσειδῶν
- Ποσειδάω

On avait ça :

Ποσειδ[ῶά][νω]

Mais en fait on peut l'écrire comme ça :

Ποσειδ[ῶάνω]{2}

# Les expressions régulières ?

Pour récapituler :

- Une classe [ ] représente la boîte de choix des caractères disponibles pour l'ordinateur,
- Un quantificateur { } spécifie combien de fois ces caractères sont disponibles.

# Les expressions régulières ?

Donc, si je reviens à mon expression  
du début :

- Ποσειδῶν
  - Ποσειδάωντι
  - Ποσειδ????
- > Ποσειδ[ῶάων]<sup>ι</sup>{2,4}

# Les expressions régulières ?

Notons que si nous devons écrire explicitement tous les choix dans les classes, cela va vite être trop long.

C'est pour cela qu'existent les **intervalles** de classe.

# Les expressions régulières ?

Les **intervalles** de classe sont des classes de choix préconstruites.

Au lieu de dire explicitement "je veux tel ou tel caractère", vous pouvez dire "je veux toute lettre de a à ω". Ou "tout chiffre de 0 à 9".

On l'écrit comme ça :

[a-ω]

[0-9]

# Les expressions régulières ?

D'autre part, vous pouvez cumuler les intervalles dans une classe, ou ajouter des caractères qui ne seraient pas dans votre intervalle.

Exemple :

[a-ω0-9;!]

Ici on cherche 1 caractère qui peut être

- toute lettre grecque entre a et ω
- tout chiffre entre 0 et 9
- un ;
- un !

# Les expressions régulières ?

Rappel, si vous souhaitez que cette classe recouvre plusieurs inconnues, vous devez mettre un quantificateur.

Exemple :

[a-zA-Z0-9;!]

Ici on cherche 1 caractère

[a-zA-Z0-9;!]{6}

Ici on cherche 6 caractères

# Les expressions régulières ?

## PROBLÈME

Nous sommes en grec, et les caractères polytoniques ne sont pas inscrits par défaut dans les intervalles.

On va donc avoir recours à des intervalles préconçus spécifiquement pour les caractères unicode.

Ces classes préconstruites pour unicode s'écrivent comme ça :

\p{code}

# Les expressions régulières ?

\p{code}

- \p indique à l'ordinateur qu'on va utiliser une classe unicode
- pour {code}, code est la lettre qui va correspondre à la catégorie de caractère qu'on cherche.

Voici quelques codes de catégories :

- L --> \p{L} --> n'importe quelle lettre (y compris grecque polytonique)
- P --> \p{P} --> n'importe quelle ponctuation
- Braille --> \p{Braille} --> n'importe quel caractère en Braille

# Les expressions régulières ?

\p{code}

Voici un lien où vous trouverez pas mal  
de codes :

<https://www.regular-expressions.info/unicode.html>

# Les expressions régulières ?

Donc, pour reprendre la recherche  
précédente :

- Ποσειδῶν
  - Ποσειδάωντι
  - Ποσειδ????
- > Ποσειδ\p{L}{2,4}

# Les expressions régulières ?

Il existe d'autres classes préconstruites qui ne sont pas spécifiquement unicode.

Voilà les plus importantes :

- `.` --> n'importe quel caractère (y compris les espaces)
- `\d` --> n'importe quel chiffre
- `\s` --> n'importe quel espace
- `\n` --> n'importe quel saut de ligne
- `\w` --> n'importe quelle lettre  
(attention, ne fonctionne pas sur tous les caractères unicode)
- `\b` --> limite de mot (début ou fin de mot)

# Les expressions régulières ?

De même qu'il existe des classes préconstruites, il existe aussi des quantificateurs préconstruits. Voilà les plus importants :

- **\*** --> de 0 à infini (équivalent de {0, })
- **+** --> de 1 à infini (équivalent de {1, })
- **?** --> de 0 à 1 (équivalent de {0, 1})

# Les expressions régulières ?

Donc, si on reprend l'exemple précédent :

- Ποσειδῶν
  - Ποσειδάωντι
  - Ποσειδ????
- > Ποσειδ\p{L}+

# Les expressions régulières ?

## QUELQUES PIÈGES :

- quand un métacaractère est dans une classe, il devient littéral

exemple :

"." cherche tout caractère, [.] cherche explicitement le .

- quand un caractère littéral est précédé d'un \ , il peut devenir un métacaractère

exemple :

"d" cherche explicitement d, mais "\d" cherche un chiffre.

# / (reg)ex/

ν εωκε, τοι ουρανόδαν γαιήοχος ἀσκεχόλωται, δὸν ὁφθαλμολύφημον, δου κράτιώπεσσι· Θόωσα δέ με θυγάτηρ, ἀλὸς ἀτρυγγαλφυροῖσι **Ποσειδόν** θυμοσῆja **Ποσειδάνων** ἐπικτείνει, πλάζει δὲ ἡμεῖς οἶδε περιφριώς ἔλθησι· **Ποσειδόν** οὐ μὲν γάρ τι δυνάκεκτηι θεῶν ἔριδαινείβετ' ἔπειτα θεὰ ν  
nd: \bΠοσειδ\p{L}+\b

[^]\*?@[^]\*?\.[^]\*

ένη ξείνω, Ταφίων ἡγήτορι, Μέντρον' ἄρα μνηστῆρας ἀγήνορας· οἱ μὲν ἔπεισι προπάροιθε θυράων θυμὸν ἔτερον, οὐδὲν δέ τινοῖσι βοῶν, οὓς ἔκτανον αὐτοῖς· εἰς δὲ αὐτοῖσι καὶ δύτρηροι θεράποντες ν ἄρ' οἶνον ἔμισγον ἐνὶ κρητῆρσι καὶ αὔτε σπόργυοισι πολυτρήτοισι τραπέζας καὶ πρότιθεν, τοὶ δὲ κρέα πολλὰ δατεῖε πολὺ πρώτος ἦδε Τηλέμαχος θεοειδῆς· γὰρ ἐν μνηστῆρσι φίλον τετιημένος ἦτοι ενος πατέρ' ἐσθλὸν ἐνὶ φρεσίν, εἴ ποθιρῶν τῶν μὲν σκέδασιν κατὰ δώματα θεῖον δέ αὐτὸς ἔχοι καὶ κτίμασιν οἴσιν ἀνά-

Find: [ . ; ]

## Exercice sur les regex

<https://etudes.ens-lyon.fr/mod/quiz/view.php?id=166087>

# Exercice sur les regex

Reprenez l'interface de Transkribus et votre transcription.

Vous allez maintenant en faire un export en plain text (.txt).

# Exercice sur les regex

Reprenez l'interface de Transkribus et votre transcription.

Vous allez maintenant en faire un export en plain text (.txt).

# Exercice sur les regex

Transkribus®

demo &gt; odyssey &gt; #1

In Progress ▾

11.1.2024, 11:31

## Download

### Image

Available sizes

S

M

L

XL

### Text

Prima Page Content XML

[Download](#)

TXT

[Download](#)

Region 1

ΟΑΣΣΣΕΙΑΣ ΟΜΗ

Region 2

POY PAVDIAIAE.

άτερὶ οδεοπίδι καλλίδι εάρος.

μοι γῆνεπε Μοῦσα πολύ.

ρόπον, ὃς μᾶλα πολλὰ

θη, ἐπεὶ Τροίης ιερὸν τῆς

λιεθρον ἔπρσε.

λῶν δ' ἀνθρώπων ἵδον ἄ

σεα, καὶ νόον ἔχνω.

Region 3

τοντῷ πάθον ἄλγεα, ὃν κατὰ θυμὸν

ήν τε ψυχὴν, καὶ νόστον ἐταιρων.

ἐτάρους ερήμυταπίέμενός πορ,

φτόρησιν ἀτασθαλίησιν ὅλοντο

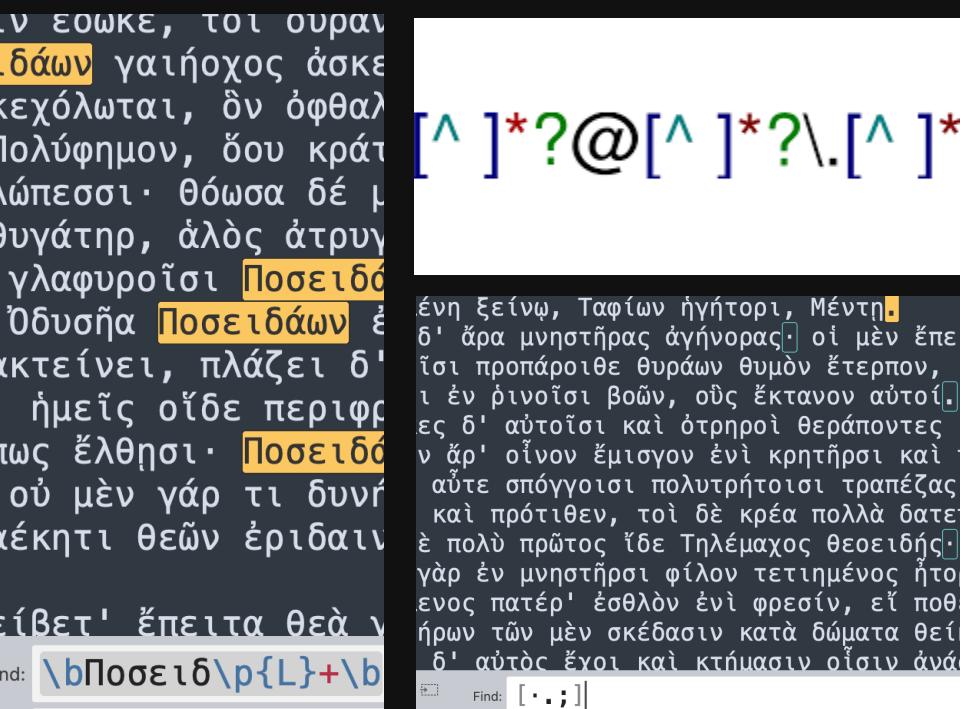
Νηπιτοι, οι κατὰ βοῦς ἐπερίονος ἀελίοιο

# Exercice sur les regex

[https://etudes.ens-  
lyon.fr/mod/quiz/view.php?id=166087](https://etudes.ens-lyon.fr/mod/quiz/view.php?id=166087)

/ (reg)ex/

# Pourquoi on a fait ça ?



# Pourquoi les regex pour les lettres classiques ?

- parce que la majorité des bases de données qu'on va voir ensuite peuvent être cherchables avec des regex
- parce que sans programmer vous pouvez déjà trouver des éléments génériques, comme le nombre de phrases, les noms propres (même avec erreurs), les paronymes, etc.

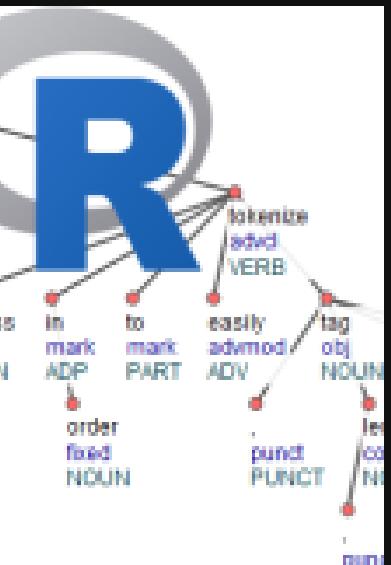
# Pourquoi ça n'est pas suffisant ?

- parce que ça nécessite une certaine maîtrise et une certaine logique
- parce que c'est un pis-aller de la recherche sémantique : vous ne pouvez pas, rien qu'avec ça, compter tous les verbes, tous les noms propres, etc.

# your Ancient Greek

and more information on this specific model here

paste your text here



[^]\*?@[^]\*?\\.[^]\*?



Un première  
représentation de  
vos textes

# D'abord les textes

Alors deux solutions :

- soit vos textes sont prêts, en sortie d'HTR, et presque propres
- soit vous pouvez reprendre le texte de l'*Odyssée* pour la démo.

# Deucalion et UDPipe

RDV [ici](#) pour une première analyse avec Deucalion, si vous analysez du grec, du latin ou du français de toute époque.

RDV [ici](#) pour une autre tentative, pour toutes les langues.

# Demo Deucalion

https://dh.charles.psl.eu/deucalion/api/grc/

token	case	degree	gend	lemma	mood	num	pers	pos	tense	voice	treated
Άνδρα	-	-	m	άνήρ	-	s	-	n	-	-	Άνδρα
μοι	-	-	m	ἐγώ	-	s	-	p	-	-	μοι
ἔννεπε	-	-	-	ἔνέπω	-	s	-	v	-	-	ἔννεπε
,	-	-	-	,	-	-	-	u	-	-	,
Μοῦσα	-	-	-	Μοῦσος	-	s	-	a	-	-	Μοῦσα
,	-	-	-	,	-	-	-	u	-	-	,
πολύτροπον	-	-	-	πολύτροπος	-	-	-	s	-	a	πολύτροπον
,	-	-	-	,	-	-	-	u	-	-	,
δς	-	-	-	δς	-	s	-	a	-	-	δς
μάλα	-	-	-	μάλα	-	-	-	a	-	-	μάλα
πολλὰ	-	-	-	πολύς	-	s	-	a	-	-	πολλὰ
πλάγχθη	-	-	-	πλάσσω	-	s	-	a	-	-	πλάγχθη
,	-	-	-	,	-	-	-	u	-	-	,
ἔπεὶ	-	-	-	ἔπει	-	-	-	n	-	-	ἔπεὶ
Τροίης	-	-	-	Τροία	-	s	-	n	-	-	Τροίης
ἱερὸν	-	-	-	ἱερός	-	s	-	n	-	-	ἱερὸν
πτολίεθρον	-	-	-	πτολίεθρον	-	-	-	s	-	n	πτολίεθρον
Ἐπερσε	-	-	-	πέρθω	-	s	-	v	-	-	Ἐπερσε
,	-	-	-	,	-	-	-	u	-	-	,
πολλῶν	-	-	m	πολύς	-	-	-	a	-	-	πολλῶν
δ	-	-	-	δέ	-	-	-	n	-	-	δ
,	-	-	-	,	-	-	-	u	-	-	,
ἀνθρώπων	-	-	-	ἄνθρωπος	-	-	-	n	-	-	ἀνθρώπων
ἴδεν	-	-	-	εἰδον	-	s	-	v	-	-	ἴδεν
ἄστεα	-	-	-	ἄστυ	-	s	-	n	-	-	ἄστεα
καὶ	-	-	-	καί	-	-	-	d	-	-	καὶ
νόον	-	-	-	νόος	-	s	-	n	-	-	νόον
ἔγνω	-	-	-	γιγνώσκω	-	s	-	v	-	-	ἔγνω
,	-	-	-	,	-	-	-	u	-	-	,
πολλὰ	-	-	-	πολύς	-	-	-	a	-	-	πολλὰ
δ	-	-	-	δέ	-	-	-	n	-	-	δ
,	-	-	-	,	-	-	-	u	-	-	,
ὅ	-	-	-	ὅς	-	s	-	p	-	-	ὅ

# Demo Deucalion

Enregistrez la page en faisant clic-droit + "Enregistrer sous", et changez le nom du fichier, en veillant bien à ce qu'il se termine par **.csv**

The screenshot shows a web browser window with the URL <https://dh.charthes.psl.eu/deucalion/api/grc/>. The page displays a table of Greek morphological analysis results. A download dialog box is open, prompting to save the file as "test\_odyssee.csv". The file explorer window shows the saved CSV file in the "Documents" folder of the user "mrebu01".

token	case	degree	gend	lemma	mood	num	pers	pos	tense	voice	treated
Ἄνδρα	-	-	m	ἀνήρ	-	s	-	n	-	-	Ἄνδρα
μοι	-	-	m	ἐγώ	-	s	-	p	-	-	μοι
Ἐννεπε	-	-	-	ἐνέπω	-	s	-	v	-	-	Ἐννεπε
,	-	-	-	,	-	-	-	u	-	-	,
Μοῦσα	-	-	-	Μοῦσος	-	s	-	-	-	-	Μοῦσα
,	-	-	-	,	-	-	-	-	-	-	,
πολύτροπον	-	-	-	πολύτροπος	-	-	-	-	-	-	πολύτροπον
,	-	-	-	,	-	-	-	-	-	-	,
δς	-	-	-	δς	-	s	-	-	-	-	δς
μάλα	-	-	-	μάλα	-	-	-	-	-	-	μάλα
πολλά	-	-	-	πολύς	-	s	-	-	-	-	πολλά
πλάγχθη	-	-	-	πλάσσω	-	s	-	-	-	-	πλάγχθη
,	-	-	-	,	-	-	-	-	-	-	,
ἐπεί	-	-	-	ἐπεί	-	-	-	-	-	-	ἐπεί
Τροίης	-	-	-	Τροία	-	s	-	-	-	-	Τροίης
ιερὸν	-	-	-	ιερός	-	s	-	-	-	-	ιερὸν
πτολίεθρον	-	-	-	πτολίεθρον	-	-	-	-	-	-	πτολίεθρον
Ἐπερσε	-	-	-	πέρθω	-	s	-	-	-	-	Ἐπερσε
,	-	-	-	,	-	-	-	-	-	-	,
πολλῶν	-	-	m	πολύς	-	-	-	-	-	-	πολλῶν
δ	-	-	-	δέ	-	-	-	-	-	-	δ

Gardez le fichier de côté pour l'instant.

# Demo UDPipe

The service is freely available for testing. Respect the CC BY-NC-SA licence of the models – explicit written permission of the **exploitation of the system**. If you use the service, you agree that data obtained by us during such use can be used for further comments and reactions are welcome.

**Model:**  UD 2.12 ([description](#))  UD 2.10 ([description](#))  UD 2.6 ([description](#))  PDT-C 1.0 ([description](#))

 ancient\_greek-perseus-ud-2.12-230717

**Actions:**  Tag and Lemmatize  Parse

**Advanced Options**

**UDPipe version:**  UDPipe 2  UDPipe 1

**Input [?]:**  Tokenize plain text  CoNLL-U  Horizontal  Vertical

**Tokenizer [?]:**  Normalize spaces  Presegmented input  Save token ranges

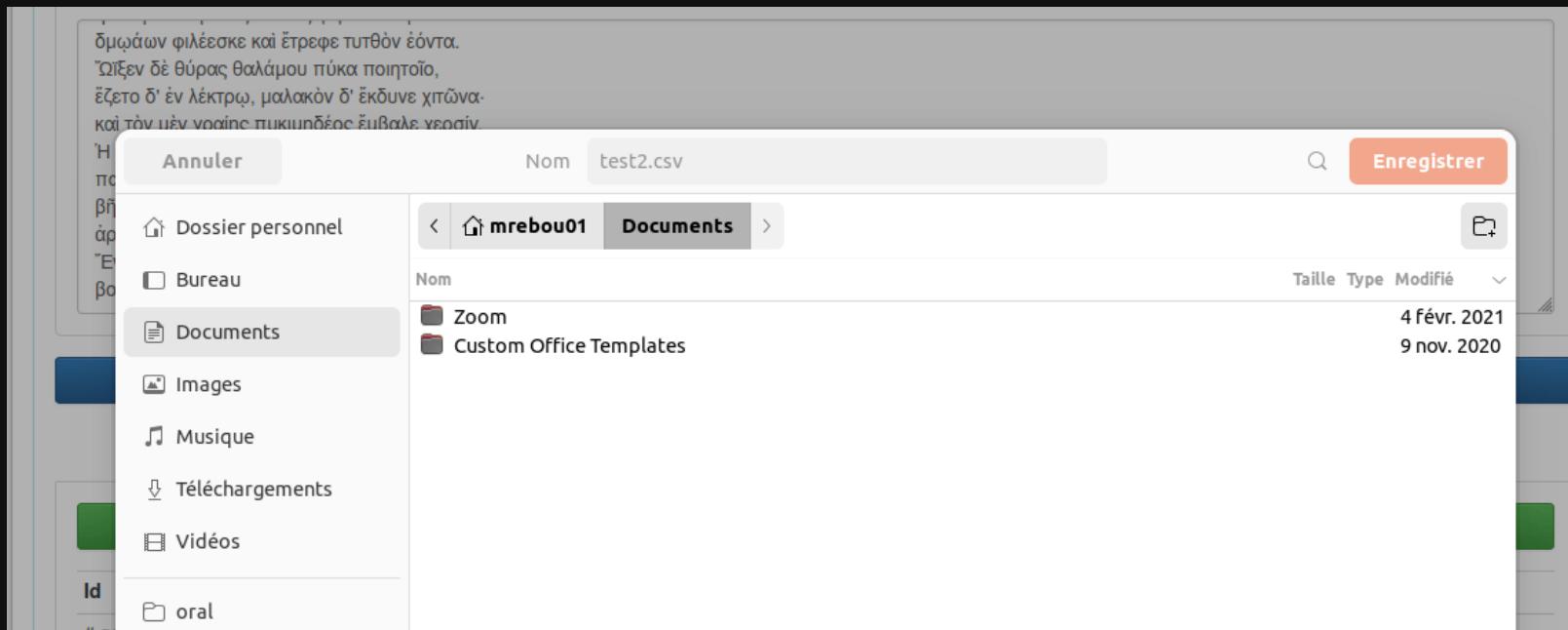
**A Input Text**

δμωάων φιλέεσκε καὶ ἔτρεφε τυτθὸν ἐόντα.  
"Ωἶξεν δὲ θύρας θαλάμου πύκα ποιητοῖο,  
ἔζετο δ' ἐν λέκτρῳ, μαλακὸν δ' ἔκδυνε χιτῶνα·  
καὶ τὸν μὲν γραῖς πυκιμπόεος ἔμβαλε χερσίν.  
Ἡ μὲν τὸν ππύξασα καὶ ἀσκήσασα χιτῶνα,  
πασσάλω ἀγκρεμάσσασα παρὰ τρητοῖσι λέχεσσιν,  
βῆ δὲ ἵνεν ἐκ θαλάμοιο, θύρην δ' ἐπέρυσσε κορώνῃ  
ἀργυρέῃ, ἐπὶ δὲ κληῆδ̄ ἐτάνυσσεν ἴμαντι.  
"Ενθ' ὅ γε παννύχιος, κεκαλυμμένος οἰός ἀώτῳ,  
βούλευε φρεσιν ἥσιν ὁδόν, τὴν πέφραδ' Αθήνη.

**Process Input**

# Demo UDPipe

Enregistrez le fichier avec "Save output file",  
et changez le nom du fichier, en veillant bien  
à ce qu'il se termine par **.csv**



Même chose, conservez le fichier.

# Préparer pour Voyant Tools

Normalement Voyant Tools permet de mettre directement un csv en import, mais souvent ça bugge. Donc on va plutôt copier les colonnes et les coller.

The screenshot shows a LibreOffice Calc spreadsheet titled "test\_odyssee.csv". The spreadsheet contains two columns of Greek words, numbered from 4066 to 4082. The first column (A) contains the words in their original form, and the second column (B) contains the same words with diacritics removed. A third column (C) contains the transliteration of the words. A fourth column (D) is highlighted in orange. An "Ajouter des textes" (Add texts) dialog box is open over the spreadsheet, containing a list of Greek characters and words, with a "Révéler" (Reveal) button at the bottom right.

	A	B	C	D
4066	παννύχιος	-	m	παννύ
4067	,	-	-	,
4068	κεκαλυμμένος	-	m	καλύπτ
4069	οίδις	-	m	δις
4070	άώτω	-	m	άωτον
4071	,	-	-	,
4072	βούλευε	-	-	βουλεύ
4073	φρεσίν	-	-	φρήν
4074	ῆσιν	-	f	ης
4075	օδόν	-	f	οδός
4076	,	-	-	,
4077	τὴν	-	f	τὸ
4078	πέφραδ	-	-	φράζω
4079	,	-	-	,
4080	Ἀθήνη	-	f	Ἀθήνη
4081	.	-	-	.
4082				

# Vous obtenez d'abord ceci :

Voyant Tools interface showing analysis results for the document "ανήρ εγώ ενέπια, Μουσος, πολύτροπος, ...".

**Visualizations:**

- Termes:** A word cloud visualization where words are colored by frequency.
- Tendances:** A line graph showing the relative frequency of terms across 10 segments of the document. The y-axis ranges from 0.0000 to 0.0060. The x-axis shows segments 1 to 10. The graph includes lines for τε (blue), ος (green), ο' (magenta), και (purple), and δε (cyan).
- Contextes:** A table showing collocations in the document. The columns are Document, Gauche, Terme, and Droite.

**Textual Analysis Summary:**

Ce corpus contient 1 document, 3,167 mots et 997 formes verbales uniques. Il a été créé il y a environ 5 minutes.

Densité du vocabulaire : 0.315

Readability Index: 4.140

Moyenne des mots par phrase : 73.7

Mots les plus fréquents :

- εγώ (121); οντός (93); και (76); ου (61); οντός (59)

En ce moment, VoyantTools a des problèmes avec le grec, donc si vous faites du grec, faites ça, sinon passez cette étape.

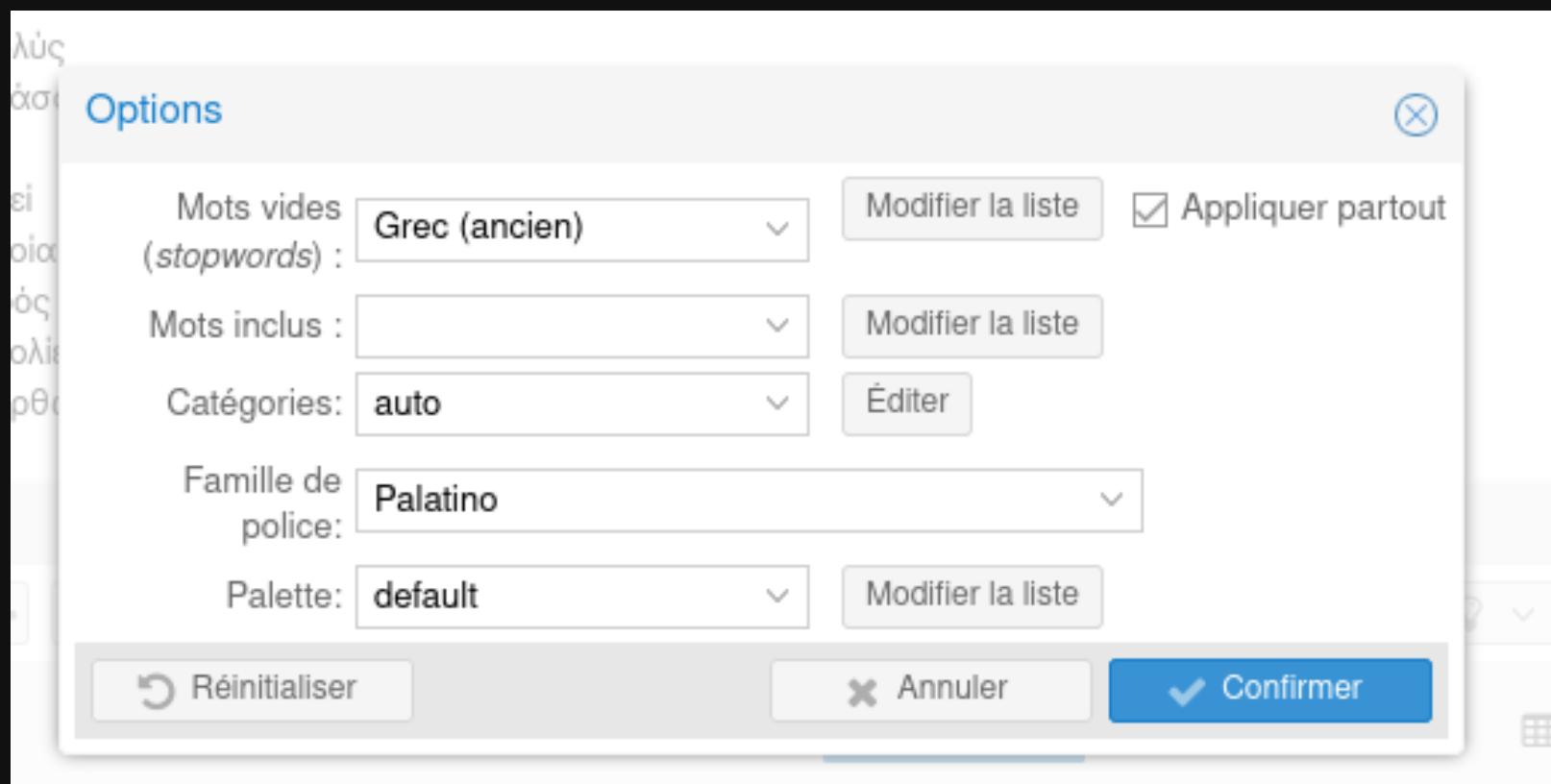
Copiez votre texte pareil, mais passez-le avant sur ce site pour enlever les accents :

<https://miniwebtool.com/remove-accent/>

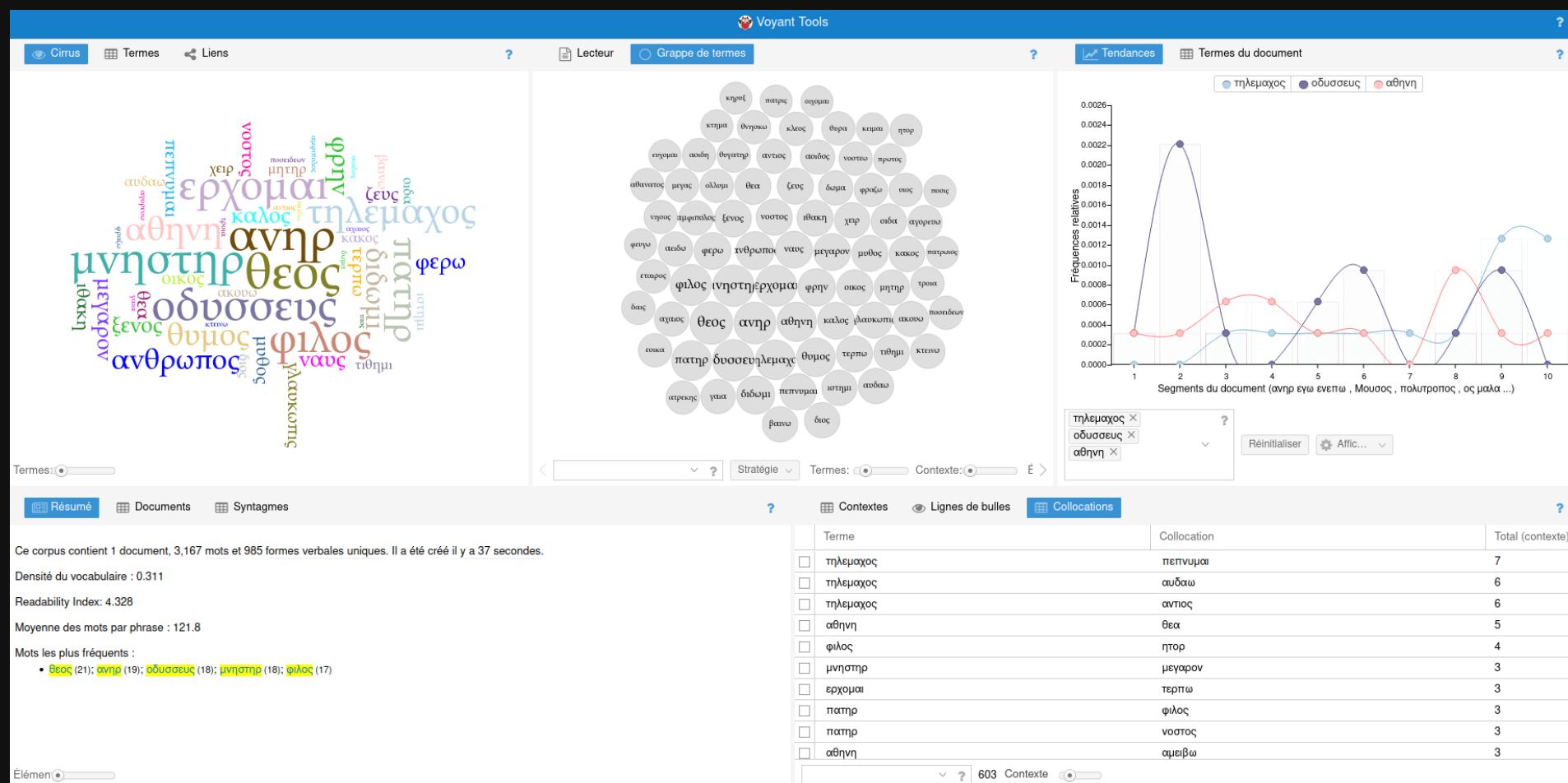
Même chose, vous aurez besoin d'une liste de mots-outils personnalisée [disponible ici](#).

Copiez-la et enlevez-lui les accents à la même adresse que précédemment.

# On modifie les stopwords sur chaque module



Cela fait, vous allez enfin pouvoir voir certains éléments



# À vous de voir quel lemmatiseur vous convient le mieux.

Ensuite, faites l'exercice suivant :

[https://etudes.ens-  
lyon.fr/mod/quiz/view.php?id=167798](https://etudes.ens-lyon.fr/mod/quiz/view.php?id=167798)

Normalement, à l'issue de ça, vous pourrez  
à minima vous débrouiller avec quelques  
textes.



Open Greek and Perseus Digital Library

# Scaife V

Browse Library    Text Search

or try the Iliad or Plato's Apology

2,414 works in 3,195 editions and transcriptions  
69.8 million words (32.2 million tokens)

 Scaife Viewer    Browse Library    Text Search

## Library

TEXT GROUPS AND WORKS

Find a text group or work... (e.g. Plato or Apology)

expand all | collapse all

- > Acta Barnabae
- > Acta Joannis
- > Acta Philippi
- > Acta Thomae
- > Adamantius Judaeus
- > Adamnan, Saint
- > Adrianus of Tyre
- > Aelian
- > Aelius Herodianus
- > Aelius Theon
- > Aeneas Tacticus
- > Aeschines
- > Aeschylus
- > Aesop



Διογένης

environment for reading and searching through texts in Latin and Greek. It consists of two connected projects, the Diogenes desktop application, which has been available for nearly 20 years, and the new DiogenesWeb webapp.

**Diogenes Desktop**

**DiogenesWeb**

# Les bases de données grecques

# Perseus Digital Library

<https://www.youtube.com/embed/WE4zg8TCMiQ?enablejsapi=1>

# Perseus Digital Library

<https://www.youtube.com/embed/WE4zg8TCMiQ?enablejsapi=1>

# Perseus Digital Library

<https://www.youtube.com/embed/MC5FYrsASyQ?enablejsapi=1>

# Perseus Digital Library

La recherche en bétacode,  
pas très pratique :  
**tentative ici**

## Autre tentative utile

Et surtout aucune recherche par lemme.

# Perseus Digital Library

Un petit exercice sur la Perseus Digital  
Library

[https://etudes.ens-  
lyon.fr/mod/quiz/view.php?id=167809](https://etudes.ens-lyon.fr/mod/quiz/view.php?id=167809)

# La nouvelle version : ScaifeViewer

[https://www.youtube.com/embed/HfUDdKLc\\_Gc?enablejsapi=1](https://www.youtube.com/embed/HfUDdKLc_Gc?enablejsapi=1)

# Les possibilités de recherche dans ScaifeViewer



Scaife Viewer   Browse Library   Text Search

## *Text Search*

### SEARCH GUIDE

- + signifies AND operation
- | signifies OR operation
- negates a single token
- " wraps a number of tokens to signify a phrase for searching

- \* at the end of a term signifies a prefix query
- ( and ) signify precedence
- num after a word signifies edit distance (fuzziness)
- num after a phrase signifies slop amount

Search...

Search Type:  Form  Lemma (Greek only)

# Les possibilités de recherche dans ScaifeViewer

Petit exercice avec ce moteur de recherche :

[https://etudes.ens-  
lyon.fr/mod/quiz/view.php?id=167822](https://etudes.ens-lyon.fr/mod/quiz/view.php?id=167822)

# Perseus Under Philologic

GREEK Arist. Eth. Nic. 1094b

CONCORDANCE KWIC COLLOCATES

? πόλις, lemma:πόλις, pos:n--s--fa-, urn:cts:greekLit:tlg0012.tlg001

Clear Show search

Aeschylus **Agamemnon**  
[Aesch. Ag.]

◀ 1 ▶ English

Click on a word for parses and a link to Logeion.

1 θεοὺς μὲν αἰτώ τῶνδ' ἀπαλλαγὴν πόνων  
φρουράς ἐτείσας μῆκος, ἦν κοιμώμενος  
στέγαις Ἀτρειδῶν ἄγκαθεν, κυνὸς δύκην,  
ἄστρων κάποια νυκτέρων ὄμήγηριν,

# Perseus Under Philologic

PERSEUS HOME PHILologic4

GREEK Arist. Eth. Nic. 1094b

CONCORDANCE Kwic

Search Terms:

Refine your search with the following options and fields:

Search Terms  Approximate match

Search Terms  Lemma match

Search Terms  Within  
 Exactly  
 In the same sentence

words in the same sentence  
words in the same sentence

Author:  (e.g., Xenophon)

Title:  (e.g., Anabasis)

Genre:  (e.g., prose)

Language:  (e.g., Greek)

Speaker:  (e.g., "Οδυσσεύς")

Cts\_urn:  (e.g., "urn:cts:greekLit:tlg0532.tlg001")

Sort results by

Results per page:

# Perseus Under Philologic

Exercice sur requêtes simples :

[https://etudes.ens-  
lyon.fr/mod/quiz/view.php?id=166080](https://etudes.ens-lyon.fr/mod/quiz/view.php?id=166080)

# Perseus Under Philologic

- Look for words by part of speech: use "pos:" e.g., pos:v-sppafn-, present active participle feminine nominative singular. Or, pos:v-..p..f.., feminine participles. See the bottom of this page for a quick overview (or look in word facets what code 'masculine accusative singular' produces)
- A quick overview of part-of-speech codes:
  - slot 1: major part of speech: verb, noun, adjective, pronoun, adverb, conjunction, preposition, g= particle,..
  - slot 2: minor part of speech: ne proper name (also: ae, de for proper-name adjectives and adverbs based on proper names), ps possessive pronoun, gm modal particle, pa definite article. Further examples: .c =copula .d = demonstrative, .i=interrogative, .r = relative, .s= possessive, .x=indefinite.
  - slot 3: person (1, 2, 3)
  - slot 4: number (s, p, dual)
  - slot 5: tense (p, imperfect, aorist, future, perfect, pluperfect, future perfect)
  - slot 6: mood (i, s, o, imperative, participle, infinitive)
  - slot 7: voice (a, m, p, e: middle-passive)
  - slot 8: gender (m, f, n, c: common gender)
  - slot 9: case (n, g, d, a, v)
  - slot 10: degree (c, s)

# TLG : Thesaurus Linguae Graecae

Vous devez passer par le proxy de la bibliothèque (qui fonctionne aléatoirement)

Vous avez le lien [ici](#)

# TLG : Thesaurus Linguae Graecae

Cela fait, vous devez vous créer un login.

# TLG : Thesaurus Linguae Graecae

Si tout fonctionne, vous devez voir quelque chose comme ça une fois loggés :

The screenshot shows the 'STATISTICS' page of the TLG website. At the top, there are search filters: 'From 8 B.C.' dropdown, 'to A.D. 20' dropdown, an unchecked 'Exclude Long Ranges' checkbox, and a checked 'Include Varia and Incerta' checkbox. Below the filters, the 'SUMMARY' section indicates there are 10,985 documents in the corpus, totaling 124,322,117 words. To the right, a chart titled 'Authors by century' (Based on TLG data) shows the distribution of authors across centuries. The chart has four bars: the first bar is light blue and the last three are dark blue. The x-axis is labeled '400'. The chart area includes a legend with Greek words: ἄπος, τρόπος, βιούλομε, γυνή, ἔρχομαι, μάλλον, καλός, δύναμις, εἶδος, γράφω, ήμέρα, μέγας.

[My Account \(inst\)](#) Logout

Access provided by:  
Ecole Normale Supérieure de Lyon