

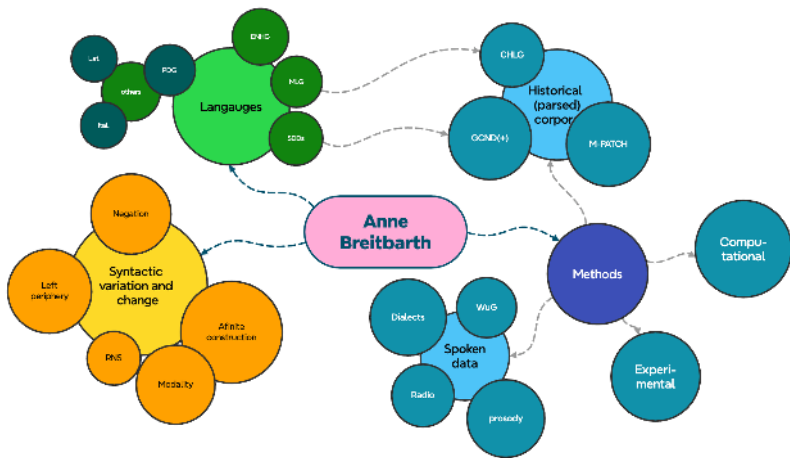
MYTH MAKING METHODS (I)

Bad data, (slightly) better data, and how to find them

Anne Breitbarth

ETHL Summer School 22–26 June 2026

WHO I AM AND WHAT I'M DOING HERE



BAD DATA

THE KNOWABILITY OF THE PAST

“The task of historical linguistics is to explain the differences between the past and the present; but to the extent that the past was different from the present, there is no way of knowing how different it was.” (Labov 1994: 21)

HOW?

- We need historical sources that allow us to form a realistic image of the past.
- What is the problem?

“Historical linguistics may be characterized as the art of making the best use of bad data.” (Labov 1982: 20)

WHY ARE HISTORICAL LANGUAGE DATA BAD?

- 1 Modality
- 2 Scarcity
- 3 Representativeness (1)
- 4 Representativeness (2)
- 5 Mediation

MODALITY

- Depending on the period we're looking at, data are most likely **written** (or **not available** at all) — it's slightly better for more recent stages, but caution with **data validity** should be universal
- Written language tends to be **conservative** and often **distant**

MODALITY

- (1) Als Jch gar nahe von jugendt auff sondere begürde gehabt [_]/ in ferre Landschafften zuo ziehen/ fürnemlich aber inn die Morgenlaender/ alß die für andere mehr beruembt vnd fruchtbar seind/ welliche die aelteste Voelcker erbawet [_]/ auch die mechtigisten Potentaten vnd Monarchen der Welt/ vor Jaren jnnengehabt/ vnnnd bewohnet [_]/ nicht allein derselbigen Jnnwoner leben/ sitten vnd gebraeuch/ warzune=men/ sonder auch vnd vilmehr die schoene Gewaechs v=n Kreüter/ vom Theophrasto, Dioscoride, Auicenna, Serapione 'c. beschriben/ an den enden vnnnd orten/ da sie wachßen/ zuerkundigen/ vnnnd zuerkennen/ damit mir jre beschreibungen inn ansehung vnd betrachtung deren/ sonderlich aber der mehr frembden vnd vnbekandten/ zum thail bekandtlicher vnd verstendtlicher wurden/ zum thail auch den Apoteckern vrsach gebe/ ferrner nach denen/ die jhnen zu haben von noeten / zu trachten/ **Suochet ich diß mein fürne=men ins werck zusetzen/ mittel vnd weg/ ...** (Rauwolf, *Morgenlaender* 1582)

SCARCITY

- Again, depending on the period we're looking at, we might be lucky to even have **written** data; often, data are **not available** at all
- The older, the worse: **material witnesses** from, say, the Middle Ages are only preserved \pm by **chance** (floods, fires, recycling etc.).
- **Suvivorship bias**: We tend to assume that other texts will have been similar to the ones we know (Kestemont et al. 2022)



■ Gaps in transmission.

10. Jahrhundert [\[Bearbeiten | Quelltext bearbeiten \]](#)

Im 10. Jahrhundert verstummt die deutsche Literatur fast völlig,

(Wikipedia entry on OHG literature)

	Jespersen's cycle in Low German		
c.	<i>ni/ne/en</i>	<i>ne/en...nicht</i>	<i>nicht</i>
9th	654 (100%)	-	-
14th	2 (0,4%)	248 (52,4%)	223 (47,1%)
16th	1 (0,1%)	132 (18,2%)	593 (81,7%)

(Breitbarth 2014)

■ No transmission at all.

(Hagemeijer 2011)

120 Tjerk Hagemeijer

Table 4. Liquids in the GGCs.

PR	ST	ANG	FA	Portuguese	English
<i>riku</i>	<i>liku</i>	<i>liku</i>	<i>liku</i>	rico	rich
<i>kura</i>	<i>kula</i>	<i>kula</i>	<i>kula</i>	curar	cure
<i>arê</i>	<i>alê</i>	<i>alê</i>	<i>alê</i>	rei	king
<i>ranha</i>	<i>lenha</i>	<i>lenha</i>	<i>lanha</i>	rainha	queen

Table 8. Clause internal function words in the GGCs, Edo and Portuguese.

	Ptg.	Edo	ST	PR	ANG	FA
relative marker	<i>que</i>	<i>nè</i>	<i>ku</i>	<i>ki</i>	<i>ki/ma</i> ^a	<i>ku</i>
Wh-marker ^b	<i>é que</i>	<i>nè</i>	<i>ku</i>	<i>ki</i>	<i>ki/ma</i>	<i>ku</i>
Complementizer	<i>que</i>	<i>vbènè</i>	<i>kuma</i>	<i>ya</i>	<i>ma</i>	<i>fa</i>
Focus marker	<i>é que</i>	<i>ore</i>	<i>so~soku</i>	<i>èli~eli ki</i>	<i>tho</i>	<i>se</i>
Clause linker	<i>então, aí 'so'</i>	<i>ore</i>	<i>so~soku</i>	<i>èli~eli ki</i>	<i>tho</i>	<i>se</i>

UNREPRESENTATIVE (1):

- **People who could write:** for most of the written history of European languages, that's **men** from a higher social class (nobility, clerics, later clerks, ...).
- Labov (1990; 2001): **Women tend to lead linguistic change.**

Language Variation and Change, 2 (1990), 205–254. Printed in the U.S.A.
© 1991 Cambridge University Press 0954-3945/91 \$5.00 + .00

The intersection of sex and social class in the course of linguistic change

WILLIAM LABOV
University of Pennsylvania

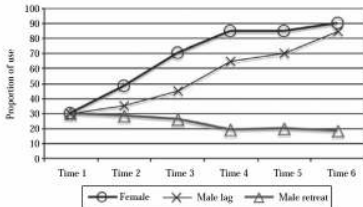


Figure 2.7 Idealized pattern of female-led linguistic change.

(Tagliamonte 2012:33)

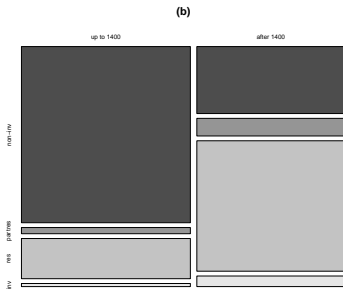
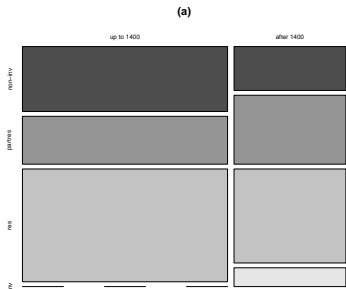
UNREPRESENTATIVE (1):

“... there is no way of knowing how different [the past] was”:

The conservative behavior of women is therefore limited by particular historical and cultural histories of the regions involved. It appears that where women have not traditionally played a major role in public life, cultural expectations will lead them to react less strongly to the linguistic norms of the dominant culture, and that this cultural force will apply even to women who are individually active and successful. It remains true that within the dominant pattern of the West, women behave in two opposite ways; they are more conservative in regard to established linguistic norms, and less conservative in regard to new linguistic changes in progress.
(Labov 1982: 78–79)

UNREPRESENTATIVE (2):

- Things that were written down: contracts | narrative texts | letters | ...
- Genre influences what we can know about syntactic change (Breitbarth to app.)



MEDIATION

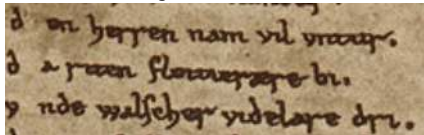
■ What is a “text”?

- Multiple manuscript witnesses, lots of variation
- Diplomatic vs. critical editions?
- Multitext/parallel editions?

Leitzmann/MHDBDB

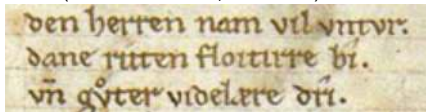
10 den herren nam vil untûr,
11 dâ **enriten** floitierre bî
12 und guoter videlaere drî.

ms. G (München Cgm 19, [1228-1236](#))



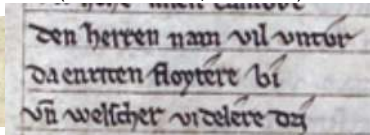
d on herren nam vil untûr.
d a guoten floitierre bi.
ynde walscher videlare dri.

ms. D (St. Gallen cod 857, [after 1250](#))



den herren nam vil untûr.
dane ritzen floitierre bi.
vñ gûter videlare dri.

ms. T (Wien Vind. 2708, [after 1270](#))



den herren nam vil untûr
Da enriten floitierre bi
vñ welscher videlare dri

(Wolfram von Eschenbach, *Parzifal*)

- **Editions:** complicated relationship with material witnesses (and variation in them), also tendency = the older, the worse

(Vanderputten/Breitbarth 2026)

DIE TERTIA

▲

CAPUT I.

Contracti sanati, disrupta ferrea vincula, caecis lumen redditum & muta loquela; pejerans punitus.

CAD. 1. **E**O namque tempore, quo fabrica altaris ecclesiae nostrae * agebatur, contigit adesse ibi quendam Ioseph nomine, qui ita curvus erat, ut supra genas incubens caput sursum erigere non posset: sed nec penitus immobile id uno in loco tenere diutius valebat. Qualiter autem illi acciderit, referebat: quia dum esset solus positus cum suo equo in campo vicini sui, eoque herbam injuste depalceret, à fulgure ignis * divinitus percussus ita exustus est, ut mox capite deposito curvus redderetur. Quadam vero nocte, qua Dominica dicitur, dum fratres ad divinum Officium persolvendum devoti insisterent, ecce idem debilis in pavimento templi prorvens, magnis clamoribus vocum * ecclesiam replevit, in tantum ut praeter nimio pavore hi, qui ad vigiliis observandas aderant, veloci cursu fugientes loca altaris irreverenter irumperent, necquid illis agere conveniret.

↑ Ms. Brussels KBR II 1180

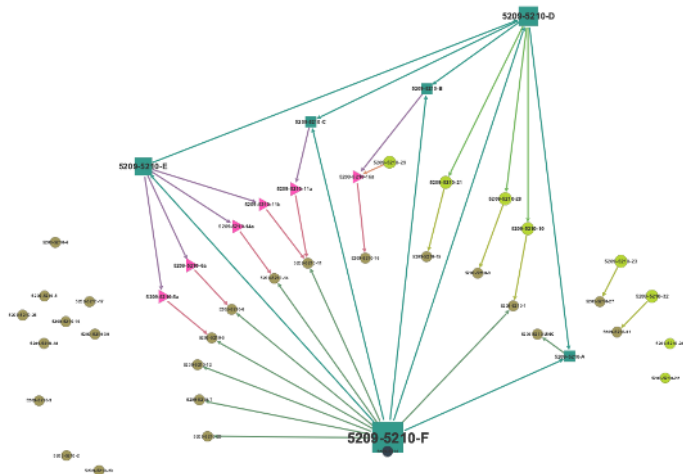
↑ Ms. Boll. 137.8

← AASS 09 Sept. tomus 01 (1746)

MEDIATION

■ editorial ecologies

(Vanderputten/Breitbarth in prep.):



SUMMARY

- **Epistemological** as well as **methodological** problems:
 - 1 What can we know about the past, based on available evidence?
 - 2 How can we assess the quality of the evidence?
 - 3 Accepting historical data are bad (if available at all!), what can we know about the past even in the absence of available (direct) evidence?

MAKING HISTORY (OR MYTHS)

MAKING HISTORY

*"The claim that the past only exists insofar as it is mediated by a historiographic act seems on the face of it perverse. After all, the monuments of the past surround us, the evidence for it is apparently collected in buildings, archaeological sites, documents, traditions, works of art, fossils, the configurations of the earth and the universe. Let us call these things collectively "witnesses". But now observe that **the witnesses themselves are neither "history" nor "the past"**. It is only human intellectual acts that gather these disjecta membra into coherent sequences and produce rational accounts of them. By themselves, **in the absence of theory, they are obtuse and unenlightening**, not parts of anything in particular. It is **only when we interpret them**, on the basis of **theories** of what they might mean, and supply, also through theory, the more important missing ones, that **a rational past comes into existence**. I am claiming that historiography is essentially **CONSTITUTIVE**, and that "history" or "the past" is what it constructs."*

(Lass 1980: 49)

MAKING HISTORY

- Labov (1974/2013): “Only when we are thoroughly at home in that everyday world [=the present], can we expect to be at home in the past.”
- Lass (1980: 53): “[T]he foundation of a theory of history [...] comes from the present [...].”
- Lass (1980: 54): “For historical argumentation to have any epistemic force it must in fact be essentially synchronic, for both psychological and logical reasons.”

HOW?

NO MIRACLES

- Lass (1997: 25): “There are no miracles.”

Uniformitarian Principle

“Knowledge of processes that operated in the past can be inferred by observing ongoing processes in the present.”

(Labov 1994, 21, citing Christy 1983, ix)

- Important **methodological precept** (Walkden 2019) – no one would claim the past was *exactly* like the present, but similar principles governed the expression of variables, and preferences for directions of shifts (e.g. organisation of inventories, hierarchical structure, directions of inference...)
- Restriction of hypothetically possible states, hence stronger theory (\pm Popper)

MAKING HISTORY

- “Facts” often need to be inferred, but can be based on things we know about the world (with a margin of error)



(Smilodon; Zoological Museum Copenhagen)



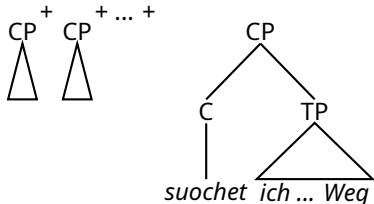
(Giant Panda skull; @Akkadian1979 on Flickr)

USING UNIFORMITY

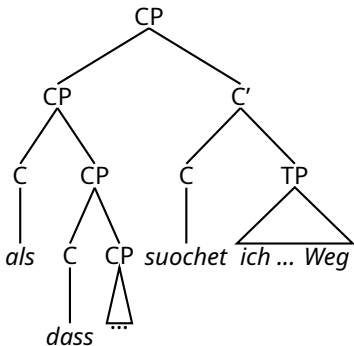
- (2) Als Jch gar nahe von jugendt auff sondere begürde gehabt [_]/ in ferre Landschafften zuo ziehen/ fürnemlich aber inn die Morgenlaender/ alß die für andere mehr beruembt vnd fruchtbar seind/ welliche die aelteste Voelcker erbawet [_]/ auch die mechtigisten Potentaten vnd Monarchen der Welt/ vor Jaren jnnengehabt/ vnnnd bewohnet [_]/ nicht allein derselbigen Jnnwoner leben/ sitten vnd gebraech/ warzune=men/ sonder auch vnd vilmehr die schoene Gewaechs v=n Kreüter/ vom Theophrasto, Dioscoride, Auicenna, Serapione 'c. beschriben/ an den enden vnnnd orten/ da sie wachßen/ zuerkundigen/ vnnnd zuerkennen/ damit mir jre beschreibungen inn ansehung vnd betrachtung deren/ sonderlich aber der mehr frembden vnd vnbekandten/ zum thail bekandtlicher vnd verstendtlicher wurden/ zum thail auch den Apoteckern vrsach gebe/ ferrner nach denen/ die jhnen zu haben von noeten [_]/ zu trachten/ **Suochet ich diß mein fürne=men ins werck zusetzen/ mittel vnd weg/ ...** (Rauwolf, *Morgenlaender* 1582)

USING UNIFORMITY

(a) Probably not:



(b) More likely:



- Assumption: Syntactic structure had the same hierarchical organisation as now.
- Again: this is a matter of probability.

MAKING SENSE

LINKING HYPOTHESES

It is only when we interpret them, on the basis of theories of what they might mean, and supply, also through theory, the more important missing ones, that a rational past comes into existence. (Lass 1997: 49)

Clearly, our inferences about the past are only as good as our reconstructions, which are necessarily hypotheses. (Epps 2015: 379)

LINKING HYPOTHESES

- Data are noise without a theory to interpret them
- **Linking hypotheses:** link linguistic theory to the observed patterns in the data (bad as they may be). Hypotheses about linguistic structure should make clear, testable predictions.
- **Testable:** clear quantitative patterns in usage data
 - Example: VO vs. OV+extraposition+V(P)R in ENHG (Bies 1996)
 - **Expected if OV:** No extraposition of weak pronouns or separable verbal particles

A linguistic variable must be defined under strict conditions if it is to be a part of a linguistic structure; otherwise, one would simply be opening the door wide to rules in which "frequently," "occasionally," or "sometimes" apply. Quantitative evidence for covariation between the variable in question and some other linguistic or extralinguistic element provides a necessary condition for admitting such a structural unit. (WLH 1968:169)

IMPROVING DATA

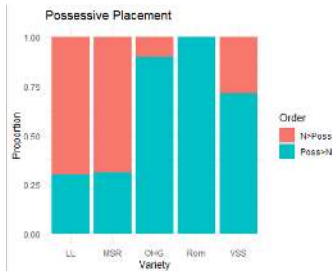
- How to mitigate the bad data problem, besides using the present?
 - ↳ “Use all the data!” (Lauersdorf 2019; 2021)
- Example: hagiographic texts from Lotharingia from the Long Tenth Century (c. 880–1030): very fluid genre; lots of variation (why?), very few other (surviving) sources
- By then, Latin was no longer acquired as a native language, but kept being used in the context of religious (and political) institutions by speakers of various vernacular languages
 - What can we know about Medieval Latin from that time?
Based on what data?
 - Given the scarcity of vernacular data from that period, can we indirectly also learn something about the vernacular languages through those sources?
- As bad as it gets...

USING ALL THE DATA

- Hagiographic texts = very local cults, still widespread genre; extremely fluid texts (editing / rewriting / expanding) – large amount also of syntactic variation (Vanderputten/Breitbarth 2026)
- **Questions:**
 - Is that variation superficial / stable, i.e., is ML still a unified language, or are there diachronic and diatopic developments?
 - What are the determinants of this variation?
- Work in progress (PhD F. Digiario): syntactic differences, still unclear why – influence of vernacular area (and which influence?) / institutional practices?

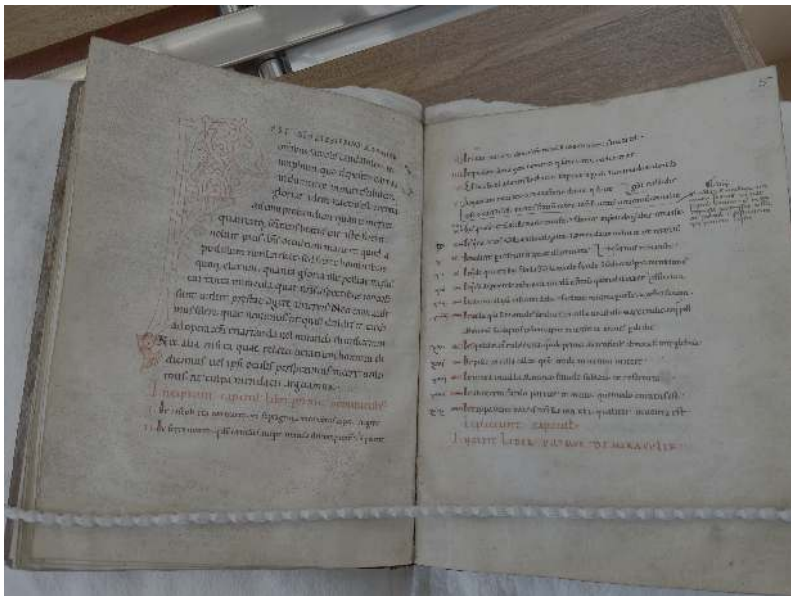
USING ALL THE DATA

BHL	10th c. mss.
7121	ad persolendum diuinum officium
7122	seruabatur honore
7127	fuissemus reuersi
7137	erant infigendae precio sui meriti
BHL	lost 11th c. ms.
7121	ad diuinum officium persolendum
7122	honore servabatur
7127	reuersi fuissemus
7137	infigendae erant sui meriti precio



- Ideally, look at the entire material transmission ($\pm 1,500$ manuscript witnesses of ± 200 texts between 10th and 18th c.) ...

USING ALL THE DATA



c MIRACULA

*Ex Ms. Colonienſi collato cum
aliis quinque Mss. & edi-
tione Mabillonii.*

LIBER I.

auctore eodem anonymo
Stabulensi

PROLOGUS.

*Auctor non
ſcribit niſi
viſa, vel à
ſide dignis
audita.*
d

POſt glorioliſſimum & agonis omnitus ſacu-
lis laudabilem triumphum, quo, deſiſto
carnali indumento, immarceſſibilem gloriæ idem
nactus eſt coronam, ad comprobandum quanti
meriti, quantæque ſanctitatis beatus Vir iſte

fuerit; noluit pius Dominus occultum manere, **D**
quod apud illum non latebat, ſed fieret * ho- *** al. fieri**
minibus quoque clarum, quantâ gloriâ ille pol-
leat in cœlis, cui tanta miracula quæ noſtris
aſpectibus conceſſa ſunt videre, præſtat egere
in terris. Non enim audent ſilere quæ novi-
mus, ne quaſi deſides & tardi ad opera Chriſti
enarranda vel mirandâ dijudicemur. Nec * alia, *** in aliis**
niſi ea quæ reſta veracium hominum didici- **Verum neq**
mus, vel ipſi oculis perſpeximus, dicere volu-
mus, ne culpa mendacii arguamur **b.**

ANNOTATA.

a *In aliis omnibus legitur, Poſt expletum
namque (vel itaque) glorioliſſimi agonis &c.*

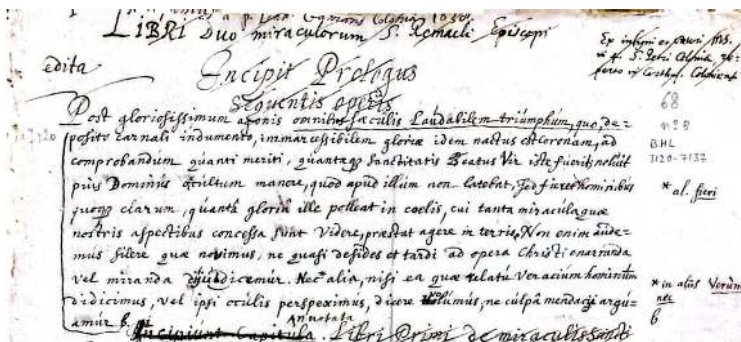
b *Item additur, Unum ergo miraculorum
illius in exemplum ponimus (Mabill, referamus,
) ut poſt facilius reliqua credibilia eſſe
comprobemus. Et mox: Eo namque tempore
&c.*

ANTIQUÆ DIVISIONIS CAPITA,
QUALIA SUNT IN MS.

E

(Acta Sanctorum 09 Septembris Tomus 01, 1746)

USING ALL THE DATA

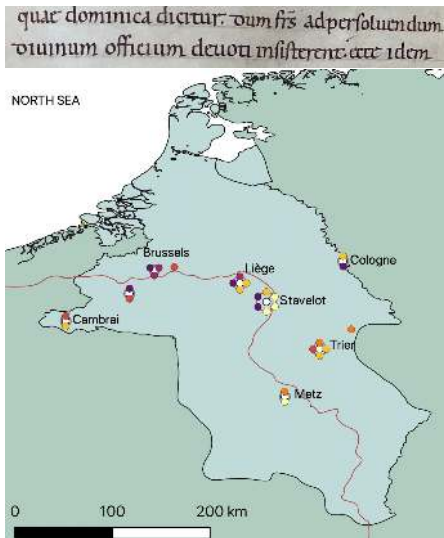


(Manuscript Bollandien 137.08)

USING ALL THE DATA

- Use all the (meta)data:
 - institution, diocese
 - author (if known),
destinatory ...
 - material metadata
- Use additional sources:
 - historical multilingualism
 - language acquisition and
~ teaching
 - any information on
contemporary vernacular
languages

⇒ innovative vs. more
classical patterns;
±Romance vernacular
influence (± vertical
continuum?)



SUMMING UP

- The “past” as such needs to be **created** by interpreting witnesses, guided by linking hypotheses
 - 1 Using Uniformitarianism ← linking hypotheses based on better data / the present
 - 2 Using all the data (not just linguistic witnesses)
- Apply care: blind spots are **expected!**

- Three case studies:
research questions, linking hypotheses, data, hypothesis testing
 - 1 How to use the past to understand the present?
 - 2 How to use the present to understand the past?
 - 3 How to use space to access time?