

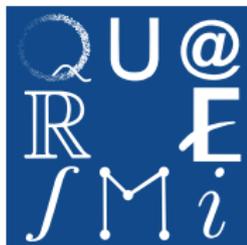
Réunion du groupe de contact FNRS

« Analyse critique et amélioration de la qualité de l'information numérique »

Méthodes quantitatives, analyse critique de données et histoire médiévale : des données contemporaines aux sources anciennes

Bruxelles, le 30 janvier 2019

Sébastien de Valeriola



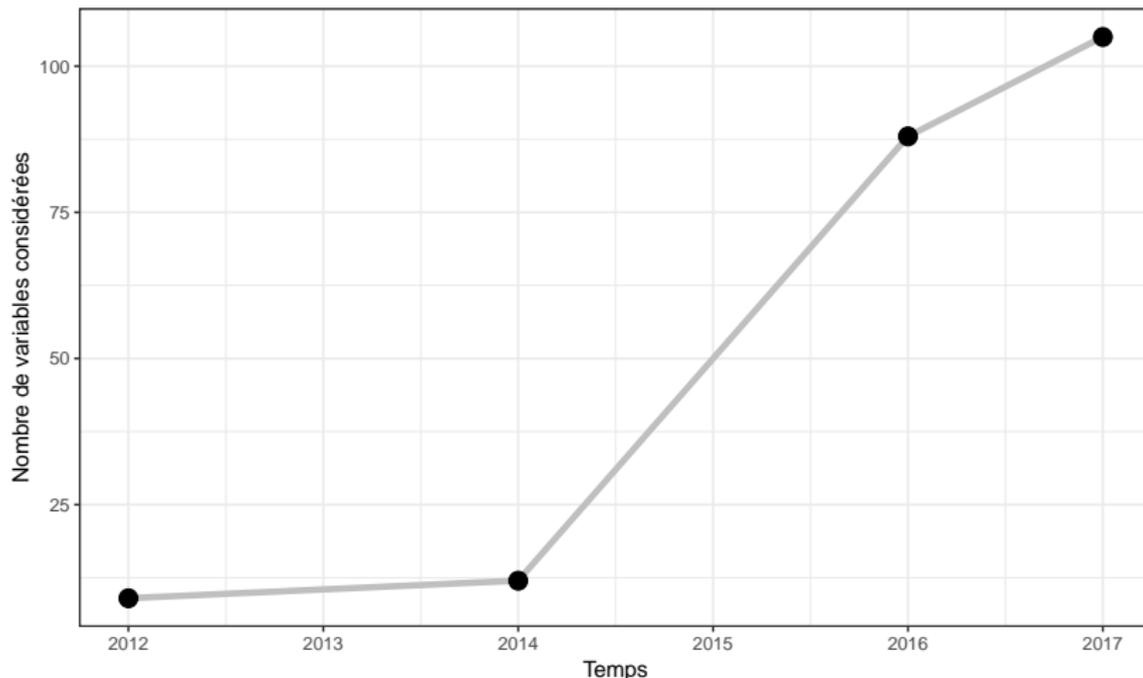
- 1 Introduction
- 2 Identification automatique des sources d'une encyclopédie médiévale
 - Contexte
 - Pré-traitement
 - Éléments d'analyse critique
- 3 Réseaux de crédit
 - Contexte
 - Pré-traitement
 - Éléments d'analyse critique
- 4 Analyse automatique des clauses des dettes
 - Contexte
 - Pré-traitement
 - Éléments d'analyse critique
- 5 Conclusion

Introduction

Une évolution très sensible

- La modélisation dans l'industrie assurancière belge a connu une évolution très sensible au cours des dernières années :

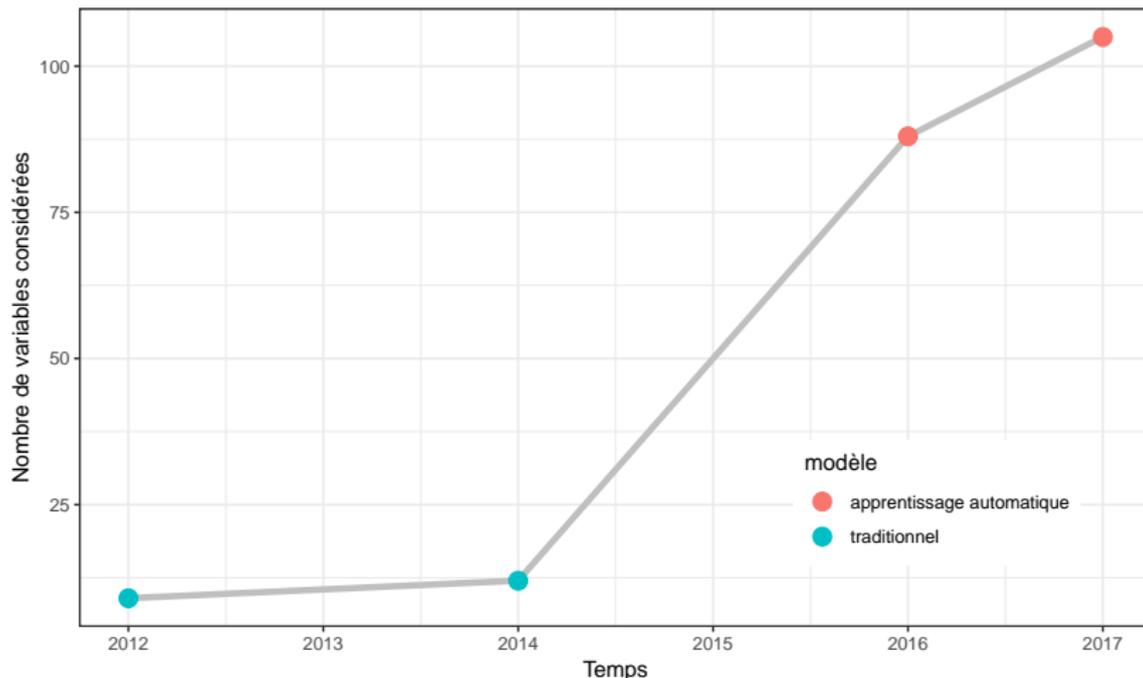
Évolution d'un modèle de tarification actuarielle



Une évolution très sensible

- La modélisation dans l'industrie assurancière belge a connu une évolution très sensible au cours des dernières années :

Évolution d'un modèle de tarification actuarielle



Trois dimensions de la modélisation

- Cette évolution concerne trois dimensions de la modélisation :
 - 1 les modèles utilisés ([ré]apparition des modèles *data-driven*) ;
 - 2 les données utilisées (quantité et complexité) ;
 - 3 les machines utilisées (puissance de calcul).

Trois dimensions de la modélisation

- Cette évolution concerne trois dimensions de la modélisation :
 - 1 les modèles utilisés ([ré]apparition des modèles *data-driven*) ;
 - 2 les données utilisées (quantité et complexité) ;
 - 3 les machines utilisées (puissance de calcul).
- Parmi ces modèles, on compte les modèles d'apprentissage automatique (*machine learning*), qui sont aujourd'hui à la mode :
 - leur philosophie est de laisser l'ordinateur « apprendre par lui-même » ;
 - ils font moins d'hypothèses (notamment sur la distribution des données) que les modèles « traditionnels » ;
 - ils permettent de prendre plus facilement en compte des données « complexes ».

Trois dimensions de la modélisation

- Cette évolution concerne trois dimensions de la modélisation :
 - 1 les modèles utilisés ([ré]apparition des modèles *data-driven*) ;
 - 2 les données utilisées (quantité et complexité) ;
 - 3 les machines utilisées (puissance de calcul).
- Parmi ces modèles, on compte les modèles d'apprentissage automatique (*machine learning*), qui sont aujourd'hui à la mode :
 - leur philosophie est de laisser l'ordinateur « apprendre par lui-même » ;
 - ils font moins d'hypothèses (notamment sur la distribution des données) que les modèles « traditionnels » ;
 - ils permettent de prendre plus facilement en compte des données « complexes ».
- Néanmoins, un *caveat* important doit être fait ici : la majorité de ces modèles...
 - ne sont pas tout neufs ;
 - ne se passent pas complètement d'hypothèses ;
 - ne sont pas miraculeux ;
 - ne rendent pas les modèles « traditionnels » obsolètes

(en tout cas dans le contexte considéré ici).

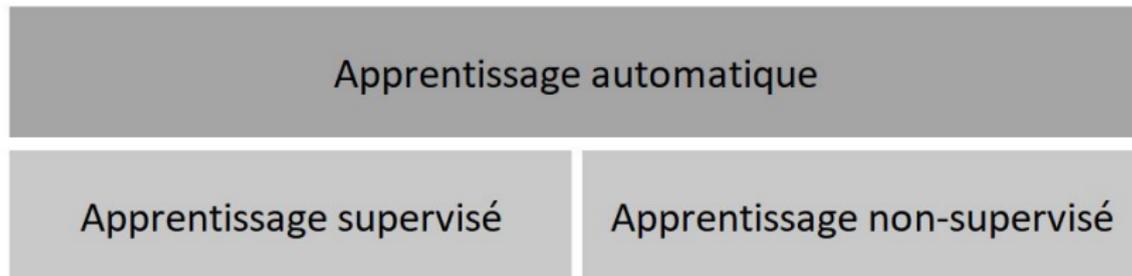
L'apprentissage automatique

- On regroupe généralement les méthodes d'apprentissage automatique en deux grandes familles :

Apprentissage automatique

L'apprentissage automatique

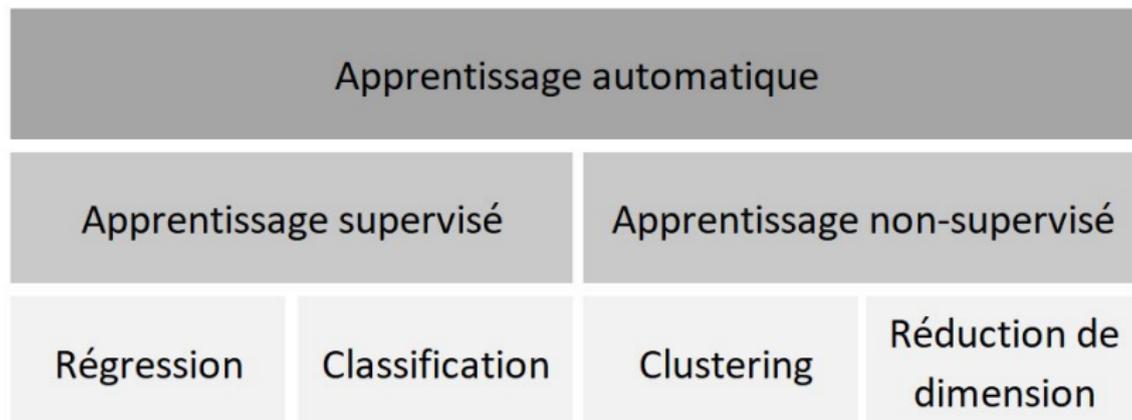
- On regroupe généralement les méthodes d'apprentissage automatique en deux grandes familles :



- 1 l'apprentissage supervisé (on effectue les calculs sur une base de données pour lesquelles on connaît la « réponse ») ;
- 2 l'apprentissage non-supervisé (on effectue les calculs sur base de données « toutes nues »).

L'apprentissage automatique

- On regroupe généralement les méthodes d'apprentissage automatique en deux grandes familles :



- 1 l'apprentissage supervisé (on effectue les calculs sur une base de données pour lesquelles on connaît la « réponse ») ;
- 2 l'apprentissage non-supervisé (on effectue les calculs sur base de données « toutes nues »).

- Cette évolution possède un corollaire important :

- Cette évolution possède un corollaire important :

La qualité des données est plus importante que jamais !

- Cette évolution possède un corollaire important :

La qualité des données est plus importante que jamais !

Work With New Electronic 'Brains' Opens Field for Army Math Experts

WASHINGTON — BIZMAC
UNIVAC, GARBAGE IN-GAR-
BAGE OUT—all new terms in
the Army

They haven't replaced GOV-
ERNMENT ISSUE or BY THE
NUMBERS and probably won't,
but in a small, select circle
they're heard more frequently
than the old expressions used
by millions of veterans of World
War II and Korea



The Hammond Times, 1957, p. 65

- Cette évolution possède un corollaire important :

La qualité des données est plus importante que jamais !

Work With New Electronic 'Brains' Opens Field for Army Math Experts

WASHINGTON — BIZMAC
UNIVAC, GARBAGE IN-GAR-
BAGE OUT—all new terms in
the Army

They haven't replaced GOV-
ERNMENT ISSUE or BY THE
NUMBERS and probably won't,
but in a small, select circle
they're heard more frequently
than the old expressions used
by millions of veterans of World
War II and Korea



The Hammond Times, 1957, p. 65

- Cela signifie donc :
 - les opérations de traitement des données sont devenues centrales ;
 - il est encore plus important d'être critique à l'égard des données et des résultats obtenus.

- Le domaine des sciences actuarielles n'est bien entendu pas le seul à avoir subi ces transformations.
- Il en va de même des sciences humaines en général, et de l'histoire en particulier :

- Le domaine des sciences actuarielles n'est bien entendu pas le seul à avoir subi ces transformations.
- Il en va de même des sciences humaines en général, et de l'histoire en particulier :
 - la quantité de données disponibles sous forme numérique a explosé ;
 - certaines de ces données peuvent être considérées comme des données complexes ;
 - les outils de traitement de ces données sont devenus très facilement accessibles.

- Le domaine des sciences actuarielles n'est bien entendu pas le seul à avoir subi ces transformations.
- Il en va de même des sciences humaines en général, et de l'histoire en particulier :
 - la quantité de données disponibles sous forme numérique a explosé ;
 - certaines de ces données peuvent être considérées comme des données complexes ;
 - les outils de traitement de ces données sont devenus très facilement accessibles.
- Dans la suite de cet exposé, je vais vous présenter trois questions de recherche sur lesquelles je travaille, et pour lesquelles il m'a été donné de manipuler de grandes quantités de données :
 - contexte ;
 - pré-traitement ;
 - éléments d'analyse critique.

- Le domaine des sciences actuarielles n'est bien entendu pas le seul à avoir subi ces transformations.
- Il en va de même des sciences humaines en général, et de l'histoire en particulier :
 - la quantité de données disponibles sous forme numérique a explosé ;
 - certaines de ces données peuvent être considérées comme des données complexes ;
 - les outils de traitement de ces données sont devenus très facilement accessibles.
- Dans la suite de cet exposé, je vais vous présenter trois questions de recherche sur lesquelles je travaille, et pour lesquelles il m'a été donné de manipuler de grandes quantités de données :
 - contexte ;
 - pré-traitement ;
 - éléments d'analyse critique.
- Je ne parlerai que de données textuelles, mais il faut noter qu'il est aussi possible de travailler sur d'autres types de données (notamment des images, par exemple des photographies de chartes).

Identification automatique des sources d'une encyclopédie médiévale

Contexte

Le problème de l'identification automatique

- Les XII et XIII^e siècles forment ce qu'on appelle parfois l'« âge d'or » des encyclopédies médiévales : c'est une période très féconde pour la littérature didactique.
- De nombreux auteurs de cette période compilent l'ensemble des connaissances de leur époque sous la forme d'encyclopédies, dans des domaines extrêmement variés.

Le problème de l'identification automatique

- Les XII et XIII^e siècles forment ce qu'on appelle parfois l'« âge d'or » des encyclopédies médiévales : c'est une période très féconde pour la littérature didactique.
- De nombreux auteurs de cette période compilent l'ensemble des connaissances de leur époque sous la forme d'encyclopédies, dans des domaines extrêmement variés.
- Les renseignements utilisés proviennent d'une multitude de sources différentes, chrétiennes et païennes, antiques et médiévales ; c'est à cette époque qu'arrivent en Occident beaucoup de traductions latines de textes grecs ou arabes.
- Une question naturelle se pose : celle de l'identification des sources utilisées par les encyclopédistes médiévaux.

Le problème de l'identification automatique

- Les XII et XIII^e siècles forment ce qu'on appelle parfois l'« âge d'or » des encyclopédies médiévales : c'est une période très féconde pour la littérature didactique.
- De nombreux auteurs de cette période compilent l'ensemble des connaissances de leur époque sous la forme d'encyclopédies, dans des domaines extrêmement variés.
- Les renseignements utilisés proviennent d'une multitude de sources différentes, chrétiennes et païennes, antiques et médiévales ; c'est à cette époque qu'arrivent en Occident beaucoup de traductions latines de textes grecs ou arabes.
- Une question naturelle se pose : celle de l'identification des sources utilisées par les encyclopédistes médiévaux.
- Certaines des encyclopédies qui résultent de ces initiatives sont très imposantes, avec plusieurs dizaines de milliers de « notices ».
- Avec Céline Engelbeen et Étienne Cuvelier, nous nous sommes donc demandés s'il était possible d'identifier **automatiquement** les sources utilisées.

Vincent de Beauvais et le *Speculum maius*

- Parmi les encyclopédistes médiévaux les plus influents, une triade d'auteurs du XIII^e siècle se dégage : le dominicain Thomas de Cantimpré, le franciscain Barthélémy l'Anglais et le dominicain Vincent de Beauvais.

Vincent de Beauvais et le *Speculum maius*

- Parmi les encyclopédistes médiévaux les plus influents, une triade d'auteurs du XIII^e siècle se dégage : le dominicain Thomas de Cantimpré, le franciscain Barthélémy l'Anglais et le dominicain Vincent de Beauvais.
- Nous avons choisi de travailler sur l'œuvre de ce dernier, qui est l'auteur du *Speculum maius* : avec ses 4 500 000 mots, c'est la plus considérable encyclopédie du monde occidental jusqu'au milieu du XVIII^e siècle.

Vincent de Beauvais et le *Speculum maius*

- Parmi les encyclopédistes médiévaux les plus influents, une triade d'auteurs du XIII^e siècle se dégage : le dominicain Thomas de Cantimpré, le franciscain Barthélémy l'Anglais et le dominicain Vincent de Beauvais.
- Nous avons choisi de travailler sur l'œuvre de ce dernier, qui est l'auteur du *Speculum maius* : avec ses 4 500 000 mots, c'est la plus considérable encyclopédie du monde occidental jusqu'au milieu du XVIII^e siècle.
- Plus précisément, nous avons sélectionné les livres zoologiques d'une des sous-parties de cette œuvre, le *Speculum naturale* :

16 : Oiseaux (171 ch.)

18 : Animaux domestiques (98 ch.)

20 : Reptiles et insectes (179 ch.)

22 : Activités des animaux (68 ch.)

17 : Monde marin (146 ch.)

19 : Animaux sauvages (139 ch.)

21 : Anatomie des animaux (66 ch.)

Vincent de Beauvais et le *Speculum maius*

- Parmi les encyclopédistes médiévaux les plus influents, une triade d'auteurs du XIII^e siècle se dégage : le dominicain Thomas de Cantimpré, le franciscain Barthélémy l'Anglais et le dominicain Vincent de Beauvais.
- Nous avons choisi de travailler sur l'œuvre de ce dernier, qui est l'auteur du *Speculum maius* : avec ses 4 500 000 mots, c'est la plus considérable encyclopédie du monde occidental jusqu'au milieu du XVIII^e siècle.
- Plus précisément, nous avons sélectionné les livres zoologiques d'une des sous-parties de cette œuvre, le *Speculum naturale* :

16 : Oiseaux (171 ch.)

17 : Monde marin (146 ch.)

18 : Animaux domestiques (98 ch.)

19 : Animaux sauvages (139 ch.)

20 : Reptiles et insectes (179 ch.)

21 : Anatomie des animaux (66 ch.)

22 : Activités des animaux (68 ch.)

- Ce choix est motivé par la façon dont il compose son ouvrage : il recopie assez fidèlement ses sources plutôt que de les paraphraser, compléter ou corriger comme d'autres encyclopédistes.

- Vincent de Beauvais a indiqué par des marqueurs quelles sont ses sources :

De bestis **VI** **De bestis**

Languinē sicut Variā q̄ dolores se-
dat cuius eisdem simi cū bellis et
oculis. addito melle modico et omnia
in flammata et supputatōes. Dodagē
q̄ p̄dest et articularis morbz emari-
bz p̄cipue utuli. **De bestis ysi.**



Bestiarū **conu.**
uocabulū. ipe
conuenit leoni-
bz. pardis ty-
grulz. lupis ac
uulpibz. canis
et similibz. ac ce-
teris q̄ ul' ore
uel unguibz seuiunt exceptis ser-
pentibz. **B**estie autē dicte sūt. au-
qua seuiunt. Ecce ū sunt appellate
co q̄ desiderio suo frantur natura

meuola. Quia ip̄ oportet ut curentur
morbus eoz et uulnus unguinū eoz i
p̄mis cū attractione. deinde cū cōsol-
datione. **S**i meoz quidē attractione
sufficit res parua. Sutor bestiaz ue-
neuosus ē et ex ei' potu fit facies uir-
dis et apostematatur et curat ex cor-
pore sutor fecidus et ex astellis. **C**ura
ei' est uomitus cū aqua tepida. et ty-
riaca. et cū sigillata. **haly.** **V**rina ū
bestiaz ardetas ē remedium. **actoz.**
Quedam autē sunt aialia: que q̄
mediū locū int' bestias et pecora trāe
ul' ad utriūq̄ genus cōmunit' p̄tine-
uerūt. Nam uerbi gr̄a cōstat apriū
maxime dentibz seuire. quem tamē
iuxta ysidorū int' pecora cōputauim.
quia hominibz scitus est et ueluti

- Vincent de Beauvais a indiqué par des marqueurs quelles sont ses sources :

li utentes libertate. Libere ei huc
illuc q̄ uagant. ⁊ quo animus dix
it eo feruntur. **aristoteles.** Pug
na bestiarū nō est. nisi p̄pter cibum
et p̄ mansionē et p̄ feminam. **ex**
libro de natura rerū. Sunt in aīa
lib; bella. uij. de causis. Videtur p̄
naturalem superbiam. p̄t̄ abū. p̄ co
ritū. et p̄ fetū saluet̄ reſtendūm.
Augustina in iij. canone. Bestia
rū quedā morſu grauant̄ etiam si
ue rabie q̄ uenenum in dentibus
hūt. ut canis et cocodrillus ⁊ hī
Quedam ⁊ in rabiem incidunt ut
canis. Si etiam qd̄ iulpes rabiosa
ſit. et mustela ſimilit̄. **ouli** q̄ ſcđm
quosdam rabiosi ſcđ sunt. ⁊ caninos
ſuos momorderūt et manūm que
et reliquis rabie; accidit in turce
runt. **Leo** pardus aut̄ ⁊ leo ſiſa ſipa
tes et eis ſimiles. nō ſunt ſicut ca
nes ſalu et hoīes. unū dentes cor

quid dicitur. apud eu ad uentem
Ougem q̄ plinius feram uocat et
ieronimus. que tam̄ teuto nomio
liber ut animalia munda et ad uelā
dū apta computat. **Plura** q̄ talia re
pūnt̄ aīalia. Vide plinius. ut tam̄
est ſupra. **Dam** me inquit caro mā
ſuelant. ⁊ tam̄ fere dia uox nō poſ
ſunt. **Complura** namq; ſūt nec pla
cda nec fera. ſ; medie int̄ utrūq; na
ture. ut in campo apri in mari delphi
ni. **Dico** ḡ de pecorib; iuxta ordinem
lit̄arū alphabeti. cōſequent̄ de beſtiis
iuxta eūdem ordinem p̄ ſequi teo ſi
uāte intendimus. **ii. de abant ⁊ al
che ⁊ aloy. ex libro de natura rerū.**

Ahane ſcđm aristotelem est ani
mal magnitudine cerui. cōt̄
conſuetudinē omnū animalium
habens fel in auce. ſimile felle hoīs
in colore. Et est amax ualre. cauſa
tūū uehementis ut ferocitatis q̄
belnandus lib. xxii. **A**ches ē

- Ces marqueurs médiévaux nous permettent de faire appel à une méthodologie qui rappelle l'apprentissage supervisé.

Apprentissage supervisé

- Ces marqueurs médiévaux nous permettent de faire appel à une méthodologie qui rappelle l'apprentissage supervisé.
- Nous développons des techniques permettant de relier les notices de Vincent à leurs sources sans faire appel à ces marqueurs.
- Nous pouvons ensuite « vérifier » quelles sont les attributions correctes et quelles sont les attributions erronées.

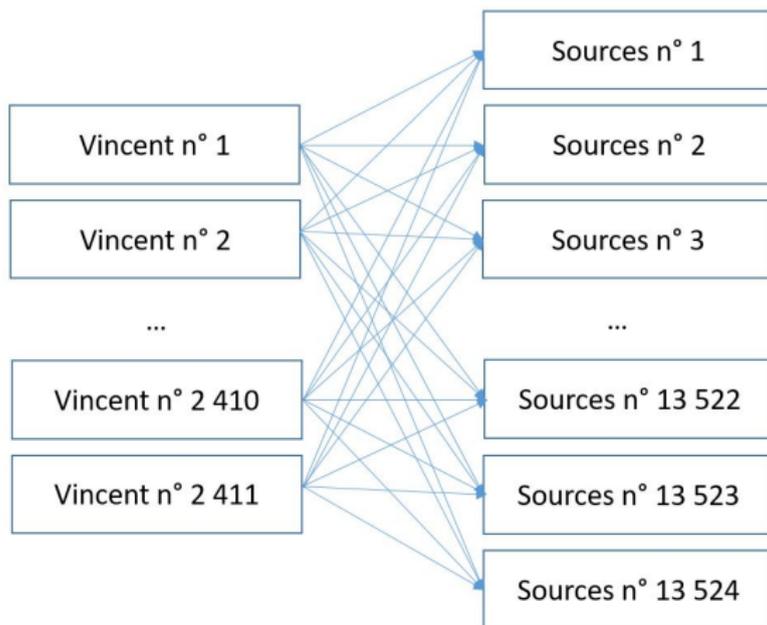
- Ces marqueurs médiévaux nous permettent de faire appel à une méthodologie qui rappelle l'apprentissage supervisé.
- Nous développons des techniques permettant de relier les notices de Vincent à leurs sources sans faire appel à ces marqueurs.
- Nous pouvons ensuite « vérifier » quelles sont les attributions correctes et quelles sont les attributions erronées.
- On peut alors évaluer l'efficacité d'une technique d'attribution de sources en calculant le pourcentage de notices correctement attribuées et donc comprendre si une technique est meilleure que l'autre → progrès successifs.

- Ces marqueurs médiévaux nous permettent de faire appel à une méthodologie qui rappelle l'apprentissage supervisé.
- Nous développons des techniques permettant de relier les notices de Vincent à leurs sources sans faire appel à ces marqueurs.
- Nous pouvons ensuite « vérifier » quelles sont les attributions correctes et quelles sont les attributions erronées.
- On peut alors évaluer l'efficacité d'une technique d'attribution de sources en calculant le pourcentage de notices correctement attribuées et donc comprendre si une technique est meilleure que l'autre → progrès successifs.
- Mais :
 - Les marqueurs de Vincent ne sont pas terriblement précis (le plus souvent, seulement l'auteur ; dans le cas de Pline, généralement le numéro du livre).
 - Vincent mentionne parfois des sources fantaisistes.

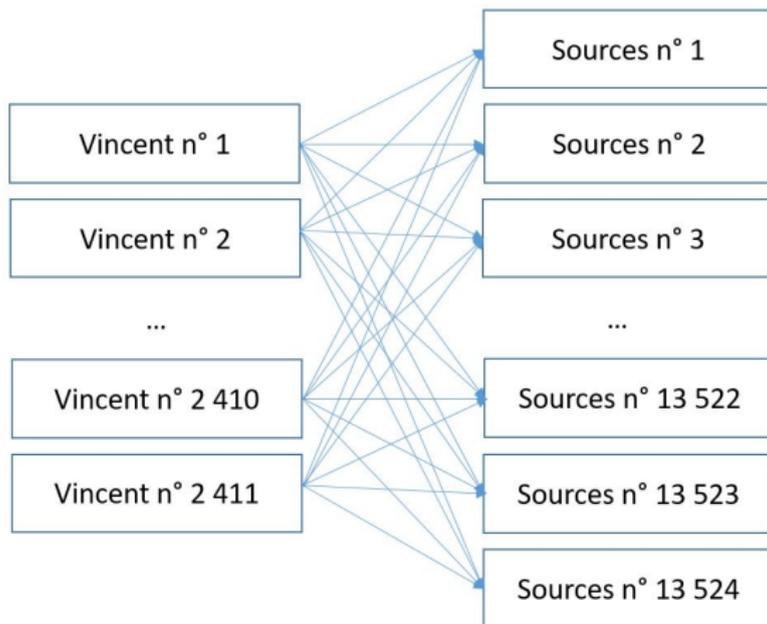
Sources retenues

auteur	total	auteur	total	auteur	total
Plinius	979	Pythagoras	10	Rufus	2
Liber de Nat. rer.	405	Radulphus	10	Strabus	2
Aristoteles	376	Andromachus	7	Valerius Maximus	2
Isidorus	309	Comestor	8	Zeno	2
[section acéphale]	146	Helinandus	7	Sacra Historia	1
Auctor	182	Glossa [Aristote]	6	Aemilius	1
Ambrosius	113	Gregorius	6	Plautus	1
Avicenna	208	Vegetius	6	Liber empyricus	1
Solinus	91	Hieronymus	5	Cicero	1
Physiologus	86	Papias	5	Suetonius	1
Glossa [Bible]	64	Physicus	5	Guillelm. de Conchis	1
Dioscorides	60	Achildes	4	Basilius	1
Palladius	56	Belbetus	4	Serapion	1
Isaac	47	Platearius	4	Bali.	1
Hali	44	Seneca	4	Rabanus	1
Iorath	40	Virgilius	4	Iuvenalis	1
Aesculapius	38	Aviarium	3	Polybius	1
Razi	17	Ovidius	3	Algazel	1
Philosophus	13	Albertus	2	Orosius	1
Alexander	19	Fulgentius	2	Origenes	1
Constantinus	15	Horatius	2	Gregorius Nissenus	1
Lucanus	15	Hesychius	2	Democritus	1
Augustinus	10	Martialis	2		

- L'idée de base de la méthodologie est de comparer chaque notice de Vincent avec chaque notice-source potentielle :



- L'idée de base de la méthodologie est de comparer chaque notice de Vincent avec chaque notice-source potentielle :



- Cela fait $2\,411 \cdot 13\,524 = 32\,606\,364$ couples à tester...

- Il s'agit donc de comparer les deux notices de chacun de ces couples (par exemple entre une notice de Vincent et une notice de Pline) :

- Il s'agit donc de comparer les deux notices de chacun de ces couples (par exemple entre une notice de Vincent et une notice de Plin) :
 - 1 on calcule les **signatures** de deux notices :
 - a. on enregistre les mots qui constituent la notice de Vincent ;
 - b. on calcule la distance entre ces mots dans la notice de Vincent : ceci nous donne un vecteur, la *signature* de la notice de Vincent ;

- Il s'agit donc de comparer les deux notices de chacun de ces couples (par exemple entre une notice de Vincent et une notice de Pline) :
 - 1 on calcule les **signatures** de deux notices :
 - a. on enregistre les mots qui constituent la notice de Vincent ;
 - b. on calcule la distance entre ces mots dans la notice de Vincent : ceci nous donne un vecteur, la *signature* de la notice de Vincent ;
 - c. on cherche ces mots-là dans la notice de Pline ;
 - d. on effectue le même calcul de distance au sein de Pline : ceci nous donne la *signature* de la notice de Pline ;

- Il s'agit donc de comparer les deux notices de chacun de ces couples (par exemple entre une notice de Vincent et une notice de Pline) :
 - 1 on calcule les **signatures** de deux notices :
 - a. on enregistre les mots qui constituent la notice de Vincent ;
 - b. on calcule la distance entre ces mots dans la notice de Vincent : ceci nous donne un vecteur, la *signature* de la notice de Vincent ;
 - c. on cherche ces mots-là dans la notice de Pline ;
 - d. on effectue le même calcul de distance au sein de Pline : ceci nous donne la *signature* de la notice de Pline ;
 - 2 on compare les deux signatures en calculant la **similarité cosinus** entre ces deux vecteurs, qui correspond à l'angle entre eux.
- Après avoir effectué ce calcul pour toutes les notices-sources potentielles, on choisit celle qui a donné le score de similarité le plus haut, c'est-à-dire le plus petit angle.

Exemple

Vincent	Pisces dicti sunt a pascendo.
Isidore	Pisces dicti unde et pecus, a pascendo scilicet.
Pline	Est paruus admodum piscis adsuetus petris,

Exemple

Vincent	piscis dictus sum pasco
Isidore	piscis dictus unde pecus pasco scilicet
Pline	sum paruus admodum piscis assuetus petra

Exemple

	1	2	3	4	5	6
Vincent	piscis	dictus	sum	pasco		
Isidore	piscis	dictus	pecus	pasco	scilicet	
Pline	sum	paruus	admodum	piscis	assuetus	petra

Exemple

	1	2	3	4	5	6
Vincent	piscis	dictus	sum	pasco		
Isidore	piscis	dictus	pecus	pasco	scilicet	
Pline	sum	paruus	admodum	piscis	assuetus	petra

	Vincent				Isidore				Pline					
	piscis	dictus	sum	pasco	piscis	dictus	sum	pasco	piscis	dictus	sum	pasco		
piscis	0	1	2	3	piscis	0	1	-	3	piscis	0	-	3	-
dictus	1	0	1	2	dictus	1	0	-	2	dictus	-	-	-	-
sum	2	1	0	1	sum	-	-	-	-	sum	3	-	0	-
pasco	3	2	1	0	pasco	3	2	-	0	pasco	-	-	-	-

Exemple

	1	2	3	4	5	6
Vincent	piscis	dictus	sum	pasco		
Isidore	piscis	dictus	pecus	pasco	scilicet	
Pline	sum	paruus	admodum	piscis	assuetus	petra

	Vincent				Isidore				Pline					
	piscis	dictus	sum	pasco	piscis	dictus	sum	pasco	piscis	dictus	sum	pasco		
piscis	0	1	2	3	piscis	0	1	-	3	piscis	0	-	3	-
dictus	1	0	1	2	dictus	1	0	-	2	dictus	-	-	-	-
sum	2	1	0	1	sum	-	-	-	-	sum	3	-	0	-
pasco	3	2	1	0	pasco	3	2	-	0	pasco	-	-	-	-

Vincent	(1 ; 2 ; 3 ; 1 ; 2 ; 1)	→	(0,98 ; 0.92 ; 0.83 ; 0.98 ; 0.92 ; 0.98)
Isidore	(1 ; - ; 3 ; - ; 2 ; -)	→	(0.98 ; 0.00 ; 0.84 ; 0.00 ; 0.92 ; 0.00)
Pline	(- ; 3 ; - ; - ; - ; -)	→	(0.00 ; 0.88 ; 0.00 ; 0.00 ; 0.00 ; 0.00)

Exemple

- Nous pouvons alors calculer des coisnus et donc des angles :

$$\text{angle}(\text{notice de Vincent}, \text{notice de Isidore}) = 46^\circ$$

$$\text{angle}(\text{notice de Vincent}, \text{notice de Pline}) = 66^\circ$$

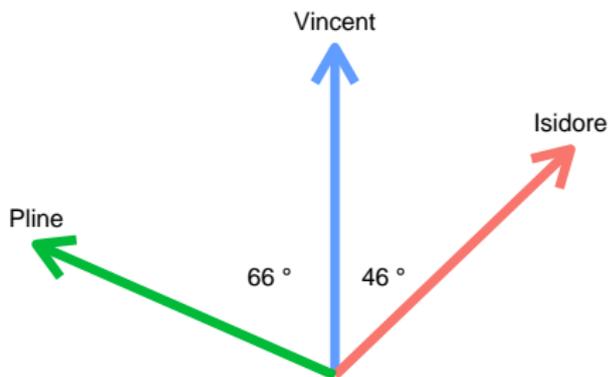
Exemple

- Nous pouvons alors calculer des cosinus et donc des angles :

$$\text{angle}(\text{notice de Vincent}, \text{notice de Isidore}) = 46^\circ$$

$$\text{angle}(\text{notice de Vincent}, \text{notice de Pline}) = 66^\circ$$

- Nous obtenons donc des vecteurs dont une représentation simplifiée est :



- La notice d'Isidore est donc plus proche de celle de Vincent que celle de Pline !

Problème de calculabilité

- La méthode que nous avons présentée est la plus efficace de toutes celles que nous avons testées.
- Mais elle est coûteuse en temps : 0,027 seconde par couple...

Problème de calculabilité

- La méthode que nous avons présentée est la plus efficace de toutes celles que nous avons testées.
- Mais elle est coûteuse en temps : 0,027 seconde par couple... pour traiter les 32 millions de couples de notices, l'ordinateur travaille 244 heures !

Problème de calculabilité

- La méthode que nous avons présentée est la plus efficace de toutes celles que nous avons testées.
- Mais elle est coûteuse en temps : 0,027 seconde par couple... pour traiter les 32 millions de couples de notices, l'ordinateur travaille 244 heures !
- Il faut donc mettre en place des techniques de réduction de temps de calcul :
 - implémentation d'une partie de la routine dans un langage compilé (ici C++);
 - parallélisation du calcul (ici sur 8 CPUs).

Problème de calculabilité

- La méthode que nous avons présentée est la plus efficace de toutes celles que nous avons testées.
- Mais elle est coûteuse en temps : 0,027 seconde par couple... pour traiter les 32 millions de couples de notices, l'ordinateur travaille 244 heures !
- Il faut donc mettre en place des techniques de réduction de temps de calcul :
 - implémentation d'une partie de la routine dans un langage compilé (ici C++);
 - parallélisation du calcul (ici sur 8 CPUs).
- L'utilisation de ces deux techniques permet de diminuer très significativement le temps de calcul : pour l'ensemble des 32 millions de couples de notices, l'ordinateur ne travaille plus que 2 heures (plutôt que 244 heures) !

Problème de calculabilité

- La méthode que nous avons présentée est la plus efficace de toutes celles que nous avons testées.
- Mais elle est coûteuse en temps : 0,027 seconde par couple... pour traiter les 32 millions de couples de notices, l'ordinateur travaille 244 heures !
- Il faut donc mettre en place des techniques de réduction de temps de calcul :
 - implémentation d'une partie de la routine dans un langage compilé (ici C++);
 - parallélisation du calcul (ici sur 8 CPUs).
- L'utilisation de ces deux techniques permet de diminuer très significativement le temps de calcul : pour l'ensemble des 32 millions de couples de notices, l'ordinateur ne travaille plus que 2 heures (plutôt que 244 heures) !
- Il est important que le temps de calcul total soit raisonnable, par exemple pour des raisons de calibration.

Problème de calculabilité

- La méthode que nous avons présentée est la plus efficace de toutes celles que nous avons testées.
- Mais elle est coûteuse en temps : 0,027 seconde par couple... pour traiter les 32 millions de couples de notices, l'ordinateur travaille 244 heures !
- Il faut donc mettre en place des techniques de réduction de temps de calcul :
 - implémentation d'une partie de la routine dans un langage compilé (ici C++);
 - parallélisation du calcul (ici sur 8 CPUs).
- L'utilisation de ces deux techniques permet de diminuer très significativement le temps de calcul : pour l'ensemble des 32 millions de couples de notices, l'ordinateur ne travaille plus que 2 heures (plutôt que 244 heures) !
- Il est important que le temps de calcul total soit raisonnable, par exemple pour des raisons de calibration.
- L'utilisation de la similarité cosinus permet d'atteindre un taux global d'identifications correctes de 87,34 %.

Identification automatique des sources d'une encyclopédie médiévale

Pré-traitement

- Travailler ainsi sur les mots ne donne pas de bons résultats !

Nécessité de la lemmatisation

- Travailler ainsi sur les mots ne donne pas de bons résultats !
- En effet, lorsqu'il copie sa source, Vincent change parfois quelque peu la tournure de la phrase, ce qui en latin peut transformer complètement les mots :

Physiologus dicit	aquilam	hanc	habere	naturam [...]
	ACC	ACC	INF	
	aquila	haec	habet	naturam [...]
	NOM	NOM	IND	

Nécessité de la lemmatisation

- Travailler ainsi sur les mots ne donne pas de bons résultats !
- En effet, lorsqu'il copie sa source, Vincent change parfois quelque peu la tournure de la phrase, ce qui en latin peut transformer complètement les mots :

Physiologus dicit	aquilam	hanc	habere	naturam [...]
	ACC	ACC	INF	
	aquila	haec	habet	naturam [...]
	NOM	NOM	IND	

- Puisque comparer des notices en considérant les mots pose problème, il faut passer aux lemmes !

Physiologus dico	aquila	hic	habeo	natura [...]
------------------	--------	-----	-------	--------------

- Cependant, la lemmatisation est une tâche compliquée, et encore plus lorsque la langue est « très fléchie », comme le latin.

- Cependant, la lemmatisation est une tâche compliquée, et encore plus lorsque la langue est « très fléchie », comme le latin.
- Pensons par exemple au mot *quo*, qui peut correspondre :
 - au relatif *qui* (« lequel ») à l'ablatif ;
 - à l'interrogatif *quis* (« quel ? ») à l'ablatif ;
 - à l'adverbe de lieu *quo* (« où, là où ») ;
 - de la conjonction *quo* (« en conséquence de quoi ») ;
 - etc.

- Cependant, la lemmatisation est une tâche compliquée, et encore plus lorsque la langue est « très fléchie », comme le latin.
- Pensons par exemple au mot *quo*, qui peut correspondre :
 - au relatif *qui* (« lequel ») à l'ablatif ;
 - à l'interrogatif *quis* (« quel ? ») à l'ablatif ;
 - à l'adverbe de lieu *quo* (« où, là où ») ;
 - de la conjonction *quo* (« en conséquence de quoi ») ;
 - etc.
- Dans ce cas, le lemmatiseur affecte en général des probabilités aux différentes possibilités.

Comment l'ordinateur lemmatise-t-il ?

- Ouvrons le capot du processus de lemmatisation : comment l'ordinateur associe-il un lemme à chaque mot ?

Comment l'ordinateur lemmatise-t-il ?

- Ouvrons le capot du processus de lemmatisation : comment l'ordinateur associe-il un lemme à chaque mot ?
- Nous l'avons vu, le lemme correct dépend de la nature du mot !
- La question est donc : comment l'ordinateur fait-il pour déterminer la nature d'un mot (en anglais : *POS tagging*) ?

Comment l'ordinateur lemmatise-t-il ?

- Ouvrons le capot du processus de lemmatisation : comment l'ordinateur associe-il un lemme à chaque mot ?
- Nous l'avons vu, le lemme correct dépend de la nature du mot !
- La question est donc : comment l'ordinateur fait-il pour déterminer la nature d'un mot (en anglais : *POS tagging*) ?
- Nous nous plaçons dans un contexte d'apprentissage supervisé : l'ordinateur va apprendre lui-même à effectuer cette opération en « étudiant » les lemmes qui correspondent aux mots d'une série de textes pour lesquels l'exercice est déjà effectué.
- Heureusement, de tels corpus existent pour le latin et le latin médiéval, comme *Les métamorphoses* d'Ovide :

In	noua	fert	animus	mutatas	dicere	formas	corpora	.
PRE	QLF	VBE	SUB	VBE	VBE	SUB	SUB	SENT

Comment l'ordinateur lemmatise-t-il ?

- Ouvrons le capot du processus de lemmatisation : comment l'ordinateur associe-il un lemme à chaque mot ?
- Nous l'avons vu, le lemme correct dépend de la nature du mot !
- La question est donc : comment l'ordinateur fait-il pour déterminer la nature d'un mot (en anglais : *POS tagging*) ?
- Nous nous plaçons dans un contexte d'apprentissage supervisé : l'ordinateur va apprendre lui-même à effectuer cette opération en « étudiant » les lemmes qui correspondent aux mots d'une série de textes pour lesquels l'exercice est déjà effectué.
- Heureusement, de tels corpus existent pour le latin et le latin médiéval, comme *Les métamorphoses* d'Ovide :

In	noua	fert	animus	mutatas	dicere	formas	corpora	.
PRE	QLF	VBE	SUB	VBE	VBE	SUB	SUB	SENT

- L'ordinateur travaille donc sur ces corpus déjà lemmatisés, et apprend comment se succèdent généralement les natures de mots.

Exemple du processus général

Quinta die ornavit Deus aerem volatilibus et aquam natatilibus.



quinta		ADJ :NUM	quintus
die		N :abl	dies
ornavit		V :IND	orno
deus		N :nom	Deus
aerem	→	N :acc	aer
volatilibus		ADJ	volatilis
et		CC	et
aquam		N :acc	aqua
natatilibus		ADJ	natatilis
.		SENT	.



quintus dies orno Deus aer volatilis et aqua natatilis .

Les limites de la lemmatisation

- Néanmoins, lemmatiser un texte zoologique en latin médiéval pose plusieurs types de problèmes :

mot	POS	lemme	remarque
<i>iactantior</i>	V :IND	<unknown>	$j \rightarrow i$

Les limites de la lemmatisation

- Néanmoins, lemmatiser un texte zoologique en latin médiéval pose plusieurs types de problèmes :

mot	POS	lemme	remarque
<i>iactantior</i>	V :IND	<unknown>	$j \rightarrow i$
<i>aque</i>	N :acc	<unknown>	digramme $ae \rightarrow e$

Les limites de la lemmatisation

- Néanmoins, lemmatiser un texte zoologique en latin médiéval pose plusieurs types de problèmes :

mot	POS	lemme	remarque
<i>iactantior</i>	V :IND	<unknown>	$j \rightarrow i$
<i>aque</i>	N :acc	<unknown>	digramme <i>ae</i> \rightarrow <i>e</i>
<i>utrorumque</i>	N :acc	<unknown>	enclitique <i>-que</i>

Les limites de la lemmatisation

- Néanmoins, lemmatiser un texte zoologique en latin médiéval pose plusieurs types de problèmes :

mot	POS	lemme	remarque
<i>iactantior</i>	V :IND	<unknown>	$j \rightarrow i$
<i>aque</i>	N :acc	<unknown>	digramme <i>ae</i> \rightarrow <i>e</i>
<i>utrorumque</i>	N :acc	<unknown>	enclitique <i>-que</i>
<i>queque</i>	ADJ	<unknown>	enclitique <i>-que</i> et digramme <i>ae</i>

Les limites de la lemmatisation

- Néanmoins, lemmatiser un texte zoologique en latin médiéval pose plusieurs types de problèmes :

mot	POS	lemme	remarque
<i>iactantior</i>	V :IND	<unknown>	$j \rightarrow i$
<i>aque</i>	N :acc	<unknown>	digramme <i>ae</i> \rightarrow <i>e</i>
<i>utrorumque</i>	N :acc	<unknown>	enclitique <i>-que</i>
<i>queque</i>	ADJ	<unknown>	enclitique <i>-que</i> et digramme <i>ae</i>
<i>fruticumve</i>	V :IND	<unknown>	enclitique <i>-ve</i>

Les limites de la lemmatisation

- Néanmoins, lemmatiser un texte zoologique en latin médiéval pose plusieurs types de problèmes :

mot	POS	lemme	remarque
<i>iactantior</i>	V :IND	<unknown>	$j \rightarrow i$
<i>aque</i>	N :acc	<unknown>	digramme <i>ae</i> \rightarrow <i>e</i>
<i>utrorumque</i>	N :acc	<unknown>	enclitique <i>-que</i>
<i>queque</i>	ADJ	<unknown>	enclitique <i>-que</i> et digramme <i>ae</i>
<i>fruticumve</i>	V :IND	<unknown>	enclitique <i>-ve</i>
<i>econverso</i>	V :PTC	<unknown>	latin médiéval inexistant en latin classique

Les limites de la lemmatisation

- Néanmoins, lemmatiser un texte zoologique en latin médiéval pose plusieurs types de problèmes :

mot	POS	lemme	remarque
<i>iactantior</i>	V :IND	<unknown>	$j \rightarrow i$
<i>aque</i>	N :acc	<unknown>	digramme <i>ae</i> \rightarrow <i>e</i>
<i>utrorumque</i>	N :acc	<unknown>	enclitique <i>-que</i>
<i>queque</i>	ADJ	<unknown>	enclitique <i>-que</i> et digramme <i>ae</i>
<i>fruticumve</i>	V :IND	<unknown>	enclitique <i>-ve</i>
<i>econverso</i>	V :PTC	<unknown>	latin médiéval inexistant en latin classique
<i>Olympto</i>	V :IND	<unknown>	nom propre

Les limites de la lemmatisation

- Néanmoins, lemmatiser un texte zoologique en latin médiéval pose plusieurs types de problèmes :

mot	POS	lemme	remarque
<i>iactantior</i>	V :IND	<unknown>	$j \rightarrow i$
<i>aque</i>	N :acc	<unknown>	digramme <i>ae</i> \rightarrow <i>e</i>
<i>utrorumque</i>	N :acc	<unknown>	enclitique <i>-que</i>
<i>queque</i>	ADJ	<unknown>	enclitique <i>-que</i> et digramme <i>ae</i>
<i>fruticumve</i>	V :IND	<unknown>	enclitique <i>-ve</i>
<i>econverso</i>	V :PTC	<unknown>	latin médiéval inexistant en latin classique
<i>Olympo</i>	V :IND	<unknown>	nom propre
<i>seleutides</i>	N :voc	<unknown>	zonyme (séleucide = passereau)

Les limites de la lemmatisation

- Néanmoins, lemmatiser un texte zoologique en latin médiéval pose plusieurs types de problèmes :

mot	POS	lemme	remarque
<i>iactantior</i>	V :IND	<unknown>	$j \rightarrow i$
<i>aque</i>	N :acc	<unknown>	digramme <i>ae</i> \rightarrow <i>e</i>
<i>utrorumque</i>	N :acc	<unknown>	enclitique <i>-que</i>
<i>queque</i>	ADJ	<unknown>	enclitique <i>-que</i> et digramme <i>ae</i>
<i>fruticumve</i>	V :IND	<unknown>	enclitique <i>-ve</i>
<i>econverso</i>	V :PTC	<unknown>	latin médiéval inexistant en latin classique
<i>Olympo</i>	V :IND	<unknown>	nom propre
<i>seleutides</i>	N :voc	<unknown>	zoonyme (séleucide = passereau)
<i>papagabio</i>	ADJ	<unknown>	vernaculaire (« <i>Psittacus qui vulgo papagabio</i> »)

Les limites de la lemmatisation

- Néanmoins, lemmatiser un texte zoologique en latin médiéval pose plusieurs types de problèmes :

mot	POS	lemme	remarque
<i>iactantior</i>	V :IND	<unknown>	$j \rightarrow i$
<i>aque</i>	N :acc	<unknown>	digramme <i>ae</i> \rightarrow <i>e</i>
<i>utrorumque</i>	N :acc	<unknown>	enclitique <i>-que</i>
<i>queque</i>	ADJ	<unknown>	enclitique <i>-que</i> et digramme <i>ae</i>
<i>fruticumve</i>	V :IND	<unknown>	enclitique <i>-ve</i>
<i>econverso</i>	V :PTC	<unknown>	latin médiéval inexistant en latin classique
<i>Olympo</i>	V :IND	<unknown>	nom propre
<i>seleutides</i>	N :voc	<unknown>	zonyme (séleucide = passereau)
<i>papagabio</i>	ADJ	<unknown>	vernaculaire (« <i>Psittacus qui vulgo papagabio</i> »)
<i>alphoraz</i>	N :acc	<unknown>	zonyme en « arabe phonétique latinisé »

Les limites de la lemmatisation

- Néanmoins, lemmatiser un texte zoologique en latin médiéval pose plusieurs types de problèmes :

mot	POS	lemme	remarque
<i>iactantior</i>	V :IND	<unknown>	$j \rightarrow i$
<i>aque</i>	N :acc	<unknown>	digramme <i>ae</i> \rightarrow <i>e</i>
<i>utrorumque</i>	N :acc	<unknown>	enclitique <i>-que</i>
<i>queque</i>	ADJ	<unknown>	enclitique <i>-que</i> et digramme <i>ae</i>
<i>fruticumve</i>	V :IND	<unknown>	enclitique <i>-ve</i>
<i>econverso</i>	V :PTC	<unknown>	latin médiéval inexistant en latin classique
<i>Olympo</i>	V :IND	<unknown>	nom propre
<i>seleutides</i>	N :voc	<unknown>	zonyme (séleucide = passereau)
<i>papagabio</i>	ADJ	<unknown>	vernaculaire (« <i>Psittacus qui vulgo papagabio</i> »)
<i>alphoraz</i>	N :acc	<unknown>	zonyme en « arabe phonétique latinisé »

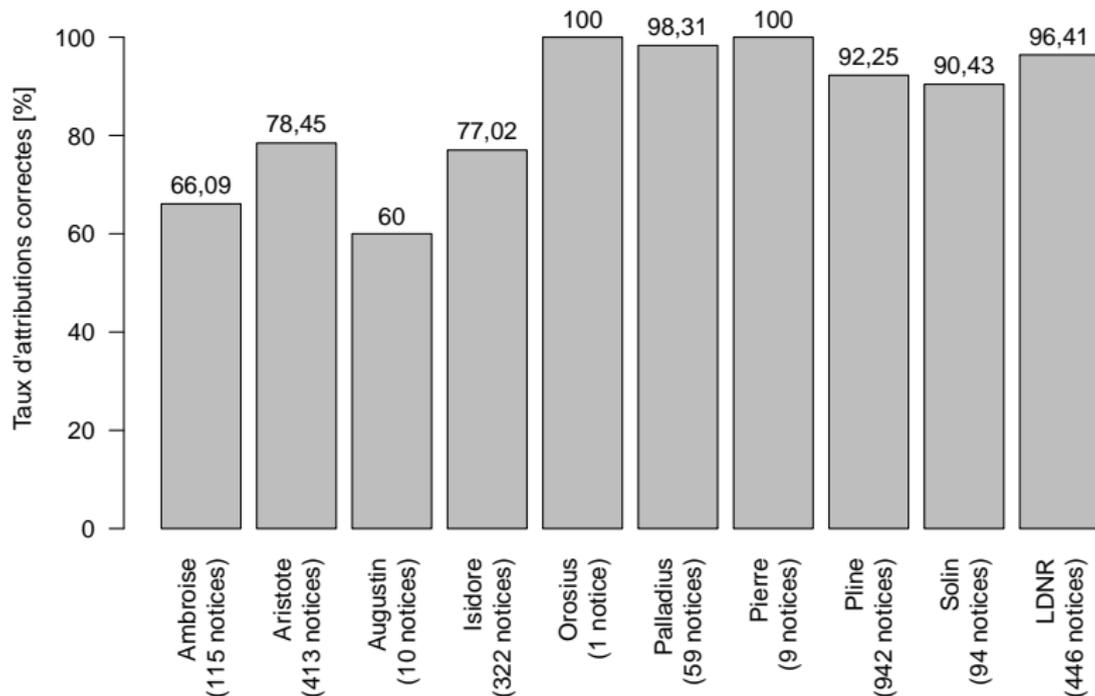
- Certains d'entre eux sont contournables, d'autres impossibles à traiter autrement qu'à la main.

Identification automatique des sources d'une encyclopédie médiévale

Éléments d'analyse critique

Analyse des identifications incorrectes

- Parmi d'autres constatations, il est intéressant de remarquer que tous les auteurs de référence ne sont pas logés à la même enseigne :



Analyse des identifications incorrectes

- On peut ainsi aller voir quels auteurs « attirent » les notices incorrectement attribuées :

	Ambroise	Aristote	Augustin	Isidore	Orosius	Palladius	Pierre	Pline	Solin	LDNR
Ambroise	76	2				1		1		36
Aristote	2	324						3		84
Augustin		1	6							3
Isidore	4	9	1	248	1		1	2	4	52
Orosius					1					
Palladius						58		1		
Pierre							9			
Pline		9		5		5	1	870	2	49
Solin	1	1						1	85	6
LDNR		13	1		1			1		430

Analyse des identifications incorrectes

- On remarque qu'un nombre important de notices extraites d'Ambroise sont attribuées au *Liber de natura rerum* :

	Ambroise	Aristote	Augustin	Isidore	Orosius	Palladius	Pierre	Pline	Solin	LDNR
Ambroise	76	2				1		1		36
Aristote	2	324						3		84
Augustin		1	6							3
Isidore	4	9	1	248	1		1	2	4	52
Orosius					1					
Palladius						58		1		
Pierre							9			
Pline		9		5		5	1	870	2	49
Solin	1	1						1	85	6
LDNR		13	1		1			1		430

Analyse des identifications apparemment incorrectes

- La raison tient dans le fait que le LDNR est lui-même une encyclopédie, qui puise dans les mêmes sources que Vincent.
- Dans beaucoup de ces cas-là, la « seconde attribution » est correcte :

Vincent	LDNR (Thomas) angle = 41, 97°	Ambroise angle = 41, 99°
<i>In exiguo cicadarum gutture dulcis est cantilena, quarum cantibus estu medio rumpunt arbusta, eo quod magis canent meridianis caloribus; quo priorem aerem spiritu attrahunt, eo cantus clariores reddunt.</i>	[...] <i>et hoc quidem dulcis est in exiguo cycadis gutture cantilena. Quarum cantibus, ut Ambrosius dicit, in meridiano estu arbusta rumpuntur, eo quod magis canore meridianis caloribus; quo puriorem aerem per id temporis attrahit spiritu, eo cantus resonant clariores. [...]</i>	<i>Quam dulcis etiam in exiguo cicadis gutture cantilena, quarum cantibus medio aestu arbusta rumpuntur, eo quod magis canorae meridianis caloribus, quo puriorem aerem id temporis attrahunt spiritu, eo cantus resonant clariores. [...]</i>

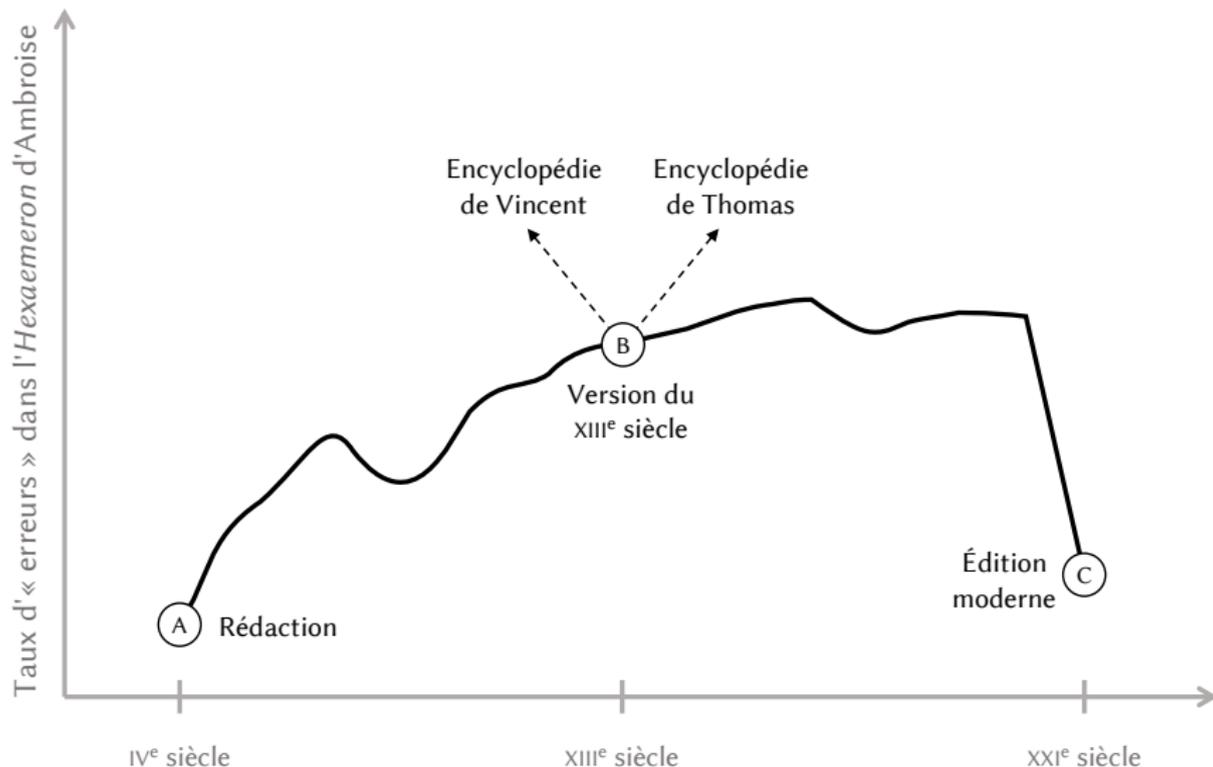
- Comment expliquer que la similarité entre le texte de Vincent et celui d'Ambroise soit plus basse que la similarité entre le texte de Vincent et celui du LDNR ?

- Comment expliquer que la similarité entre le texte de Vincent et celui d'Ambroise soit plus basse que la similarité entre le texte de Vincent et celui du LDNR ?
 - 1 Vincent s'est servi du LDNR plutôt que de puiser directement dans Ambroise (source de « deuxième main »), en indiquant tout de même Ambroise dans le marqueur ;

- Comment expliquer que la similarité entre le texte de Vincent et celui d'Ambroise soit plus basse que la similarité entre le texte de Vincent et celui du LDNR ?
 - 1 Vincent s'est servi du LDNR plutôt que de puiser directement dans Ambroise (source de « deuxième main »), en indiquant tout de même Ambroise dans le marqueur ;
 - 2 Vincent et l'auteur du LDNR ont tous les deux travaillé d'après Ambroise, mais la version du texte qu'ils ont utilisée est différente de celle que nous connaissons aujourd'hui (qui a certainement été normalisée par les philologues) : il ne faut pas oublier que plus de huit siècles séparent le père de l'Église et les deux encyclopédistes !

Analyse des identifications apparemment incorrectes

- On serait alors face à la situation suivante :



- Dans certains cas, la source identifiée automatiquement ne correspond pas au marqueur médiéval, mais leur examen permet de se rendre compte que c'est le marqueur qui est imparfait :

- Dans certains cas, la source identifiée automatiquement ne correspond pas au marqueur médiéval, mais leur examen permet de se rendre compte que c'est le marqueur qui est imparfait :
 - il ne donne qu'une des sources multiples utilisées ;
 - il semble se tromper complètement.

- Dans certains cas, la source identifiée automatiquement ne correspond pas au marqueur médiéval, mais leur examen permet de se rendre compte que c'est le marqueur qui est imparfait :
 - il ne donne qu'une des sources multiples utilisées ;
 - il semble se tromper complètement.
- Bien entendu, ces « erreurs apparentes » sont peut-être explicables par des éléments qui ne nous sont pas connus (on pourrait ainsi imaginer que Vincent disposait de manuscrits « hybrides » contenant des textes d'auteurs différents mélangés et donc difficiles à distinguer).

Analyse des identifications apparemment incorrectes

- Dans certains cas, la source identifiée automatiquement ne correspond pas au marqueur médiéval, mais leur examen permet de se rendre compte que c'est le marqueur qui est imparfait :
 - il ne donne qu'une des sources multiples utilisées ;
 - il semble se tromper complètement.
- Bien entendu, ces « erreurs apparentes » sont peut-être explicables par des éléments qui ne nous sont pas connus (on pourrait ainsi imaginer que Vincent disposait de manuscrits « hybrides » contenant des textes d'auteurs différents mélangés et donc difficiles à distinguer).
- Ces erreurs amènent parfois des absurdités : par exemple, le marqueur d'une notice indique qu'elle est inspirée de Pline, tandis que le texte mentionne un nom de poisson arabisé (*alphoraz*, « poisson-écume »), alors qu'on sait que l'œuvre de Pline est parvenue à l'Occident médiéval sans jamais passer par l'arabe (au contraire d'autres ouvrages zoologiques, comme l'*Historia animalium* d'Aristote).

Réseaux de crédit

Contexte

La gestion du risque de défaut

- Je m'intéresse à la gestion du *risque de défaut* durant le XIII^e siècle et le début du XIV^e siècle en Flandre et dans le nord de la France.

La gestion du risque de défaut

- Je m'intéresse à la gestion du *risque de défaut* durant le XIII^e siècle et le début du XIV^e siècle en Flandre et dans le nord de la France.
- Dans cette région durant cette période, le processus de conceptualisation du risque n'a pas encore abouti : le risque n'existe pas d'un point de vue théorique, il n'est pas encore un « objet intellectuel ».
- Cependant, les acteurs économiques médiévaux font face au risque tous les jours : comment le gèrent-ils ?

La gestion du risque de défaut

- Je m'intéresse à la gestion du *risque de défaut* durant le XIII^e siècle et le début du XIV^e siècle en Flandre et dans le nord de la France.
- Dans cette région durant cette période, le processus de conceptualisation du risque n'a pas encore abouti : le risque n'existe pas d'un point de vue théorique, il n'est pas encore un « objet intellectuel ».
- Cependant, les acteurs économiques médiévaux font face au risque tous les jours : comment le gèrent-ils ?
- Je m'intéresse plus particulièrement au *risque de défaut*, c'est-à-dire au risque que le débiteur ne rembourse pas son créancier lorsque le terme de la dette est venu.

La gestion du risque de défaut

- Je m'intéresse à la gestion du *risque de défaut* durant le XIII^e siècle et le début du XIV^e siècle en Flandre et dans le nord de la France.
- Dans cette région durant cette période, le processus de conceptualisation du risque n'a pas encore abouti : le risque n'existe pas d'un point de vue théorique, il n'est pas encore un « objet intellectuel ».
- Cependant, les acteurs économiques médiévaux font face au risque tous les jours : comment le gèrent-ils ?
- Je m'intéresse plus particulièrement au *risque de défaut*, c'est-à-dire au risque que le débiteur ne rembourse pas son créancier lorsque le terme de la dette est venu.
- Les garanties contre ce risque peuvent être regroupées en deux catégories :
 - les sûretés réelles : c'est un bien matériel qui sert de garantie ;
 - les sûretés personnelles : c'est une personne qui sert de garantie.
- Les deux types de garanties se retrouvent dans les sources médiévales.

- C'est dans un contexte urbain que j'étudie la gestion du risque de défaut ; j'ai donc cherché des villes pour lesquelles des corpus de sources étaient disponibles.

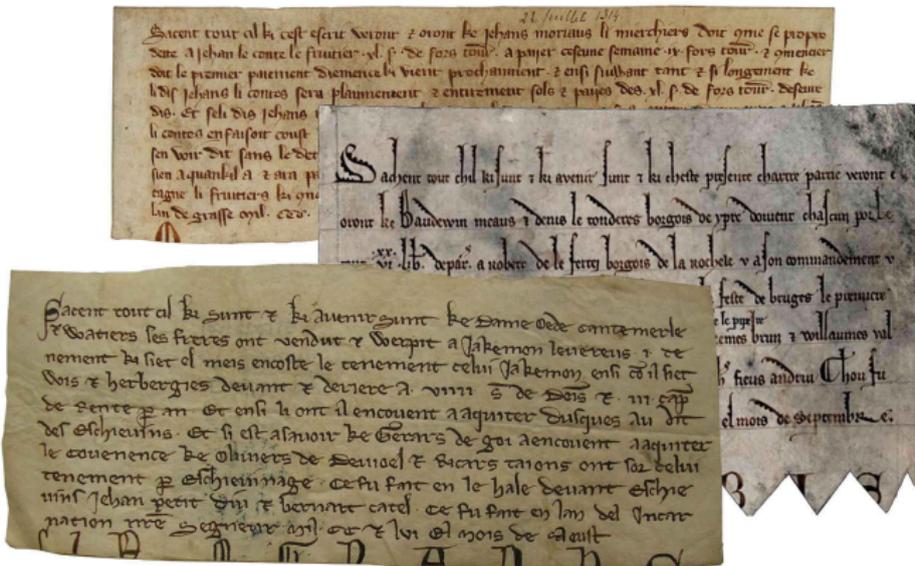
Trois villes

- C'est dans un contexte urbain que j'étudie la gestion du risque de défaut ; j'ai donc cherché des villes pour lesquelles des corpus de sources étaient disponibles.



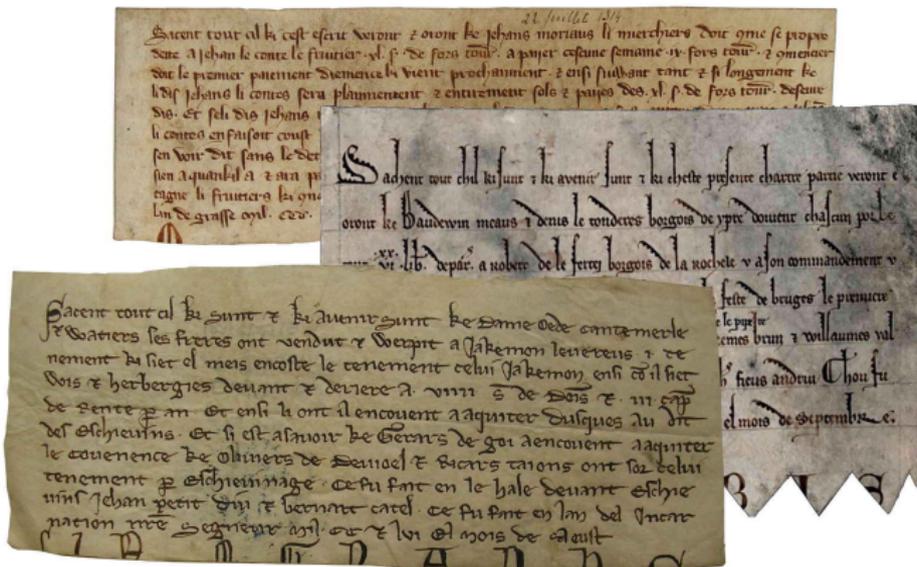
Des chirographes

- Dans les trois villes, les reconnaissances de dette prennent la forme de chirographes.



Des chirographes

- Dans les trois villes, les reconnaissances de dette prennent la forme de chirographes.



- Les trois corpus sont cependant très différents en termes d'état de conservation.

Le principe des chirographes

- Le processus de confection d'un chirographe est simple :



Le principe des chirographes

- Le processus de confection d'un chirographe est simple :

Sacent tout cil ki sunt et ki avenir sunt ke Mehaus de Mauricourt
doit et a enconvent a Watier Rousiel le moelekitijer .lxv. £ et .vi.
deniers de parisis de boene dete de loial

.....
.....

Le principe des chirographes

- Le processus de confection d'un chirographe est simple :

Sacent tout cil ki sunt et ki avenir sunt ke Mehaus de Mauricourt
doit et a enconvent a Watier Rousiel le moekeitijer .I.xv. £ et .vi.
deniers de parisis de boene dete de loial

Sacent tout cil ki sunt et ki avenir sunt ke Mehaus de Mauricourt
doit et a enconvent a Watier Rousiel le moekeitijer .I.xv. £ et .vi.
deniers de parisis de boene dete de loial

Le principe des chirographes

- Le processus de confection d'un chirographe est simple :

Sacent tout cil ki sunt et ki avenir sunt ke Mehaus de Mauricourt
doit et a enconvient a Watier Rousiel le moelekitijer .lxv. £ et .vi.
deniers de parisis de boene dete de loial

C H I R O G R A P H U M

Sacent tout cil ki sunt et ki avenir sunt ke Mehaus de Mauricourt
doit et a enconvient a Watier Rousiel le moelekitijer .lxv. £ et .vi.
deniers de parisis de boene dete de loial

Le principe des chiromaphes

- Le processus de confection d'un chiromaphes est simple :

Sacent tout cil ki sunt et ki avenir sunt ke Mehaus de Mauricourt
doit et a enconvent a Watier Rousiel le moelekitijer .lxv. £ et .vi.
deniers de parisis de boene dete de loial

.....

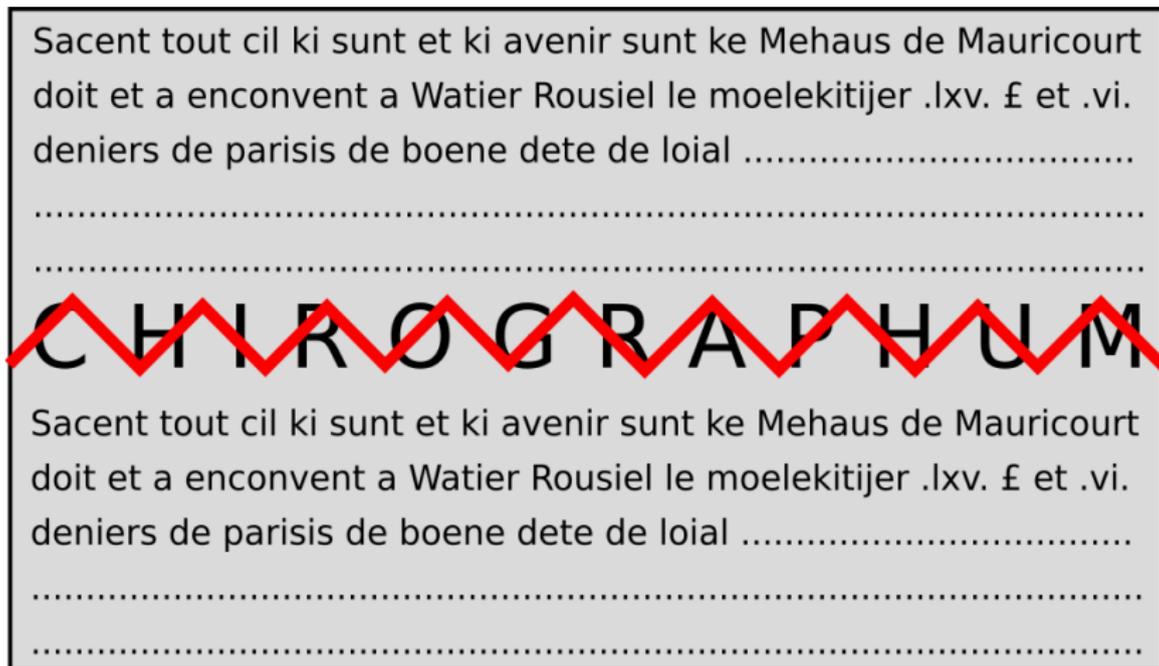
.....

C H I R O G R A P H U M

Sacent tout cil ki sunt et ki avenir sunt ke Mehaus de Mauricourt
doit et a enconvent a Watier Rousiel le moelekitijer .lxv. £ et .vi.
deniers de parisis de boene dete de loial

.....

.....



Le principe des chirographes

- Le processus de confection d'un chirographe est simple :

Sacent tout cil ki sunt et ki avenir sunt ke Mehaus de Mauricourt
doit et a enconvent a Watier Rousiel le moelekitijer .lxv. £ et .vi.
deniers de parisis de boene dete de loial

C H I L D O G R A P H U M

C T R O G A P U M

Sacent tout cil ki sunt et ki avenir sunt ke Mehaus de Mauricourt
doit et a enconvent a Watier Rousiel le moelekitijer .lxv. £ et .vi.
deniers de parisis de boene dete de loial

Exemple simple de reconnaissance de dette

- La plupart des reconnaissances de dette sont structurées de la même façon :

Chirographe yprois

Sacent tout chil ki sunt et ki avenir sunt et ki cheste presente chartre partie verront et orront ke Jehans Loerin li mauniers, bourgeois d'Ypre, doit à Pieron Bardonc, bourgeois d'Ypre, 8 lb. d'ar. à lui ou à son commandement ki ceste presente chartre partie apportera. De chou li doit il paier le moiet el mois d'aoust prochain venant et l'autre moiet à la Saint Remi prochain apries. A ceste counissance furent eschevin d'Ypre, Pieres de Lo et Lambiers de Scotés. Ce fu fait en l'an del incarnation .M.CC. et .IIII^{XX}. el mois de march lendemain del jour del behordich.

[Au dos :] Cest Pieron Bardonc, bourgeois d'Ypre.

Exemple simple de reconnaissance de dette

- La plupart des reconnaissances de dette sont structurées de la même façon :

Chirographe yprois

Sacent tout chil ki sunt et ki avenir sunt et ki cheste presente chartre partie verront et orront ke Jehans Loerin li mauniers, bourgeois d'Ypre, doit à Pieron Bardonc, bourgeois d'Ypre, 8 lb. d'ar. à lui ou à son commandement ki ceste presente chartre partie apportera. De chou li doit il paier le moiet el mois d'aoust prochain venant et l'autre moiet à la Saint Remi prochain apries. A ceste counissance furent eschevin d'Ypre, Pieres de Lo et Lambiers de Scotès. Ce fu fait en l'an del incarnation .M.CC. et .IIII^{XX}. el mois de march lendemain del jour del behordich.

[Au dos :] Cest Pieron Bardonc, bourgeois d'Ypre.

Exemple simple de reconnaissance de dette

- La plupart des reconnaissances de dette sont structurées de la même façon :

Chirographe yprois

Sacent tout chil ki sunt et ki avenir sunt et ki cheste presente chartre partie verront et orront ke **Jehans Loerin li mauniers, bourgeois d'Ypre, doit à Pieron Bardonc, borgois d'Ypre, 8 lb. d'ar.** à lui ou à son commandement ki ceste presente chartre partie apportera. De chou li doit il paier le moiet el mois d'aoust prochain venant et l'autre moiet à la Saint Remi prochain apries. A ceste counissance furent eschevin d'Ypre, Pieres de Lo et Lambiers de Scotés. Ce fu fait en l'an del incarnation .M.CC. et .III^{XX}. el mois de march lendemain del jour del behordich.

[Au dos :] Cest Pieron Bardonc, borgois d'Ypre.

Exemple simple de reconnaissance de dette

- La plupart des reconnaissances de dette sont structurées de la même façon :

Chirographe yprois

Sacent tout chil ki sunt et ki avenir sunt et ki cheste presente chartre partie verront et orront ke Jehans Loerin li mauniers, bourgeois d'Ypre, doit à Pieron Bardonc, bourgeois d'Ypre, 8 lb. d'ar. **à lui ou à son commandement ki ceste presente chartre partie apportera.** De chou li doit il paier le moiet el mois d'aoust prochain venant et l'autre moiet à la Saint Remi prochain apries. A ceste counissance furent eschevin d'Ypre, Pieres de Lo et Lambiers de Scotès. Ce fu fait en l'an del incarnation .M.CC. et .IIII^{XX}. el mois de march lendemain del jour del behordich.

[Au dos :] Cest Pieron Bardonc, bourgeois d'Ypre.

Exemple simple de reconnaissance de dette

- La plupart des reconnaissances de dette sont structurées de la même façon :

Chirographe yprois

Sacent tout chil ki sunt et ki avenir sunt et ki cheste presente chartre partie verront et orront ke Jehans Loerin li mauniers, bourgeois d'Ypre, doit à Pieron Bardonc, bourgeois d'Ypre, 8 lb. d'ar. à lui ou à son commandement ki ceste presente chartre partie apportera. **De chou li doit il paier le moiet el mois d'aoust prochain venant et l'autre moiet à la Saint Remi prochain apries.** A ceste counissance furent eschevin d'Ypre, Pieres de Lo et Lambiers de Scotés. Ce fu fait en l'an del incarnation .M.CC. et .IIII^{XX}. el mois de march lendemain del jour del behordich.

[Au dos :] Cest Pieron Bardonc, bourgeois d'Ypre.

Exemple simple de reconnaissance de dette

- La plupart des reconnaissances de dette sont structurées de la même façon :

Chirographe yprois

Sacent tout chil ki sunt et ki avenir sunt et ki cheste presente chartre partie verront et orront ke Jehans Loerin li mauniers, bourgeois d'Ypre, doit à Pieron Bardonc, bourgeois d'Ypre, 8 lb. d'ar. à lui ou à son commandement ki ceste presente chartre partie apportera. De chou li doit il paier le moiet el mois d'aoust prochain venant et l'autre moiet à la Saint Remi prochain apries. **A ceste counissance furent eschevin d'Ypre, Pieres de Lo et Lambiers de Scotés.** Ce fu fait en l'an del incarnation .M.CC. et .III^{XX}. el mois de march lendemain del jour del behordich.

[Au dos :] Cest Pieron Bardonc, bourgeois d'Ypre.

Exemple simple de reconnaissance de dette

- La plupart des reconnaissances de dette sont structurées de la même façon :

Chirographe yprois

Sacent tout chil ki sunt et ki avenir sunt et ki cheste presente chartre partie verront et orront ke Jehans Loerin li mauniers, bourgeois d'Ypre, doit à Pieron Bardonc, bourgeois d'Ypre, 8 lb. d'ar. à lui ou à son commandement ki ceste presente chartre partie apportera. De chou li doit il paier le moiet el mois d'aoust prochain venant et l'autre moiet à la Saint Remi prochain apries. A ceste counissance furent eschevin d'Ypre, Pieres de Lo et Lambiers de Scotés. **Ce fu fait en l'an del incarnation .M.CC. et .III^{xx}. el mois de march lendemain del jour del behordich.**

[Au dos :] Cest Pieron Bardonc, bourgeois d'Ypre.

Exemple simple de reconnaissance de dette

- La plupart des reconnaissances de dette sont structurées de la même façon :

Chirographe yprois

Sacent tout chil ki sunt et ki avenir sunt et ki cheste presente chartre partie verront et orront ke Jehans Loerin li mauniers, bourgeois d'Ypre, doit à Pieron Bardonc, bourgeois d'Ypre, 8 lb. d'ar. à lui ou à son commandement ki ceste presente chartre partie apportera. De chou li doit il paier le moiet el mois d'aoust prochain venant et l'autre moiet à la Saint Remi prochain apries. A ceste counissance furent eschevin d'Ypre, Pieres de Lo et Lambiers de Scotés. Ce fu fait en l'an del incarnation .M.CC. et .IIII^{XX}. el mois de march lendemain del jour del behordich.

[Au dos :] Cest Pieron Bardonc, bourgeois d'Ypre.

Trois corpus différents

- Les chirographes des trois corpus sont dans des « états de conservation » très différents :

	Originaux	Édités	Résumés
Douai	491		
Tournai	468	190	
Ypres	1	157	5.362

- La grande majorité des actes d'Ypres et de Douai ont brûlé lors des deux guerres mondiales.

Trois corpus différents

- Les chirographes des trois corpus sont dans des « états de conservation » très différents :

	Originaux	Édités	Résumés
Douai	491		
Tournai	468	190	
Ypres	1	157	5.362

- La grande majorité des actes d'Ypres et de Douai ont brûlé lors des deux guerres mondiales.
- Un grand nombre de reconnaissances de dette yproises ont été résumées :

1 N°	2 Date	3 Créancier	4 Origine	5 Débiteur	6 Origine	7 Montant de la dette	8 Échéance	9 Sûreté		10 Échevins
								a. personnelle	b. réelle	
36	oct.	Michel Bonevente	Ypres	Joh. Bonevente le juvenes fil Joh. Bonevente	Ypres	6 lb. 19 s. art.	foire de Lengi	—	—	Joh. Firtons Joh. Anguille Joh. Brun Joh. Anguille Jehans Firtons Tiersis le Maedes
37	oct.	Nicoles de le Bouke, clers de Sen Martin	—	Joh. li Fazieres des cars	Ypres	18 lb. art.	avant août	Hug Baudri	—	Joh. Brun Joh. Anguille Tiersis le Maedes
38	oct.	Michiel de Cassiel, son frère	Ypres	Jehans de Cassiel	Ypres	300 lb. par.	foire d'Ypres	—	maison	Jehans Firtons Joh. Brun Jeh. Paldine
39	4 oct.	Pierre de la Vie	La Rochelle	Bauduins Mirus Will. Trevoius dame Marote ki femme fu Denis le Tondeur	Ypres	170 lb. art.	foire d'Ypres	—	—	Jeh. Paldine

Trois corpus différents

- Les chirographes des trois corpus sont dans des « états de conservation » très différents :

	Originaux	Édités	Résumés
Douai	491		
Tournai	468	190	
Ypres	1	157	5.362

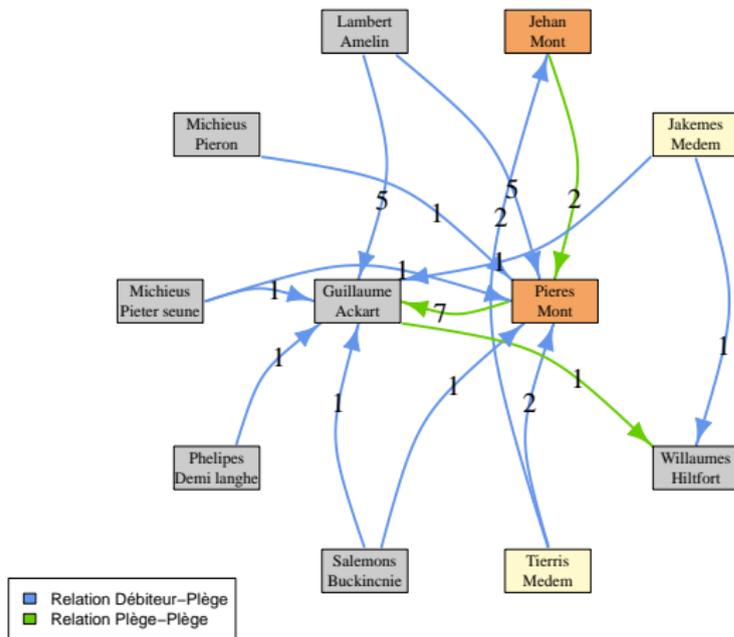
- La grande majorité des actes d'Ypres et de Douai ont brûlé lors des deux guerres mondiales.
- Un grand nombre de reconnaissances de dette yproises ont été résumées :

1 N°	2 Date	3 Créancier	4 Origine	5 Débiteur	6 Origine	7 Montant de la dette	8 Échéance	9 Sûreté		10 Échevins
								a. personnelle	b. réelle	
36	oct.	Michel Bonevente	Ypres	Joh. Bonevente le juvenes fil Joh. Bonevente	Ypres	6 lb. 19 s. art.	foire de Lengi	—	—	Joh. Firtons Joh. Anguille Joh. Brun
37	oct.	Nicoles de le Bouke, clers de Sen Martin	—	Joh. li Fazieres des cars	Ypres	18 lb. art.	avant août	Hug Baudri	—	Joh. Anguille Joh. Firtons
38	oct.	Michiel de Cassiel, son frère	Ypres	Jehans de Cassiel	Ypres	300 lb. par.	foire d'Ypres	—	maison	Tierris li Maedes Jehans Firtons
39	4 oct.	Pierre de la Vie	La Rochelle	Bauduins Mirus Will. Trevaus dame Marote ki femme fu Denis le Tondeur	Ypres	170 lb. art.	foire d'Ypres	—	—	Joh. Brun Jeh. Paldine

- Dans ce cas, l'information disponible pour chaque acte est réduite : date, noms et origines des contractants, montant, terme, sûretés et noms des échevins témoins.

L'analyse des réseaux sociaux

- La forme sous laquelle ce corpus est disponible suggère la mise en œuvre de l'analyse des réseaux sociaux.
- L'idée de celle-ci est de représenter les protagonistes des chirographes et leurs relations commerciales par un objet mathématique : un graphe.



Réseaux de crédit

Pré-traitement

Nécessité de normaliser les anthroponymes

- L'utilisation des données brutes des résumés des chiromorphes donne un réseau avec beaucoup d'acteurs :
11.311 individus et 4.887 familles

Nécessité de normaliser les anthroponymes

- L'utilisation des données brutes des résumés des chirographes donne un réseau avec beaucoup d'acteurs :

11.311 individus et 4.887 familles

- Un problème se pose toutefois au sein de celui-ci, puisqu'on y trouve ces individus :

Jehans	de Neuveglise
Jean	de Noeveglise
Jan	de Nieukerke
Jehannet	de Nueveglise
Jehan	Noveglise

... ..

- Il est pourtant probable qu'il s'agisse du même homme !

Nécessité de normaliser les anthroponymes

- L'utilisation des données brutes des résumés des chirographes donne un réseau avec beaucoup d'acteurs :

11.311 individus et 4.887 familles

- Un problème se pose toutefois au sein de celui-ci, puisqu'on y trouve ces individus :

Jehans	de Neuveglise
Jean	de Noeveglise
Jan	de Nieukerke
Jehannet	de Nueveglise
Jehan	Noveglise

... ..

- Il est pourtant probable qu'il s'agisse du même homme !
- Lorsqu'on « normalise » les anthroponymes, le réseau est bien différent :

5.467 individus et 2.444 familles

Nécessité de normaliser les anthroponymes

- L'utilisation des données brutes des résumés des chirographes donne un réseau avec beaucoup d'acteurs :

11.311 individus et 4.887 familles

- Un problème se pose toutefois au sein de celui-ci, puisqu'on y trouve ces individus :

Jehans de Neuveglise
Jean de Noeveglise
Jan de Nieukerke
Jehannet de Nueveglise
Jehan Noveglise

... ..

- Il est pourtant probable qu'il s'agisse du même homme !
- Lorsqu'on « normalise » les anthroponymes, le réseau est bien différent :

5.467 individus et 2.444 familles

- Il est donc absolument nécessaire d'effectuer ce traitement.

- Nous voudrions donc obtenir une sorte de « lemme » pour chaque prénom et pour chaque nom de famille ; cependant, il n'est pas question de lemmatiser, puisque les variations de graphie sont pour la plupart imprévisibles !

Méthodologie de normalisation

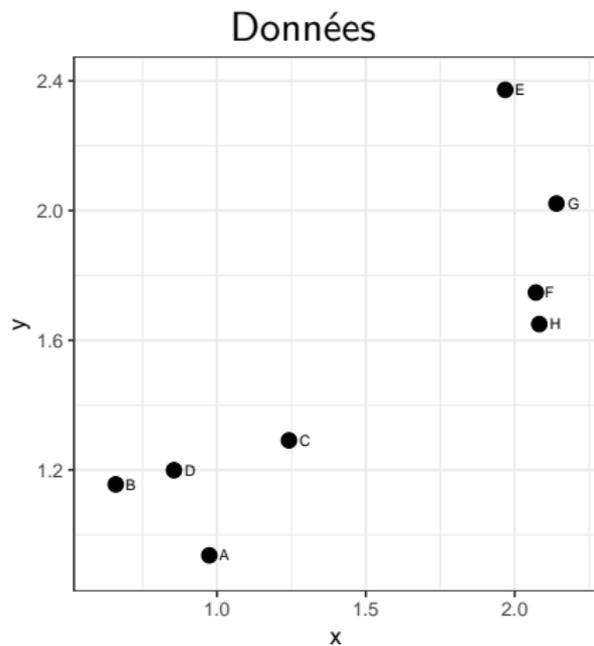
- Nous voudrions donc obtenir une sorte de « lemme » pour chaque prénom et pour chaque nom de famille ; cependant, il n'est pas question de lemmatiser, puisque les variations de graphie sont pour la plupart imprévisibles !
- Nous pouvons néanmoins appliquer un autre genre de méthode, en remarquant que les différentes graphies d'un même anthroponyme sont sans doute assez semblables les unes aux autres.
- Nous pourrions dès lors tenter de « regrouper » les graphies les plus « proches » et assigner à chacun des groupes ainsi constitués une graphie normalisée.

- Nous voudrions donc obtenir une sorte de « lemme » pour chaque prénom et pour chaque nom de famille ; cependant, il n'est pas question de lemmatiser, puisque les variations de graphie sont pour la plupart imprévisibles !
- Nous pouvons néanmoins appliquer un autre genre de méthode, en remarquant que les différentes graphies d'un même anthroponyme sont sans doute assez semblables les unes aux autres.
- Nous pourrions dès lors tenter de « regrouper » les graphies les plus « proches » et assigner à chacun des groupes ainsi constitués une graphie normalisée.
- Pour ce faire, nous devons comprendre ce que signifient :
 - 1 « regrouper » : comment peut-on regrouper des éléments sur base de la distance qui les sépare ?
 - 2 « proches » : quand deux noms sont-ils proches l'un de l'autre ?

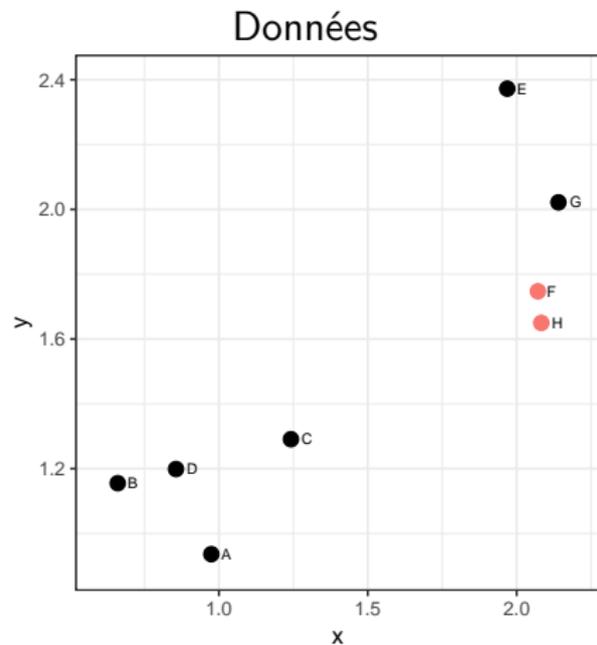
Méthodologie de normalisation

- Nous voudrions donc obtenir une sorte de « lemme » pour chaque prénom et pour chaque nom de famille ; cependant, il n'est pas question de lemmatiser, puisque les variations de graphie sont pour la plupart imprévisibles !
- Nous pouvons néanmoins appliquer un autre genre de méthode, en remarquant que les différentes graphies d'un même anthroponyme sont sans doute assez semblables les unes aux autres.
- Nous pourrions dès lors tenter de « regrouper » les graphies les plus « proches » et assigner à chacun des groupes ainsi constitués une graphie normalisée.
- Pour ce faire, nous devons comprendre ce que signifient :
 - 1 « regrouper » : comment peut-on regrouper des éléments sur base de la distance qui les sépare ?
 - 1 Il faut une méthode de regroupement d'éléments d'après la distance qui les sépare.
 - 2 « proches » : quand deux noms sont-ils proches l'un de l'autre ?
 - 2 Il faut une notion de distance entre deux noms.

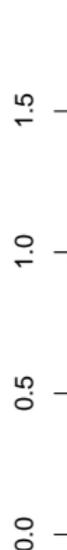
Clustering hiérarchique



Clustering hiérarchique

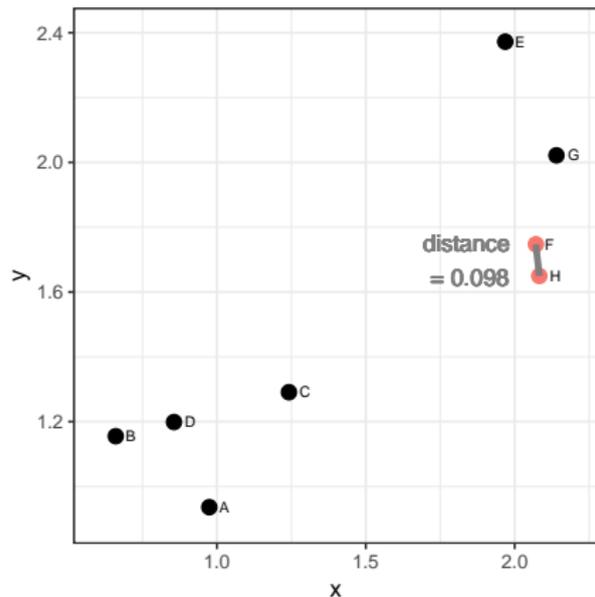


Dendrogramme

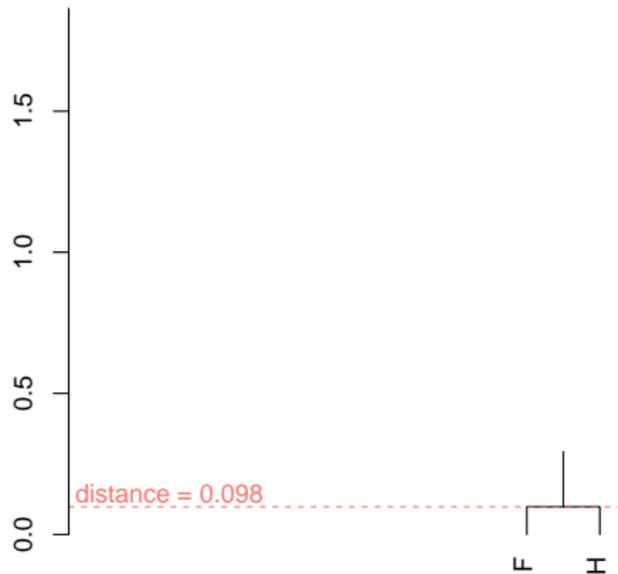


Clustering hiérarchique

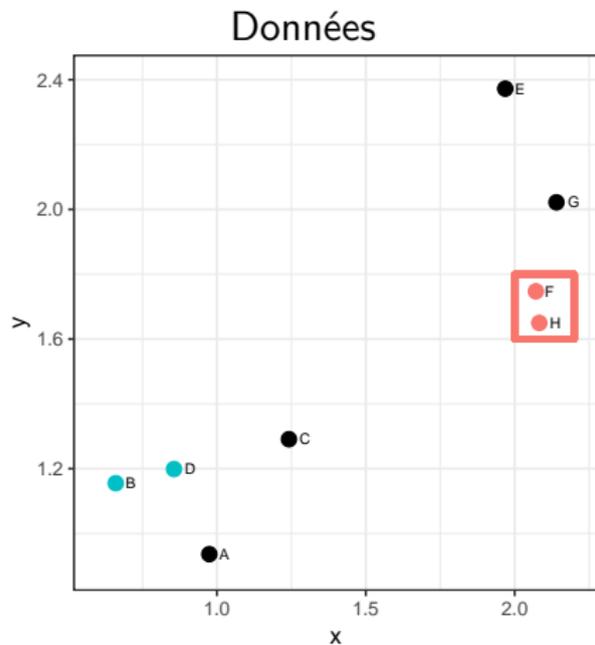
Données



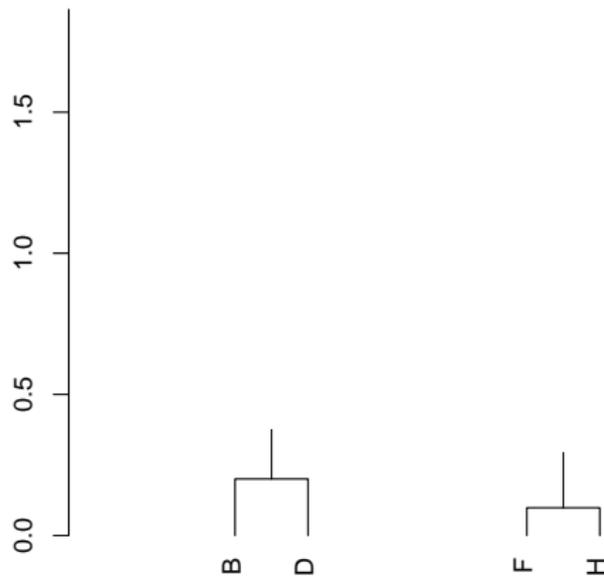
Dendrogramme



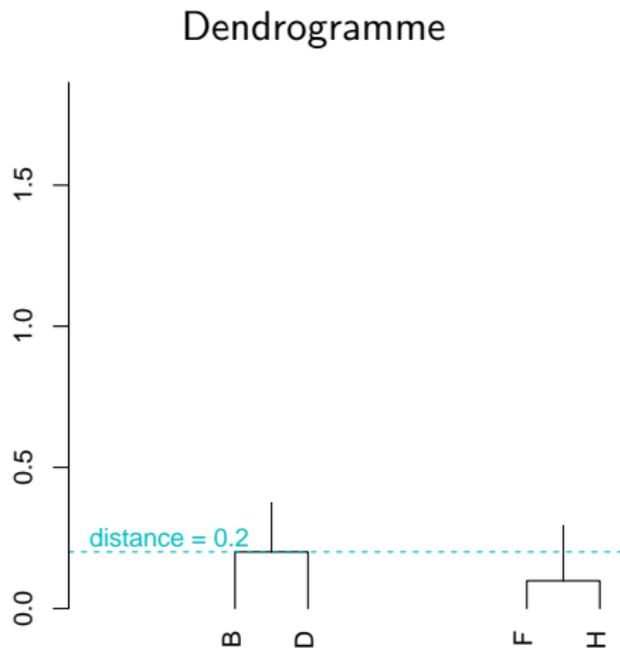
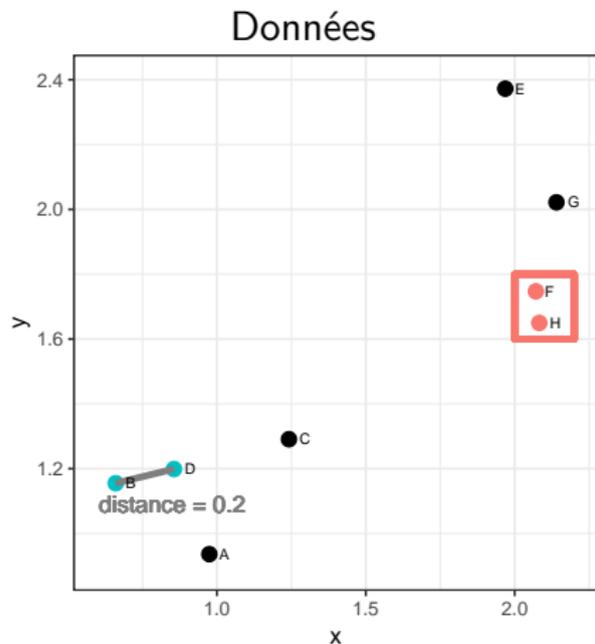
Clustering hiérarchique



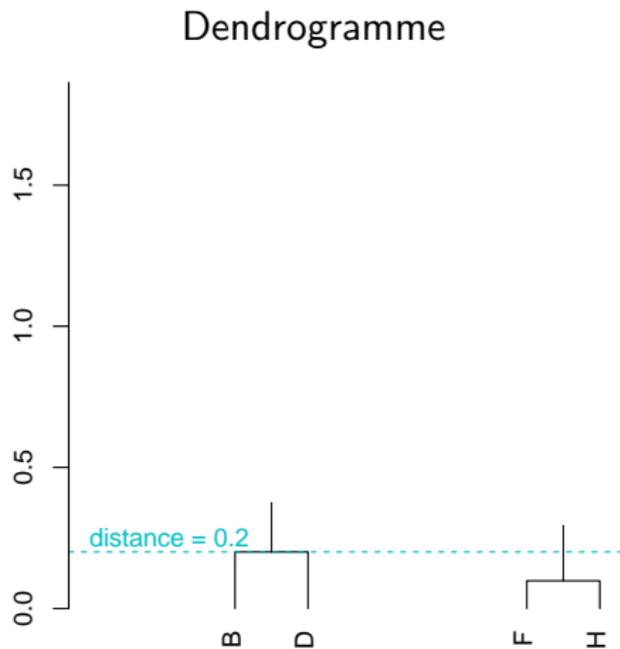
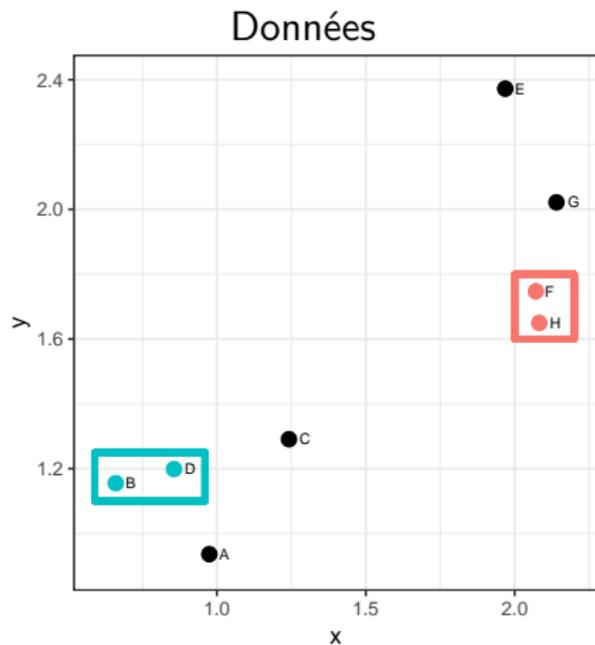
Dendrogramme



Clustering hiérarchique

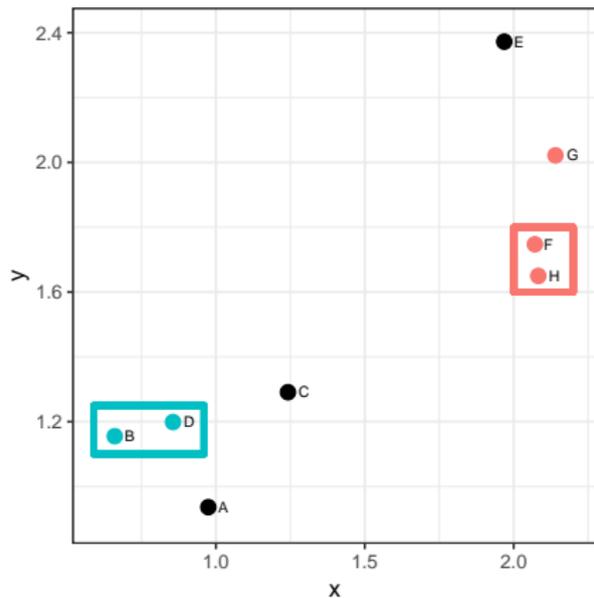


Clustering hiérarchique

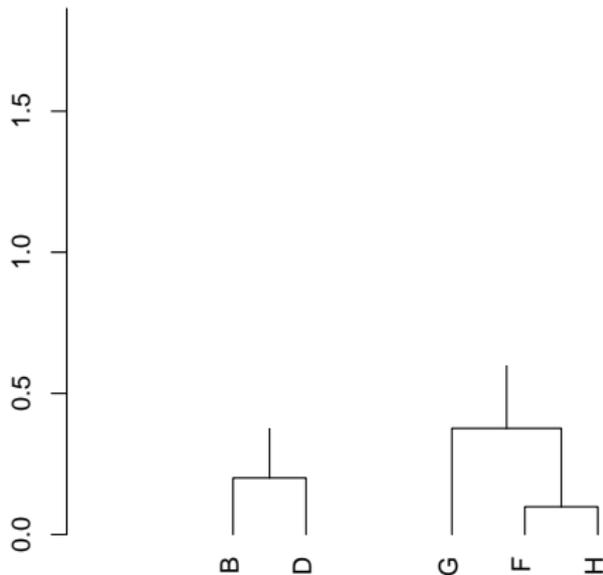


Clustering hiérarchique

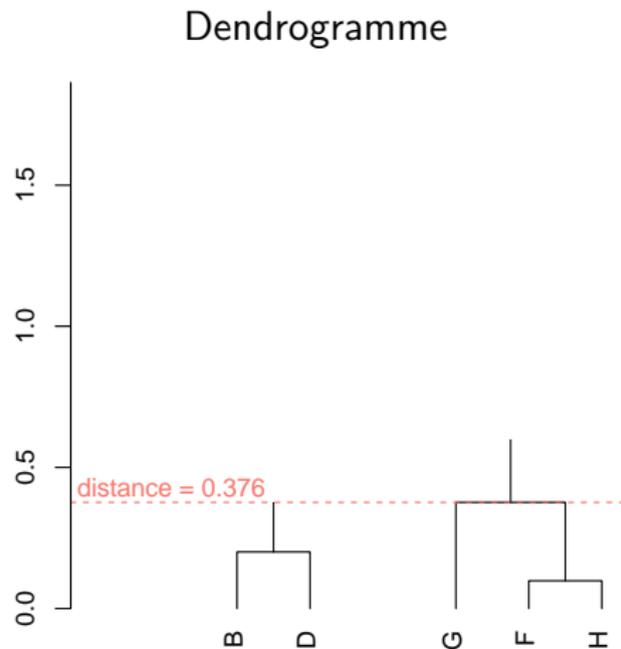
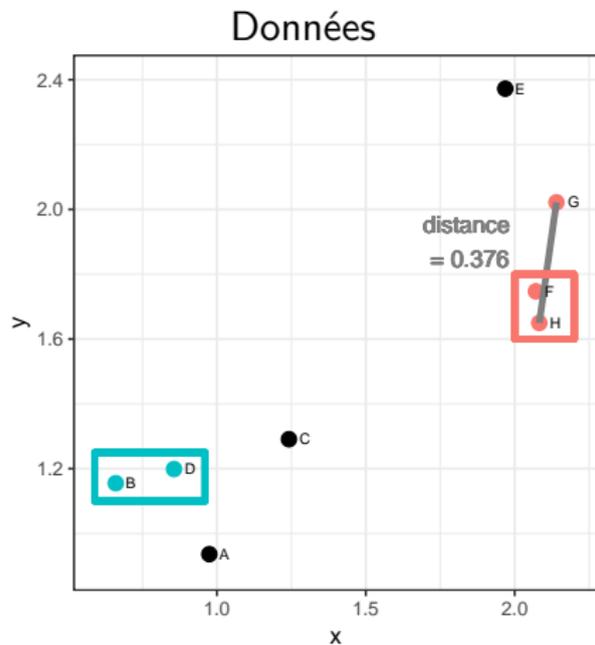
Données



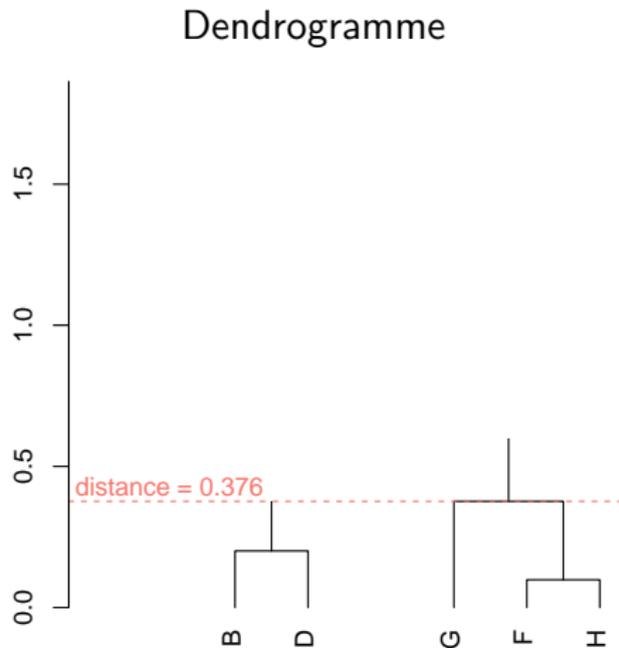
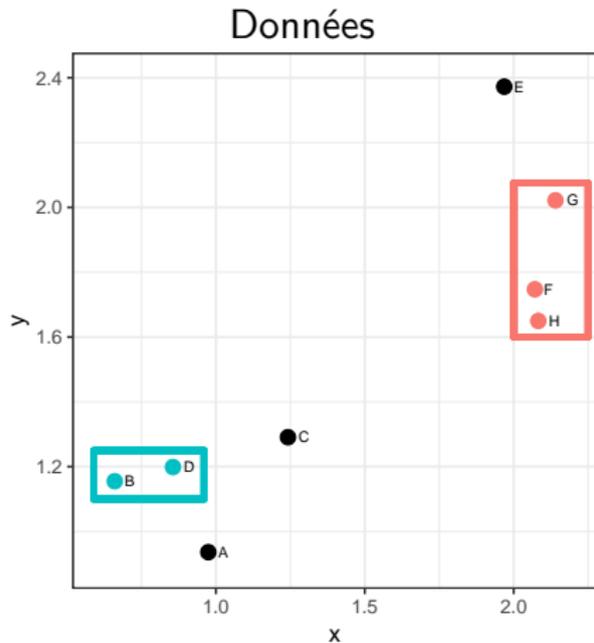
Dendrogramme



Clustering hiérarchique

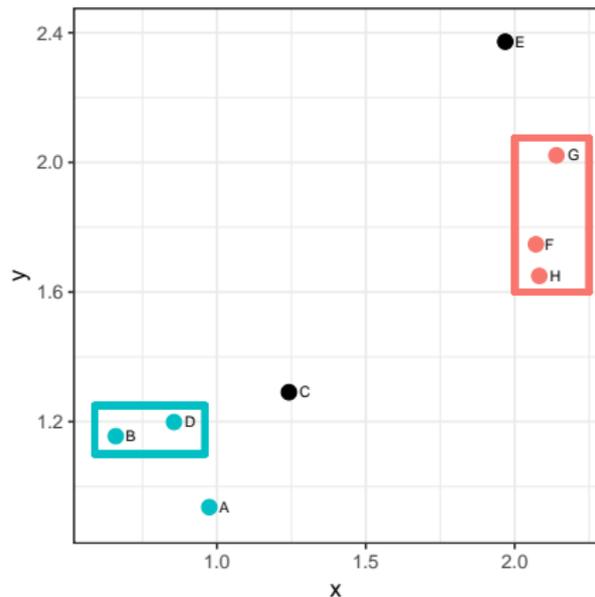


Clustering hiérarchique

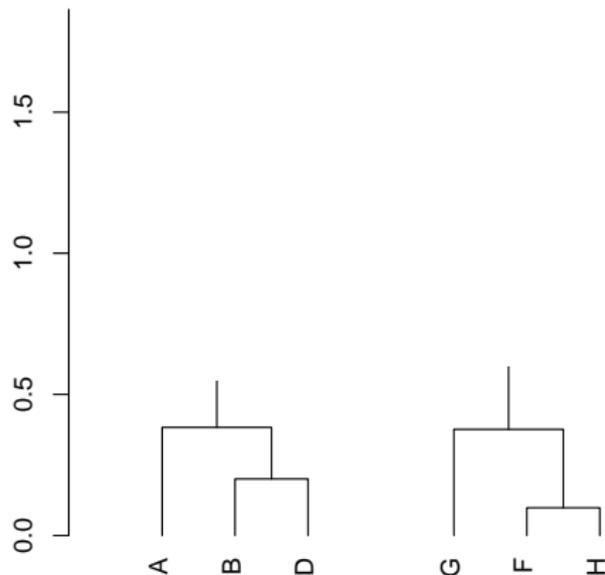


Clustering hiérarchique

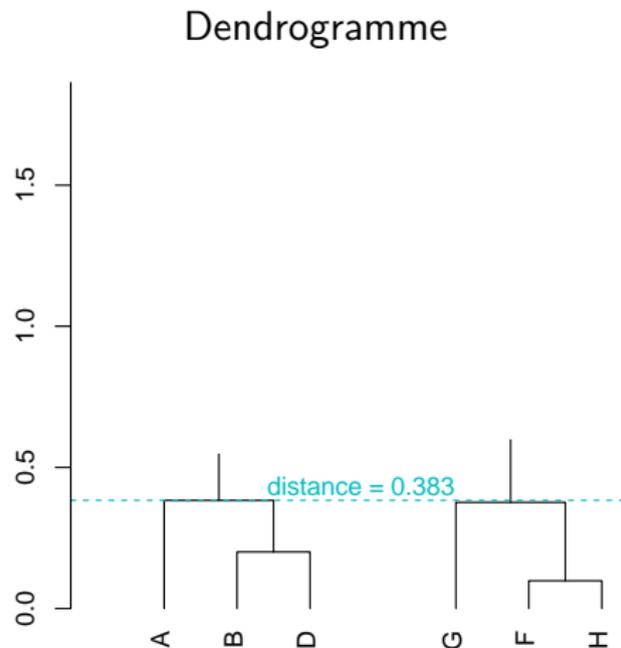
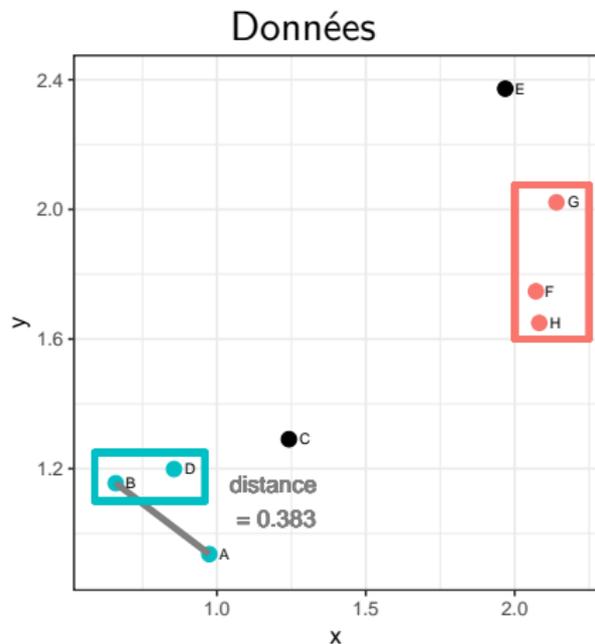
Données



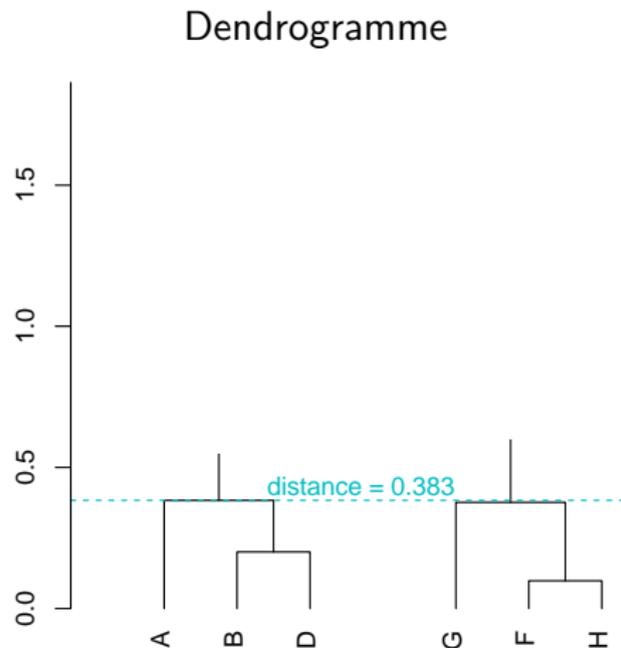
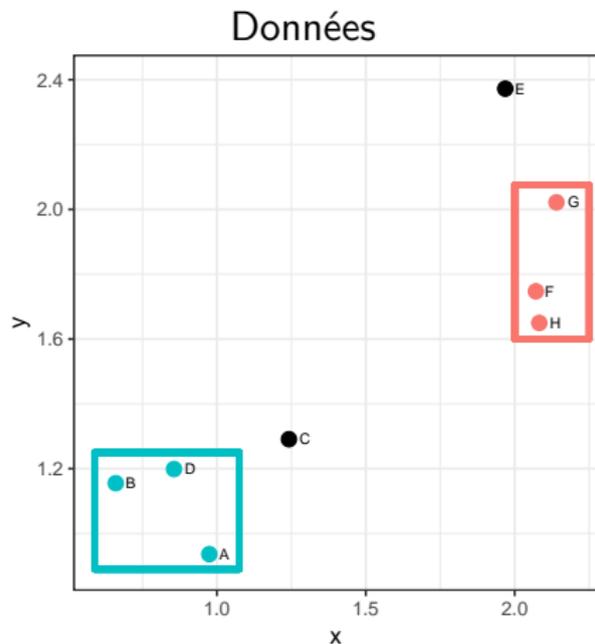
Dendrogramme



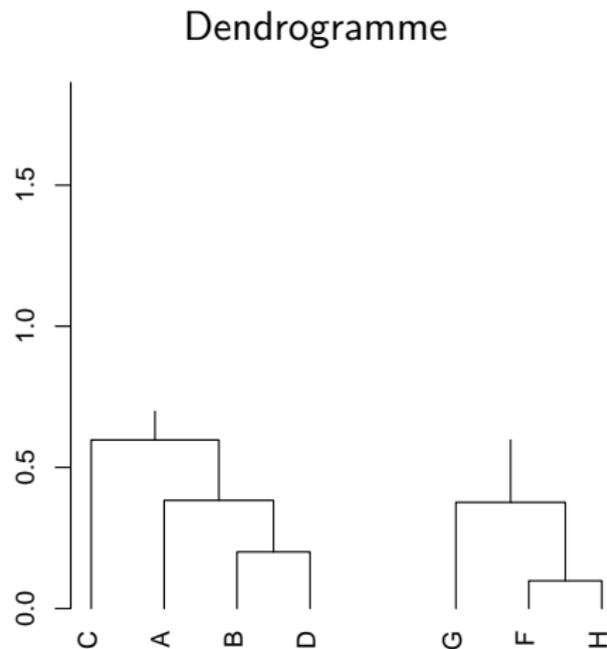
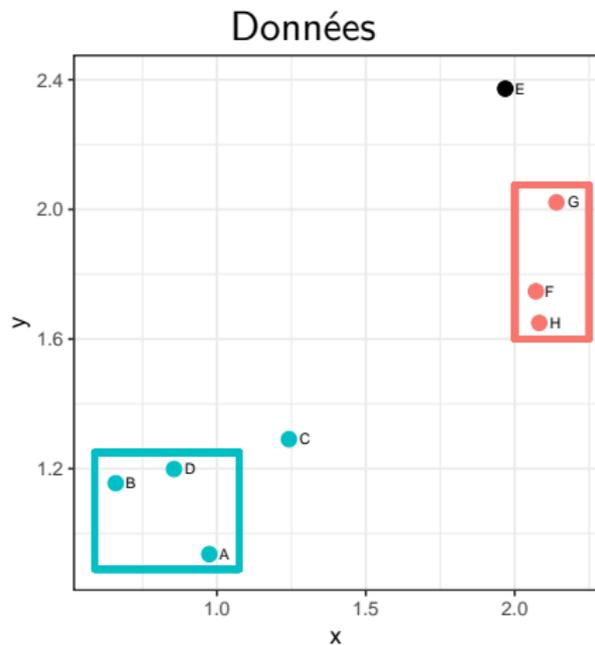
Clustering hiérarchique



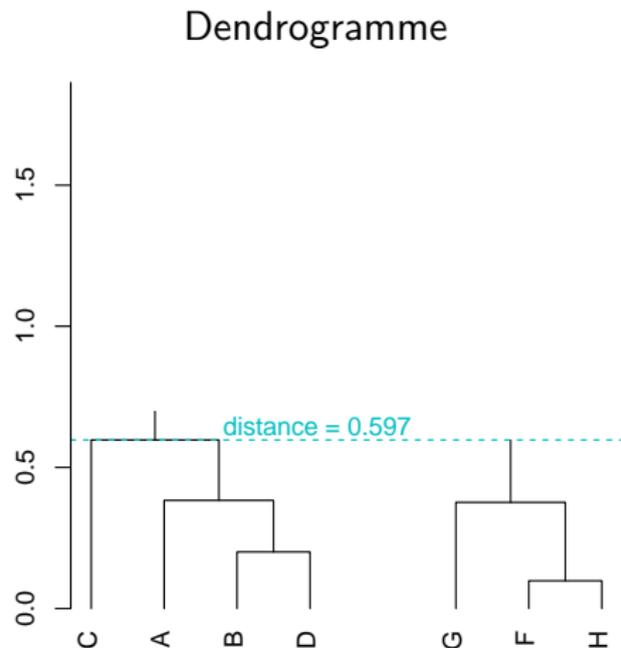
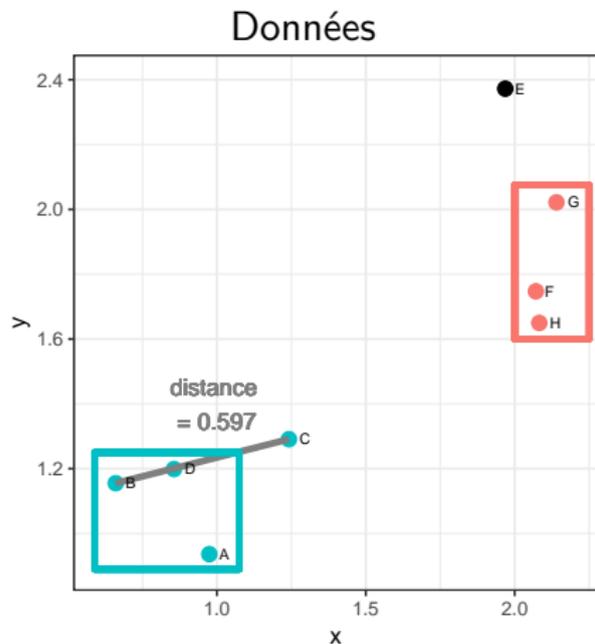
Clustering hiérarchique



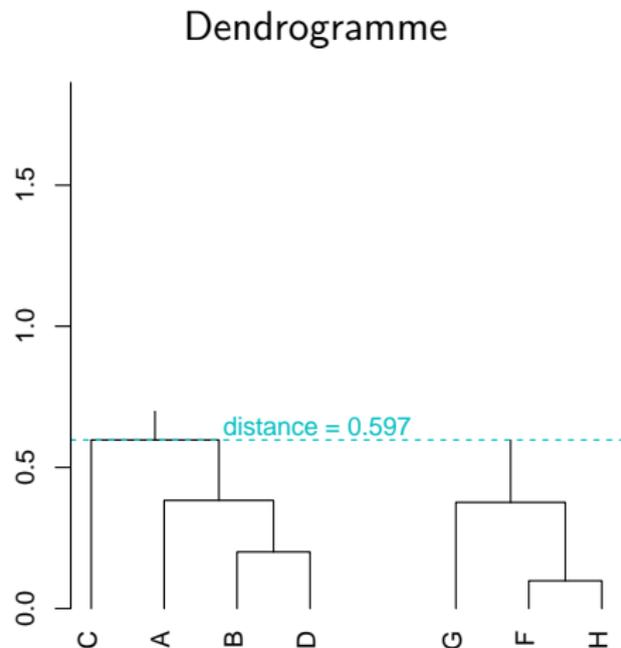
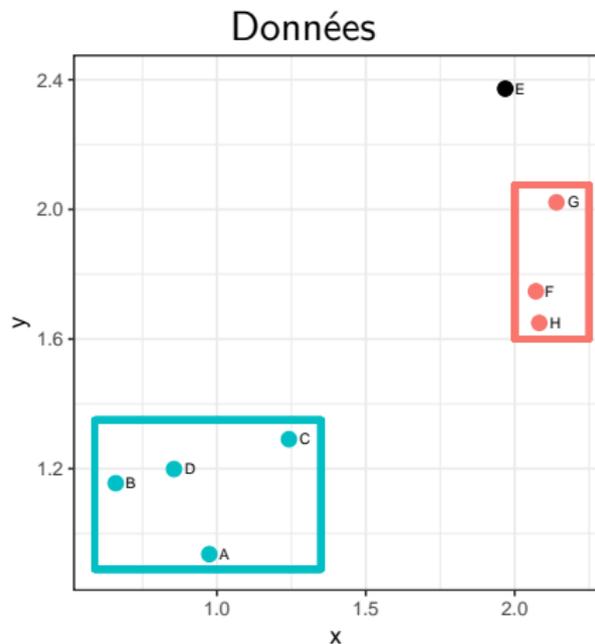
Clustering hiérarchique



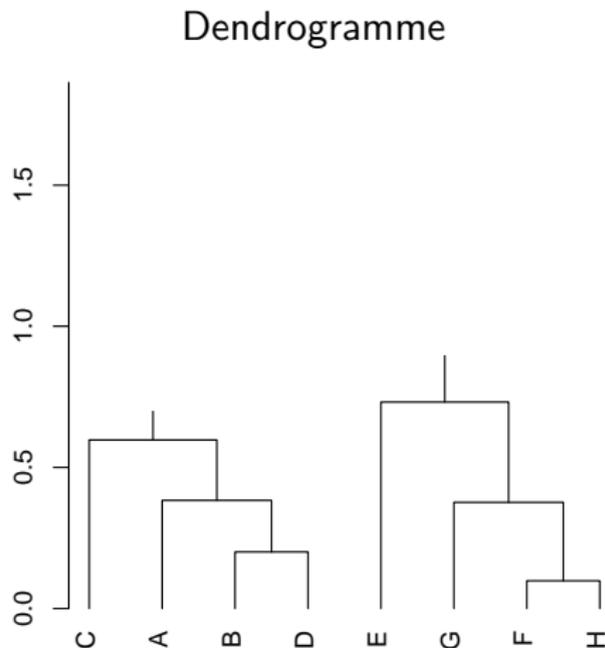
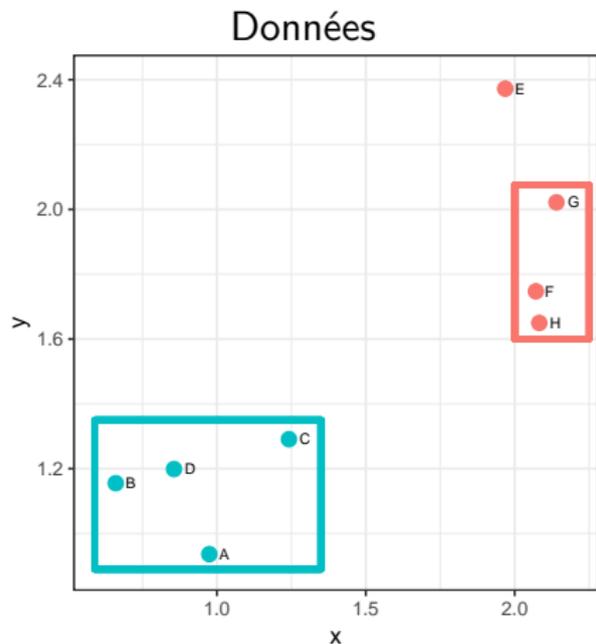
Clustering hiérarchique



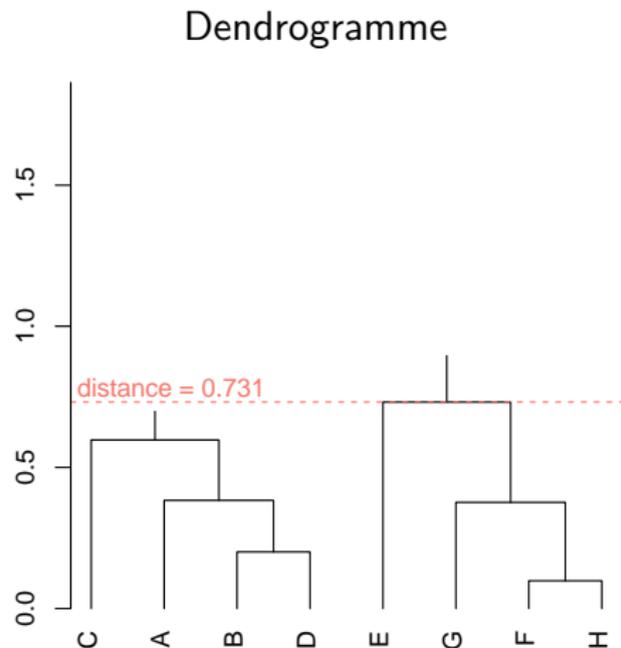
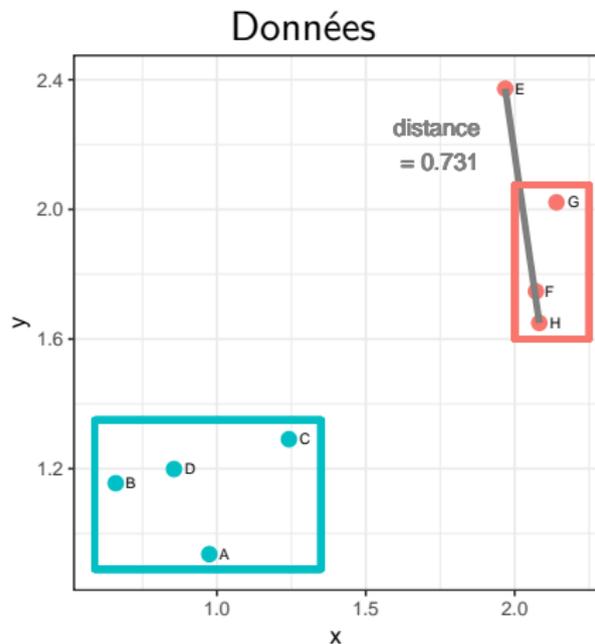
Clustering hiérarchique



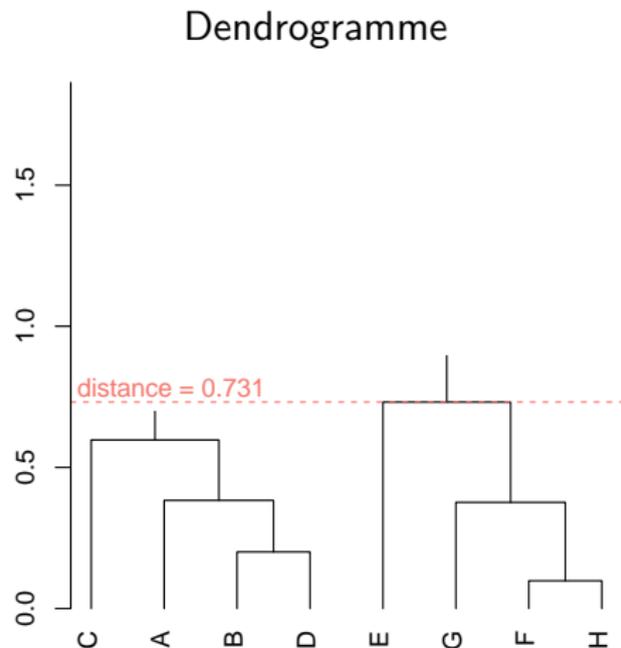
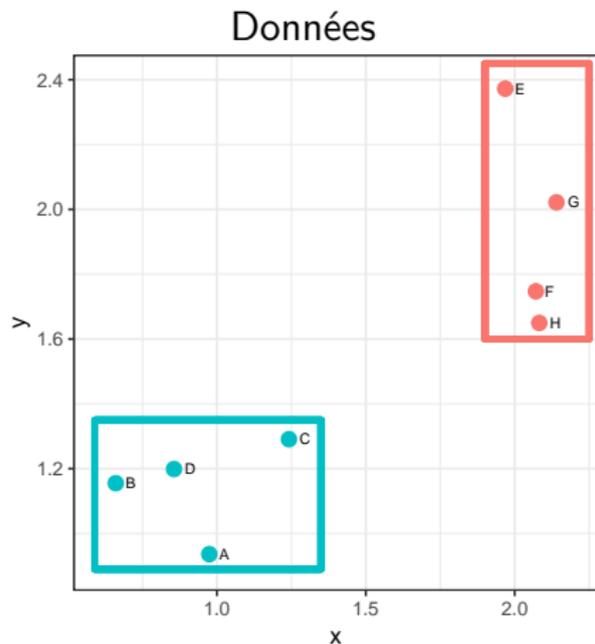
Clustering hiérarchique



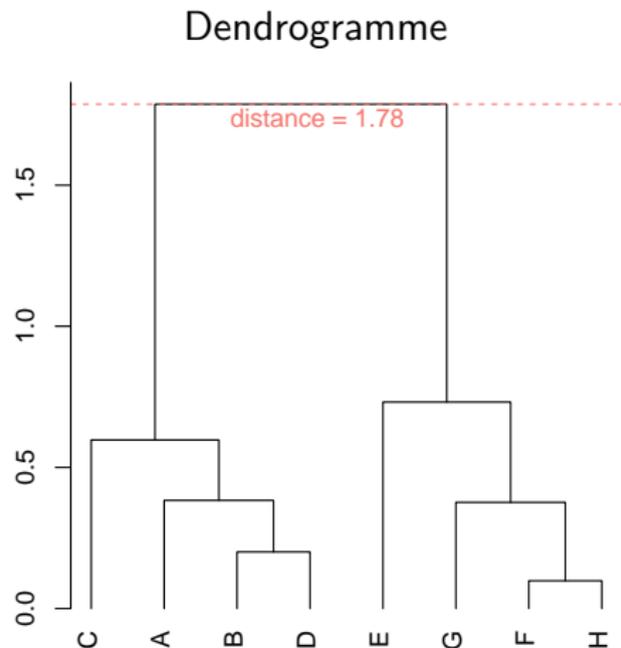
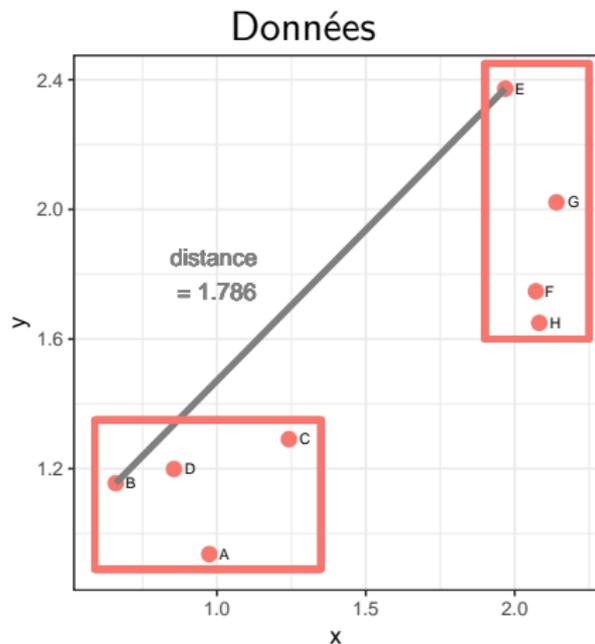
Clustering hiérarchique



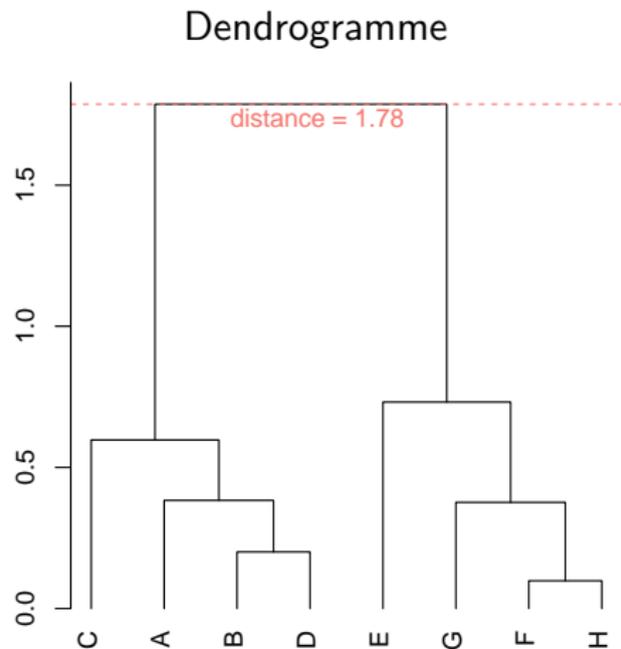
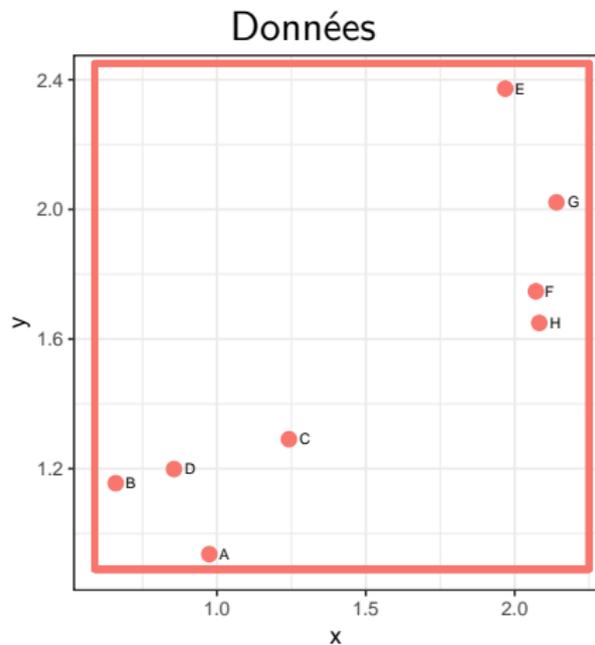
Clustering hiérarchique



Clustering hiérarchique



Clustering hiérarchique



Distance sur les chaînes de caractère

- Pour deux mots m_1 et m_2 , on définit la distance de Damerau-Levenshtein comme le nombre minimal d'opérations à effectuer sur les lettres de m_1 pour obtenir m_2 .
- Ces opérations s'appliquent :
 - à un seul caractère : insertion, suppression ou substitution ;
 - à deux caractères adjacents : transposition.

Distance sur les chaînes de caractère

- Pour deux mots m_1 et m_2 , on définit la distance de Damerau-Levenshtein comme le nombre minimal d'opérations à effectuer sur les lettres de m_1 pour obtenir m_2 .
- Ces opérations s'appliquent :
 - à un seul caractère : insertion, suppression ou substitution ;
 - à deux caractères adjacents : transposition.
- Exemples :
 - $\text{dist}(\text{"Nueveglise"}, \text{"Noeveglise"}) = 1$ car
1 "Nueveglise" \rightarrow "Noeveglise"

Distance sur les chaînes de caractère

- Pour deux mots m_1 et m_2 , on définit la distance de Damerau-Levenshtein comme le nombre minimal d'opérations à effectuer sur les lettres de m_1 pour obtenir m_2 .
- Ces opérations s'appliquent :
 - à un seul caractère : insertion, suppression ou substitution ;
 - à deux caractères adjacents : transposition.
- Exemples :
 - $\text{dist}(\text{"Nueveglise"}, \text{"Noeveglise"}) = 1 \text{ car}$
 - 1 "Nu~~e~~veglise" → "No~~e~~veglise"
 - $\text{dist}(\text{"Nueveglise"}, \text{"Noveglise"}) = 2 \text{ car}$
 - 1 "Nu~~e~~veglise" → "Neveglise"
 - 2 "Ne~~v~~veglise" → "No~~e~~veglise"

Distance sur les chaînes de caractère

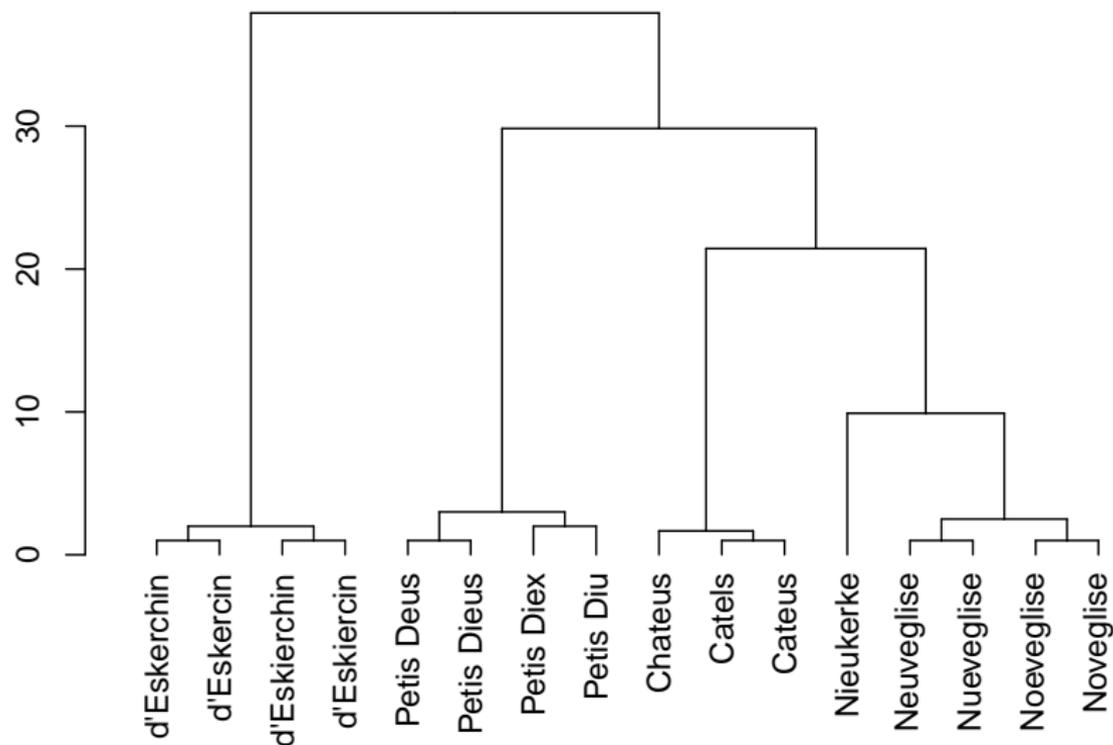
- Pour deux mots m_1 et m_2 , on définit la distance de Damerau-Levenshtein comme le nombre minimal d'opérations à effectuer sur les lettres de m_1 pour obtenir m_2 .
- Ces opérations s'appliquent :
 - à un seul caractère : insertion, suppression ou substitution ;
 - à deux caractères adjacents : transposition.
- Exemples :
 - $\text{dist}(\text{"Nueveglise"}, \text{"Noeveglise"}) = 1$ car
 - 1 "Nu~~e~~veglise" \rightarrow "No~~e~~veglise"
 - $\text{dist}(\text{"Nueveglise"}, \text{"Noveglise"}) = 2$ car
 - 1 "Nu~~e~~veglise" \rightarrow "Neveglise"
 - 2 "Ne~~v~~veglise" \rightarrow "No~~e~~veglise"
 - $\text{dist}(\text{"Nueveglise"}, \text{"Nieukerke"}) = 7$

Distance sur les chaînes de caractère

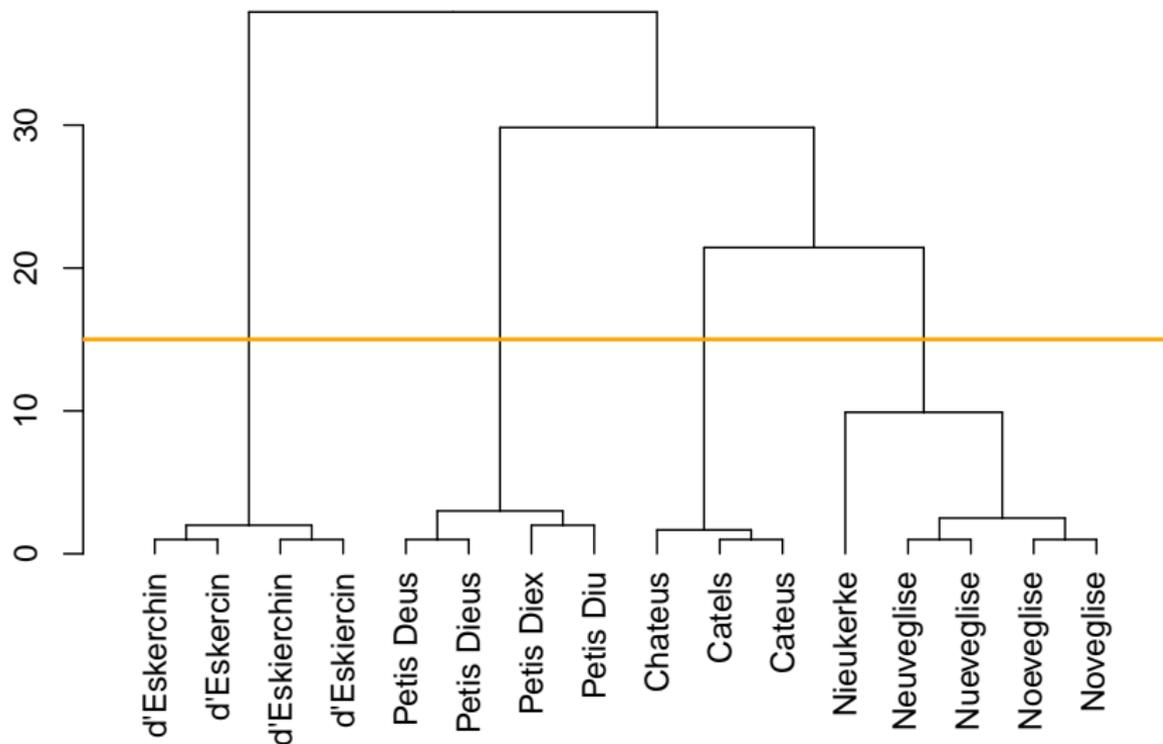
- Pour deux mots m_1 et m_2 , on définit la distance de Damerau-Levenshtein comme le nombre minimal d'opérations à effectuer sur les lettres de m_1 pour obtenir m_2 .
- Ces opérations s'appliquent :
 - à un seul caractère : insertion, suppression ou substitution ;
 - à deux caractères adjacents : transposition.
- Exemples :
 - $\text{dist}(\text{"Nueveglise"}, \text{"Noeveglise"}) = 1$ car
 - 1 "Nueveglise" \rightarrow "Noeveglise"
 - $\text{dist}(\text{"Nueveglise"}, \text{"Noveglise"}) = 2$ car
 - 1 "Nueveglise" \rightarrow "Neveglise"
 - 2 "Neveglise" \rightarrow "Noeveglise"
 - $\text{dist}(\text{"Nueveglise"}, \text{"Nieukerke"}) = 7$
- D'autres distances peuvent être définies, qui sont mieux adaptées au contexte linguistique des anthroponymes traités : on peut ainsi diminuer le « coût » des substitutions qui sont naturelles :

ae \equiv aa ; -rt \equiv -rd ; ch \equiv c ; gh \equiv ck

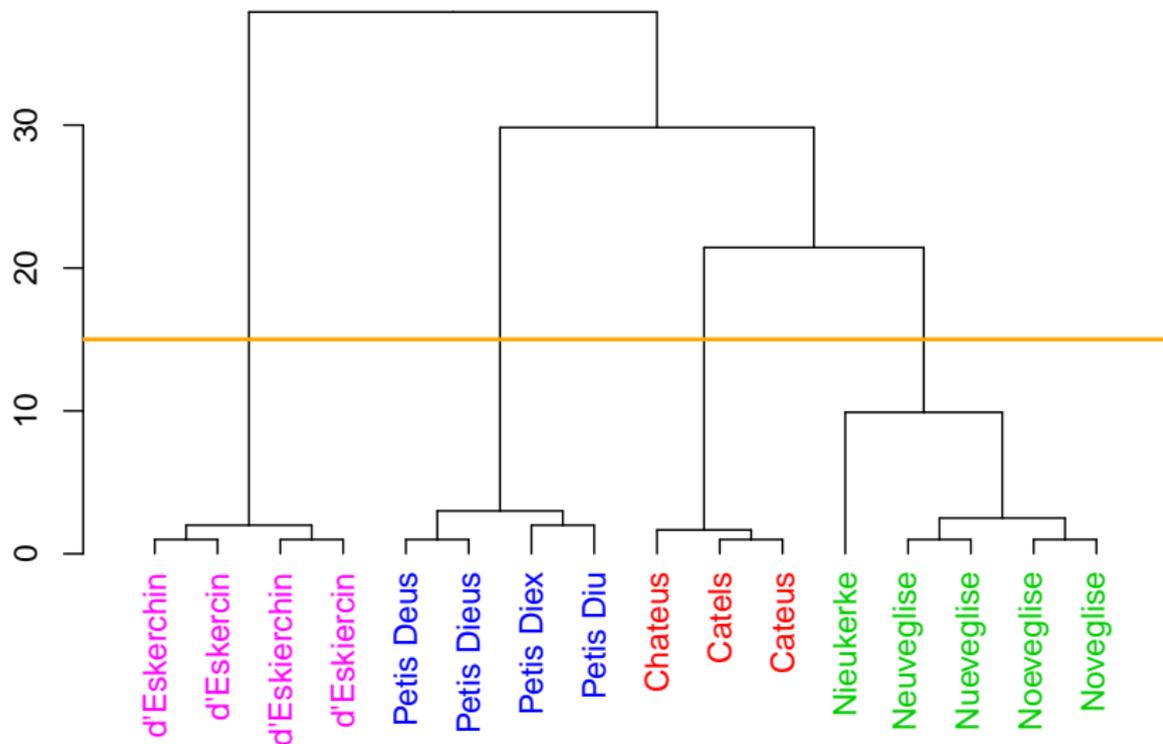
- Le dendrogramme produit permet de normaliser les anthroponymes :



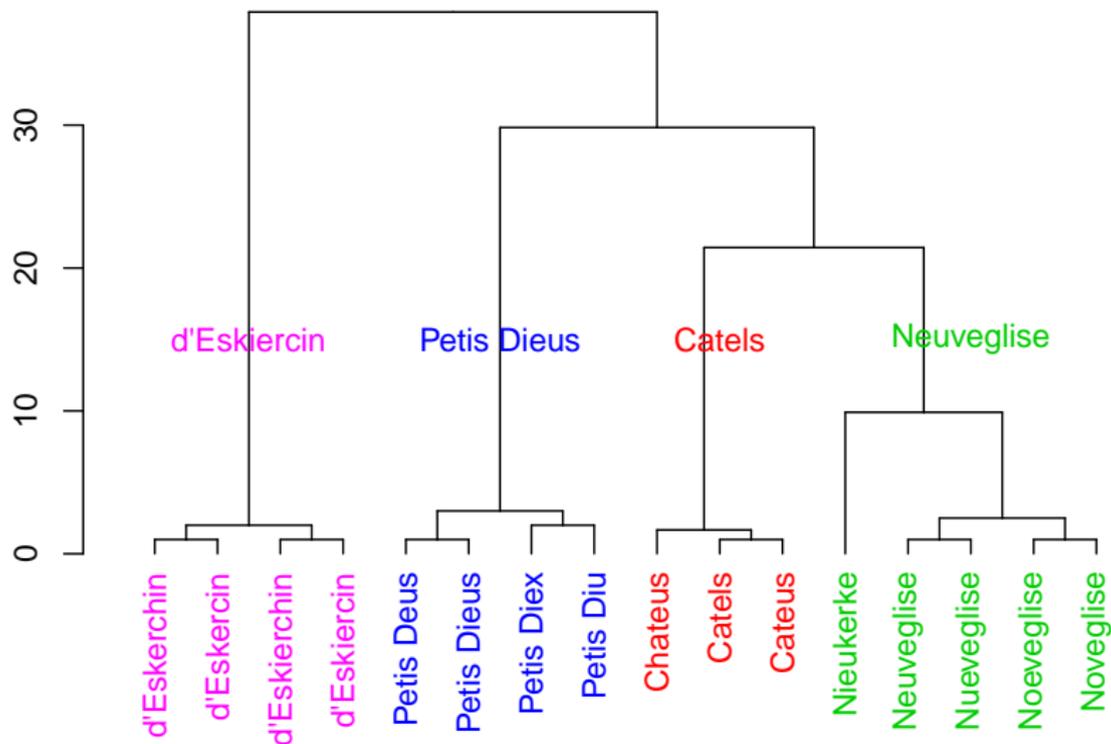
- Le dendrogramme produit permet de normaliser les anthroponymes :



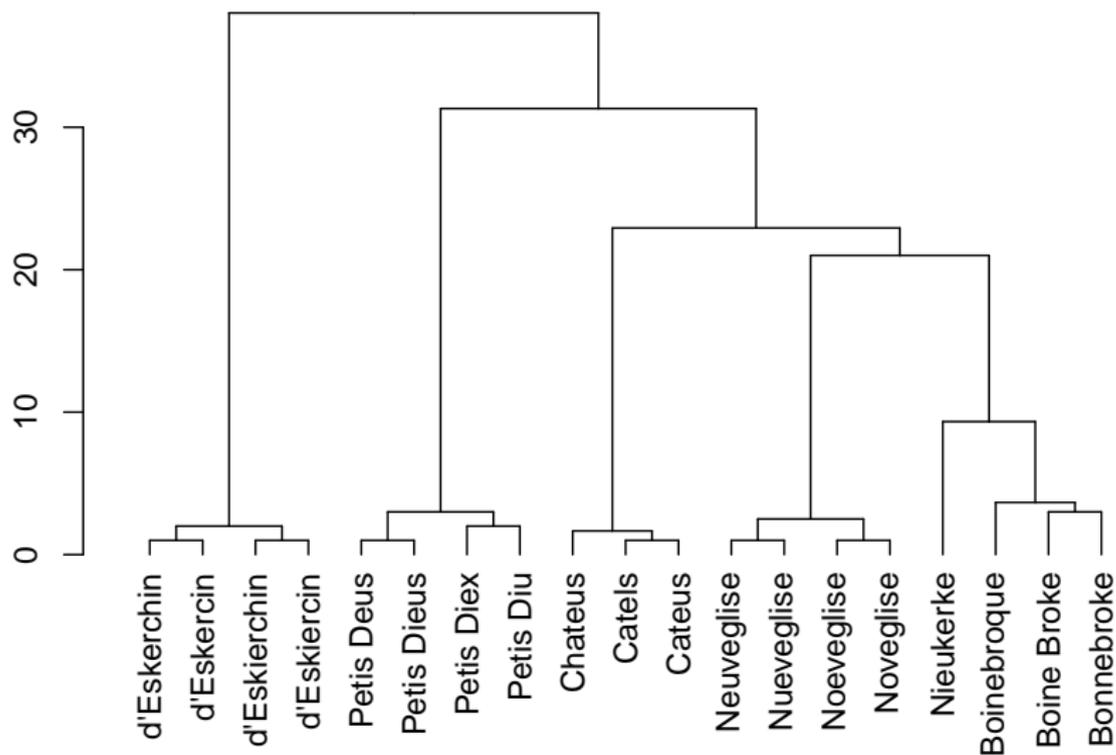
- Le dendrogramme produit permet de normaliser les anthroponymes :



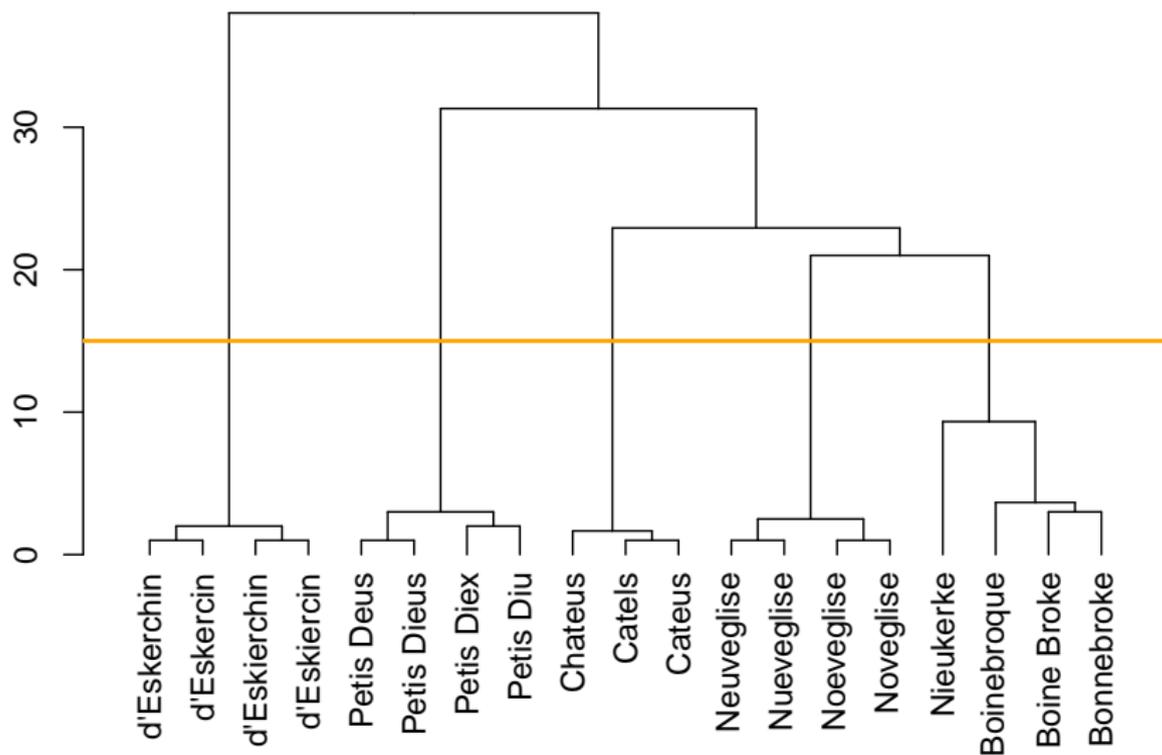
- Le dendrogramme produit permet de normaliser les anthroponymes :



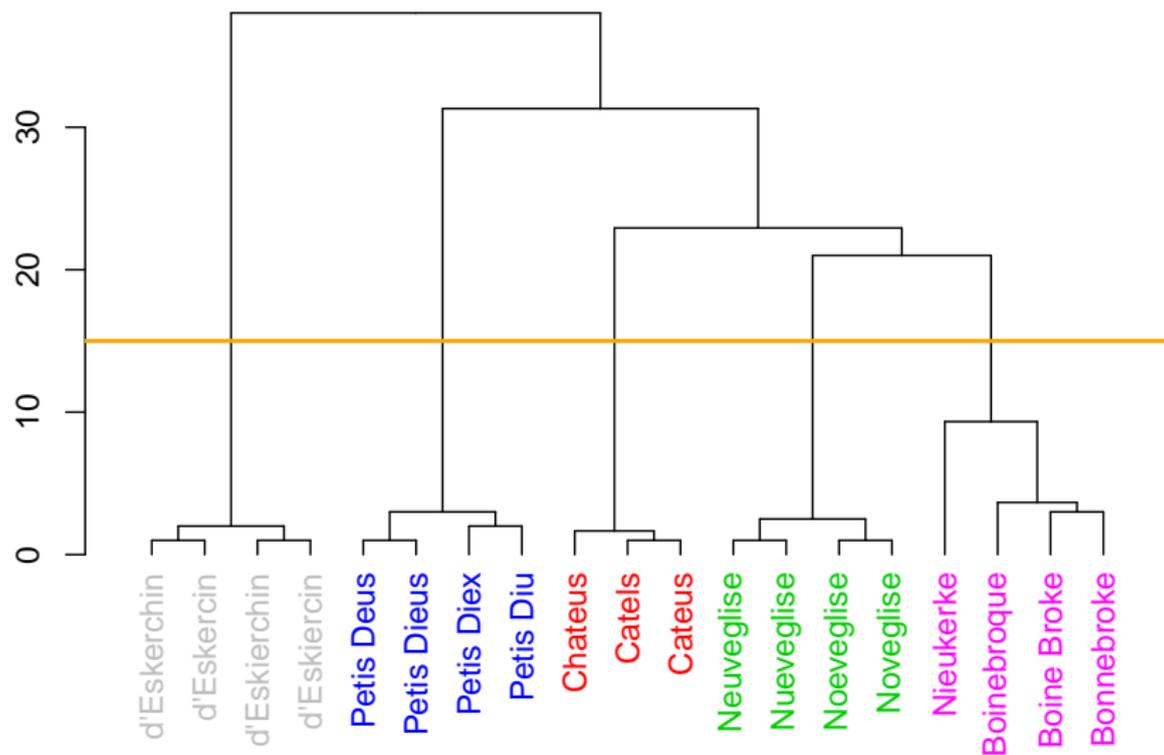
- Certaines normalisations sont néanmoins incorrectes :



- Certaines normalisations sont néanmoins incorrectes :



- Certaines normalisations sont néanmoins incorrectes :



Réseaux de crédit

Éléments d'analyse critique

- La représentation du réseau de crédit yprois permet de poser plusieurs questions à propos des cautions personnelles, appelées « plèges » dans les sources :

- La représentation du réseau de crédit yprois permet de poser plusieurs questions à propos des cautions personnelles, appelées « plèges » dans les sources :
 - au sein de quelle « sphère sociale » sont-ils sélectionnés ?
 - leur « solvabilité » est-elle prise en considération ?
 - quel place tiennent-ils au sein du réseau de crédit ?

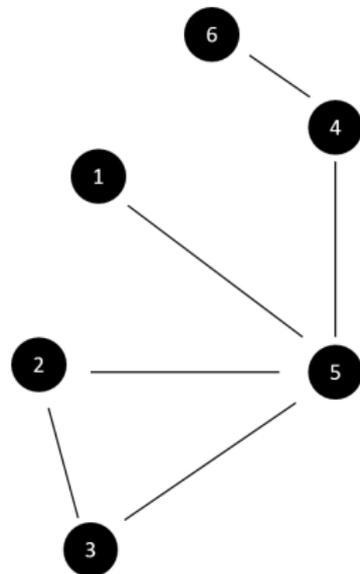
- La représentation du réseau de crédit yprois permet de poser plusieurs questions à propos des cautions personnelles, appelées « plèges » dans les sources :
 - au sein de quelle « sphère sociale » sont-ils sélectionnés ?
 - leur « solvabilité » est-elle prise en considération ?
 - quel place tiennent-ils au sein du réseau de crédit ?
- Pour répondre à cette dernière question, on peut s'interroger sur les propriétés mathématiques des noeuds du graphe qui représentent les individus qui jouent le rôle de garant.

- La représentation du réseau de crédit yprois permet de poser plusieurs questions à propos des cautions personnelles, appelées « plèges » dans les sources :
 - au sein de quelle « sphère sociale » sont-ils sélectionnés ?
 - leur « solvabilité » est-elle prise en considération ?
 - quel place tiennent-ils au sein du réseau de crédit ?
- Pour répondre à cette dernière question, on peut s'interroger sur les propriétés mathématiques des noeuds du graphe qui représentent les individus qui jouent le rôle de garant.
- On peut notamment se demander si ces noeuds occupent une place « centrale » dans le graphe.
- Pour cela, on calcule une métrique de centralité :

$$\text{centralité de proximité } (s) = \frac{1}{\sum_{t \in G} \text{dist}(s, t)}$$

La centralité : exemple

- Effectuons ce calcul sur un exemple très simple :

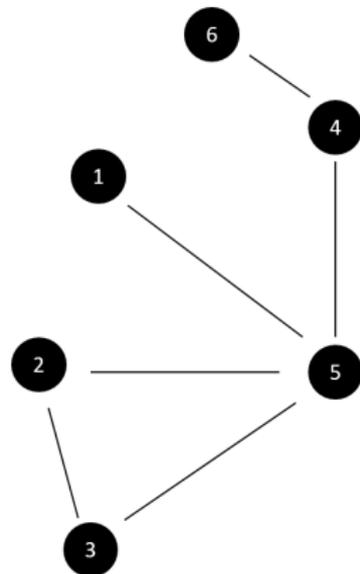


La centralité : exemple

- Effectuons ce calcul sur un exemple très simple :

La centralité du sommet 6 vaut :

$$\frac{1}{\text{dist}(6, 1) + \text{dist}(6, 2) + \text{dist}(6, 3) + \text{dist}(6, 4) + \text{dist}(6, 5)}$$
$$= \frac{1}{3 + 3 + 3 + 1 + 2} = \frac{1}{12}$$



La centralité : exemple

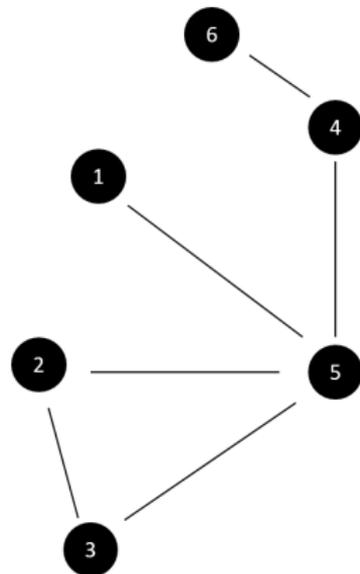
- Effectuons ce calcul sur un exemple très simple :

La centralité du sommet 6 vaut :

$$\frac{1}{\text{dist}(6,1) + \text{dist}(6,2) + \text{dist}(6,3) + \text{dist}(6,4) + \text{dist}(6,5)}$$
$$= \frac{1}{3 + 3 + 3 + 1 + 2} = \frac{1}{12}$$

La centralité du sommet 5 vaut :

$$\frac{1}{\text{dist}(5,1) + \text{dist}(5,2) + \text{dist}(5,3) + \text{dist}(5,4) + \text{dist}(5,6)}$$
$$= \frac{1}{1 + 1 + 1 + 1 + 2} = \frac{1}{6}$$



La centralité : exemple

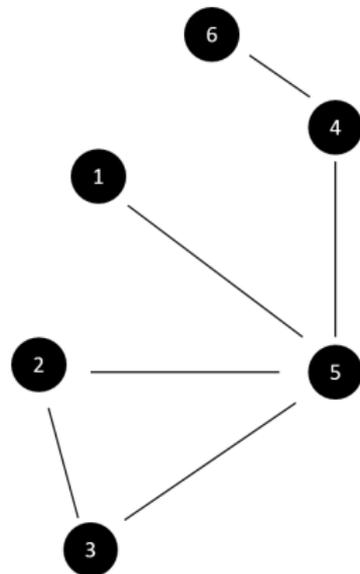
- Effectuons ce calcul sur un exemple très simple :

La centralité du sommet 6 vaut :

$$\frac{1}{\text{dist}(6,1) + \text{dist}(6,2) + \text{dist}(6,3) + \text{dist}(6,4) + \text{dist}(6,5)}$$
$$= \frac{1}{3 + 3 + 3 + 1 + 2} = \frac{1}{12}$$

La centralité du sommet 5 vaut :

$$\frac{1}{\text{dist}(5,1) + \text{dist}(5,2) + \text{dist}(5,3) + \text{dist}(5,4) + \text{dist}(5,6)}$$
$$= \frac{1}{1 + 1 + 1 + 1 + 2} = \frac{1}{6}$$



- Donc le sommet 5 est plus « central » que le sommet 6, conformément à l'intuition.

- La centralité de proximité n'est pas la seule mesure de ce type, et il est bon d'effectuer le calcul selon plusieurs points de vue.

Des pivots au sein du réseau

- La centralité de proximité n'est pas la seule mesure de ce type, et il est bon d'effectuer le calcul selon plusieurs points de vue.
- Comparons la centralité au sein du réseau de crédit des individus qui apparaissent en tant que plèges par rapport aux autres :

Mesure	Plèges	Autres	Différence
Proximité	$6,29 \cdot 10^{-8}$	$5,94 \cdot 10^{-8}$	significative
Intermédiarité	0,09%	0,01%	significative
Centralité spectrale	4,39%	1,56%	significative

Des pivots au sein du réseau

- La centralité de proximité n'est pas la seule mesure de ce type, et il est bon d'effectuer le calcul selon plusieurs points de vue.
- Comparons la centralité au sein du réseau de crédit des individus qui apparaissent en tant que plèges par rapport aux autres :

Mesure	Plèges	Autres	Différence
Proximité	$6,29 \cdot 10^{-8}$	$5,94 \cdot 10^{-8}$	significative
Intermédiarité	0,09%	0,01%	significative
Centralité spectrale	4,39%	1,56%	significative

- Nous pouvons dès lors répondre à la question posée :

Les plèges jouent le rôle de pivots au sein du réseau de crédit.

Des pivots au sein du réseau

- La centralité de proximité n'est pas la seule mesure de ce type, et il est bon d'effectuer le calcul selon plusieurs points de vue.
- Comparons la centralité au sein du réseau de crédit des individus qui apparaissent en tant que plèges par rapport aux autres :

Mesure	Plèges	Autres	Différence
Proximité	$6,29 \cdot 10^{-8}$	$5,94 \cdot 10^{-8}$	significative
Intermédiarité	0,09%	0,01%	significative
Centralité spectrale	4,39%	1,56%	significative

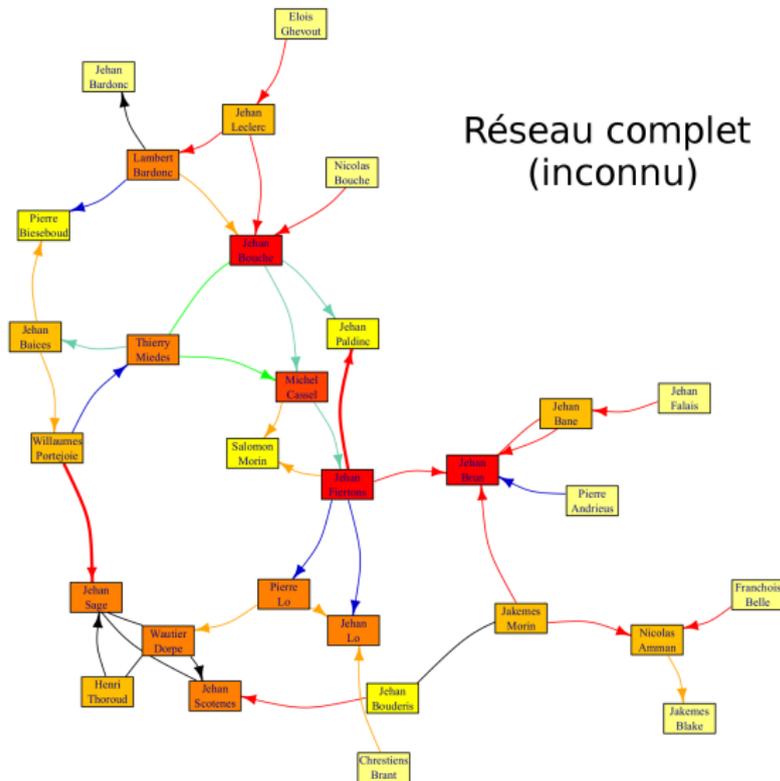
- Nous pouvons dès lors répondre à la question posée :

Les plèges jouent le rôle de pivots au sein du réseau de crédit.

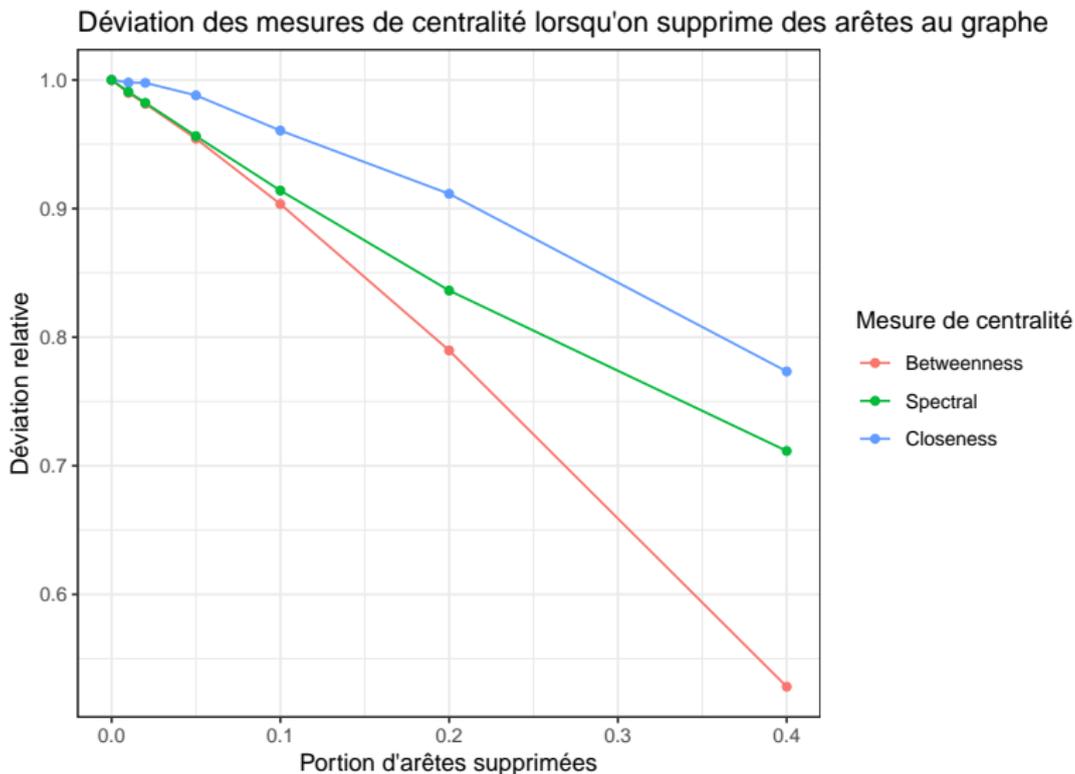
- Cela signifie donc que les débiteurs ne présentent pas un plège riche mais inconnu, ils présentent comme garant un individu qui est connu de tous et qui a sans doute la confiance des acteurs du réseau.

Critique

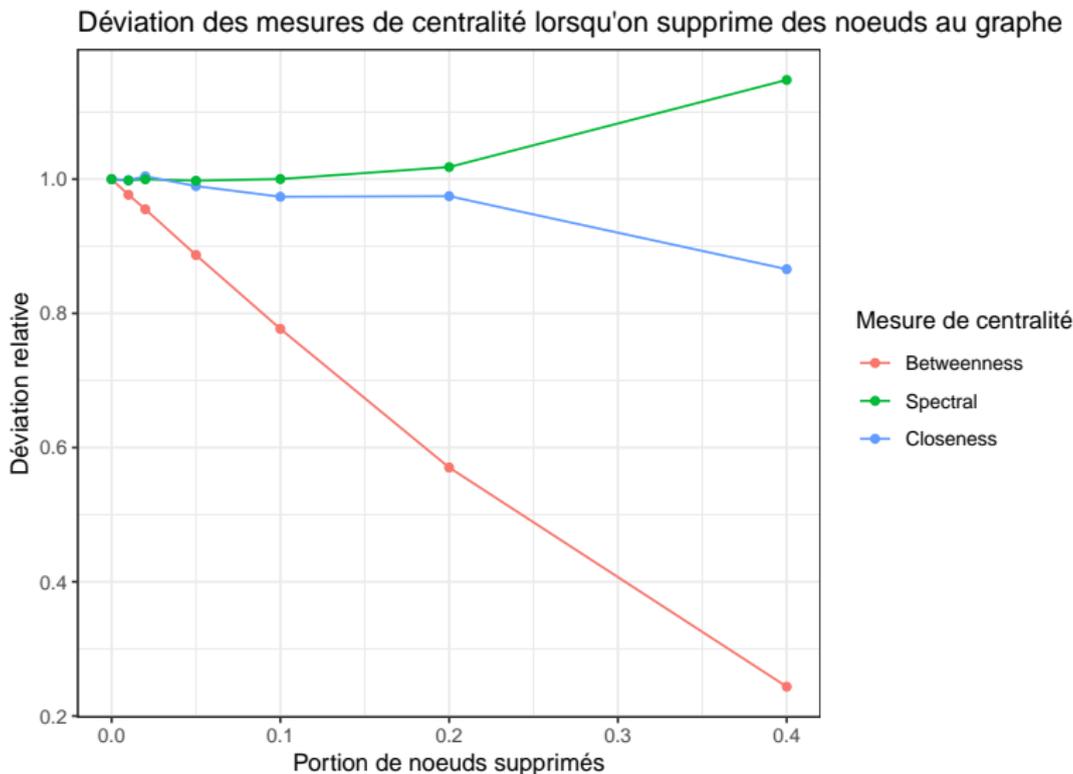
- Il ne faut pas perdre de vue le fait que nous ne disposons évidemment pas de l'entièreté du réseau de crédit yprois du XIII^e siècle.
- Les documents qui nous sont parvenus ne donnent qu'une image incomplète de la situation, comme très souvent en histoire.



- Il est dès lors très important de sélectionner des mesures robustes !



- Il est dès lors très important de sélectionner des mesures robustes !



Analyse automatique des clauses des dettes

Contexte

Nécessité d'analyser les reconnaissances de dette

- Outre les actes dont ne subsistent que des résumés, nous avons accès à un nombre important d'actes dont le texte est disponible :

	Texte disponible	Résumé disponible
Ypres	157	5.362
Tournai	658	
Douai	491	

- Parmi les questions qui peuvent être posées à ces chirographes :
 - des questions d'analyse des réseaux ;
 - des questions à propos des clauses qui les composent ;
 - ...

- L'analyse globale des reconnaissances de dette des trois villes a mis en évidence une importante variabilité dans les clauses, qui s'exprime à trois niveaux :
 - 1 au niveau des villes : les garanties associées aux dettes sont très différentes à Ypres, à Douai et à Tournai, tant du point de vue de leur fréquence que des clauses qui les signifient ;

- L'analyse globale des reconnaissances de dette des trois villes a mis en évidence une importante variabilité dans les clauses, qui s'exprime à trois niveaux :
 - 1 au niveau des villes : les garanties associées aux dettes sont très différentes à Ypres, à Douai et à Tournai, tant du point de vue de leur fréquence que des clauses qui les signifient ;
 - 2 au niveau des communautés marchandes (?) : les dettes douaisiennes qui trouvent leur origine dans le commerce du vin sont très différentes des autres, notamment en termes de garanties ;

- L'analyse globale des reconnaissances de dette des trois villes a mis en évidence une importante variabilité dans les clauses, qui s'exprime à trois niveaux :
 - 1 au niveau des villes : les garanties associées aux dettes sont très différentes à Ypres, à Douai et à Tournai, tant du point de vue de leur fréquence que des clauses qui les signifient ;
 - 2 au niveau des communautés marchandes (?) : les dettes douaisiennes qui trouvent leur origine dans le commerce du vin sont très différentes des autres, notamment en termes de garanties ;
 - 3 au niveau des individus :
 - ▶ à l'échelle de l'acte : quelles garanties sont exigées (?) par quels créanciers, pour quels débiteurs ?
 - ▶ à l'échelle de la clause : même lorsqu'on considère des actes qui comportent des clauses similaires, certains éléments diffèrent d'une dette à l'autre.

La fouille des textes (*text mining*)

- Pour comprendre la variabilité qui existe au niveau des clauses, j'ai effectué un balisage xml (semi-automatique) des (sous-)clauses des actes.

La fouille des textes (*text mining*)

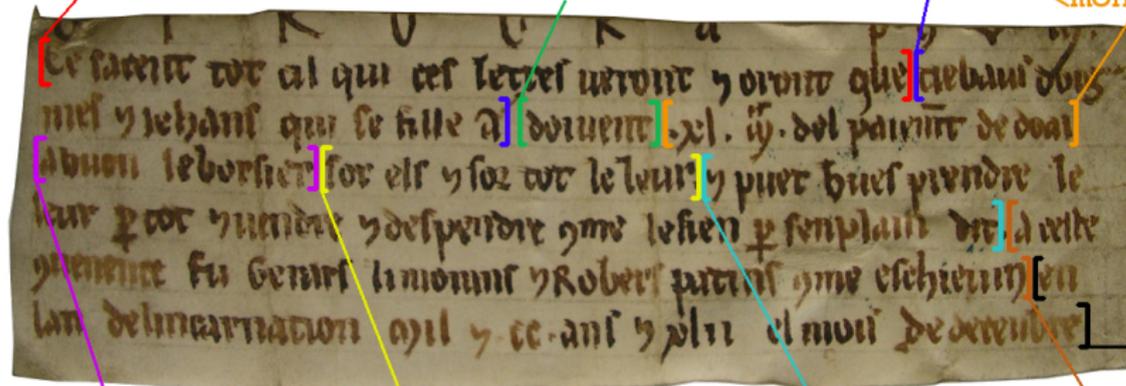
- Pour comprendre la variabilité qui existe au niveau des clauses, j'ai effectué un balisage xml (semi-automatique) des (sous-)clauses des actes.
- Chaque acte a dès lors été découpé en clauses et sous-clauses :

<notification>

<debiteur>

<action>

<montant>



<creancier>

<obligationGenerale>

<conditionsObligation>

<date>

<listeEchevins>

La fouille des textes (*text mining*)

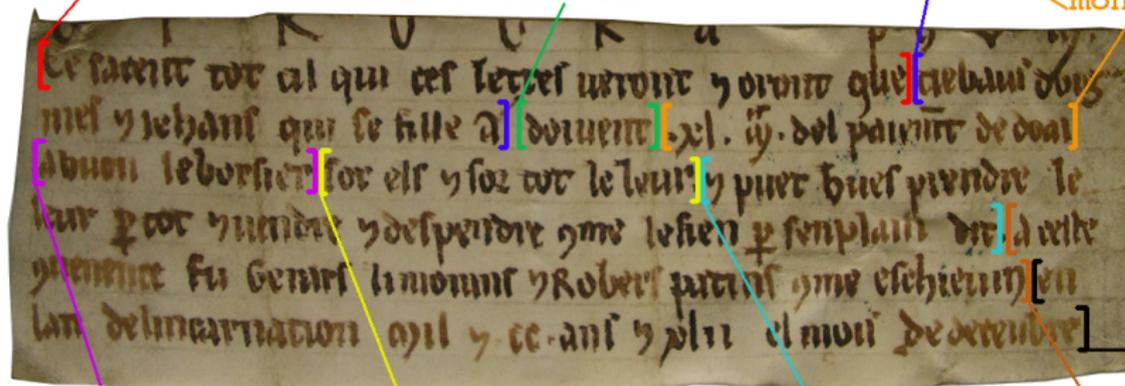
- Pour comprendre la variabilité qui existe au niveau des clauses, j'ai effectué un balisage xml (semi-automatique) des (sous-)clauses des actes.
- Chaque acte a dès lors été découpé en clauses et sous-clauses :

<notification>

<debiteur>

<action>

<montant>



<creancier>

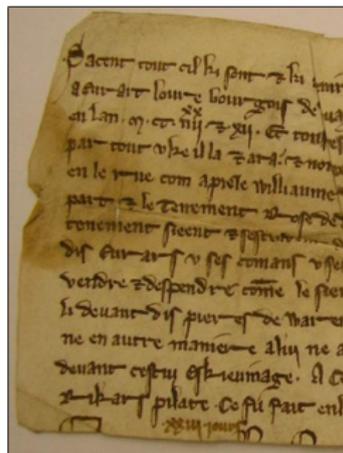
<obligationGenerale>

<conditionsObligation>

<listeEchevins>

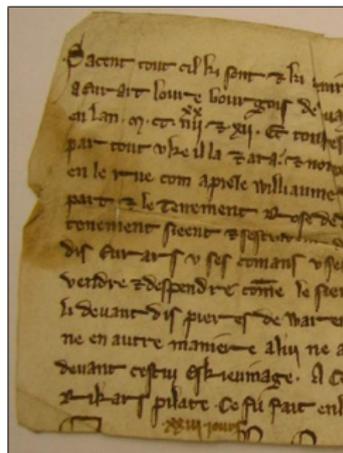
- La base de données ainsi constituée peut être utilisée telle quelle pour étudier les différents niveaux de variabilité.

De l'acte à la base de données...



Sacent tout cil ki sont et
Pieres de Waremons doit et
propre dete |a Evrart Loure
[livres] de par[isis] de bo
paiier au jour de Toussains
.m.cc.iiii*xx. et .xii. ;
devant dis a Pieres devant
et au sien |partout u ke il
nomeement soursen hiretage
tenement kil a a le Miessie
Willlaume de saint Aubin ki
Crestienne Vrient d'une |pa
de Dourges d'autre part Tou
hiretages et tout cil |tene
estendent Wit et Hierbo

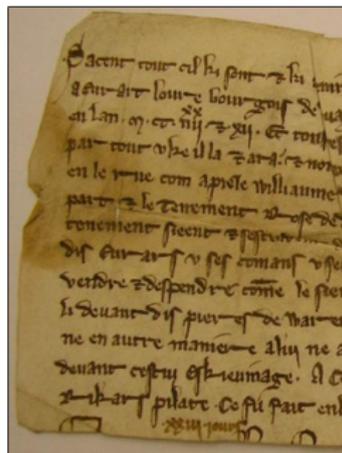
De l'acte à la base de données...



Sacent tout cil ki sont et
Pieres de Waremons doit et
propre dete |a Evrart Loure
[livres] de par[isis] de bo
paiier au jour de Toussains
.m.cc.iiii*xx. et .xii. ;
devant dis a Pieres devant
et au sien |partout u ke il
nomeement soursen hiretage
tenement kil a a le Miessis
Willlaume de saint Aubin ki
Crestienne Vrient d'une |pa
de Dourges d'autre part Tou
hiretages et tout cil |tene
estendent Wit et Hierbo

```
<acte>  
<cote>FF666/6415</cote>  
<origine>Inedit</origine>  
<ClDebut>  
  <notification>Sacent t  
  venir sont ke</notific  
<debiteur>Pieres de Wa  
<action>doit et a enc  
<propredette>co[u]me s  
dete</propredette>  
<creancier>|a Evrart I  
Douay</creancier>  
<montant> .c. [livres]  
<detteloyale>de boine
```

De l'acte à la base de données...



Sacent tout cil ki sont et
 Pieres de Waremons doit et
 propre dete |a Evrart Loure
 [livres] de par[isis] de bo
 paiier au jour de Toussains
 .m.cc.iiii*xx. et .xii. ;
 devant dis a Pieres devant
 et au sien |partout u ke i
 nomeement soursen hiretage
 tenement kil a a le Miessie
 Willaume de saint Aubin ki
 Crestienne Vrient d'une |pa
 de Dourges d'autre part Tou
 hiretages et tout cil |tene
 estendent Wit et Hierbo

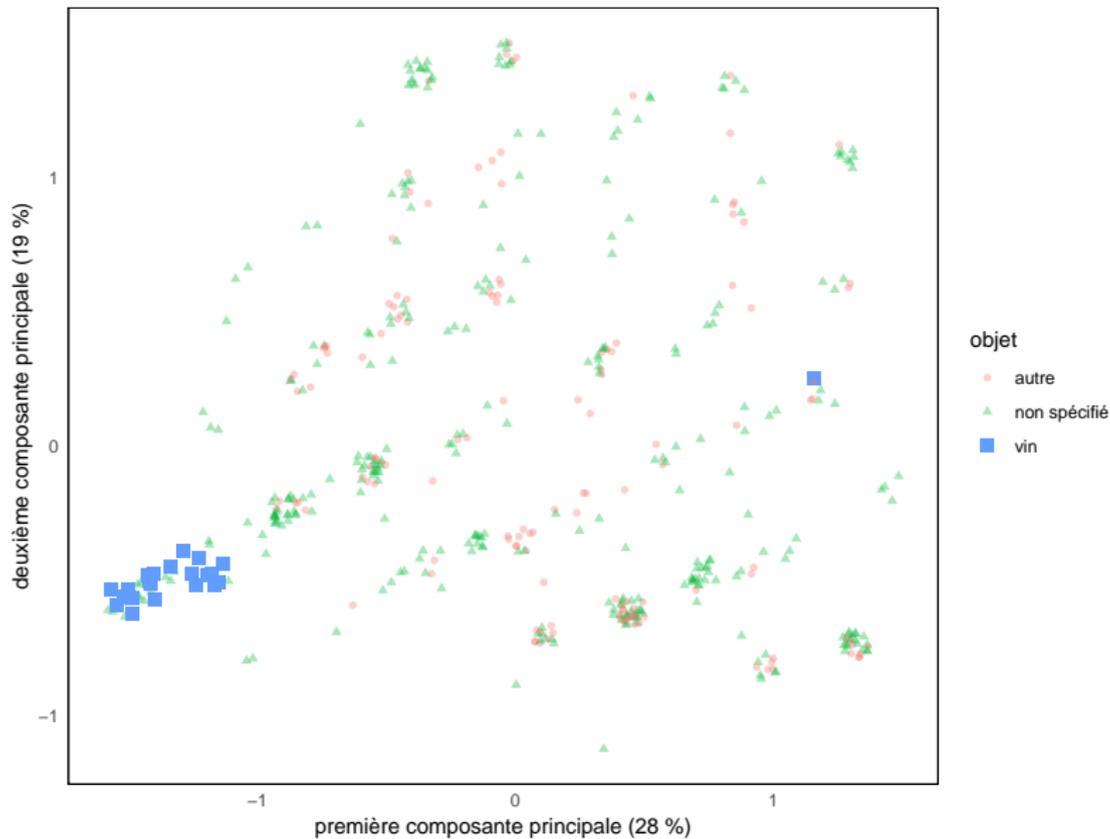
```
<acte>
  <cote>FF666/6415</cote>
  <origine>Inedit</origine>
  <ClDebut>
    <notification>Sacent t
  venir sont ke</notific
  <debiteur>Pieres de Wa
  <action>doit et a encc
  <propredette>co[u]me s
  dete</propredette>
  <creancier>|a Evrart I
  Douay</creancier>
  <montant>.c. [livres]
  <detteloyale>de boine
```

	cote	ClDebut.notification	ClDebut.debiteur	ClDebut.action
300	FF666/6413	Sachent tout cil ki sunt et ki avenir sunt ke	Andrius de Chassiel	doit et a encouvent
301	FF666/6414	Sacent tout chil q[ui] sunt et q[ui] a venir sunt q[ue]	Jehans Potins, Jehans Hakins et Estienes Poules d...	doivent et ont encouvent
302	FF666/6415	Sacent tout cil ki sont et ki a venir sont ke	Pieres de Waremons	doit et a encouvent
303	FF666/6416	Sachent tout cil ki sunt et ki avenir sunt ke	Wätiers li Roys et Martins dou Mares	doivent et ont encouvent.
304	FF666/6417	Sacent tout chil ki sont et qui avenir sont que	Jehans Crakelins, Baudes li Wärmiers, Jehans Potin...	doivent et ont encouvent.
305	FF666/6418	Sacent tout cil ki sont et ki a venir sont ke	Pieres Blavos	doit et encouvent
306	FF666/6419	Sacent tout cil ki sunt et ki a venir sunt ke	Wätiers Juaille ki maint a Vilers Vieus Kaaignicort	doit et a encouvent
307	FF666/6420	Sacent tout cil ki sont et ki a venir sont ke	Biernars Honcles	doit et a en(couv)ent

- On obtient donc une base de données dont
 - chacune des 491 lignes correspond à un acte ;
 - chacune des 143 colonnes correspond à une sous-clause.
- Il faut maintenant explorer cette base, dans l'espoir de découvrir des « motifs », par exemple des groupes d'actes partageant les mêmes caractéristiques et intégrant les mêmes clauses.

- On obtient donc une base de données dont
 - chacune des 491 lignes correspond à un acte ;
 - chacune des 143 colonnes correspond à une sous-clause.
- Il faut maintenant explorer cette base, dans l'espoir de découvrir des « motifs », par exemple des groupes d'actes partageant les mêmes caractéristiques et intégrant les mêmes clauses.
- Puisque visualiser des données dans un espace à 143 dimensions est difficile, on fait appel à des techniques de réduction de dimension, dont l'Analyse en Composantes Principales (ACP).
- Celle-ci permet de visualiser les données en deux dimensions de manière « optimale ».

Résultat de l'ACP



Analyse automatique des clauses des dettes

Pré-traitement

- Pour obtenir une telle base de données, il est bien entendu nécessaire d'effectuer des opérations à trois niveaux :
 - 1 local (actes) : découpage de chaque acte en clauses.
 - 2 interne (clauses) : extraction de l'information contenue dans certaines clauses ;
 - 3 global (villes) : normalisation des anthroponymes en considérant chaque ville dans son ensemble.

Découpage

- C'est la structure interne relativement fixe des actes qui permet leur découpage en clauses.
- Cependant, il existe de grosses variations de graphie dans les actes !

- C'est la structure interne relativement fixe des actes qui permet leur découpage en clauses.
- Cependant, il existe de grosses variations de graphie dans les actes !

Ço sacent cil ki cest escrit veront et oront que

Çou sacent tout cil ki cest escrit veront et oront que

Ce sacent tot cil ki sont et ki à venir sont et ki cest escrit veront et oront ke

Ce sacent tout cil ki cest escrit veront et oront ke

Sacent tout cil ki sunt et ki à venir sunt et cest escrit veront et oront ke

Ce sacious cil ki sunt et à venir sunt et cest escrit verunt et orunt ke

Sacent tout cil ki cest escrit verront et orront ke

Cognute chose soit à tos ceaus qui veront cest escrit ke

Cho sacent tot cil qui cest escrit oront que

Ço sacent cil ki sunt et ki à venir sunt et cest escrit veront que

Couneute cose soit à çals ki ore sunt et ki à venir sont ke

Ce sacent chil ki sunt et ki à venir sunt ke

Ce sacent tot cil ki cest escrit oront et veront que

Ce sacent cil ki sunt et ki à venir sunt et cest escrit verunt que

Les expressions régulières

- Pour effectuer le découpage, nous faisons appel aux expressions régulières, qui permettent de « chercher » de manière flexible un grand nombre de chaînes de caractères.
- Il s'agit d'une approche similaire à celle de beaucoup de moteurs de recherche, où `hist*` permet de chercher tous les mots qui commencent par `hist`.

Les expressions régulières

- Pour effectuer le découpage, nous faisons appel aux expressions régulières, qui permettent de « chercher » de manière flexible un grand nombre de chaînes de caractères.
- Il s'agit d'une approche similaire à celle de beaucoup de moteurs de recherche, où `hist*` permet de chercher tous les mots qui commencent par `hist`.
- Quelques exemples (sur base de la chaîne `abcfdrrffdsaaabdaac`) :

Les expressions régulières

- Pour effectuer le découpage, nous faisons appel aux expressions régulières, qui permettent de « chercher » de manière flexible un grand nombre de chaînes de caractères.
- Il s'agit d'une approche similaire à celle de beaucoup de moteurs de recherche, où `hist*` permet de chercher tous les mots qui commencent par `hist`.
- Quelques exemples (sur base de la chaîne `abcfdrrffdsaaabdaac`) :

- le *pipe* `|` permet de désigner plusieurs caractères à la fois :

`a|b`

`ab` cfdrrffds `aaab` d `aa` c

Les expressions régulières

- Pour effectuer le découpage, nous faisons appel aux expressions régulières, qui permettent de « chercher » de manière flexible un grand nombre de chaînes de caractères.
- Il s'agit d'une approche similaire à celle de beaucoup de moteurs de recherche, où `hist*` permet de chercher tous les mots qui commencent par `hist`.
- Quelques exemples (sur base de la chaîne `abcfdrffdsaaabdaac`) :

- le *pipe* `|` permet de désigner plusieurs caractères à la fois :

`a|b`

`ab` `cfdrffds` `aaab` `d` `aa` `c`

- les parenthèses `()` permettent de former des groupes de caractères :

`(ab)|d`

`ab` `cf` `d` `rff` `d` `saa` `abd` `aac`

Les expressions régulières

- Pour effectuer le découpage, nous faisons appel aux expressions régulières, qui permettent de « chercher » de manière flexible un grand nombre de chaînes de caractères.
- Il s'agit d'une approche similaire à celle de beaucoup de moteurs de recherche, où `hist*` permet de chercher tous les mots qui commencent par `hist`.
- Quelques exemples (sur base de la chaîne `abcfdrrffdsaaabdaac`) :

- le *pipe* `|` permet de désigner plusieurs caractères à la fois :

`a|b`

`ab cfdrrffds aaab d aa c`

- les parenthèses `()` permettent de former des groupes de caractères :

`(ab)|d`

`ab cf d rff d saa abd aac`

- le point `.` permet de désigner n'importe quel caractère :

`a.c`

`abc fdrffdsaaabd aac`

Les expressions régulières

- Pour effectuer le découpage, nous faisons appel aux expressions régulières, qui permettent de « chercher » de manière flexible un grand nombre de chaînes de caractères.
- Il s'agit d'une approche similaire à celle de beaucoup de moteurs de recherche, où `hist*` permet de chercher tous les mots qui commencent par `hist`.
- Quelques exemples (sur base de la chaîne `abcfdrrffdsaaabdaac`) :

- le `pipe |` permet de désigner plusieurs caractères à la fois :

`a|b`

`ab cfdrrffds aaab d aa c`

- les parenthèses `()` permettent de former des groupes de caractères :

`(ab)|d`

`ab cf d rff d saa abd aac`

- le point `.` permet de désigner n'importe quel caractère :

`a.c`

`abc fdrffdsaaabd aac`

- les quantificateurs permettent de spécifier combien de fois un caractère apparaît (`?` : 0 ou 1 ; `*` : 0 ou plus ; `+` : 1 ou plus ; `{2,3}` : entre 2 et 3 fois) :

`a{1,4}b`

`ab cfdrrffds aaab aa`

Découpage

- Nous désirons insérer une balise à la fin de la clause de notification.
- Pour cela, nous construisons une expression régulière qui permet de « sélectionner » en même temps la fin de toutes les clauses de notification.
- Nous procédons étape par étape, en considérant chaque fois les clauses qui ne sont pas identifiées dans les étapes suivantes :

Ço sacent cil ki cest escrit veront et oront que

Çou sacent tout cil ki cest escrit veront et oront que

Expression régulière utilisée

Découpage

- Nous désirons insérer une balise à la fin de la clause de notification.
- Pour cela, nous construisons une expression régulière qui permet de « sélectionner » en même temps la fin de toutes les clauses de notification.
- Nous procédons étape par étape, en considérant chaque fois les clauses qui ne sont pas identifiées dans les étapes suivantes :

Ço sacent cil ki cest escrit veront et oront que

Çou sacent tout cil ki cest escrit veront et oront que

Expression régulière utilisée

que

- Nous désirons insérer une balise à la fin de la clause de notification.
- Pour cela, nous construisons une expression régulière qui permet de « sélectionner » en même temps la fin de toutes les clauses de notification.
- Nous procédons étape par étape, en considérant chaque fois les clauses qui ne sont pas identifiées dans les étapes suivantes :

Ço sacent cil ki cest escrit veront et oront que

Çou sacent tout cil ki cest escrit veront et oront que

Expression régulière utilisée

oront que

Découpage

- Nous désirons insérer une balise à la fin de la clause de notification.
- Pour cela, nous construisons une expression régulière qui permet de « sélectionner » en même temps la fin de toutes les clauses de notification.
- Nous procédons étape par étape, en considérant chaque fois les clauses qui ne sont pas identifiées dans les étapes suivantes :

Ço sacent cil ki cest escrit veront et oront que

Çou sacent tout cil ki cest escrit veront et oront que

Ce sacent tout cil ki cest escrit veront et oront ke

Expression régulière utilisée

oront que

Découpage

- Nous désirons insérer une balise à la fin de la clause de notification.
- Pour cela, nous construisons une expression régulière qui permet de « sélectionner » en même temps la fin de toutes les clauses de notification.
- Nous procédons étape par étape, en considérant chaque fois les clauses qui ne sont pas identifiées dans les étapes suivantes :

Ço sacent cil ki cest escrit veront et oront que

Çou sacent tout cil ki cest escrit veront et oront que

Ce sacent tout cil ki cest escrit veront et oront ke

Expression régulière utilisée

oront (qu)|ke

- Nous désirons insérer une balise à la fin de la clause de notification.
- Pour cela, nous construisons une expression régulière qui permet de « sélectionner » en même temps la fin de toutes les clauses de notification.
- Nous procédons étape par étape, en considérant chaque fois les clauses qui ne sont pas identifiées dans les étapes suivantes :

Ço sacent cil ki cest escrit veront et oront que

Çou sacent tout cil ki cest escrit veront et oront que

Ce sacent tout cil ki cest escrit veront et oront ke

Ce sacient cil ki sunt et à venir sunt et cest escrit verunt et orunt ke

Expression régulière utilisée

oront (qu)|ke

- Nous désirons insérer une balise à la fin de la clause de notification.
- Pour cela, nous construisons une expression régulière qui permet de « sélectionner » en même temps la fin de toutes les clauses de notification.
- Nous procédons étape par étape, en considérant chaque fois les clauses qui ne sont pas identifiées dans les étapes suivantes :

Ço sacent cil ki cest escrit veront et oront que

Çou sacent tout cil ki cest escrit veront et oront que

Ce sacent tout cil ki cest escrit veront et oront ke

Ce sacient cil ki sunt et à venir sunt et cest escrit verunt et orunt ke

Expression régulière utilisée

`or.nt (qu)|ke`

Découpage

- Nous désirons insérer une balise à la fin de la clause de notification.
- Pour cela, nous construisons une expression régulière qui permet de « sélectionner » en même temps la fin de toutes les clauses de notification.
- Nous procédons étape par étape, en considérant chaque fois les clauses qui ne sont pas identifiées dans les étapes suivantes :

Ço sacent cil ki cest escrit veront et oront que

Çou sacent tout cil ki cest escrit veront et oront que

Ce sacent tout cil ki cest escrit veront et oront ke

Ce sacient cil ki sunt et à venir sunt et cest escrit verunt et orunt ke

Sacient tout cil ki cest escrit verront et orront ke

Expression régulière utilisée

`or.nt (qu)|ke`

Découpage

- Nous désirons insérer une balise à la fin de la clause de notification.
- Pour cela, nous construisons une expression régulière qui permet de « sélectionner » en même temps la fin de toutes les clauses de notification.
- Nous procédons étape par étape, en considérant chaque fois les clauses qui ne sont pas identifiées dans les étapes suivantes :

Ço sacent cil ki cest escrit veront et oront que

Çou sacent tout cil ki cest escrit veront et oront que

Ce sacent tout cil ki cest escrit veront et oront ke

Ce sacient cil ki sunt et à venir sunt et cest escrit verunt et orunt ke

Sacient tout cil ki cest escrit verront et orront ke

Expression régulière utilisée

`or{1,2}.nt (qu) |ke`

Découpage

- Nous désirons insérer une balise à la fin de la clause de notification.
- Pour cela, nous construisons une expression régulière qui permet de « sélectionner » en même temps la fin de toutes les clauses de notification.
- Nous procédons étape par étape, en considérant chaque fois les clauses qui ne sont pas identifiées dans les étapes suivantes :

Ço sacent cil ki cest escrit veront et oront que

Çou sacent tout cil ki cest escrit veront et oront que

Ce sacent tout cil ki cest escrit veront et oront ke

Ce sacient cil ki sunt et à venir sunt et cest escrit verunt et orunt ke

Sacient tout cil ki cest escrit verront et orront ke

Ce sacent tot cil ki cest escrit oront et veront que

Expression régulière utilisée

`or{1,2}.nt (qu) |ke`

- Nous désirons insérer une balise à la fin de la clause de notification.
- Pour cela, nous construisons une expression régulière qui permet de « sélectionner » en même temps la fin de toutes les clauses de notification.
- Nous procédons étape par étape, en considérant chaque fois les clauses qui ne sont pas identifiées dans les étapes suivantes :

Ço sacent cil ki cest escrit veront et oront que

Çou sacent tout cil ki cest escrit veront et oront que

Ce sacent tout cil ki cest escrit veront et oront ke

Ce sacient cil ki sunt et à venir sunt et cest escrit verunt et orunt ke

Sacient tout cil ki cest escrit verront et orront ke

Ce sacent tot cil ki cest escrit oront et veront que

Expression régulière utilisée

$(\text{or}\{1,2\}.\text{nt}) | (\text{ver}\{1,2\}.\text{nt}) (\text{qu}) | \text{ke}$

Découpage

- Nous désirons insérer une balise à la fin de la clause de notification.
- Pour cela, nous construisons une expression régulière qui permet de « sélectionner » en même temps la fin de toutes les clauses de notification.
- Nous procédons étape par étape, en considérant chaque fois les clauses qui ne sont pas identifiées dans les étapes suivantes :

Ço sacent cil ki cest escrit veront et oront que

Çou sacent tout cil ki cest escrit veront et oront que

Ce sacent tout cil ki cest escrit veront et oront ke

Ce sacient cil ki sunt et à venir sunt et cest escrit verunt et orunt ke

Sacient tout cil ki cest escrit verront et orront ke

Ce sacent tot cil ki cest escrit oront et veront que

Expression régulière utilisée

$o|(ve)r\{1,2\}.nt\ (qu)|ke$

Traitement des dates de temps

- Il est nécessaire d'extraire les dates qui apparaissent dans nos actes : non seulement les dates des actes eux-mêmes, mais aussi les dates de remboursement des dettes.

Traitement des dates de temps

- Il est nécessaire d'extraire les dates qui apparaissent dans nos actes : non seulement les dates des actes eux-mêmes, mais aussi les dates de remboursement des dettes.
- Les dates sont exprimées de plusieurs manières dans nos sources :

En absolu, avec le jour du mois

Ce fu fait l'an de l'Incarnation .MCC. et .LXVII. ou disietisme jour de Fenerec

Traitement des dates de temps

- Il est nécessaire d'extraire les dates qui apparaissent dans nos actes : non seulement les dates des actes eux-mêmes, mais aussi les dates de remboursement des dettes.
- Les dates sont exprimées de plusieurs manières dans nos sources :

En absolu, avec le jour du mois

Ce fu fait l'an de l'Incarnation .MCC. et .LXVII. ou disietisme jour de Fenerec

En l'an de l'Incarnation Jhésu-Crist .MCC. et .LXXXI., à l'entrée de Sietembre

Traitement des dates de temps

- Il est nécessaire d'extraire les dates qui apparaissent dans nos actes : non seulement les dates des actes eux-mêmes, mais aussi les dates de remboursement des dettes.
- Les dates sont exprimées de plusieurs manières dans nos sources :

En absolu, avec le jour du mois

Ce fu fait l'an de l'Incarnation .MCC. et .LXVII. ou disietisme jour de Fenerec
En l'an de l'Incarnation Jhésu-Crist .MCC. et .LXXXI., à l'entrée de Sietembre

En absolu, avec un jour particulier (saint, fête, etc.)

L'an de grasse .MCCIII^{XX}. et .XVIII., le nuit de Toutsains

Traitement des dates de temps

- Il est nécessaire d'extraire les dates qui apparaissent dans nos actes : non seulement les dates des actes eux-mêmes, mais aussi les dates de remboursement des dettes.
- Les dates sont exprimées de plusieurs manières dans nos sources :

En absolu, avec le jour du mois

Ce fu fait l'an de l'Incarnation .MCC. et .LXVII. ou disietisme jour de Fenerec
En l'an de l'Incarnation Jhésu-Crist .MCC. et .LXXXI., à l'entrée de Sietembre

En absolu, avec un jour particulier (saint, fête, etc.)

L'an de grasse .MCCIII^{XX}. et .XVIII., le nuit de Toutsains

En relatif, avec l'un ou l'autre système

Ce fut fait l'an de grasse .MCCIII^{XX}. et .XV., le darrain demierkes de May

Traitement des dates de temps

- Il est nécessaire d'extraire les dates qui apparaissent dans nos actes : non seulement les dates des actes eux-mêmes, mais aussi les dates de remboursement des dettes.
- Les dates sont exprimées de plusieurs manières dans nos sources :

En absolu, avec le jour du mois

Ce fu fait l'an de l'Incarnation .MCC. et .LXVII. ou disietisme jour de Fenerec
En l'an de l'Incarnation Jhésu-Crist .MCC. et .LXXXI., à l'entrée de Sietembre

En absolu, avec un jour particulier (saint, fête, etc.)

L'an de grasse .MCCIIII^{XX}. et .XVIII., le nuit de Toutsains

En relatif, avec l'un ou l'autre système

Ce fut fait l'an de grasse .MCCIIII^{XX}. et .XV., le darrain demierkes de May
Ce fu fait l'an de l'Incarnation .MCC. et .LXXI., el mois de octobre, le deluns
apriès les octaves de le Saint-Remi

Traitement des dates de temps

- Avant de traiter les jour et les mois, il faut traiter les millésimes.
- Lire les chiffres romains du millésime ne pose pas de problème, à condition d'expliquer à l'ordinateur comment interpréter les motifs non standards, comme « IIII^{XX} » ou « ... et ... ».

Traitement des dates de temps

- Avant de traiter les jour et les mois, il faut traiter les millésimes.
- Lire les chiffres romains du millésime ne pose pas de problème, à condition d'expliquer à l'ordinateur comment interpréter les motifs non standards, comme « IIII^{XX} » ou « ... et ... ».
- Il faut néanmoins prendre en compte le fait qu'au Moyen Âge, l'année ne commence pas toujours le premier janvier.

Traitement des dates de temps

- Avant de traiter les jour et les mois, il faut traiter les millésimes.
- Lire les chiffres romains du millésime ne pose pas de problème, à condition d'expliquer à l'ordinateur comment interpréter les motifs non standards, comme « IIII^{XX} » ou « ... et ... ».
- Il faut néanmoins prendre en compte le fait qu'au Moyen Âge, l'année ne commence pas toujours le premier janvier.
- À Ypres, Tournai et Douai aux XIII^e et XIV^e siècles, c'est à Pâques que la nouvelle année commence (style de Pâques) : en 1296, Pâques tombe le 25 mars, et donc

Style de Pâques	Nouveau style
10/10/1295	10/10/1295
25/01/1295	25/01/1296
22/03/1295	22/03/1296
15/04/1296	15/04/1296

Traitement des dates de temps

- Avant de traiter les jour et les mois, il faut traiter les millésimes.
- Lire les chiffres romains du millésime ne pose pas de problème, à condition d'expliquer à l'ordinateur comment interpréter les motifs non standards, comme « IIII^{XX} » ou « ... et ... ».
- Il faut néanmoins prendre en compte le fait qu'au Moyen Âge, l'année ne commence pas toujours le premier janvier.
- À Ypres, Tournai et Douai aux XIII^e et XIV^e siècles, c'est à Pâques que la nouvelle année commence (style de Pâques) : en 1296, Pâques tombe le 25 mars, et donc

Style de Pâques	Nouveau style
10/10/1295	10/10/1295
25/01/1295	25/01/1296
22/03/1295	22/03/1296
15/04/1296	15/04/1296

- Les règles de fixation de la date de Pâques sont compliquées, mais systématiques, et peuvent donc être implémentées sans problème.

- Une fois que le millésime est traité et que le « type » de date est détecté :
 - En absolu, avec le jour du mois : il faut lemmatiser ou normaliser les mots utilisés ;

- Une fois que le millésime est traité et que le « type » de date est détecté :
 - En absolu, avec le jour du mois : il faut lemmatiser ou normaliser les mots utilisés ;
 - En absolu, avec un jour particulier : il faut utiliser un calendrier des fêtes des saints et des autres fêtes religieuses (certaines fêtes ont bougé depuis le Moyen Âge, certaines fêtes sont très locales) ;

- Une fois que le millésime est traité et que le « type » de date est détecté :
 - En absolu, avec le jour du mois : il faut lemmatiser ou normaliser les mots utilisés ;
 - En absolu, avec un jour particulier : il faut utiliser un calendrier des fêtes des saints et des autres fêtes religieuses (certaines fêtes ont bougé depuis le Moyen Âge, certaines fêtes sont très locales) ;
 - En relatif, avec l'un ou l'autre système : il faut utiliser un calendrier capable d'identifier le jour de la semaine d'une date médiévale.

- Une fois que le millésime est traité et que le « type » de date est détecté :
 - En absolu, avec le jour du mois : il faut lemmatiser ou normaliser les mots utilisés ;
 - En absolu, avec un jour particulier : il faut utiliser un calendrier des fêtes des saints et des autres fêtes religieuses (certaines fêtes ont bougé depuis le Moyen Âge, certaines fêtes sont très locales) ;
 - En relatif, avec l'un ou l'autre système : il faut utiliser un calendrier capable d'identifier le jour de la semaine d'une date médiévale.
- Attention, il faut évidemment tenir compte du passage du calendrier julien au calendrier grégorien !

Analyse automatique des clauses des dettes

Éléments d'analyse critique

- Lorsqu'il s'agit de travailler sur les textes d'actes qui ont disparu (comme pour les chirographes d'Ypres et la plupart des chirographes de Tournai), nous devons bien entendu nous reposer sur des éditions.
- Nous sommes donc exposés au risque que les documents n'aient pas été édités correctement.

- Lorsqu'il s'agit de travailler sur les textes d'actes qui ont disparu (comme pour les chirographes d'Ypres et la plupart des chirographes de Tournai), nous devons bien entendu nous reposer sur des éditions.
- Nous sommes donc exposés au risque que les documents n'aient pas été édités correctement.
- Il est cependant possible d'effectuer des « tests de qualité » relativement à certains aspects des actes édités.
- Ces tests peuvent être automatisés, faisant gagner un temps précieux à l'historien.

- Lorsqu'il s'agit de travailler sur les textes d'actes qui ont disparu (comme pour les chirographes d'Ypres et la plupart des chirographes de Tournai), nous devons bien entendu nous reposer sur des éditions.
- Nous sommes donc exposés au risque que les documents n'aient pas été édités correctement.
- Il est cependant possible d'effectuer des « tests de qualité » relativement à certains aspects des actes édités.
- Ces tests peuvent être automatisés, faisant gagner un temps précieux à l'historien.
- Il existe plusieurs façons de vérifier la cohérence des dates éditées.
- Nous pouvons par exemple visualiser globalement les mentions des échevins devant lesquels passent les actes.

- De la même manière, dans les cas où le jour de la semaine est donné dans la date de temps, nous pouvons vérifier si celle-ci est cohérente.

Correction du millésime

- De la même manière, dans les cas où le jour de la semaine est donné dans la date de temps, nous pouvons vérifier si celle-ci est cohérente.
- Considérons par exemple :

Ce fu fait en l'an de l'Incarnation .M.CC. et .LXXXII., el mois d'Aoust, le nuit Saint-Leurenc, par .I. deluns.

Correction du millésime

- De la même manière, dans les cas où le jour de la semaine est donné dans la date de temps, nous pouvons vérifier si celle-ci est cohérente.
- Considérons par exemple :

Ce fu fait en l'an de l'Incarnation .M.CC. et .LXXXII., el mois d'Aoust, le nuit Saint-Leurenc, par .I. deluns.

- Cependant, le jour de la Saint-Laurent, le 10 août, tombe un dimanche en 1282 !
- Il est probable que l'éditeur ait ici mal compté le nombre de I de la date, et que celle-ci désignait en fait 1283, année pendant laquelle le 10 août est effectivement un lundi.

Correction du millésime

- De la même manière, dans les cas où le jour de la semaine est donné dans la date de temps, nous pouvons vérifier si celle-ci est cohérente.
- Considérons par exemple :

Ce fu fait en l'an de l'Incarnation .M.CC. et .LXXXII., el mois d'Aoust, le nuit Saint-Leurenc, par .I. deluns.

- Cependant, le jour de la Saint-Laurent, le 10 août, tombe un dimanche en 1282 !
- Il est probable que l'éditeur ait ici mal compté le nombre de I de la date, et que celle-ci désignait en fait 1283, année pendant laquelle le 10 août est effectivement un lundi.
- Il est possible de demander à l'ordinateur de tester un ensemble de modifications du millésime qui peuvent correspondre à la version « non fautive » de la lecture de l'acte.

Conclusion

- Nous avons vu trois exemples de dossiers d'histoire médiévale dans lesquels de grandes quantités de données sont utilisées.

- Nous avons vu trois exemples de dossiers d'histoire médiévale dans lesquels de grandes quantités de données sont utilisées.
- Comme en modélisation actuarielle, des méthodes quantitatives peuvent être utilisées pour chacune des trois séquences de travail sur les données :

- Nous avons vu trois exemples de dossiers d'histoire médiévale dans lesquels de grandes quantités de données sont utilisées.
- Comme en modélisation actuarielle, des méthodes quantitatives peuvent être utilisées pour chacune des trois séquences de travail sur les données :
 - 1 le traitement des données ;
 - 2 l'analyse ;
 - 3 la critique des résultats.

- Nous avons vu trois exemples de dossiers d'histoire médiévale dans lesquels de grandes quantités de données sont utilisées.
- Comme en modélisation actuarielle, des méthodes quantitatives peuvent être utilisées pour chacune des trois séquences de travail sur les données :
 - 1 le traitement des données ;
 - 2 l'analyse ;
 - 3 la critique des résultats.
- En particulier, je crois qu'il faut insister sur la phase de critique, avec laquelle les méthodes quantitatives sont en dialogue :

- Nous avons vu trois exemples de dossiers d'histoire médiévale dans lesquels de grandes quantités de données sont utilisées.
- Comme en modélisation actuarielle, des méthodes quantitatives peuvent être utilisées pour chacune des trois séquences de travail sur les données :
 - 1 le traitement des données ;
 - 2 l'analyse ;
 - 3 la critique des résultats.
- En particulier, je crois qu'il faut insister sur la phase de critique, avec laquelle les méthodes quantitatives sont en dialogue :
 - 1 l'utilisation de ces méthodes nécessite une « couche » de critique supplémentaire ;
 - 2 ces méthodes constituent de nouveaux outils pour la critique « classique ».

- Cependant, il me semble qu'il existe des différences entre ces deux contextes, notamment en termes de la nature des éléments qui affectent la qualité des données.

- Cependant, il me semble qu'il existe des différences entre ces deux contextes, notamment en termes de la nature des éléments qui affectent la qualité des données.
- Si on embrasse une approche « microscopique » et qu'on se place au niveau du script, le traitement de
 - 1 "Madme" dans une base de données actuarielle et de
 - 2 "orunt" dans un corpus de textes historiquesrevient sans doute au même.

La question de la qualité des données

- Cependant, il me semble qu'il existe des différences entre ces deux contextes, notamment en termes de la nature des éléments qui affectent la qualité des données.
- Si on embrasse une approche « microscopique » et qu'on se place au niveau du script, le traitement de
 - 1 "Madme" dans une base de données actuarielle et de
 - 2 "orunt" dans un corpus de textes historiquesrevient sans doute au même.
- Mais si on embrasse une approche « macroscopique » et qu'on inspecte le processus d'amélioration de la qualité des données dans son ensemble, ces deux exemples ne peuvent être considérés de la même façon :
 - 1 "Madme" est une « véritable erreur », tandis que
 - 2 "orunt" est une « variation naturelle » de graphie à replacer dans un contexte linguistique bien précis.



"That's all Folks!"