

Troisième réunion du groupe de contact FNRS  
« Les humanités des données / Data-driven humanities »

Les méthodes quantitatives dans les sciences humaines :  
quelles incidences sur les objets et processus de  
recherche ?

Matthieu Pichon, Guillaume Quintin et Marie Serisier

ReSIC / Ratio DH, Université libre de Bruxelles

9 décembre 2024



# Table des matières

**1** Introduction

**2** L'objet texte dans les humanités numériques

**3** La posture du chercheur

**4** Conclusion

- Trois expériences de recherche différentes.
- Une convergence autour du **texte**, sous diverses formes :
  - Des textes scientifiques : des articles de revues francophones d'une part, les thèses numérisées de l'ULB d'autre part.
  - Des textes littéraires : dix siècles de production hagiographique écrite en latin médiéval (du III<sup>e</sup> au XIII<sup>e</sup> s.).
  - Des échanges sur le web : quatre communautés affiliées à l'antiféminisme sur la plateforme Reddit depuis leur création (entre 2009 et 2015 jusqu'à 2022).
- Une démarche commune d'**humanités numériques**.

# Introduction : Humanités numériques ?

## Définition

- ***Digitized*** et ***numerical humanities*** (ROTH 2019)
- « Humanités numériques **théoriques** » et « humanités numériques **appliquées** » ? (PIOTROWSKI et XANTHOS 2020)

Notre positionnement :

- Au carrefour entre *digitized* et *numerical humanities*.
- Plutôt du côté des humanités numériques appliquées.

# Introduction : Une proposition

- Parler à trois voix de nos expériences singulières et communes, mettre l'accent sur quelques enjeux et susciter la réflexion collective, sans prétention à donner des réponses définitives.
- Autour de **deux problématiques principales** :
  - Que font les humanités numériques et les approches quantitatives à l'objet de recherche, en l'occurrence le texte ?
  - Que fait le *digital* à l'*humanist*, i.e. quels sont les effets du « quantitatif » sur la posture du/de la chercheur·se ?

## 1 Introduction

## 2 L'objet texte dans les humanités numériques

- Le propre, le sale et la qualité des données
- Que reste-t-il du texte ?

## 3 La posture du chercheur

- Le chercheur face à l'objet de recherche et sa modélisation
- La quantification : gage de scientificité, illusion, protection ?
- Le *digital humanist* se doit-il d'être expert formel **et** expert disciplinaire ?

## 4 Conclusion

# Table des matières

## 1 Introduction

## 2 L'objet texte dans les humanités numériques

- Le propre, le sale et la qualité des données
- Que reste-t-il du texte ?

## 3 La posture du chercheur

## 4 Conclusion

## Définition

Nous entendons le terme « texte » comme « Suite de signes linguistiques constituant un écrit ou une œuvre » (CNRTL)

**Première interrogation** : le *propre* et le *sale* des données textuelles, ou comment aborder la qualité des données mobilisées ?

- Sans rouvrir entièrement cette question, très présente dans la littérature et les débats en HN, quelles sont les implications dans nos recherches ?

# L'objet texte dans les humanités numériques :

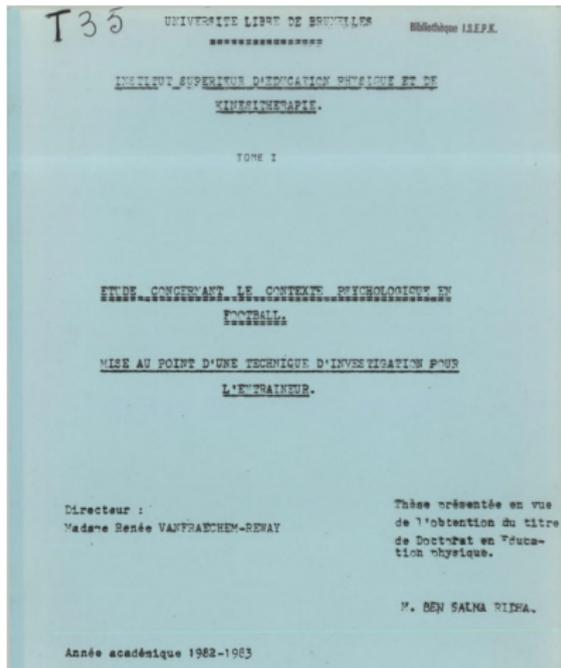
Le propre, le sale et la qualité des données

- Les erreurs d'OCR/HTR : un faux problème ?
  - *To correct or not to correct* : un débat très présent dans la littérature en HN.
  - Des résultats robustes malgré des données sales.

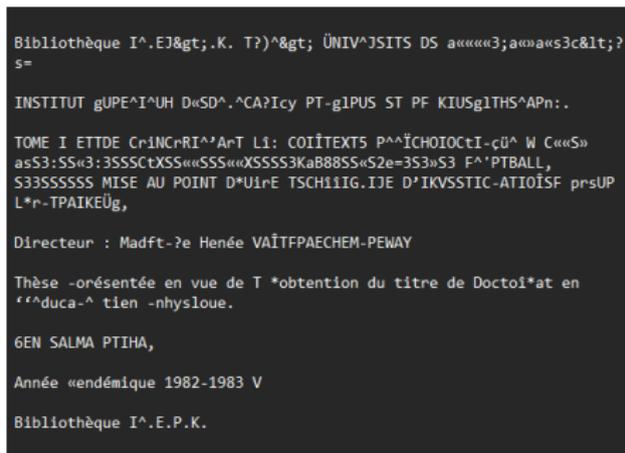
# L'objet texte dans les humanités numériques :

Le propre, le sale et la qualité des données

Les erreurs d'OCR/HTR : des cas plus problématiques ?



Pour l'oeil humain



Pour l'oeil de l'OCR

# L'objet texte dans les humanités numériques :

Le propre, le sale et la qualité des données

- Langue très spécifique : néologismes, mots tronqués, sigles et acronymes spécifiques aux communautés étudiées
- Corpus de 276 384 soumissions (annotées, avant nettoyage) soit 53 275 720 millions de tokens et 15 194 548 de commentaires

## Exemples

*I now believe ' **AWALT** ' and think that women are always taking advantage of my social status/money... (r/TheRedPill)*

*While you are working so hard, day and night to money **maxx** , your future wife is getting pounded by broke Tyrones day and night.  
(r/PurplePillDebate)*

# L'objet texte dans les humanités numériques :

Le propre, le sale et la qualité des données

- Utilisation du package UDPipe, modèle d'annotation "english web treebank" (SILVEIRA et al. 2014) + vérification
- Fiabilité mesurée sur 4000 tokens vérifiés manuellement : moyenne entre 90 et 95% selon la catégorie (token, lemme, UPOS). En comparaison, test sur un corpus web du COCA (Corpus of Contemporary American English) : presque 100% de fiabilité.
- Impossibilité de vérifier tout, accepter une inexactitude de principe.

# L'objet texte dans les humanités numériques :

Le propre, le sale et la qualité des données

Subreddit	Phrase	Terme mal-étiqueté	Erreur
r/PurplePillDebate r/TheRedPill	"They could be thief who steals your shit" "I didn't get SHITTT, everything was in her name"	thief SHITTT	thiefenately PROPN

Exemples d'erreurs d'étiquetage

# L'objet texte dans les humanités numériques :

## Le propre, le sale et la qualité des données

- Fenêtre de 5-10% qui nous donne à réfléchir, penser le traitement du corpus comme un modèle perfectible.
- Difficile d'entraîner des grands modèles statistiques sur des corpus aussi spécifiques, visée généraliste du modèle.
- Penser en dehors du modèle d'annotation : qu'est-ce qui a du sens pour qui ?

→ **Exemple** : *"How often do we see exchanges on here that go like-  
\*\*Man : \*\* "Women all go for the top (insert small percentage)"  
(r/PurplePillDebate)*

# L'objet texte dans les humanités numériques :

## Le propre, le sale et la qualité des données

- La question de la « qualité » d'un texte ne se pose pas uniquement au niveau de sa numérisation ou de son traitement quantitatif :
  - L'édition d'un texte est le fruit d'un travail philologique invisibilisé.
  - Il peut être difficile de distinguer texte et paratexte.
- La qualité des *métadonnées* est tout aussi cruciale que celle des données :
  - Les métadonnées sont la porte d'entrée vers les textes.
  - Exemple concret de la datation automatique : l'information « date » ne se trouve pas dans le texte lui-même, mais dans les métadonnées.
- Ne pas oublier que les métadonnées non plus ne sont pas « données » !

# L'objet texte dans les humanités numériques :

Que reste-t-il du texte ?

## Deuxième interrogation : que reste-t-il du texte ?

- Les approches quantitatives du texte imposent des transformations profondes de l'objet initial.
- Qu'est-ce que cela fait à la manière d'appréhender le texte, de s'en saisir, de le comprendre ?
- En particulier, l'approche quantitative impose généralement une forme de *distant reading* du texte.

# L'objet texte dans les humanités numériques :

Que reste-t-il du texte ?

## Distant reading ?

Expression proposée par MORETTI 2000, qui consiste dans les grandes lignes à aborder le texte autrement que de manière suivie, de l'appréhender à *distance*, et notamment de « *focus on units that are much smaller or much larger than the text* » (MORETTI 2013, p. 57).

- Si on peut l'envisager de manière qualitative, le *distant reading* semble néanmoins particulièrement adapté à des approches quantitatives.
- Cette lecture d'un texte à *distance* trouve des précurseurs dans la stylométrie (dès le XIX<sup>e</sup> s.) et le formalisme russe de l'entre-deux-guerres (LUTOSLAWSKI 1898 ; YARKHO 2016).

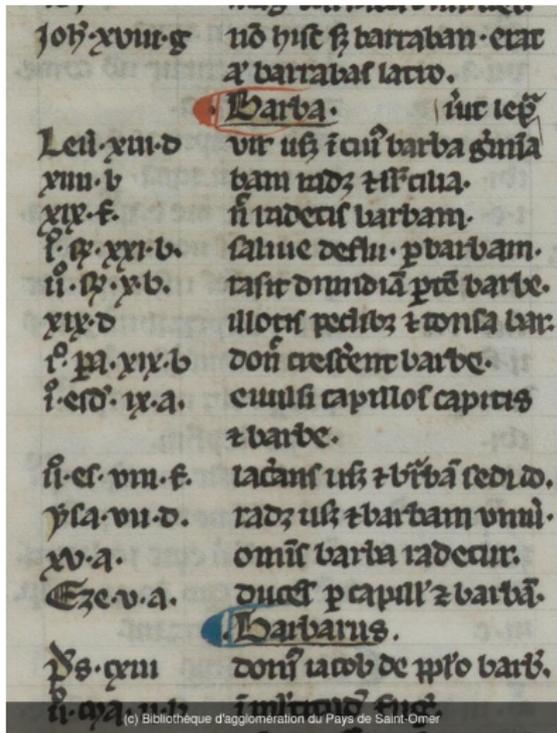
# L'objet texte dans les humanités numériques :

Que reste-t-il du texte ?

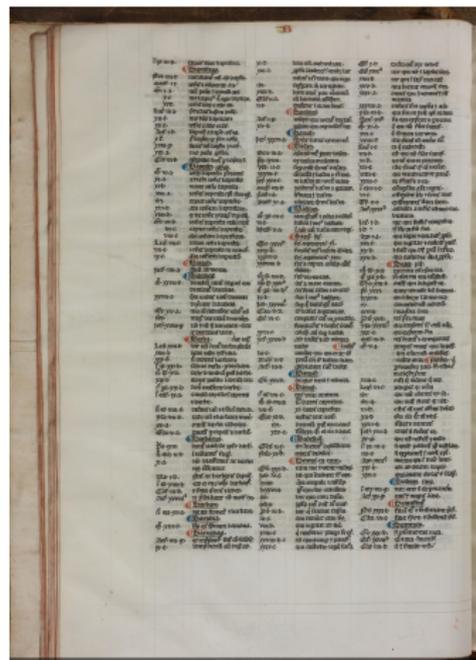
- En réalité, on peut déjà trouver des exemples de « démembrement du texte » au moins depuis l'époque médiévale.
- Un exemple : la concordance d'une œuvre, c'est-à-dire une liste alphabétique des mots qui se trouvent dans cette œuvre, accompagnés de leur entourage immédiat et d'une référence à leur position dans le texte original.
- La première concordance est une concordance de la Bible : la [Concordance dite de Saint-Jacques](#), qui date du milieu du XIII<sup>e</sup> s. (BLAIR 2010, p. 35-46).

# L'objet texte dans les humanités numériques :

Que reste-t-il du texte ?



(c) Bibliothèque d'agglomération du Pays de Saint-Omer



Une page de la Concordance de Saint-Jacques

Zoom sur la page

# L'objet texte dans les humanités numériques :

Que reste-t-il du texte ?

- Le « père spirituel » des humanités numériques, le jésuite Roberto Busa, entreprend la création d'une concordance numérique de l'oeuvre de Thomas d'Aquin.
- Commencée en 1949 sous le sponsor d'IBM, cette entreprise s'est achevée trente ans plus tard (BUSA 1980 ; SULA et HILL 2019).
- Par ailleurs, l'*Index Thomisticus* n'était pas une initiative isolée :
  - Application de l'outil informatique à la démarche philologique par FROGER, un moine bénédictin (FROGER 1961, 1970, 1968) ;
  - Création du LASLA en 1961 à l'ULiège et du CETEDOC en 1968 à l'UCLouvain.



Roberto Busa regardant  
une carte perforée

# L'objet texte dans les humanités numériques :

## Que reste-t-il du texte ?

Alphabétiquement [↵ ↻]	Fréquences [↵ ↻]	<b>Itinera Electronica</b> <i>Du texte à l'hypertexte</i>	
Nymphaeae 2 Nyctae 1 o 108 <b>ob 16</b> obducta 1 obductum 1 obeat 1	16 minus 16 nati 16 nota <b>16 ob</b> 16 omnem 16 pedes 16 pedibus	<b>Virgile, L'Énéide</b>  <b>ob</b>	

[Classement: alphabétique](#)

**Chant, vers**

1, 0	~ui superum saeuae memorem Iunonis	<b>ob</b>	iram ~multa quoque et bello
1, 40	ipsos potuit sommergere ponto ~unius	<b>ob</b>	noxam et furias Aiacis Oilei?
1, 230	quibus tot funera passis ~cunctus	<b>ob</b>	Italiam terrarum clauditur orbis? ~certe
1, 250	arcem, ~naubis infandum! amissis unius	<b>ob</b>	iram ~prodimur atque Italiam longe
2, 130	~quos illi fors et poenas	<b>ob</b>	nostra repositos ~effugia, et culpam
2, 570	ferenti. ~illa sibi infestos euersa	<b>ob</b>	Pergama Teucros ~et Danaum poenam
5, 280	Aeneas promisso munere donat ~seruatam	<b>ob</b>	nauem laetus sociosque reductos. ~olli
6, 610	quae maxima turba est) ~quique	<b>ob</b>	adulterium caesi, quique arma secuti
6, 660	siluam uoluitur amnis. ~hic manus	<b>ob</b>	patriam pugnando uulnera passi, ~quique
7, 180	aliquae ab origine reges, ~Martiaque	<b>ob</b>	patriam pugnando uulnera passi. ~multaque
10, 680	fluctuat illuc, ~an sese mucrone	<b>ob</b>	tantum dedecus amens ~induat et
10, 850	tuum maculaui crimine nomen, ~pulsus	<b>ob</b>	inuidiam solio sceptrisque paternis. ~debueram
11, 340	libertatem fandi flatusque remittat, ~cuius	<b>ob</b>	auspiciis infaustum moresque sinistros ~dicam
11, 530	subitaque animum dulcedine mouit. ~pulsus	<b>ob</b>	inuidiam regno uirisque superbas ~Priuero
12, 560	Iuppiter hac stat, ~neu quis	<b>ob</b>	inceptum subitum mihi signior ito.
12, 860	in faciem Turni se pestis	<b>ob</b>	ora ~fertque refertque sonans clipeumque

Un exemple de concordance numérique ([accessible en ligne](#))

# L'objet texte dans les humanités numériques :

Que reste-t-il du texte ?

- Travailler sur un texte numérique ou numérisé, c'est travailler sur un texte qui a déjà subi des transformations imposées par l'ordinateur.
- C'était évident à l'époque des cartes perforées, mais c'est toujours le cas aujourd'hui à travers nos écrans.
- Où placer la limite où le texte cesse d'être et devient autre chose que lui-même ? Le PDF, la concordance, la matrice de fréquences ?

→ Continuum de transformations, dont les plus extrêmes à l'égard du texte permettent de donner du sens à de très grands corpus et de les étudier dans leur globalité.

# L'objet texte dans les humanités numériques :

Que reste-t-il du texte ?

auteur	titre	paquet	a	ad	ex	in	ob	per	pro
GerAur	DeGeo	25	0.004	0.008	0.010	0.026	0.000	0.018	0.002
GerAur	DeGeo	26	0.016	0.006	0.004	0.026	0.000	0.008	0.000
GerAur	DeGeo	27	0.008	0.000	0.000	0.070	0.000	0.018	0.000
GerAur	DeGeo	28	0.000	0.002	0.002	0.074	0.000	0.022	0.002
GerAur	DeGeo	29	0.002	0.010	0.010	0.052	0.000	0.012	0.002
GerAur	DeGeo	30	0.010	0.002	0.004	0.032	0.000	0.032	0.000
GerAur	DeGeo	31	0.022	0.012	0.004	0.030	0.000	0.014	0.002
GerAur	DeNuDi	1	0.004	0.030	0.002	0.024	0.000	0.068	0.000
GerAur	DeNuDi	2	0.012	0.032	0.002	0.026	0.000	0.040	0.002
GerAur	DeRaEtR	1	0.012	0.008	0.000	0.020	0.000	0.002	0.000
GerAur	DeRaEtR	2	0.000	0.014	0.008	0.012	0.000	0.000	0.000
GerAur	DeRaEtR	3	0.006	0.004	0.022	0.022	0.000	0.000	0.000
GerAur	DeRaEtR	4	0.014	0.004	0.006	0.024	0.000	0.000	0.000
GerAur	DeRaEtR	5	0.008	0.012	0.002	0.022	0.000	0.002	0.000
GerAur	DeRaEtR	6	0.002	0.012	0.004	0.024	0.000	0.000	0.000
GerAur	DeRaEtR	7	0.008	0.004	0.000	0.012	0.000	0.004	0.000
GerAur	DeSpCo	1	0.006	0.016	0.008	0.026	0.000	0.028	0.002
GerAur	EpScAnS	1	0.006	0.010	0.000	0.010	0.000	0.000	0.002
GerAur	EpScAnS	2	0.008	0.016	0.008	0.004	0.000	0.000	0.004
GerAur	EpScAnS	3	0.004	0.014	0.000	0.020	0.000	0.006	0.000

Une matrice de fréquences des prépositions chez Gerbert d'Aurillac

# L'objet texte dans les humanités numériques :

Que reste-t-il du texte ?

- Porte ouverte vers une collecte de données immense « big data ».
  - Sorte de frénésie à vouloir tout collecter, tout analyser ?
  - Quel coûts (humain, matériel, énergétique) ?
  - Quel nouveau rapport à son objet ?



# L'objet texte dans les humanités numériques :

Que reste-t-il du texte ?

- Grands corpus : choix positionnel par rapport à son objet → choix de ne lire que des fragments de son corpus
- Pas de linéarité mais une « réorganisation informative ».
  
- Cas du *topic modeling* : extraction thématique non-linéaire
- Extraction complète de l'émotion d'un texte
- Permet tout de même de caractériser un corpus par utilisation d'un modèle probabiliste (ici LDA, BLEI et al. 2003)



# L'objet texte dans les humanités numériques :

Que reste-t-il du texte ?

- Réorganisation : arbitraire de l'association et la séparation dans le cadre d'un corpus de textes.
- Corpus de soumissions et de commentaires : comment déterminer l'unité « texte » ?
- Écriture multiple → *never-ending text*
- Nécessité d'arbitrer ce que l'on considère être une unité de texte par rapport à l'usage qui en est fait et aux modèles utilisés.

# L'objet texte dans les humanités numériques :

Que reste-t-il du texte ?

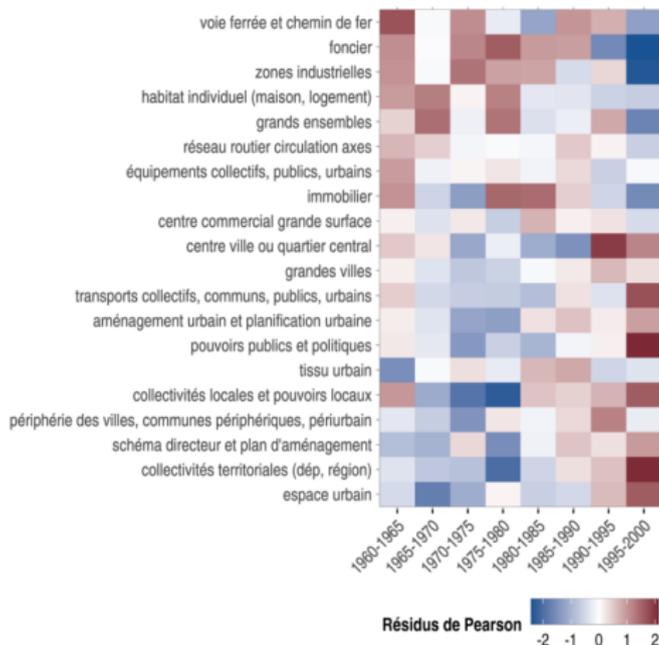
- Auto-analyse d'une recherche au prisme du couple ***distant/close reading***.
- **Une** question pour **deux** stratégies de recherche :
  - ① Close reading → forte implication du sujet-chercheur → corpus réduit et texte préservé → capturer une « littérature d'intervention »
  - ② Distant reading → moindre implication du sujet-chercheur → corpus large (voire exhaustif?) et texte démembré → capturer la « recherche ordinaire »

# L'objet texte dans les humanités numériques :

Que reste-t-il du texte ?

## Évolution de la présence des 20 syntagmes les plus fréquents

Résidus d'un test du chi-2 croisant les périodes et les syntagmes



Le texte disparu ou se servir du texte comme *proxy*

# Table des matières

## 1 Introduction

## 2 L'objet texte dans les humanités numériques

## 3 La posture du chercheur

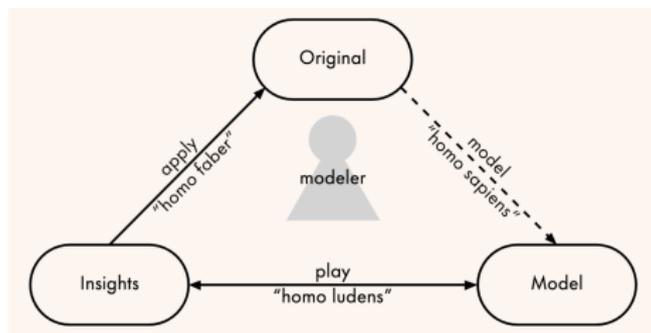
- Le chercheur face à l'objet de recherche et sa modélisation
- La quantification : gage de scientificité, illusion, protection ?
- Le *digital humanist* se doit-il d'être expert formel **et** expert disciplinaire ?

## 4 Conclusion

# La posture du chercheur :

## Le chercheur face à l'objet de recherche et sa modélisation

- Le corpus est un modèle en soi
- Distinction quali/quantitative est donc dissoute dans la question de la modélisation
- Triade du modèle (PIOTROWSKI 2022)



Triade du modèle selon PIOTROWSKI 2022

# La posture du chercheur :

## Le chercheur face à l'objet de recherche et sa modélisation

- Le corpus *en soi* n'est rien (BRAILLARD 2013), toute action posée dessus est une modélisation
- Deux types d'actions (CADIÈRE 2013) :
  - Mathématique
  - Géométrique
- Attention à la confusion entre l'original, le corpus et le modèle

*"[Mesurer] est le moment où l'auteur se ressent au plus loin de son regard familier sur la réalité observée, car elle perd l'unité que par la pensée il lui accordait. Parcelliser, catégoriser, comptabiliser... Ce qui s'observe ne vient plus du corpus lui-même, mais de l'opération qu'on effectue sur ce corpus." CADIÈRE 2013*

# La posture du chercheur :

## Le chercheur face à l'objet de recherche et sa modélisation

- Dans cette triade original-modèle-corpus : part du sujet non-anéantissable
- Subjectivité nécessaire à l'appréciation de la qualité des objets (nature, propriétés)
- Appréciation qualitative et définitoire des objets : conséquences sur le reste du processus d'analyse.

→ **Exemple 1** : Collecte d'articles sur "la ville", "l'urbain" : opération subjective.

→ **Exemple 2** : Entraînement de modèles de reconnaissance de l'ironie ?  
Sujet à débat même en dehors de la recherche !

# La posture du chercheur :

La quantification : gage de scientificité, illusion, protection ?

- Les approches quantitatives, computationnelles et/ou modélisatrices suscitent des réactions polarisées : tantôt perçues comme gage de scientificité, tantôt comme une forme de mystification.
- Problème : la représentation qu'on s'en fait charrie un certain nombre de préjugés, par habitude ou par méconnaissance.

# La posture du chercheur :

La quantification : gage de scientificité, illusion, protection ?

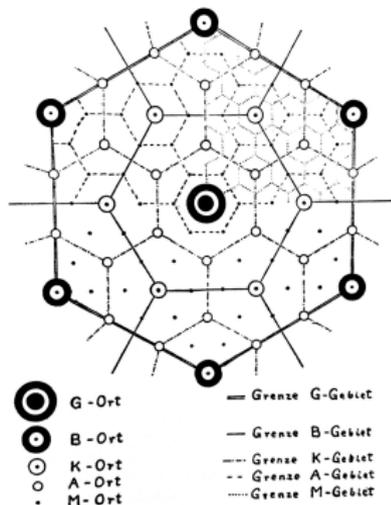
## Précautions à garder à l'esprit

- Quantifier n'est pas (nécessairement) modéliser
- Et modéliser n'est pas (nécessairement) quantifier

La décomposition du mouvement d'affaires total sur la foi de sondages bancaires systématiques a donné les résultats ci-dessous :

Commerces de détail	Alimentation 33,5 p. 100	} 45,6 p. 100
Artisanat	Autres..... 12,1	
Matériaux de construction et commerce des ferm.	.....	18,5
Commerce des porcs	.....	9,1
Commerces de gros (autres que ceux déjà nommés)	.....	-6,6
Embauche	.....	6,5
Transports	.....	6
Industries agricoles et alimentaires	.....	3
Bâtiment et travaux publics	.....	1,3
Divers	.....	0,4
		3
		100

Labasse, 1954



Christaller, 1933

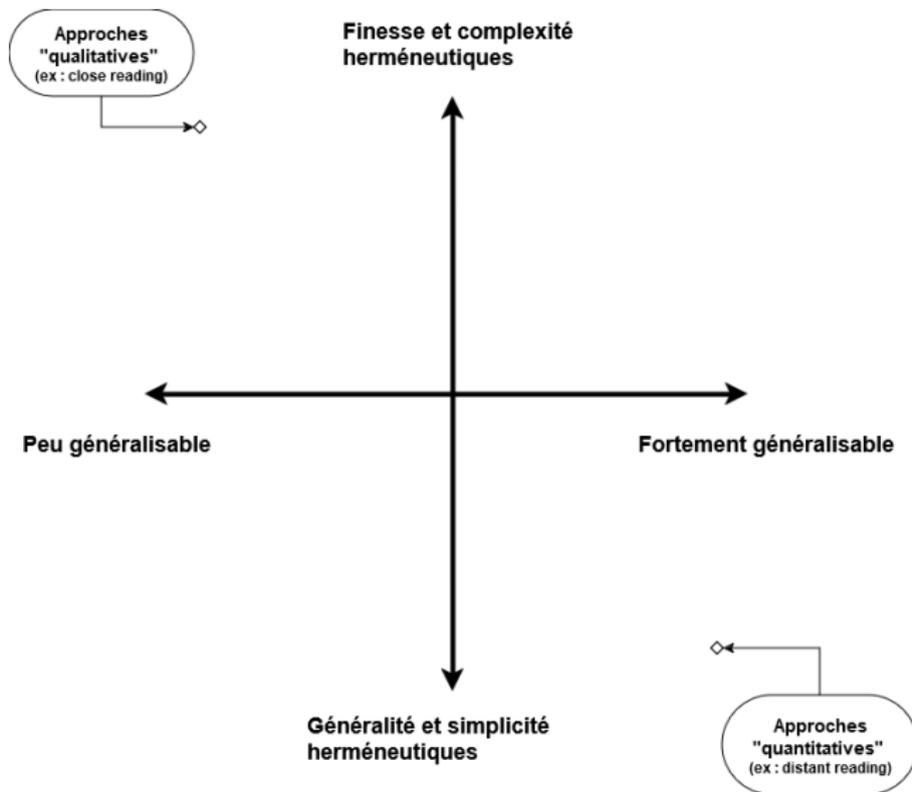
# La posture du chercheur :

La quantification : gage de scientificité, illusion, protection ?

- Une confession : un attrait pour les méthodes quantitatives / computationnelles.
- Mais aussi une invitation à se méfier...
  - ... d'un **effet d'exhaustivité** ?
  - ... d'un **effet de généralité** ?

# La posture du chercheur :

La quantification : gage de scientificité, illusion, protection ?



# La posture du chercheur :

La quantification : gage de scientificité, illusion, protection ?

- L'ordinateur n'est pas une machine « magique » qui corrige les biais ou erreurs des données d'entrée.
- Pour cette raison, il faut être critique envers les résultats obtenus, mais l'être encore plus avec les données de départ → risque d'illusion de scientificité, cachant des résultats peu solides.
- Exemple de la datation automatique : si le modèle de datation est entraîné sur des datations fausses, les résultats obtenus seront également faux !
- C'est ce que SOMMERSCHIED et al. (2023, p. 11) résumant en un « risque de circularité ».
- Nécessité de vérifier la qualité des métadonnées (du moins le champ « date »), c'est-à-dire de réaliser un travail de philologie purement traditionnelle.

# La posture du chercheur :

La quantification : gage de scientificité, illusion, protection ?

- La distance : parfois un outil de protection personnelle
- Corpus comportant énormément de données sensibles : appels au viol, appel au meurtre, discriminations etc.
- Enjeu épistémologique derrière cette protection : celui de la traduction
- Protection → formalisation → traduction langue naturelle/langue formelle

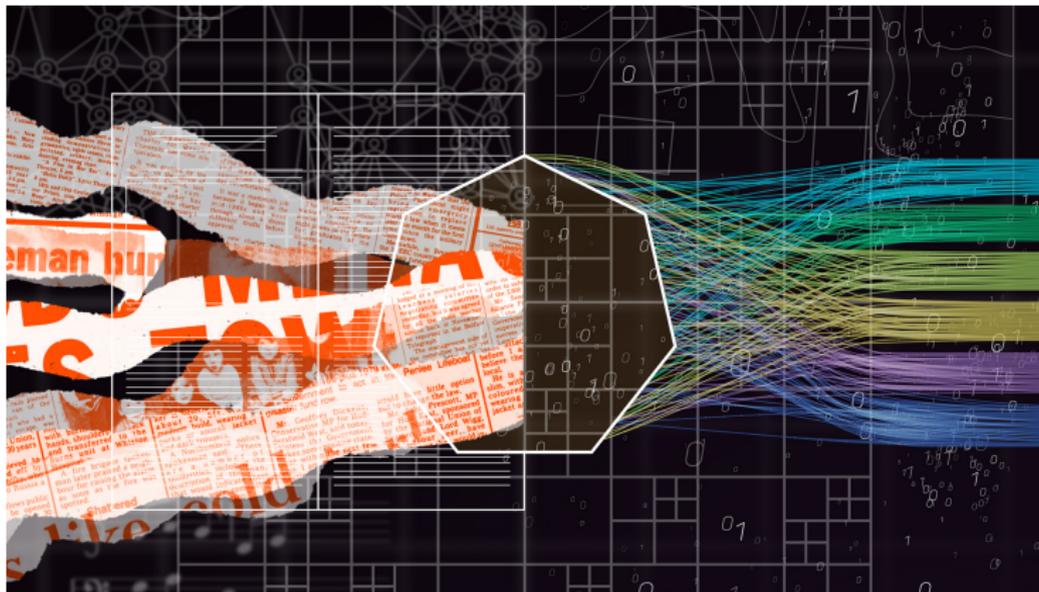
## Extrait du subreddit r/TheRedPill

*It seems to me that an Alpha should slut shame a female who would cheat on a guy she's in a relationship with... make her understand she is worthless b/c of her lack of loyalty and hypergamy. Women are that way naturally, but if guys (especially alphas and betas) discouraged the behavior, they would be forced to at least partially hold their outrageous behavior in check or become "damaged goods".*

# La posture du chercheur :

Le *digital humanist* se doit-il d'être expert formel **et** expert disciplinaire ?

Les méthodes formelles : critique/idolâtrie



# La posture du chercheur :

Le *digital humanist* se doit-il d'être expert formel **et** expert disciplinaire ?

- Importance cruciale du savoir disciplinaire dans n'importe quel projet d'humanités numériques.
- Exemple concret : un article de stylométrie portant sur la paternité de l'*Historia Augusta* (MARRIOTT 1979).
- L'auteur se base en partie sur la longueur des phrases pour faire ses calculs statistiques, ce qui pose deux problèmes :
  - Critique méthodologique : déjà en 1979, la longueur des phrases n'était plus considérée comme une métrique efficace (STOVER et KESTEMONT 2016) ;
  - Critique philologique : étudier la longueur des phrases, c'est davantage étudier les habitudes d'édition moderne que de rédaction antique (SANSONE 1990).

# La posture du chercheur :

Le *digital humanist* se doit-il d'être expert formel **et** expert disciplinaire ?

- Pour autant, le philologue ne rejette pas l'utilisation de l'ordinateur et de la statistique par la philologie.
- Selon lui, pour pouvoir critiquer les résultats, il faut comprendre la méthode (SANSONE 1990, p. 177).
- De nombreux philologues s'étaient laissés convaincre par l'article de MARRIOTT : exemple de l'illusion d'objectivité donnée par les chiffres pour un public non-statisticien ?

# La posture du chercheur :

Le *digital humanist* se doit-il d'être expert formel **et** expert disciplinaire ?

- Le *digital humanist* doit aussi garder à l'esprit que les modèles et les outils ont des présupposés théoriques et/ou des représentations relatives à l'objet de recherche qu'ils permettent d'analyser.

Extrait de P. Bourdieu et L. Wacquant, *Réponses. Pour une anthropologie réflexive*, 1992, p. 88-89

« la tâche de la science est de porter au jour la structure de la distribution des ressources [...] qui tend à déterminer la structure des prises de positions individuelles ou collectives. Dans la network analysis, l'analyse de ces structures [...] a été sacrifiée à l'analyse des liaisons particulières (entre des agents ou des institutions) [...] »

# Table des matières

- 1 Introduction
- 2 L'objet texte dans les humanités numériques
- 3 La posture du chercheur
- 4 Conclusion

## Conclusion : « Notre journal intime »

- Profils originels qualitatistes, puis saut épistémologique et disciplinaire (chacun différemment).
- Courbe d'apprentissage des processus formels (apprendre la langue de la machine) : difficile mais très enrichissant.
- Réconfort et frustration vis-à-vis de la machine : pas d'erreur possible.
- Ordinateur comme « faiseur d'ordre » : il faut créer de l'ordre et de la structure dans un matériau qui ne s'en satisfait pas toujours, qui sortira toujours en partie des cases qui lui seront imposées.
- Nécessité d'avoir un processus de travail et d'organisation rigoureux.
- Impact sur la vision des choses et des statistiques de façon générale.



Merci pour votre écoute, place aux échanges !

# Bibliographie I

Ann M. BLAIR.

Too Much to Know : Managing Scholarly Information before the Modern Age. Yale University Press, 2010.

David M. BLEI, Andrew Y. NG et Michael I. JORDAN. « Latent dirichlet allocation ». In : The Journal of Machine Learning Research 3 (2003), p. 993-1022. ISSN : 1532-4435.

Pierre-Alain BRAILLARD. « Chapitre 9. Que peut expliquer un modèle complexe et peut-on le comprendre ? » In : Modéliser & simuler – Tome 1. Sciences & philosophie. Éditions Matériologiques, 2013, p. 277-297.

R. BUSA. « The Annals of Humanities Computing : The Index Thomisticus ». In : Computers and the Humanities 14.2 (1980), p. 83-90. ISSN : 0010-4817. JSTOR : [30207304](#). (Visité le 29/04/2024).

Joël CADIÈRE. L'apprentissage de la recherche en travail social. Presses de l'EHESP, 2013. (Visité le 29/10/2024).

Jacques FROGER. « Emploi de la machine électronique dans les études médiévales ». In : Les Etudes Philosophiques 3 (1961), p. 177-188. DOI : [10.1484/j.bpm.3.19](#).

Jacques FROGER. « La critique des textes et l'ordinateur ». In : Vigiliae Christianae 24.3 (1970), p. 210-217. ISSN : 0042-6032. DOI : [10.2307/1583073](#). JSTOR : [1583073](#). (Visité le 25/10/2024).

Jacques FROGER. La critique des textes et son automatisation. Dunod, 1968.

# Bibliographie II

W. LUTOSLAWSKI. « Principes de stylométrie appliqués à la chronologie des œuvres de Platon ». In : Revue des Études Grecques 11.41 (1898), p. 61-81. ISSN : 0035-2039. JSTOR : [44283556](#). (Visité le 30/04/2024).

Ian MARRIOTT. « The Authorship of the *Historia Augusta* : Two Computer Studies ». In : The Journal of Roman Studies 69 (1979), p. 65-77. ISSN : 1753-528X, 0075-4358. DOI : [10.2307/299060](#). (Visité le 06/08/2023).

Franco MORETTI. « Conjectures on World Literature ». In : New Left Review 1 (2000), p. 54-68.

Franco MORETTI. Distant Reading. Verso, 2013.

Michael PIOTROWSKI. « Epistemological Issues in Digital Humanities ». 2022. DOI : [10.5281/zenodo.6498979](#). (Visité le 24/10/2024).

Michael PIOTROWSKI et Aris XANTHOS. « Décomposer les humanités numériques ». In : Humanités numériques 1 (2020). ISSN : 2736-2337. DOI : [10.4000/revuehn.381](#). (Visité le 03/08/2023).

Camille ROTH. « Digital, digitized, and numerical humanities ». In : Digital Scholarship in the Humanities 34.3 (2019), p. 616-632. ISSN : 2055-7671. DOI : [10.1093/llc/fqy057](#). (Visité le 13/08/2024).

David SANSONE. « The Computer and the *Historia Augusta* : A Note on Marriott ». In : The Journal of Roman Studies 80 (1990), p. 174-177. ISSN : 1753-528X, 0075-4358. DOI : [10.2307/300286](#). (Visité le 21/11/2024).

Natalia SILVEIRA et al. « A Gold Standard Dependency Corpus for English ». In : Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC) 2014.

Thea SOMMERSCHIED et al. « Machine Learning for Ancient Languages : A Survey ». In : Computational Linguistics (2023), p. 1-44. ISSN : 0891-2017, 1530-9312. DOI : [10.1162/coli\\_a\\_00481](https://doi.org/10.1162/coli_a_00481). (Visité le 09/06/2023).

Justin STOVER et Mike KESTEMONT. « The Autorship of the *Historia Augusta* : Two New Computational Studies ». In : Bulletin of the Institute of Classical Studies 59.2 (2016), p. 140-157. ISSN : 0076-0730. DOI : [10.1111/j.2041-5370.2016.12043.x](https://doi.org/10.1111/j.2041-5370.2016.12043.x). (Visité le 11/01/2023).

Chris Alen SULA et Heather V HILL. « The Early History of Digital Humanities : An Analysis of Computers and the Humanities (1966–2004) and Literary and Linguistic Computing (1986–2004) ». In : Digital Scholarship in the Humanities 34 (Supplement 1 2019), p. 190-206. ISSN : 2055-7671. DOI : [10.1093/llc/fqz072](https://doi.org/10.1093/llc/fqz072). (Visité le 13/08/2024).

Boris YARKHO. « The Elementary Foundations of Formal Analysis ». Trad. par Michael LAVERY et Igor PILSHCHIKOV. In : Studia Metrica et Poetica 3.2 (2016), p. 151-174. (Visité le 25/10/2024).