

Manuscrits perdus et extinction des œuvres écrites

Modéliser la transmission des textes comme processus de vie et de mort

Jean-Baptiste Camps¹ Julien Randon-Furling²

¹École nationale des chartes, Université PSL

²Chair of Mathematical modelling and Human and Social Sciences, Centre Borelli, ENS Paris-Saclay, UPS

M2HN: Modélisation et méthodes quantitatives
January 18, 2024



Cultural Transmission and Lost Manuscripts

ERC LostMA (2024-2029)

The Lost Manuscripts of Medieval Europe: Modelling the Transmission of Texts

- Studying the tradition of different textual genres, in different languages, regions;
- finding properties of textual genealogies;
- identifying biases in survivorship of texts, manuscripts,...

2 PhD Scholarship: 2024-2027 and 2025-2028

Related problems

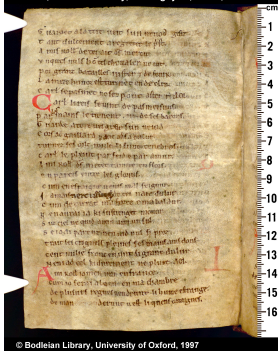
- constitution of textual canons;
- long term evolution of narrative or formal aspects.

Outline

- 1 Philology as a Data Oriented & Evolutionary Science
- 2 Bédier's problem
- 3 Solving the Bédier's problem
- 4 Current results

Philology has always been about data...

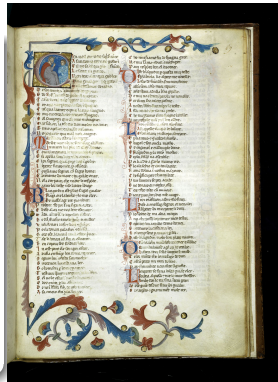
Oxford, Bodleian Library, MS. Digby 23, Part 2, fol. 52v



© Bodleian Library, University of Oxford, 1997

justification

Philology is not a purely abstract endeavour, such as mathematics or philosophy, but actually studies facts and objects from the material world, like other sciences such as biology or archaeology.

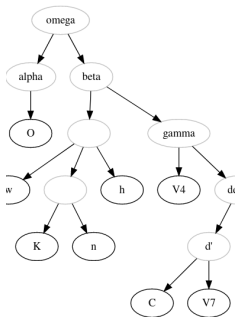


Marciana, fr. 4, fol. 69r

Digby 23, fol. 52v



... but philology is also about **models** and theory.

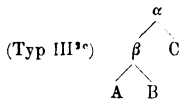
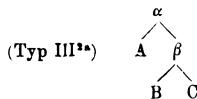


Segre's stemma for *Roland*

partial

Philology's
fundamental problem
(in my opinion)

The process of transmission
(and transformation)
of texts.

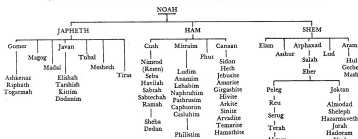


Maas (1937), *Leitfehler...*



One-to-Many Genealogies: from Babel to Darwin

GENEALOGICAL TABLE OF THE DESCENDANTS OF NOAH



Babel and the descendants of Noah

§ XXIX.

Poffet variarum lectionum ortus, per singulos codices, per paria codicum, per fyzygas minores majorisque, per familias, trius, natusque illorum, investigari & representari: & inde propinquitates diffectionesque codicum ad fchematicos quosdam reduci, & fchematiciorum aliquae concordantiae fieri; atque ita res tota per tabulam quandam quasi genealogicam oculis fubijci, ad quam tabulam quilibet varietas insignior cum agmine fuorum codicum, ad convincendos etiam tardiffimos dubitatores, exigetur. Magnam conjecturae nostra filiam habent: fed manum de tabula, ne ritum periculo exponatur veritas. Bene est, quod praetergredi montem hanc, & planiore via pervenire datur ad codices discriminandos. Datur autem per hanc regulam aequilimum: Quo fupius non modo fupjci codices, fed etiam fyzygas minores ortum vel majores, in aberrationes manifeflas tradunt; eo levius ferunt lectionum in diffeptione difficiliorum, neque magis levis ab eis diferta, tanquam genuina retrorari debet. Non est erudendum a dubijs vexatis: eas cujus aberrationes, quae culpeae minime proleantur, eligere integrum est, regulamque istam applicare; quae quidem declarationes perro aliquas admittit.

L. Vbi

Bengel (1763), *Apparatus criticus ad Novum Testamentum*, Tubingae.

S. Timpanaro (2003); C. Ginzburg (2005); also, Roelli et al. (2020).

ADDITIONS. 463

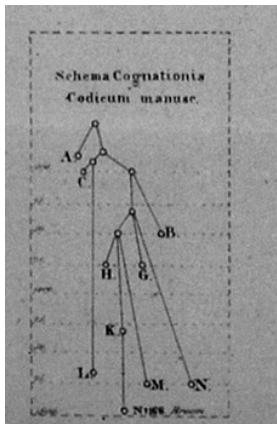
TABLEAU
Servant à montrer l'origine des différens animaux.

Vers.	Infusoires. Polypes. Râdiaires.
	Insectes. Arachnoïdes. Crustacés.
Annélides. Cirrhipèdes. Mollusques.	
	Poissons. Reptiles.
Oiseaux.	
Monotrèmes.	M. Amphibies. M. Cétacés. M. Ongulés.
	M. Onguiculés.

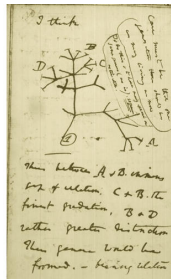
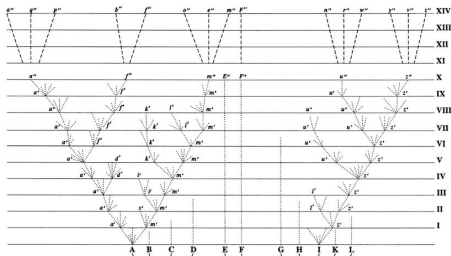
Cette série d'animaux commençant par deux

Lamarck (1809), *Philosophie zoologique*.

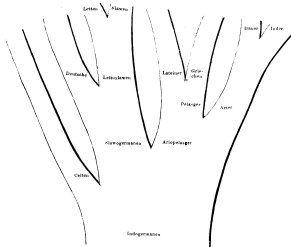
Philology, biology, linguistics: shared intuitions and models



Schlyter & Collins (1827), *Corpus iuris Sueo-Gothorum antiqui...*



Darwin (1837), *The Transmutation of species* and (1859) *The Evolution of Species*.



Schleicher (1853). *Die ersten Spaltungen...*

Descent with modification

Evolutionary sciences, individuals and tree model

philology texts → stemma

linguistics languages → language family tree/Stammbaum

biology living beings → phylogenetic trees

Trees are build from surviving or extant and recoverable specimen, based on **common innovations** (mutations, errors, changes).

→ Philology is the science of texts
and their **evolution**.

i.e., studies texts in similar ways as biology studies living organisms and linguistics language varieties.

Goal

As a science in itself (\neq as an auxiliary science), the goal of philology is not simply to provide usable texts to others, but to study and describe the processes of textual transmission.

Need to walk on two legs: data and theory.

The 'common errors' method



Gaston Paris (1839-1903)

Karl Lachmann (1793-1851)

Study the genealogy of texts to allow for **mechanical reconstruction of the original.**

Gaston Paris: *La vie de saint Alexis* (1872)

Readings are established based on the classification of manuscripts.

Paul Lejay, 1903

A family of manuscripts is constituted by their common errors, or, if one prefers this more exact term, by their common innovations (...) *The errors alone are probative.*

See also: A. Hoenen in *Handbook of*

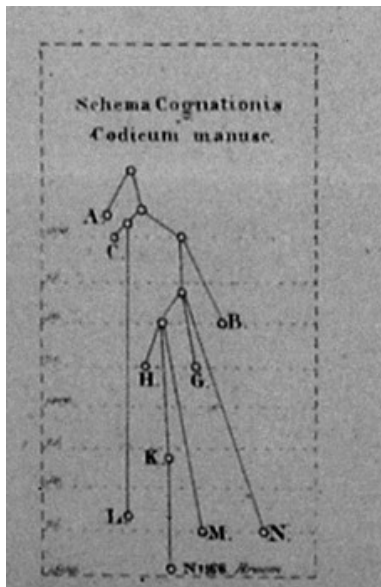
Outline

- 1 Philology as a Data Oriented & Evolutionary Science
- 2 Bédier's problem
- 3 Solving the Bédier's problem
- 4 Current results

Outline

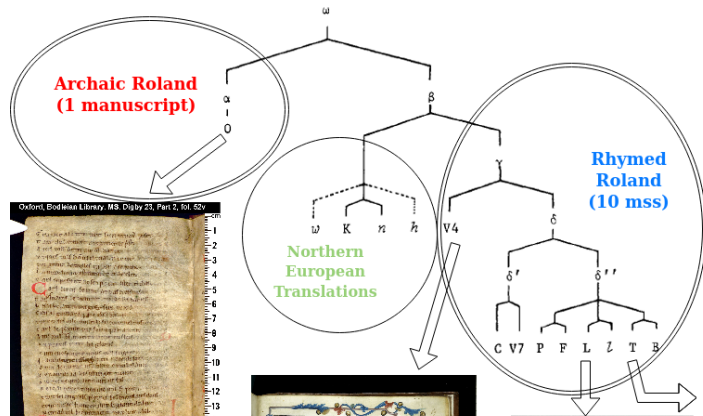
- 1 Philology as a Data Oriented & Evolutionary Science
- 2 Bédier's problem
 - Intuition
 - Nature of the problem
 - Early responses
 - Beyond bifidity
- 3 Solving the Bédier's problem
- 4 Current results

The oldest (Schlyter, 1827)

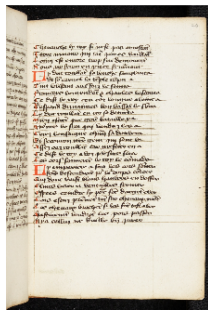
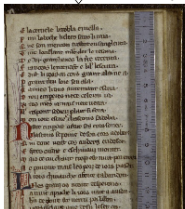
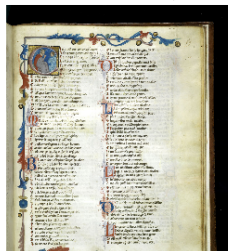
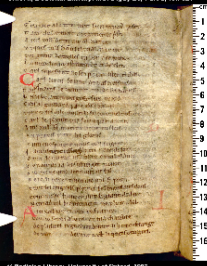


D. C. J. Schlyter & D.H.S. Collins,
Corpus iuris Sueo-Gothorum antiqui,
I (Samling af Sweriges Gamla Lagar),
Stockholm, 1827, Tab. III.

La Chanson de Roland

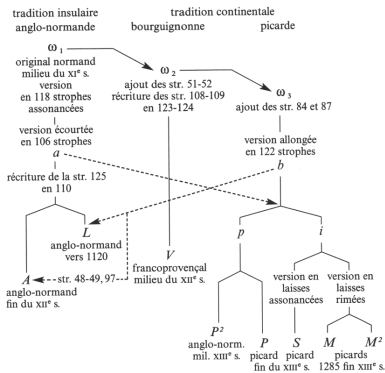
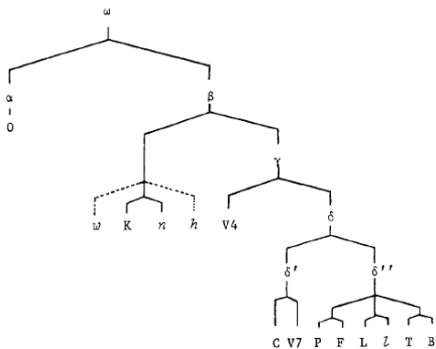


Oxford, Bodleian Library, MS. Digby 23, Part 2, fol. 52v



Two long lived Old French texts

Roland (Segre) and saint Alexis (Zufferey)



'Real trees' of the printed editions of two Renaissance texts

(Trovato & Guidi, 2004)

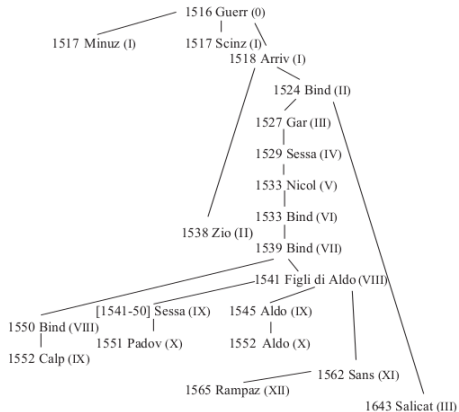


FIG. 2. Lo stemma Richardson della tradizione a stampa di Fortunio, *Regole grammaticali*. I numeri romani tra parentesi tonde indicano il rango genealogico delle varie edizioni.

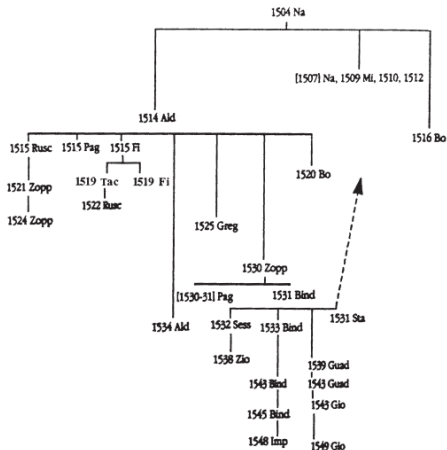
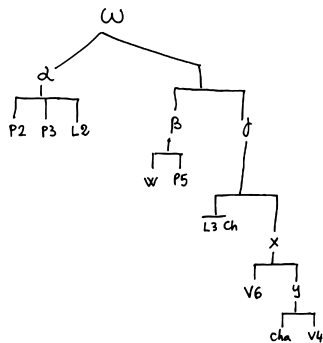
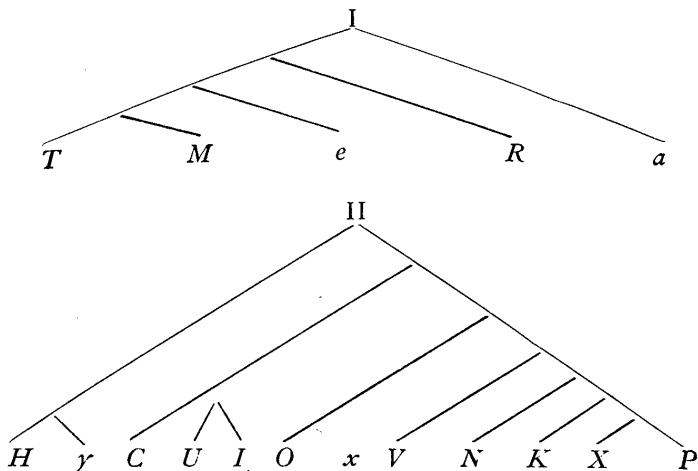


FIG. 3. Lo stemma Trovato di Sannazaro, *Arcadia* (n. redazione).

The *Chanson d'Aspremont* (Palumbo, 2013)



The lyrical songs of Conon de Béthune (Wallensköld, 1891)



The Odd properties of stemmata

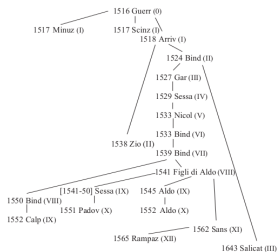
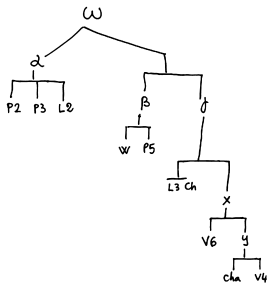
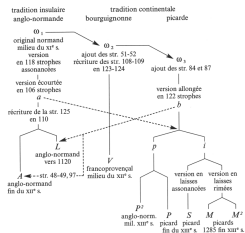
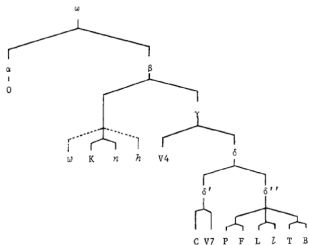


FIG. 2. Lo stemma Richardson della tradizione a stampa di Fortunio, *Regole grammaticali*. I numeri romani tra parentesi tonde indicano il rango genealogico delle varie edizioni.

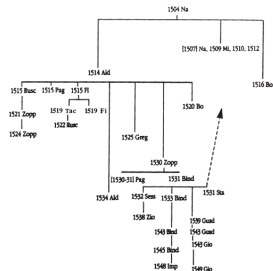
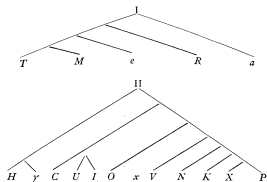
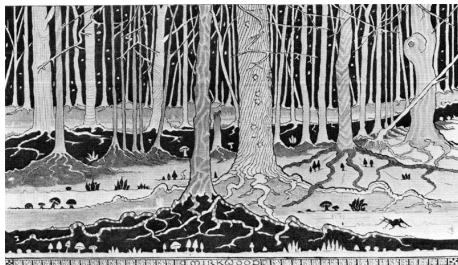


FIG. 3. Lo stemma Thivrot di Sanzauro, *Arcaica* (a redazione).



Deeper into the ominous forest

The Paris-Bédier observation



JRR Tolkien, *Mirkwood*, 1937.

G. Paris (1872, p. 10)

Nowhere is the immensity of the losses we have suffered in medieval manuscripts more evident than when we possess several texts of the same work. Indeed, it is infinitely rare for one of these texts to be copied from the other; **almost always they are the extreme shoots of perfectly distinct branches, which form a vast ramification around the stem, which we never possess.**

Bédier (1913, 1928) collected stemmata with tracing paper. A vast majority ($> 95\%$) of trees show a root bifurcation.

Silva portentosa

Methodological bias?

My PS: Bédier forgets **all the texts in a single manuscript.**

Why it's a problem?

The original!

Keep in mind the main goal at the time was:

having a way to reconstruct the original (or the archetype at least)
in the less subjective and most mechanical way

Two branches → no majority principle → return to qualitative assessment...

Bédier's (*bona fide?*) conclusion

philologists want to choose, break the mechanics.

His suggestion

Stop reconstructing the original (sometimes understood as: let go of the genealogical approach entirely).

Why it's a problem?

The original!

Keep in mind the main goal at the time was:

having a way to reconstruct the original (or the archetype at least)
in the less subjective and most mechanical way

Two branches → no majority principle → return to qualitative assessment...

Bédier's (*bona fide*?) conclusion

philologists want to choose, break the mechanics.

His suggestion

Stop reconstructing the original (sometimes understood as: let go of the genealogical approach entirely).

A law?

A fact of bipartition in the history of the transmission of texts is not surprising, but the law of bipartition is surprising, and the astonishment grows (...) the more it assumes the constancy, the majesty, the necessity of a law of nature.

Necessary and, by the same token, absurd, it cannot possibly have governed the destiny of texts in ancient times.

or a methodological bias?

as long as a schema retains three branches, it is an automaton that has been delegated once and for all to the task of establishing the text. (...) apart from the four or five naive people who simply accepted slavery, the editors of old French texts have all found out how to manage, some, the semi-skilled, to weaken the automaton, others, the skilled, to paralyse it completely.

Previous work on bifidity: approaches

Descriptive quantify the properties of observed stemmata, from Bédier's 95.5% of bifidity; to values c.70%-83% (Shepard, 1930; Castellani, 1957; Haugen, 2015);

Combinatorial Maas (1937) ... Hoenen (2017): how many root bifurcations for n nodes, assuming all equally likely

Decimation/loss-based

- Greg (1931) “**Prevalent dichotomy is merely due to decimation**”;
- Fourquet (1945) ... Trovato and Guidi (2004) static approach, decimation applied to preexisting trees.

My two thoughts on that

- ① Need an open collection of stemma on which to establish quantification;
- ② Study of stemma properties need to account for the **dynamic** nature of transmission (cf. Weitzmann).

Observations

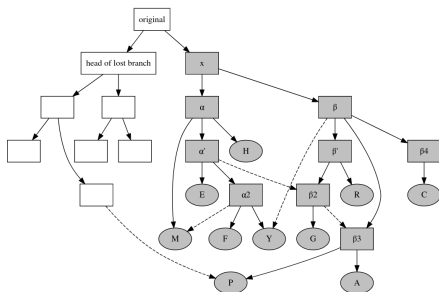
A Pareto-Like World

- prevalence of root bifurcation (cf. Bédier);
- asymmetry/inequality of branches;
- extinct and **almost extinct branches** vs single productive branch;
- archetype as root rather than the original;
- the new version drives out the old (perhaps).

Beyond bifidity: some properties to monitor

That can all reflect the **extent of losses** as well as the **dynamics of transmission** and the **randomness or deviation from chance**.

F



Guy de Warewic (Ewert, 1932) with lateral transmission (contamination)

- Root and node outdegree (2 or not);
- Root (LCA) as archetype or original;
- asymmetry in branches / 'biodiversity' of the branches;
- (direct) filiation between two witnesses;
- Contamination and extra-stemmatic contamination.

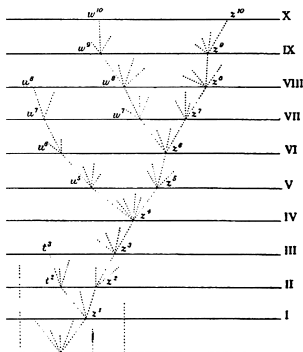
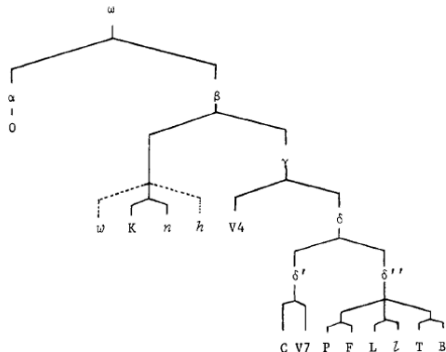
'imbalanced tree' problem

Biology

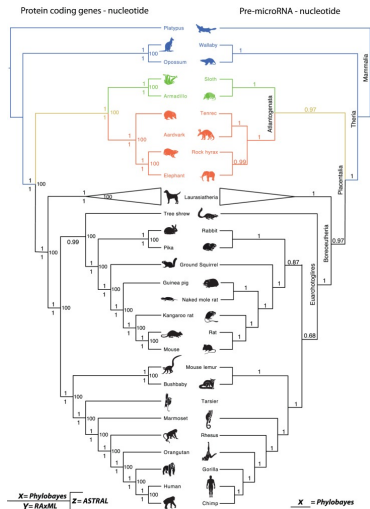
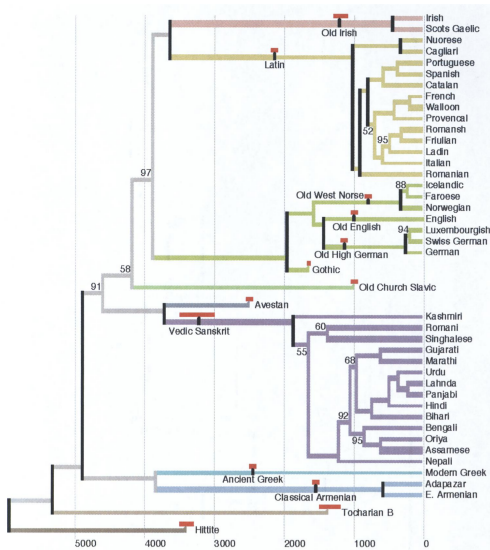
Yule (1925): long-tailed distribution of species per *genus* (Homo, Ornythorinchus,... 1; Agrilus, 2265)

Aldous (2001): 'imbalanced trees'.

Methodological bias, or effect of drift, selection?



Biology and linguistics



Outline

- 1 Philology as a Data Oriented & Evolutionary Science
- 2 Bédier's problem
- 3 Solving the Bédier's problem
 - Framework
 - Methodological approach
 - History of this research
- 4 Current results

Main question and sub-questions

- ① Is bifidity (and other structural properties) **A**, the result of methodological bias or **B**, a real feature resulting from the transmission of texts?
- ② If **B**, is it due to an essentially random process or to a type of selection (drift vs selection)?
- ③ If selection, what type (content based, reinforcing, ...)?
- ④ What can it teach us about other aspects of transmission (extent of losses, dates, ...)?

Antagonist tendencies to...

Variation (diversity increases)

- accidents in the transmission (copying errors, lost folio, ...);
- intentional innovations.

Favoured by

- evolution of language and taste;
- “active tradition” of less authoritative texts (Vàrvaro);
- variance supposedly consubstantial with vernacular works (Cerquiglino);
- *varietas tollit fastidium*.

Fixation (diversity decreases)

Extinction is the rule.

This operates as a force of homogenisation of traditions.

Caused by...

- Random birth and death processes?
- “Biases” induced by the participants (scribes, patrons, audiences,...)?

Three perspectives to articulate

We need a **common framework** that can account for observations and investigations in these three areas

Population dynamics,
Loss & productivity

Distribution of “birth”
and “death” rates of
manuscript
populations;

Loss of manuscripts,
works, ...
cf. Buringh, Cisne,
etc.

Phylogenetics /
Stemmatology

Structural Properties
of textual **genealogies**.

Ecosystemic approach /
Reception

Selective/adaptive value in
a given “ecosystem”;
biodiversity of texts

- historical and literary context;
- **taste** (literary, aesthetic);
- **fashions**;
- **canon** and vulgate texts.

A methodology with three components

Generative models

Models of textual transmission

- analytic models + simulations;
- Null-model;
- geotemporal models?

Data collection and analysis

- collect stemmata;
- collect data on textual traditions;
- quantify those properties.

Models vs data

- compare simulations with data on observable properties;
- infer non observable properties;
- quantification + machine learning.

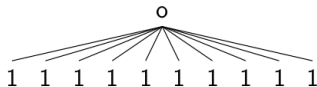
Step 1: 2013, the initial idea

Des premières tentatives à la stémato-logie « basée sur l'ordinateur »

De la critique de Joseph Bédier aux tentatives de Dom Quentin
Dom Froger

Mettons que nous partions de 10 copies de l'original, et qu'un manuscrit ait

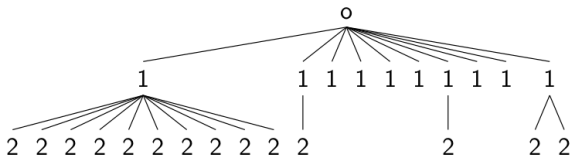
- 50% de chances de n'être pas copié du tout ;
- 40% de chances d'être copié entre 1 et 4 fois (10% 1 fois, 10% 2 fois,...) ;
- 10% de chances d'être copié entre 5 et 10 fois (2% 5 fois, 2% 6 fois ...)



Step 1: 2013, the initial idea

Des premières tentatives à la stémattologie « assistée par ordinateur » De la critique de Joseph Bédier aux tentatives de Dom Quentin et Dom Froger

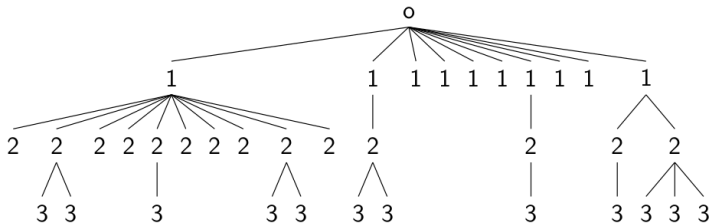
2^e génération, quatre branches actives



Step 1: 2013, the initial idea

ordinateur » Dom Froger

3^e génération, quatre branches actives (dont une quasi inactive)

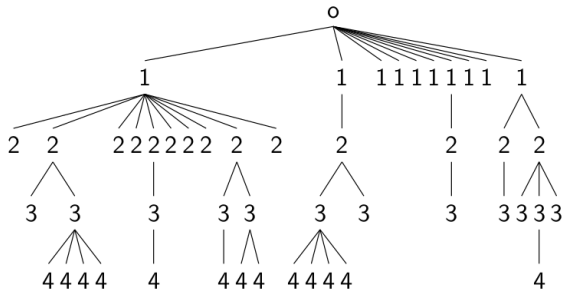


Step 1: 2013, the initial idea

Des premières tentatives à la stémato-logie « assistée par ordinateur »

De la critique de Joseph Bédier aux tentatives de Dom Quentin et Dom Froger

4^e génération, trois branches actives (dont une quasi inactive)

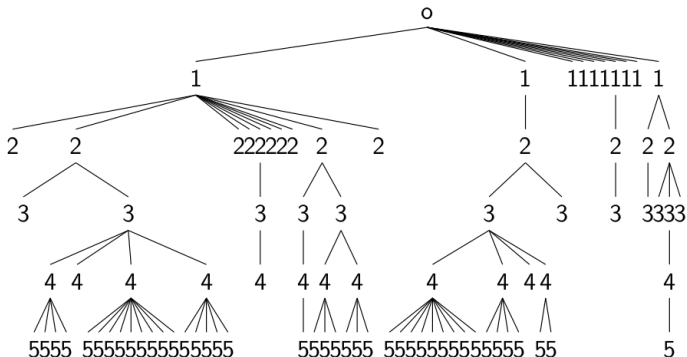


Step 1: 2013, the initial idea

Des premières tentatives à la stémattologie « assistée par ordinateur »

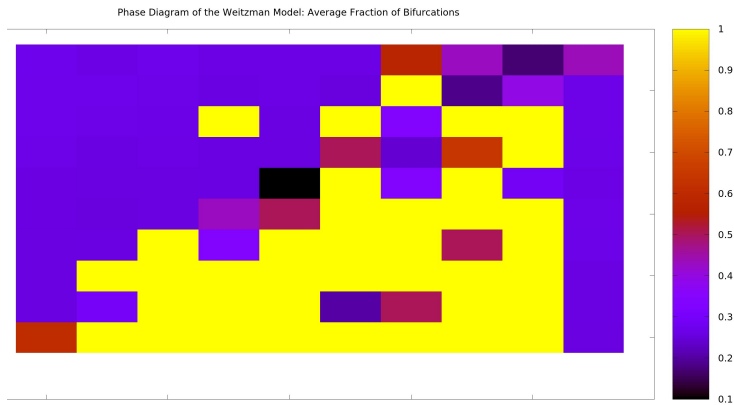
De la critique de Joseph Bédier aux tentatives de Dom Quentin et Dom Froger

5^e génération, trois branches actives (dont une quasi inactive)



I then went on to finish my PhD. In 2017, we put in place the new MA in Computational Humanities at the École des chartes. In this context, a class of «Mathematical modelling» was established, that I decided to follow. In it, Julien Randon-Furling presented a variety of modelling approaches and agent-based systems. After a quick discussion together, it became obvious that it was the right tool to approach this question: in fact, it turned out it had already been used in the 1980's in this direction, in two unnoticed papers by Weitzmann. We then started to work on a 'Dynamic Model of Manuscript Transmission''

Step 2: 2017, collaboration on a dynamic model of manuscript transmission



Step 3: 2022, proof-of-concept paper

Publ. E

JB Camps, J. Randon-Furling, "Lost Manuscripts and Extinct Texts: A Dynamic Model of Cultural Transmission", *CHR2022*.

How to model a **complex process** where

- production, loss etc. may vary substantially through time;
- manuscripts may be copied, changed, lost and retrieved;
- entire branches may go extinct almost suddenly?

Weitzman (1982, 1987)

- model of manuscript population evolution via birth & death processes;
- numerical simulations;
- case studies: Greek & Latin classics.

Canettieri et al. (2008); Hoenen (2016)

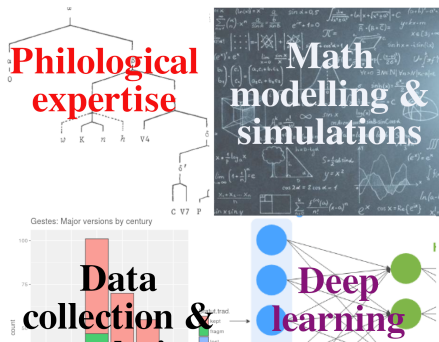
- parallels with:
gambler's ruin problem,
species extinctions,
family names extinction, ...
- semi-dynamical approaches.

Step 4: project LostMa

Proof of concept Camps & Randon-Furling, “Lost Manuscripts and Extinct Texts: *A Dynamic Model of Cultural Transmission*”, *CHR2022*.

Current Corpus 65k medieval manuscript pages; 40 million words
Langlais, Camps et al., “Corpus of French literary fictions (1050-1920)”, *DH2022*.

Multidisciplinary methodology



Impacts

Understanding human cultures

How cultural artefacts ‘lived and died’

Methodological paradigm shift

Transforming humanities methods

Philology

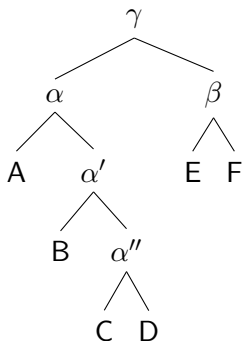
Unbalanced tree problem (Bédier) 37 / 64

Outline

- 1 Philology as a Data Oriented & Evolutionary Science
- 2 Bédier's problem
- 3 Solving the Bédier's problem
- 4 Current results
 - A null-model for manuscript transmission
 - Collecting data on textual traditions
 - Comparing model and data

Towards a Null-Model of Manuscript Transmission

- Is it possible to understand the dynamics at hand through **computer simulations and analytic methods**?
- Can a stochastic model reproducing manuscript transmission be designed?
- Can deviations from this model be measured and explained?
- Can the results of model simulations be crossed with observables of textual traditions?
- Can non observables be inferred (e.g., survival rates)?



Modelling a complex process

Joint work with Julien Randon-Furling (statistical physics, Univ. Paris Panthéon-Sorbonne)

How to model a **complex process** where

- production, loss etc. may vary substantially through time;
- manuscripts may be copied, changed, lost and retrieved;
- entire branches may go extinct almost suddenly?

Weitzman (1982, 1987)

- model of manuscript population evolution via birth & death processes;
- numerical simulations;
- case studies: Greek & Latin classics.

Canettieri et al. (2008); Hoenen (2016)

- parallels with:
gambler's ruin problem,
species extinctions,
family names extinction, ...
- semi-dynamical approaches.

Stochastic Models

- “Birth & death” process:
 - each agent disappears at (death) rate $\mu(t)$,
 - and each agent generates “descendants” at (birth) rate $\lambda(t)$.
 - (in philology see Weitzman 1982 & 1987)
 - ↪ phase diagram when μ and λ are varied across their ranges?
 - ↪ estimates from data for μ and λ ?
- Branching process:
 - an agent splits into several branches at a certain (branching) rate,
 - branches disappear at a certain rate (locally and/or globally),
 - branches may combine/recombine.
 - ↪ much used in phylogenetics
- Split-and-Drift random graphs:
 - a *null model* for speciation
 - introduced in 2019 by Bienvenu, Débarre & Lambert

Models

Weitzman 1

- individuals appear (are born) at constant rate λ ;
- and disappear (die) at constant rate μ ;
- *can be solved with equations.*

Weitzman 2

- same, but with variable rates; (*need computer simulations, for now*).

Cisne

Account for population dynamics and logistic growth

birth rate λ depend on pop size at step k_{t-1} and max support capacity (K)

$$\lambda_t = \lambda - \left(1 - \frac{k_{t-1}}{K}\right),$$

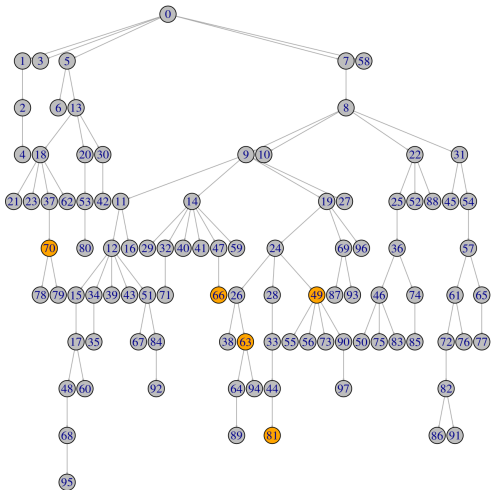
Cisne (2005), "How science survived: medieval manuscripts' demography and classic texts' extinction" *Science* 307(5713) 1305-1307

Current simulations: choosing parameter ranges

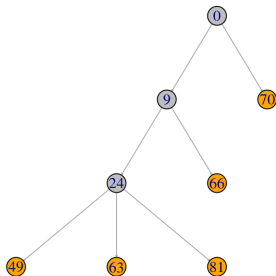
- μ loss estimates based on historical information, around -25% per century, globally around 90 or 99% for 500 years (Trovato), or 75% (Buringh). For 1% survival rate, $(1 - \mu)^{2000} = 0.01$, so $\mu = 0.002$. So we retain values of μ between 10^{-4} to 10^{-3} .
- λ can not be order of magnitudes different from λ .
- $K = 100\,000 = 10^5$: medieval population of France or Italy was in the 10^7 range; maximum saturation of this market for a given book around 1% of the population?
- T 250 pseudo-years active, 250 inactive ($\approx 1250-1499, 1500-1750$), t is ≈ 3 months (ms. copy time), so 2000 steps.

One simulation result ($\lambda = 0.0008$, $\mu = 0.0005$ and $K = 10^5$)

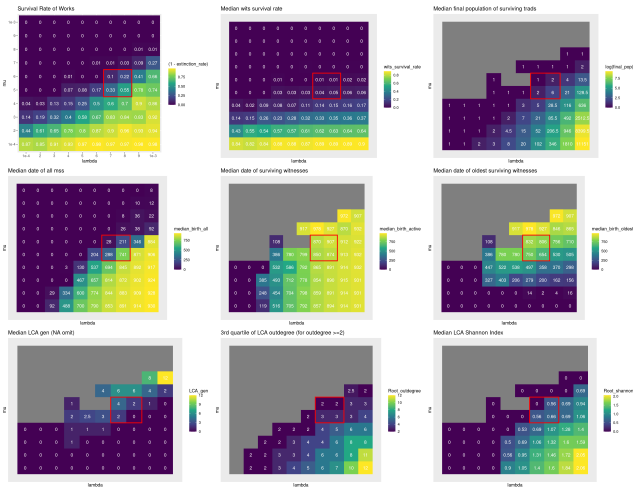
Pop_8_5_74



Pop_8_5_74

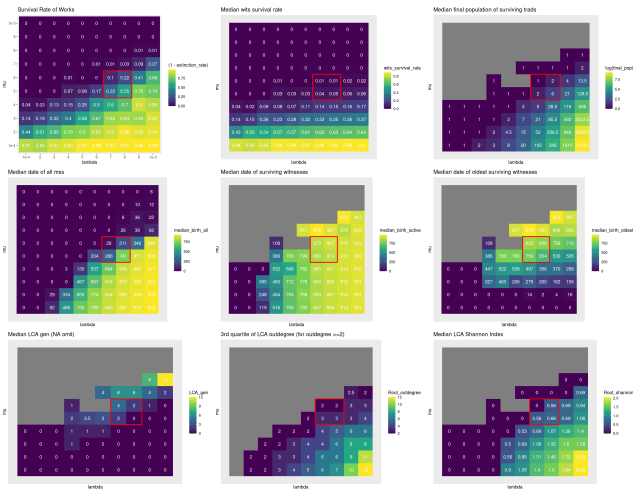


Phase diagrams



100 simulations per pair of parameter values.

Phase diagrams



results for ratio $\frac{\mu}{\lambda}$ between $\frac{5}{8}$ and $\frac{6}{7}$ surprisingly consistent with the observed properties of the traditions of medieval chivalric narratives, and particularly Old French epics (*chansons de geste*).

The OpenStemmata project

Building an open collection of stemmata

How to contribute

We welcome contributions by all, on any kind of textual tradition (any period in history, language, genre, etc.).

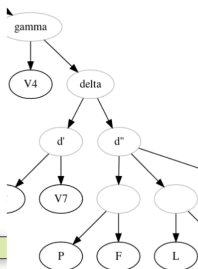
If you are new to digital philology, have a look at our [detailed guidelines](#).

Contributions can be sent as pull request on our [github repository](#). They are composed of

- A **metadata file**, that can be created using [our form](#) (**mandatory**)
- An **image** file, i.e. a scan of the source stemma (optional)
- A **DOT (Graphviz) transcription** of the stemma, that can be easily done with the help of an online editor such as [Edotor](#) (optional)

If you're a bit lost or don't know how to make a PR, send us your file by writing an [issue ticket](#).

LEARN MORE ON HOW TO CONTRIBUTE !



J.B. Camps, Gustavo Fernández Riva, Benedetta Salvati, Simon Gabay, Lena Reich, *OpenStemmata*.

Collaborative, Open Source collection of stemma reproduction, graphs and metadata.

Already **more than 150 stemmata** in the database, and a lot more to come.

Contribute at: openstemmata.github.io/

openstemmata.github.io/ Team

Core Team



J.B. Camps



Gustavo
Fernandez
Riva



You?

Role

- maintain the submission guidelines and website;
- peer review submissions;
- maintain workflow.

Requirements

Membership of
github.com/OpenStemmata/

Contributors

- Aurélien Berra
- Simon Gabay
- Lena Reich
- Benedetta Salvati



You?

Role

- Submit contributions
- report issues

Requirements

Follow submission guidelines

Scope and submissions

Scope

In principle

- any time;
- any language;
- any medium (mss, print,...);
- any type of text (music,...);
- any publ. source (arts, eds, ...).

In practice Current focus on

- Medieval
- European
- epics and romances
- and lyric poetry.

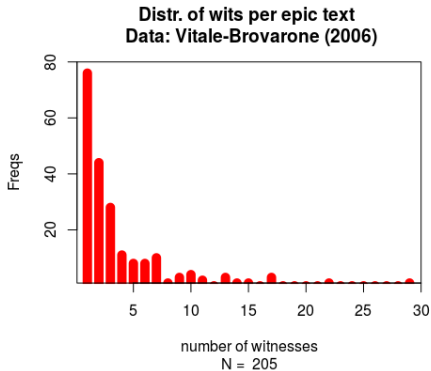
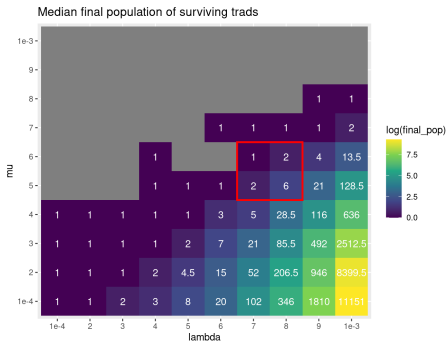
Outline of a submission

- 1 **Metadata file** (mandatory), as txt;
- 2 **image** of the stemma, as png;
- 3 **transcription** of the graph, as dot.

Looking for contributions

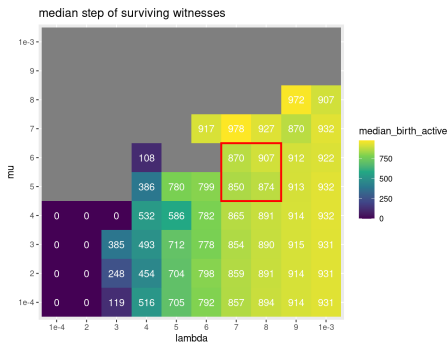
- Latin
- Medieval Italian
- Middle Welsh
- Old/Middle English
- Norse

Surviving manuscripts: final population



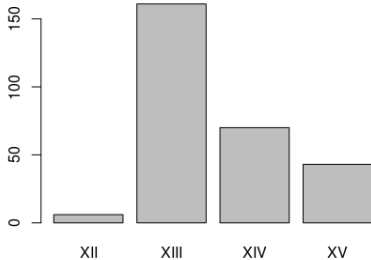
Median: 2; Mean: 3.5

Surviving manuscripts: age



(roughly 200 years after the origin in average).

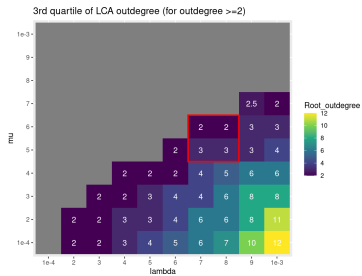
Epic mss per century



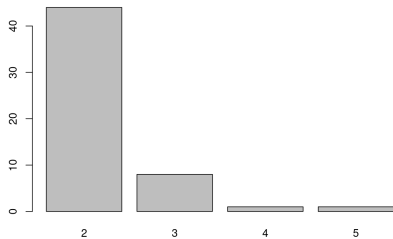
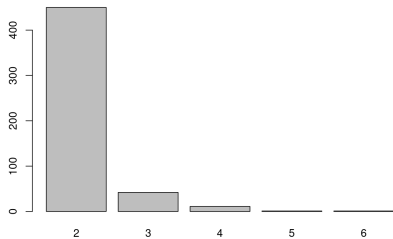
From Duggan (1982), Careri (2006) and Careri et al. (2011)

Distribution center: somewhere between mid XIIIth to early XIVth.

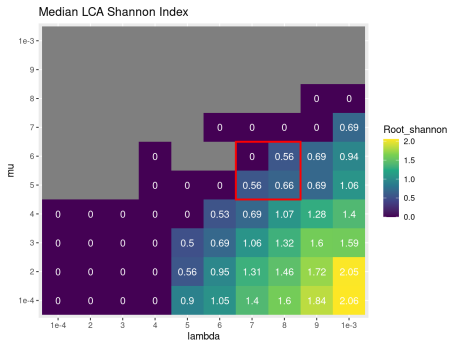
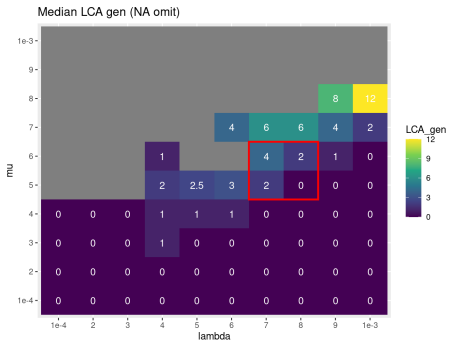
Observed stemmata: root outdegree



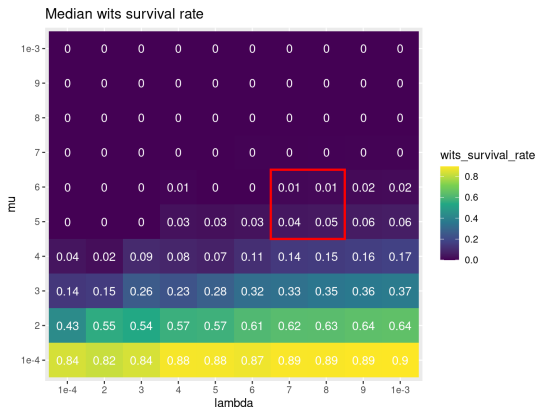
Median: 2; 3rd Qu.: 2; Mean: 2.4



Other stemmatic properties



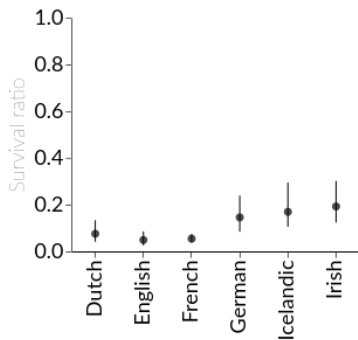
Loss of manuscripts



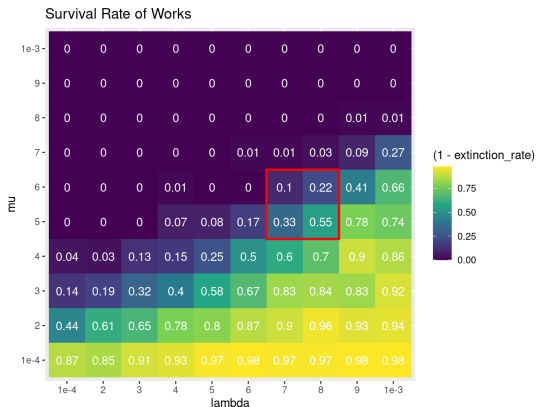
Kestemont, Karsdorp et al.
2022

Mean: **0.05**

SURVIVAL OF DOCUMENTS

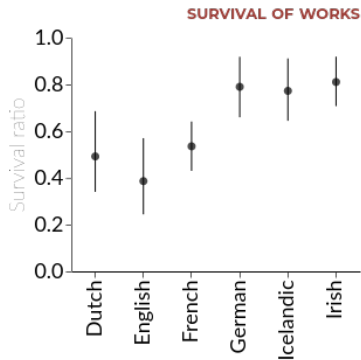


Loss of works



Kestemont, Karsdorp et al.
2022

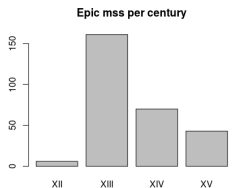
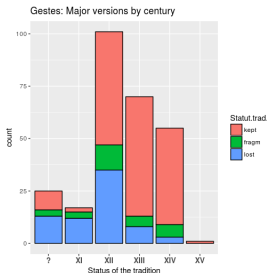
Mean: **0.54**



Current loss estimates: discussion

- estimates by Kestemont et al. certainly in range of consistent values;
- congruent with estimations on *romans*, but perhaps too low for chansons de geste, if model is right;
- for those, more likely to be:
wit. survival rate 0.01; work survival rate c. 0.25-0.33.
- (consistent with estimates by Trovato et al. on chivalric incunabula).

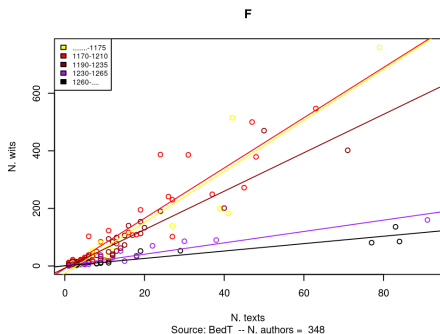
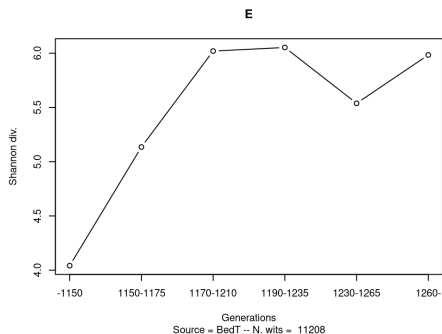
Why? Specific case of bottleneck (12th c. mss) + vulnerable book types.



From Duggan (1982), Careri (2006) and Careri et al. (2011)

The erosion of “biodiversity” of texts

The example of Troubadour Poetry



Mass extinctions

Mass extinction of the very early French literature (before 1150)?



Figure: Artist's view of the Chicxulub impact (-66M); North Pacific waste vortex (c. 2021)

Current conclusions

Current results

- Already, the ability to **calibrate** on observables, and infer non observable values (e.g., loss rates);
- cross-verify or revise existing estimates (e.g., Kestemont et al.).
- most of Bédier's observation seem to be explained by this very simple null model (so, drift?).

Future work

- **analyse other traditions**, other languages,...;
- **refine models**: variable rates, ecological 'niches', geotemporal models,...
- experiment with different types of **selection**;
- model **content change**;
- **progress on mathematical understanding** of the models, and continue simulations;
- experiment with **hybrid approaches**: simulations + machine learning.

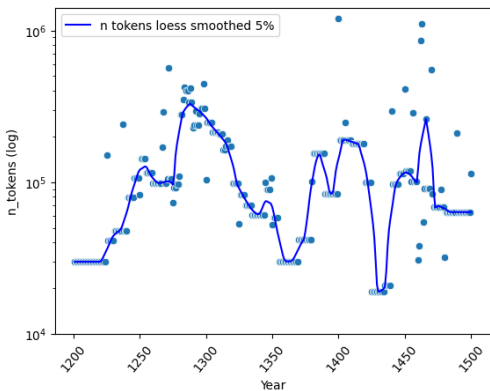
Renewed questions

- What part does drift and selection play in what we have preserved or lost? in the literary canon?
- How do content or context biases (e.g., conformity, prestige) affect the selection of variants or versions? How is a variant or a vulgate established?
- Effect of bottlenecks (e.g., transition from pre-Gothic to Gothic script)?
- How has the variety of texts, themes and forms increased or decreased over time?
- What is the effect of the existence of different regions or 'niches' (particularly more isolated or island regions)? Circulation between regions?
- In what contexts does lateral transmission (contamination) occur?
- What drives convergent evolution (e.g. polygenetic variants)?

For these macro investigations, computational methods are central.

The Corpus of Medieval French Epics and Romances

Corpus	MSS	Witnesses	Word tokens
CMFER- <i>full</i>	265	409	38.5M
CMFER- <i>select</i>	203	370	36.4M

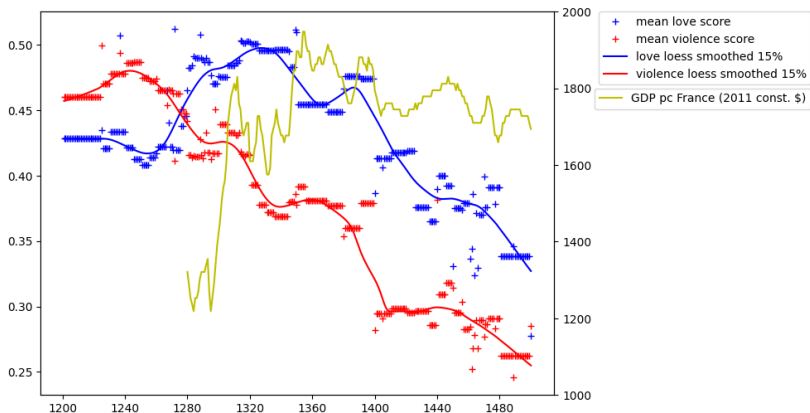


Scope

- All digitised mss and incunabula of ch. de geste and romans 'de chevalerie'
- 12th-15th c.

J.B. Camps *et al.*, "Make Love or War? Monitoring the Thematic Evolution of Medieval French Narratives", CHR2023 Paris.

Testing DUBY hypothesis



CHR2023

J.B. Camps *et al.*, "Make Love or War? Monitoring the Thematic Evolution of Medieval French Narratives".

Beyond drift?

- How substantial is the deviation from chance in observed stemmatic properties?
- Are they biases leading to some types of selection,
 - context: prestige; conformist;...
 - content: emotional nature, informativeness,... (e.g., importance of love).

Thank you for you attention!

