Ancestral State Reconstruction

Hedvig Skirgård

Department of Linguistic & Cultural Evolution
Max Planck Institute of Evolutionary Anthropology

Me

- Born and raised in Uppsala
- Bachelor and masters degrees in linguistics from Stockholm University
- Worked as a research assistant for Dr Hammarström in Nijmegen for 1.5 years
- PhD from Australian National University
- Now
  - Postdoctoral researcher at the Department of Linguistic and Cultural Evolution at the Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.
  - Coding-coordinator for Grambank
- Research focus: evolutionary linguistics, Pacific languages & impact of social dynamics on language

Talk overview

- Part one
  - the input data
- Part two
  - the ASR

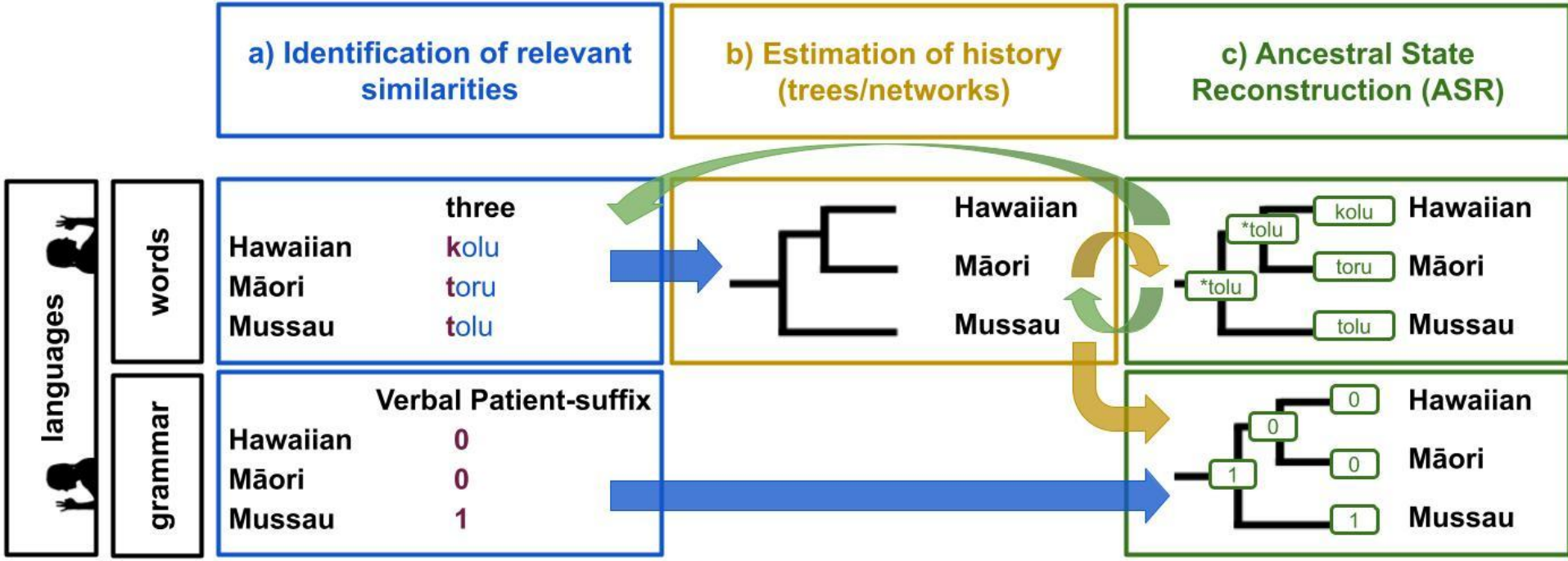
- case study: how much do computational approaches agree with traditional historical linguistics when estimating grammar of Oceanic proto-languages?

- excluded: construction of trees/identification of subgroups

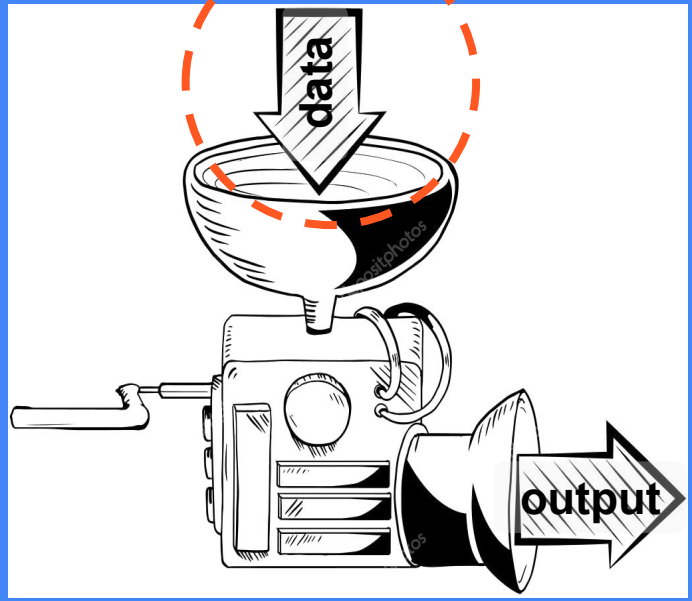
- coming at ASR from linguistic typology and evolutionary biology, generally with traditional historical linguists in mind

Part one

Part two



Part one: the input



What kind of data can we reconstruct historically?

- words
- sounds
- grammar
  - grammatical words/morphemes
  - paradigm organisation
  - word-order and other abstract features
    - derived from reading grammars and filling in questionnaires (grambank, WALS, Jazyki Mira etc)
    - derived from cross-linguistic corpora

Let's start with the usual

- ➔ words
- ➔ sounds

The core material

Three								
Pai		*tjelu*				l		
Pasih		*turu*				r		u
Puyuma		*tero*	t		e	r		o
Batak		*tulu*	t		u	l		u
Mentawai		*telu*	t		e	l		u
Roti		*telu*	t		e	l		u
Sa	*butua*							

cognates

double-cognacy

Walkden (2013)

sound correspondences

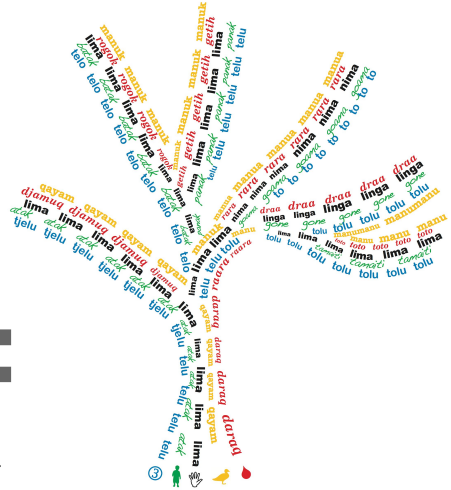
Paiwan	*tjelu*
Pasih	*turu*
Puyuma	*tero*
Batak	*tulu*
Mentawai	*telu*
Roti	*telu*
Sa	*butua*

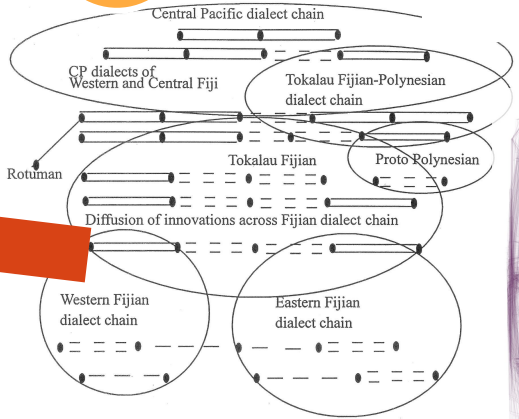
fewest changes

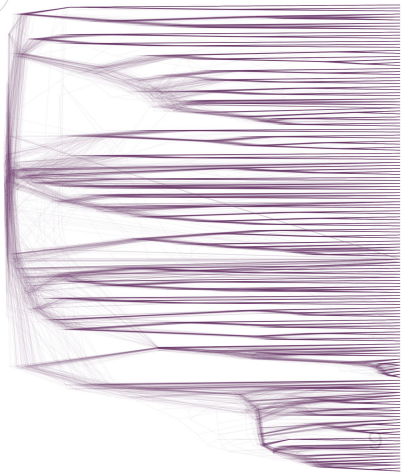
+

knowledge about region, archaeology, cultural history etc

=







ancestral states

working subgroupings

phonological changes

semantic changes

Extending to the unusual material

➔ grammar

Grammar as source data for analysis

if we consider paradigm structure and other abstract traits

- ➔ cognate loss and gain ≠ loss and gain of grammatical features
- ➔ similarity ≠ inheritance
  - ◆ dependencies (c.f. regular sound change?)
  - ◆ design space size (e.g. 2^201 vs 15^100)
  - ◆ different evolutionary constraints
    - neuro-linguistic processing
    - communicative efficiency (redundancy/robustness vs economy)
    - information uniformity (c.f. Wallenberg)
    - "pragmatic bottleneck" - multimodal, common ground, inferences etc (c.f. Levinson 2024
    - complexity/compositionality may vary with social dynamics
- ➔ unclear how to deal with most of these, generally and in ASR specifically

Fitness of grammatical data for reconstruction

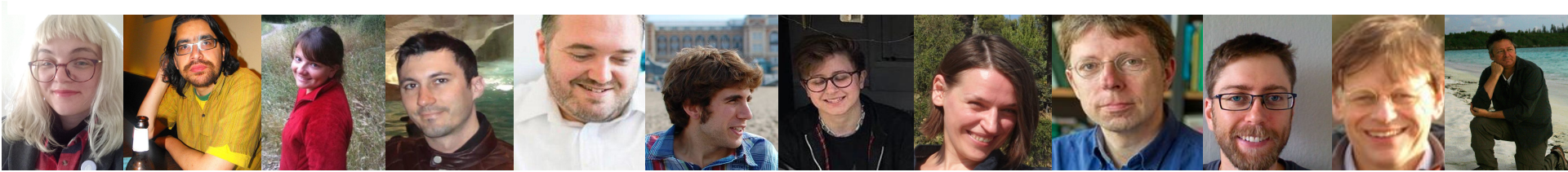
This study specifically: Grambank

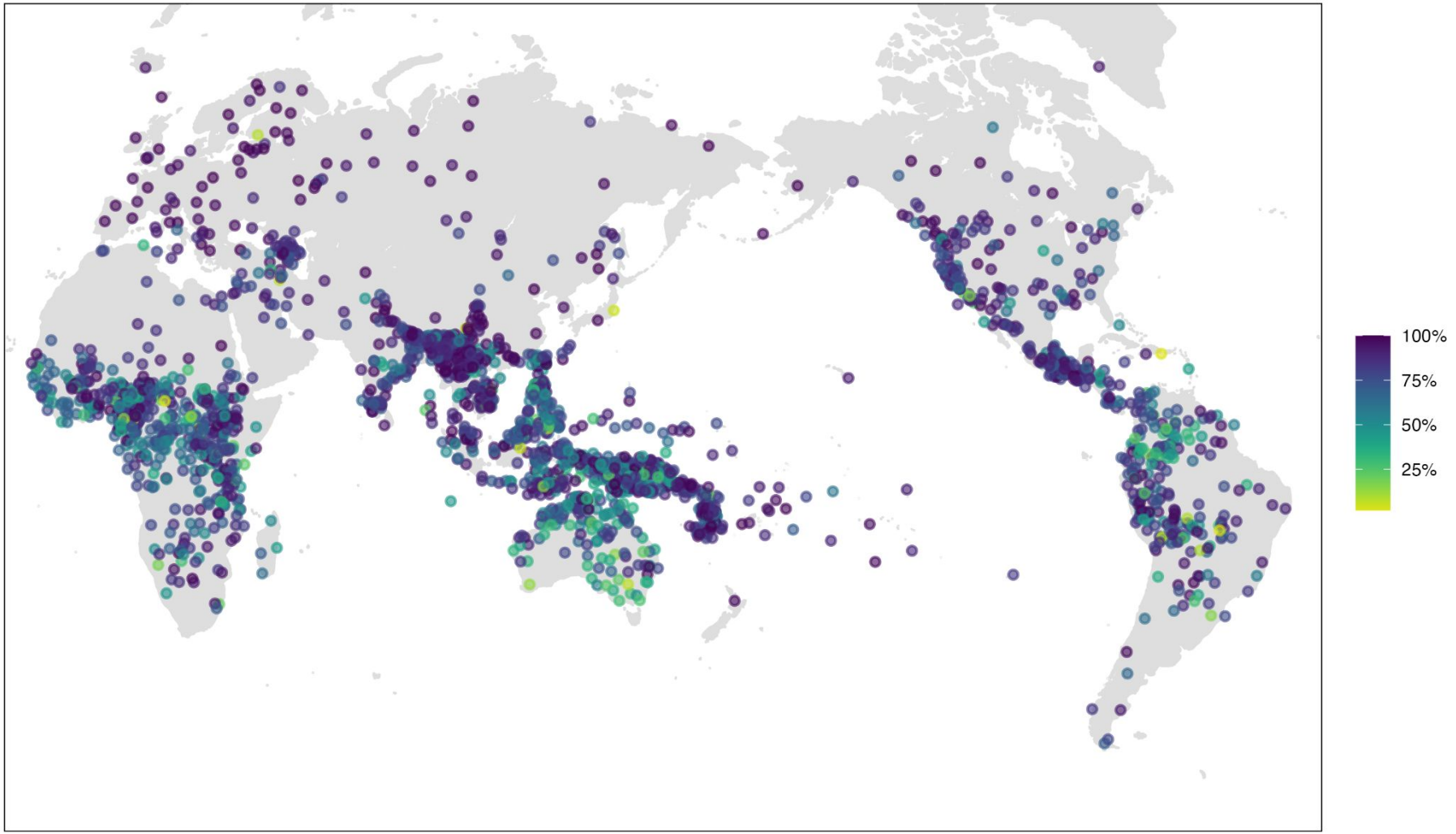
- global typological questionnaire of grammatical features
- tracks abstract features, not specific forms
- currently at over 2,000 languages in the database
- 280 Oceanic languages included

- makes possible research on cognitive constraints, contact, deep history, dynamics of evolution etc.

Grambank overview

- ★ Glottobank consortium
  - Funded and run from the Department of Linguistic and Cultural Evolution at the Max Planck Institute for Evolutionary Anthropology in Leipzig
- ★ 195 features
  - GB020 Are there definite or specific articles?
  - GB111 Are there conjugation classes?
  - GB159 Are nouns reduplicated?
- ★ based on NTS, Sahul, Pioneers and WALS-questionnaires
- ★ for more on data gathering, feature description etc see wiki

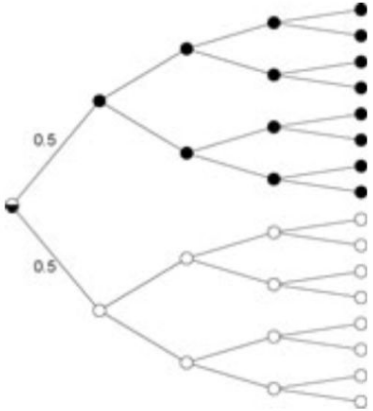
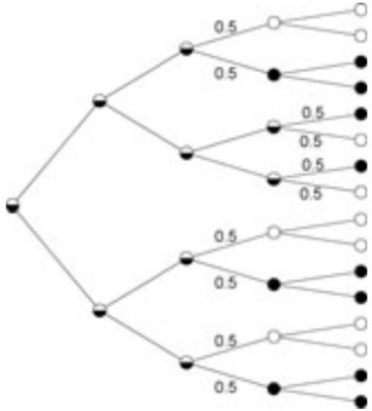
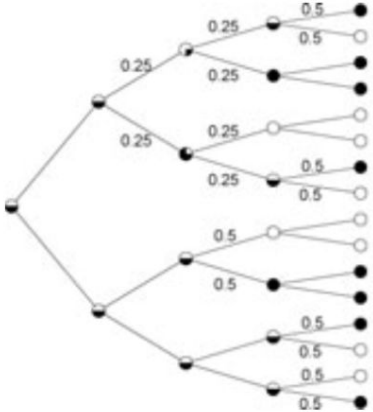
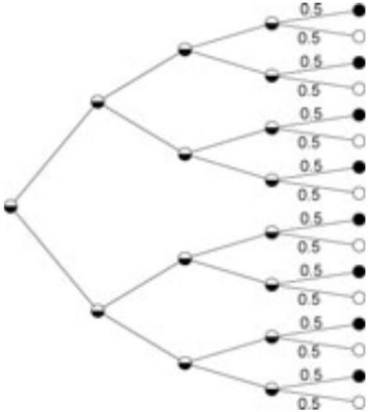



100%
75%
50%
25%

Fitness of Grambank features for ASR

- ➔ we cannot, at least not easily, investigate the double cognacy of grambank features à la Walkden (2013)
- ➔ phylogenetic signal however can be tested!

Fritz & Purvis' D-estimate

	Phylogenetic pattern[a]			
	extremely clumped	*clumped (Brownian)*	*random*	*overdispersed*
Phylogeny[b]				
Σd_{obs}[c]	1	5	6.5	8
D[d]	−2.4	0.0	1.0	1.9

Fritz & Purvis' D-estimate

For the features that historical linguists have made predictions about, over 3 trees/sets of trees.

tree	D-estimate (mean)	Proportion of features not significantly dissimilar to 0	features unfit for D-estimate	Too few tips altogether
Glottolog	0.34	47%	8	0
Gray - MCCT	0.28	58%	17	1
Gray - posteriors	-0.01	81%	22	1

Another crucial input: the trees

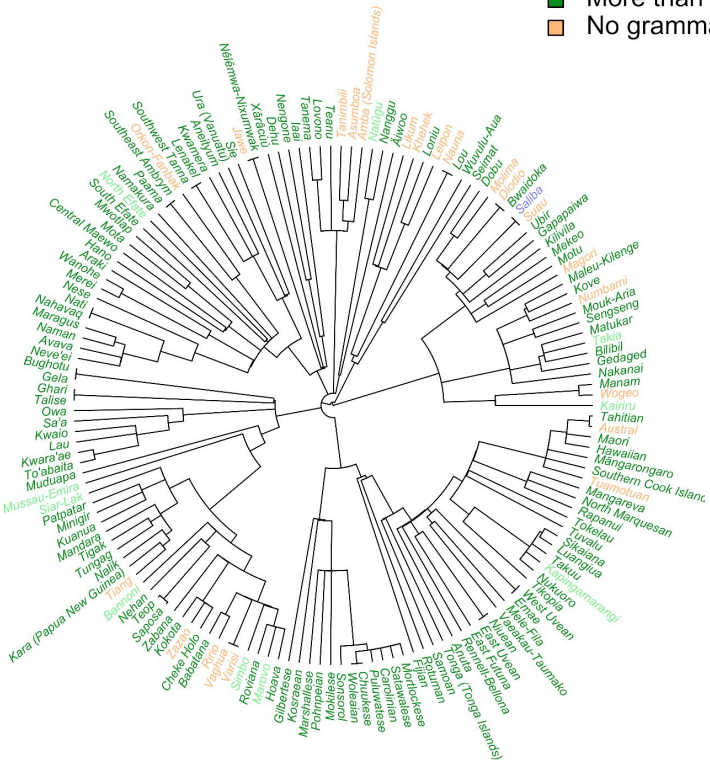
- ➔ in classical historical linguistics, the tree and the ASR are co-estimated - they're done in tandem. Often they start with broad widely accepted subgroups

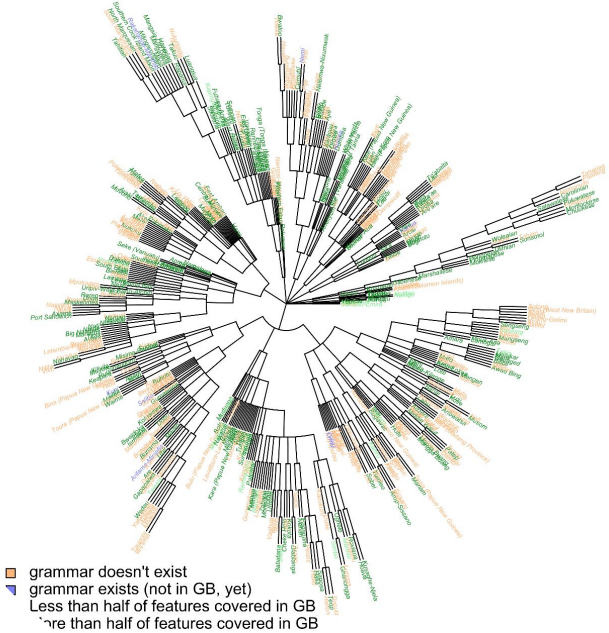
Design: Jeremy Slagle

- ➔ for many computational approaches to ASR, a particular tree or set or trees are used and the reconstructions don't affect them
  - ◆ I used 3 different trees: Glottolog, MCCT of Gray et al 2009 MCCT and random posterior of ditto
- ➔ sometimes people don't even use trees based on the same kind of data that they want to reconstruct, e.g. using lexical trees for cultural traits.

The trees

- Grammar exists, but not in GB (yet)
- Less than half of features covered in GB
- More than half of features covered in GB
- No grammar





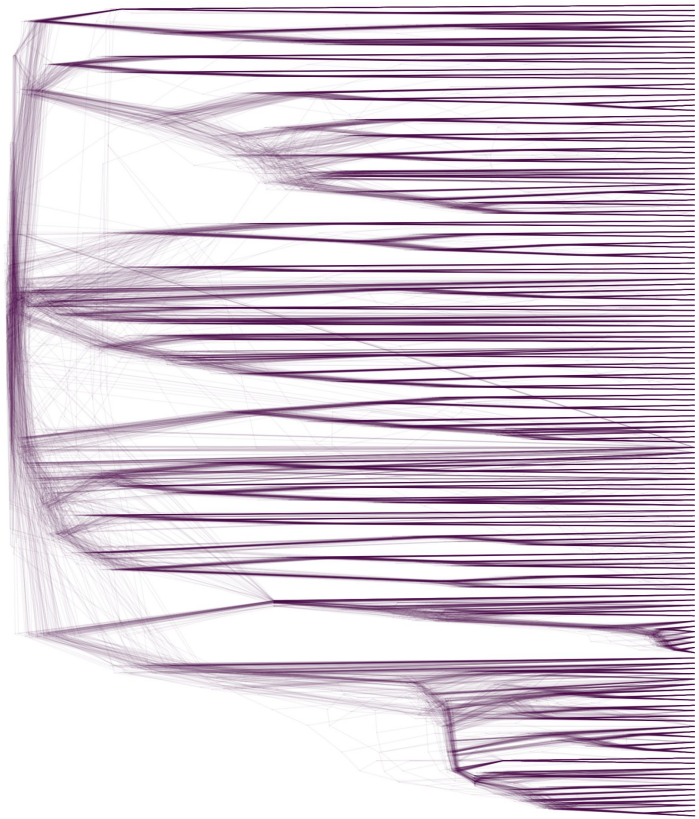
Gray et al (2009)

Glottolog 4.0

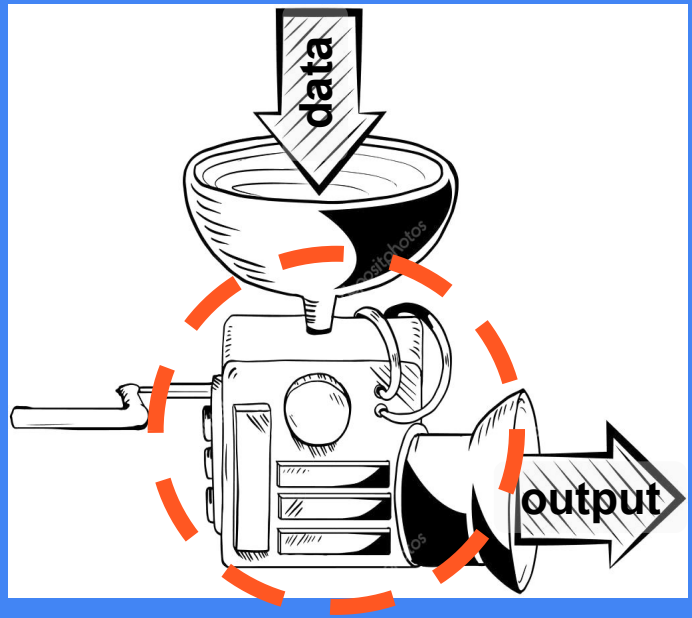
sampling a Bayesian posterior

- random sample of 100 trees in the posterior of Gray et al (2008) which contains 4,200 trees

- could perhaps work as a way of factoring in contact

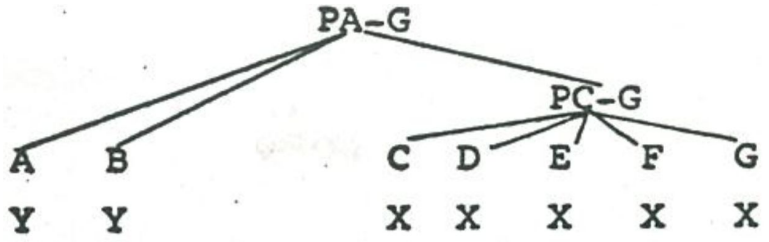


Part 2: the ASR



Classical HL reconstruction

- reconstruction is based on fewest changes possible in the tree (Max Parsimony)
- but also (in particular for reconstructions of structural traits):
  - plausibility of changes
  - plausibility of reconstructed language as a whole



Clark 1973

- What is plausible is something people often disagree on

Predictions from classical historical linguistics

- it is less common to study grammar compared to vocabulary or phonology
- at least 11 scholars have published reconstructions of grammar in Oceanic languages
- 115 data points for the four relevant proto-languages in the Oceanic subgroup

- disagreement on alignment of Proto-Polynesian & Proto-Central Pacific

	A	B	C	E	F	G
1	Feature_	Feature	Value	Source	Comment	
2	GB028	Is there a distinction between inclusive and exclusive?	1	Pawley (1973:112);		Crowley (19
3	GB023	Are there postnominal articles?	0	Pawley (1973:112);		Ross (2004
4	GB431	Can adnominal possession be marked by a prefix on the possessed noun?	0	Pawley (1973:117);		Ross (2004
5	GB105	Can the recipient in a ditransitive construction be marked like the monotransitive	0	Pawley (1973:118)		
6	GB133	Is a pragmatically unmarked constituent order verb-final for transitive clauses?	1	Pawley (1973:118)		
7	GB131	Is a pragmatically unmarked constituent order verb-initial for transitive clauses?	1	Pawley (1973:118);		Lynch, Ross
8	GB079	Do verbs have prefixes/proclitics, other than those that only mark A, S or P (do	1	Pawley (1973:142);		Ross (2007
9	GB140	Is verbal predication marked by the same negator as all of the following types of	0	Pawley (1973:143-146);		Lynch,
10	GB058	Are there possessive classifiers?	1	Pawley (1973:154);		Ross (2004
11	GB065	Is the order of possessor noun and possessed noun possessed-possessor?	3	Pawley (1973:155-156);		Ross (2
12	GB408	Is there any accusative alignment of flagging?	1	Pawley (1973:167);		Ross (2004
13	GB074	Are there prepositions?	1	Pawley (1973:167);		Ross (2004
14	GB113	Are there verbal affixes or clitics that turn intransitive verbs into transitive ones?	1	Pawley (1973:171);		Wilson (198
15	GB115	Is there a phonologically bound reciprocal marker on the verb?	1	Pawley (1973:172);		Ross (2004
16	GB059	Is the adnominal possessive construction different for alienable and inalienable	1	Ross (2004:492, 511-512);		Lync

Example: Proto-Oceanic GB coding sheet

Computational methods of reconstruction: overview

- objective and principled
- lacks human knowledge of plausibility (blessing or curse?)
- generally requires a known tree (or set of trees)

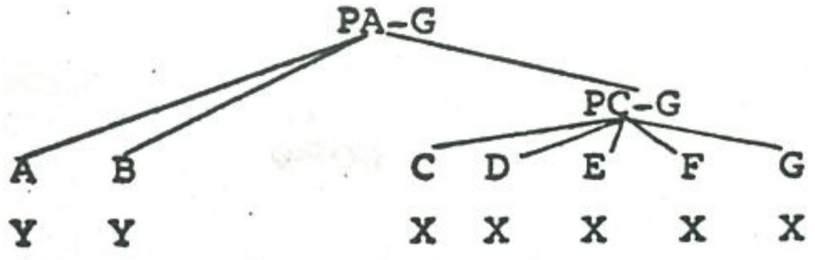
- Major methods:
  - Maximum Parsimony
  - Maximum Likelihood
  - Minimal Lateral Networks (MLN)
  - Stochastic Character Mapping (SCM)

This study

- three methods
  - Maximum Parsimony
  - Maximum Likelihood
  - Most Common (reality check)
- three trees
  - Gray et al (2009) - 2 versions
    - the Maximum Clade Credibility Tree (MCCT)
    - random sample of 100 from posterior
  - Glottolog 4.0
    - mainly based on Lynch, Ross and Crowley 2002

Maximum Parsimony (MP)

- lowest number of changes given a particular tree and particular feature distribution
- simple = good
- already core component of classic HL reconstruction
- branch lengths irrelevant - only splits are relevant
- is the solution with the fewest amount of changes really the best?



(Clark 1976)

Maximum Likelihood (ML)

- computes likelihood of all ancestral states given tree, branch lengths and feature distribution
- takes branch length into account
- fewest changes ≠ best solution

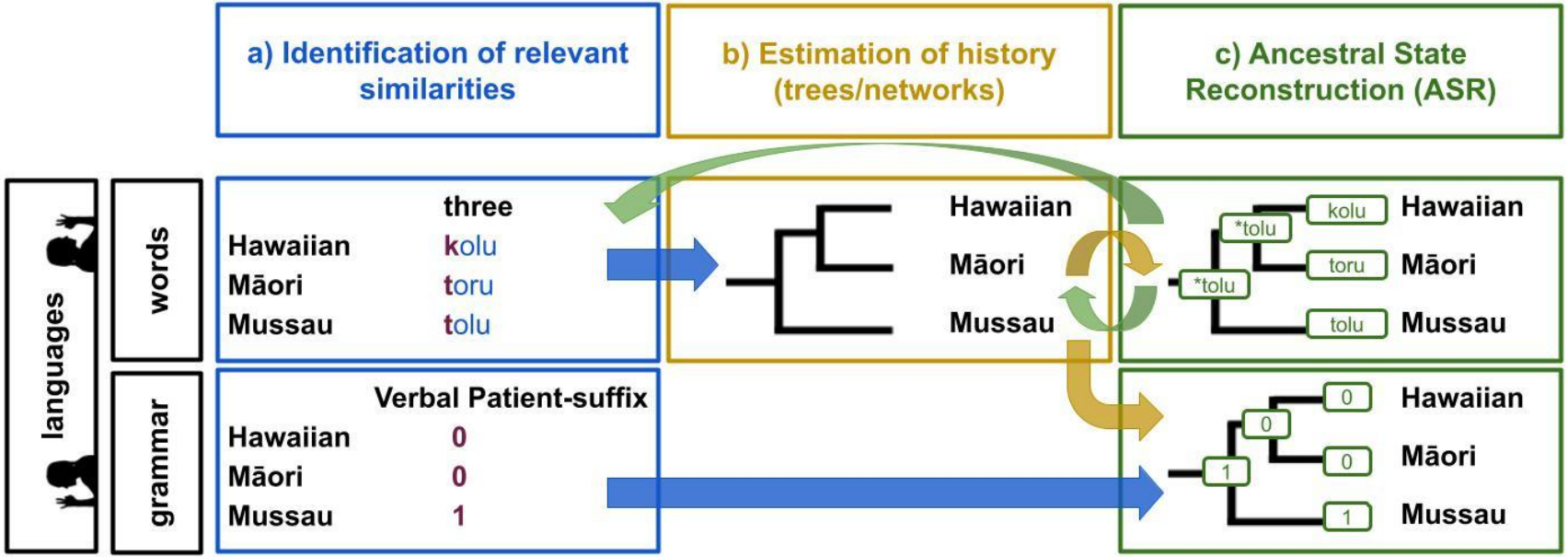
- many instances of sister pairs having different values ➔ high rate of change
  - has consequences for predictions in the entire tree

Most Common

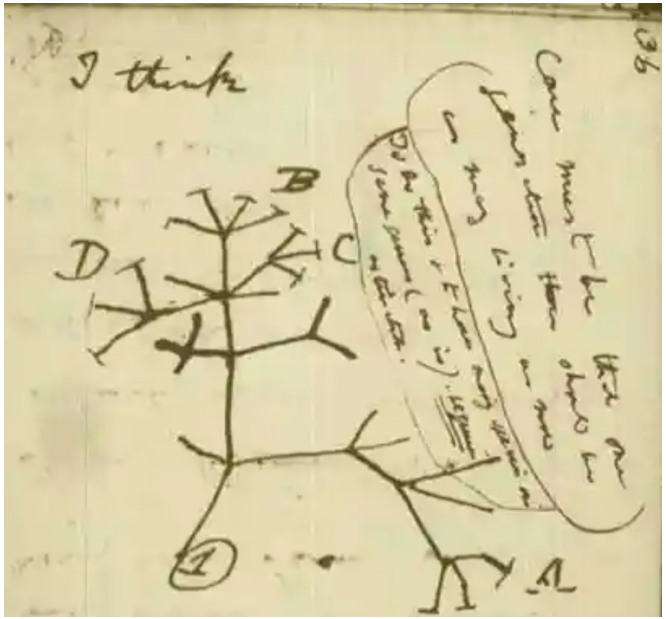
- a count of the most common state in all the daughter languages, regardless of tree structure
- similar to Maximum Parsimony but even simpler
- also known as "majority-rule frequency heuristic" (cf. Goldstein 2022)

Enter biology

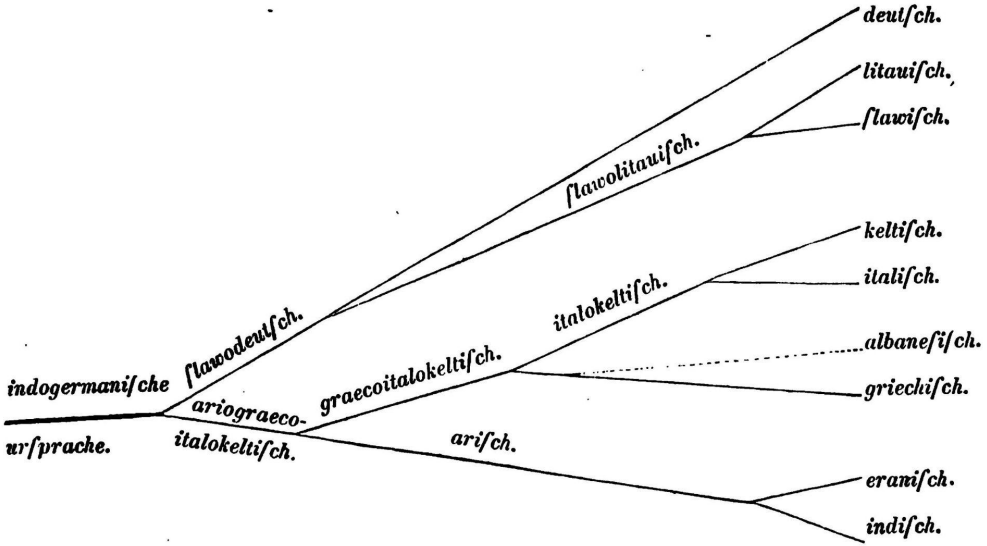
general observations


a) Identification of relevant similarities
b) Estimation of history (trees/networks)
c) Ancestral State Reconstruction (ASR)
languages
words
grammar
three
Hawaiian kolu
Māori toru
Mussau tolu
Verbal Patient-suffix
Hawaiian 0
Māori 0
Mussau 1
Hawaiian
Māori
Mussau
kolu
*tolu
toru
*tolu
tolu
Hawaiian
Māori
Mussau
0
0
0
1
1
Hawaiian
Māori
Mussau

both make trees (and linguists were first!)



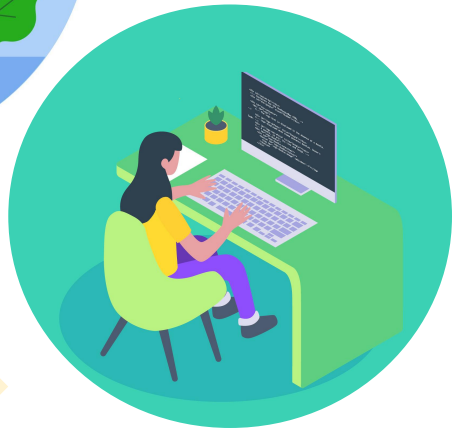
Darwin (circa 1837)



Schleicher (1861)

NB: Schlegel (1808)

ASR in linguistics vs biology

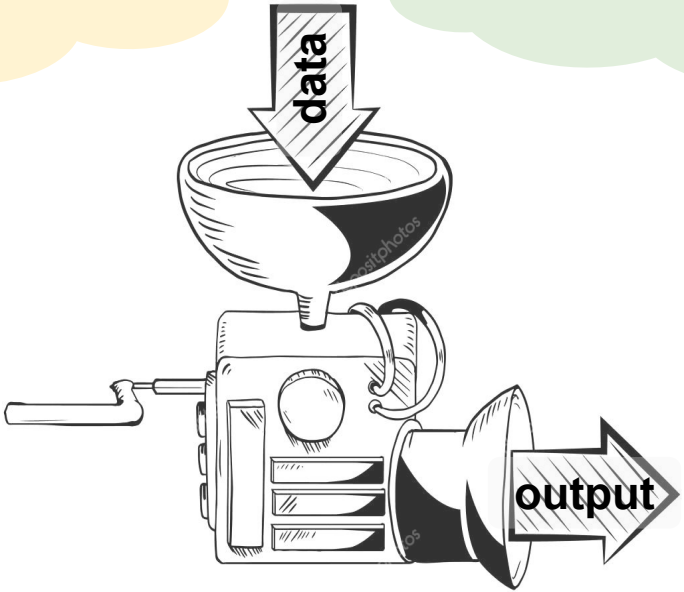


more human-based

more technology-based

What is going into the machine?

What is going on in the machine?

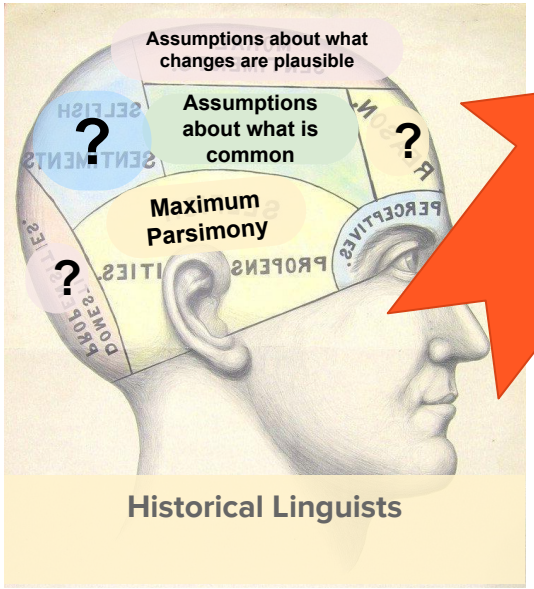


Linguistics

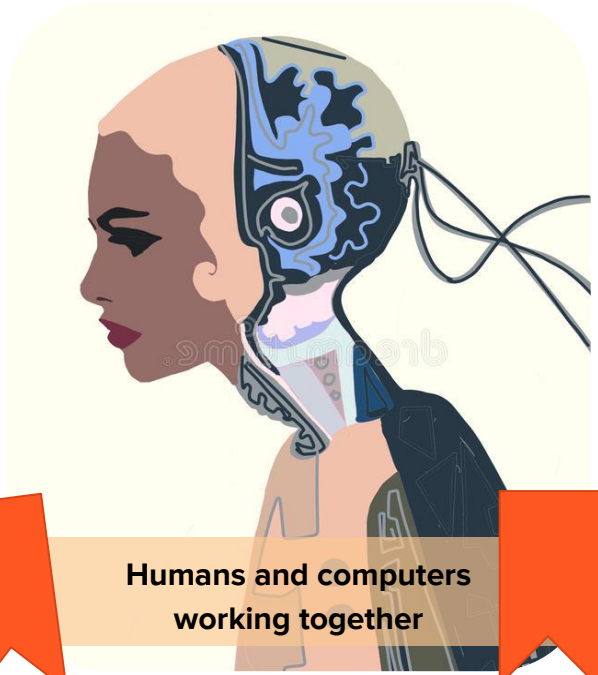
ASR

Biology

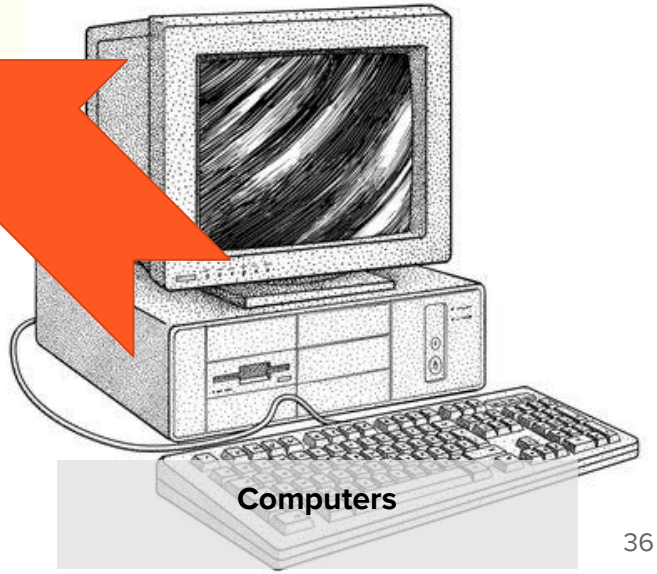
The Future



Historical Linguists



Humans and computers working together



Computers

Results of case study

Moai Statues, Rapa Nui. Dimtry Moiseenko, AirPano.com.

Concordance comparison

- 4 proto-languages
  - Proto-Oceanic, Proto-Central Pacific, Proto-Polynesian and Proto-Eastern Polynesian
- 115 data points in total to compare
- 3 contested data points (alignment)

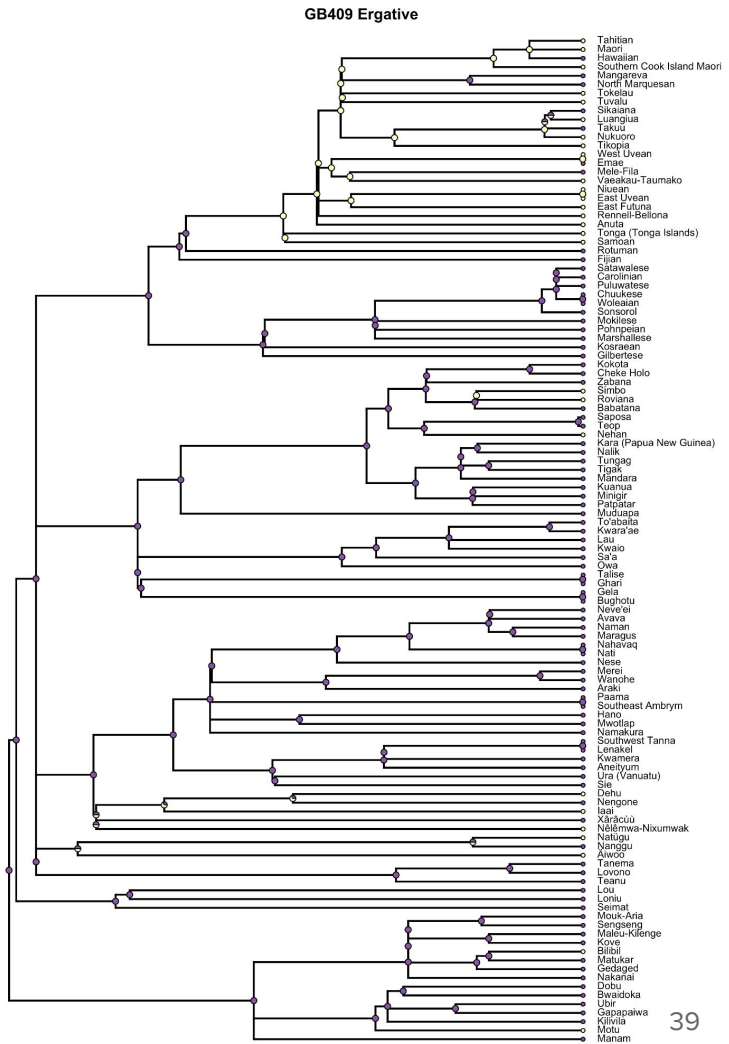
Finding in historical linguistics	Prediction by MP or ML	Result
Absence	>60% Absence	True Negative
Absence	>60% Presence	False Positive (type 1-error)
Presence	>60% Presence	True Positive
Presence	>60% Absence	False Negative (type 2-error)
Absence	40-60% Presence/Absence	Half/Half
Presence	40-60% Presence/Absence	Half/Half

Overview

- ancestral states are estimated for each ancestral language in every tree
  - for the 100 posteriors, the mean is taken

- concordance is estimated with a measure which is based on "accuracy" but awards some points for "half" states

$$\frac{\text{agree} + \frac{\text{half}}{2}}{\text{all reconstructions}}$$

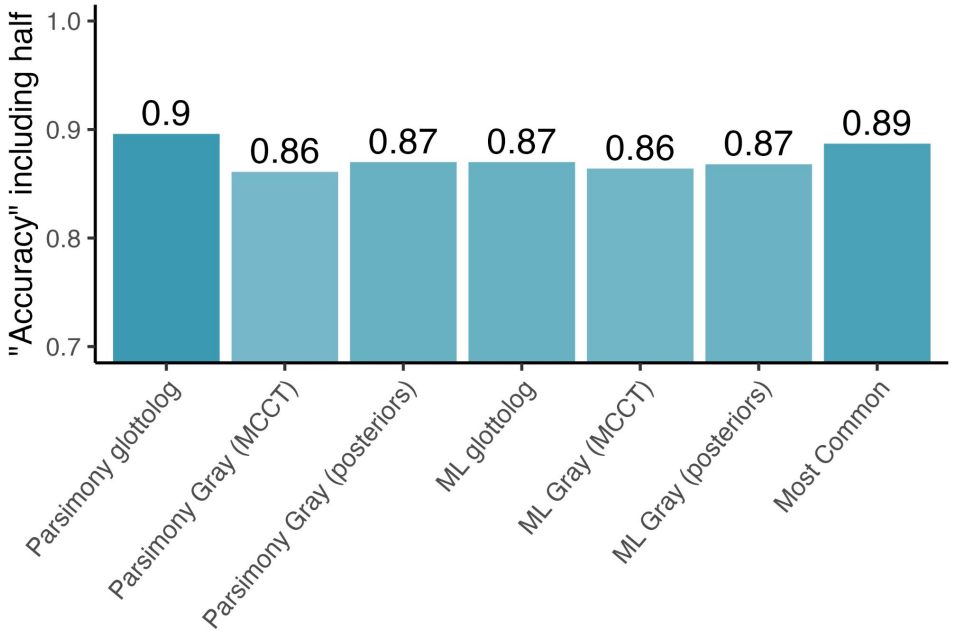


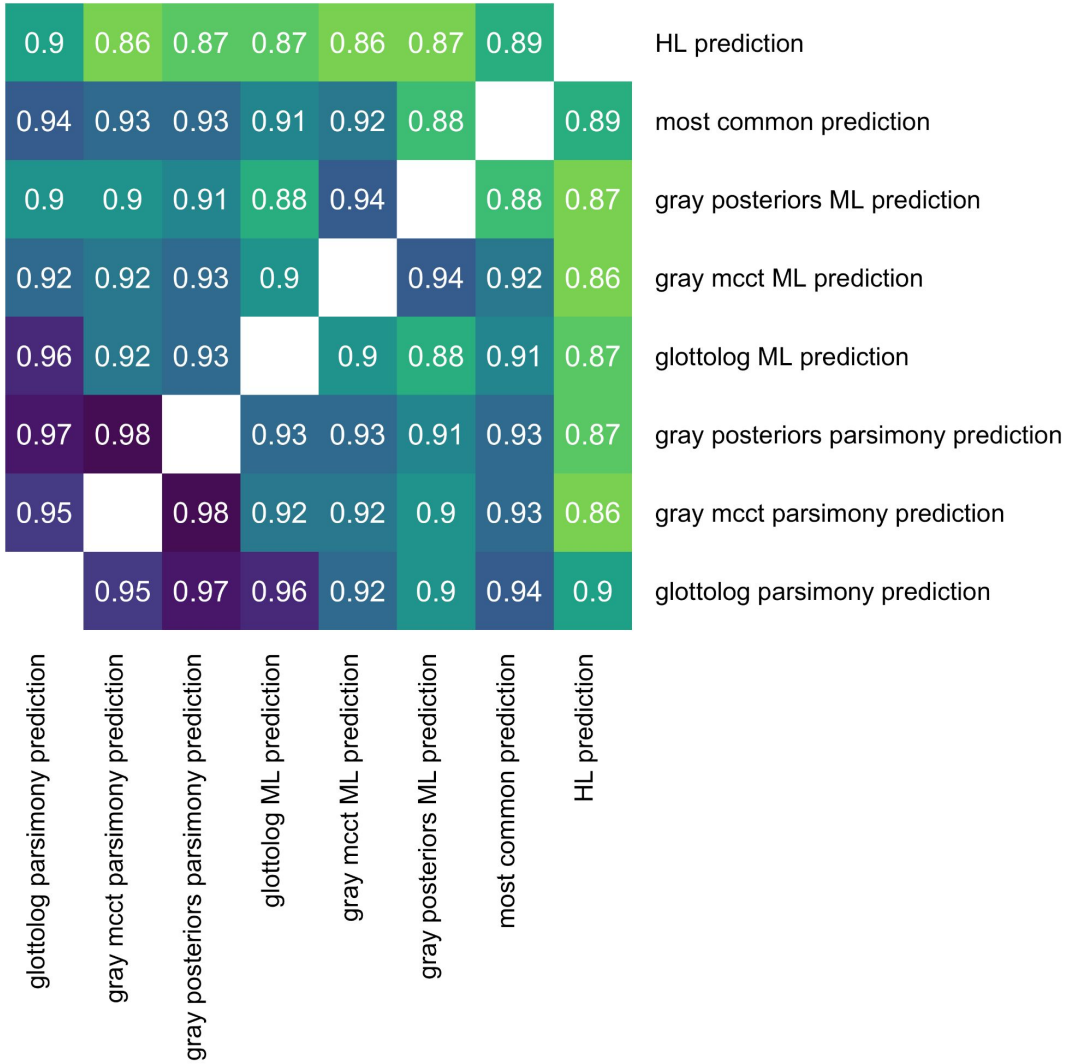
Overall counts

Method	False Negative	False Positive	Half	True Negative	True Positive	Total
ML Glottolog	10	3	4	46	52	115
ML Gray et al (2009) - MCCT	9	2	9	43	51	114
ML Gray et al (2009) - posteriors	10	1	8	44	51	114
Most common	5	0	16	46	48	115
Parsimony Glottolog	8	2	4	46	55	115
Parsimony Gray et al (2009) - MCCT	6	5	10	42	52	115
Parsimony Gray et al (2009) - posteriors	7	6	4	43	55	115

Results

- there are many ways to calculate performance
- displayed here is accuracy (incl half)
- all methods score very similar
- MP with Glottolog and Most Common score highest

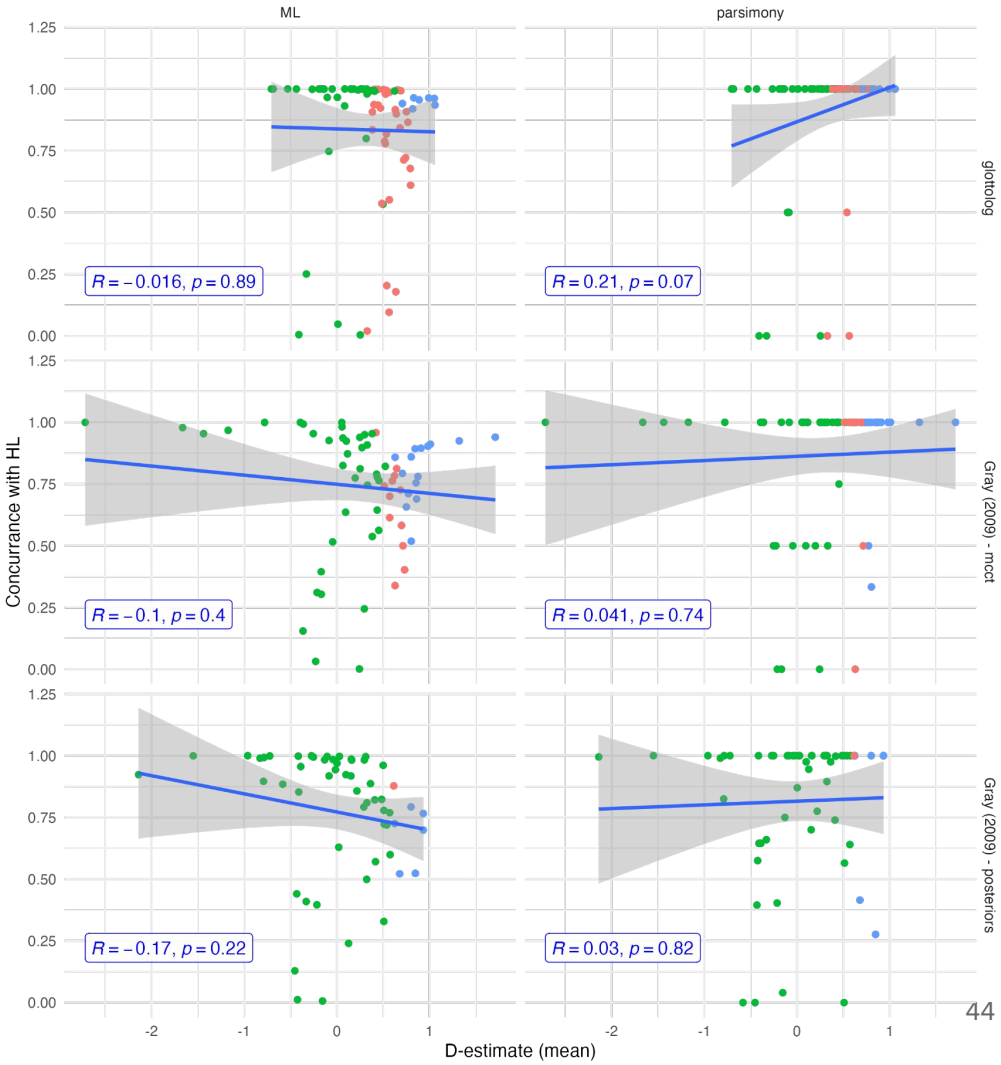



0.9 0.86 0.87 0.87 0.86 0.87 0.89
HL prediction
0.94 0.93 0.93 0.91 0.92 0.88 0.89
most common prediction
0.9 0.9 0.91 0.88 0.94 0.88 0.87
gray posteriors ML prediction
0.92 0.92 0.93 0.9 0.94 0.92 0.86
gray mcct ML prediction
0.96 0.92 0.93 0.9 0.88 0.91 0.87
glottolog ML prediction
0.97 0.98 0.93 0.93 0.91 0.93 0.87
gray posteriors parsimony prediction
0.95 0.98 0.92 0.92 0.9 0.93 0.86
gray mcct parsimony prediction
0.95 0.97 0.96 0.92 0.9 0.94 0.9
glottolog parsimony prediction
glottolog parsimony prediction
gray mcct parsimony prediction
gray posteriors parsimony prediction
glottolog ML prediction
gray mcct ML prediction
gray posteriors ML prediction
most common prediction
HL prediction

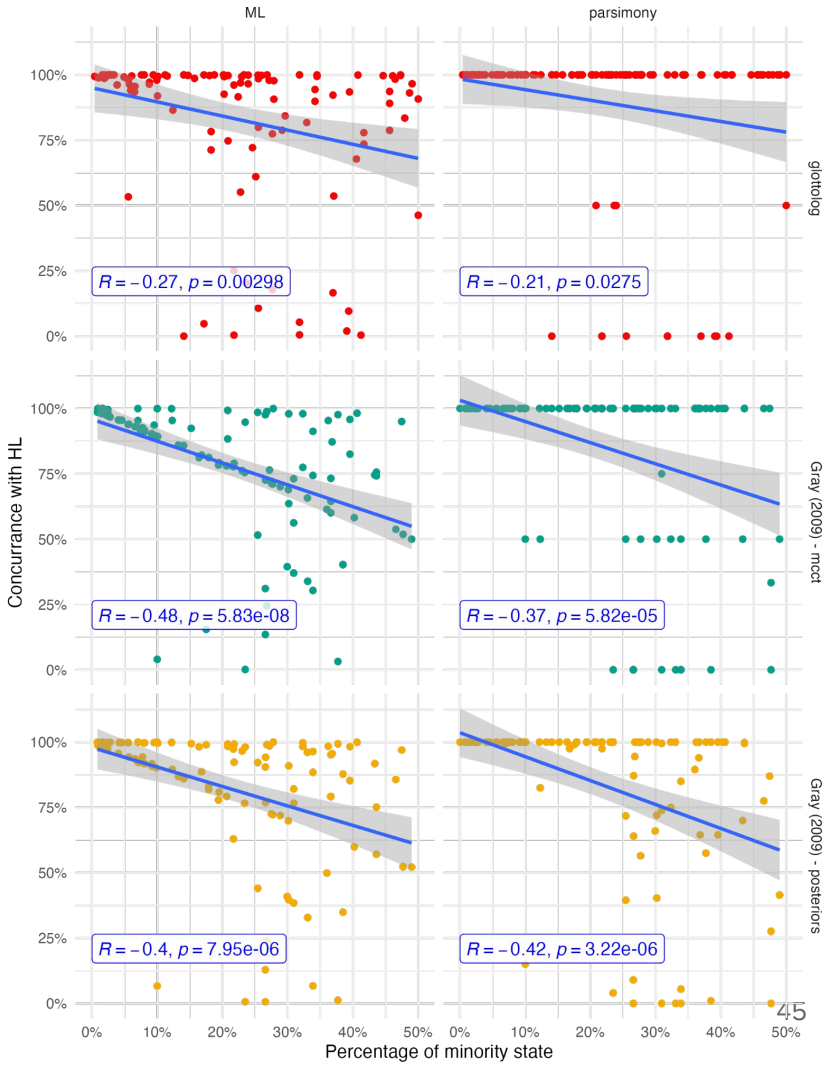
Some features reconstructed not in HL

- In total, the results include 654 predictions not in HL (afaik / yet)
- There are 111 features in the 4 Proto-languages that all MP and ML methods reconstruct as present, but which aren't predicted by historical linguists in the comparison
- some are:
  - Proto-Oceanic has inclusionary constructions and a difference between nominal conjunction and comitative ("and" vs "with")
  - Proto-Central Pacific has clusivity and dual number in pronouns
  - Proto-Polynesian has tense particles and numeral classifiers
  - Proto-Eastern Polynesian can have content interrogatives in situ and 3+ distance contrasts in demonstratives

D-estimate vs HL-concurrence



Prop vs HL-concurrence



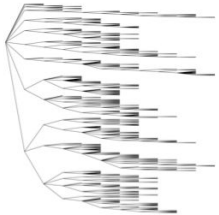
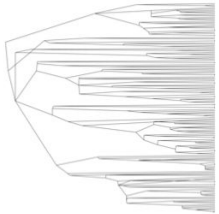
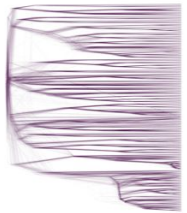
Which method is best?

- we should choose based on principles, not results
- (besides, they are mostly quite similar results-wise anyway)

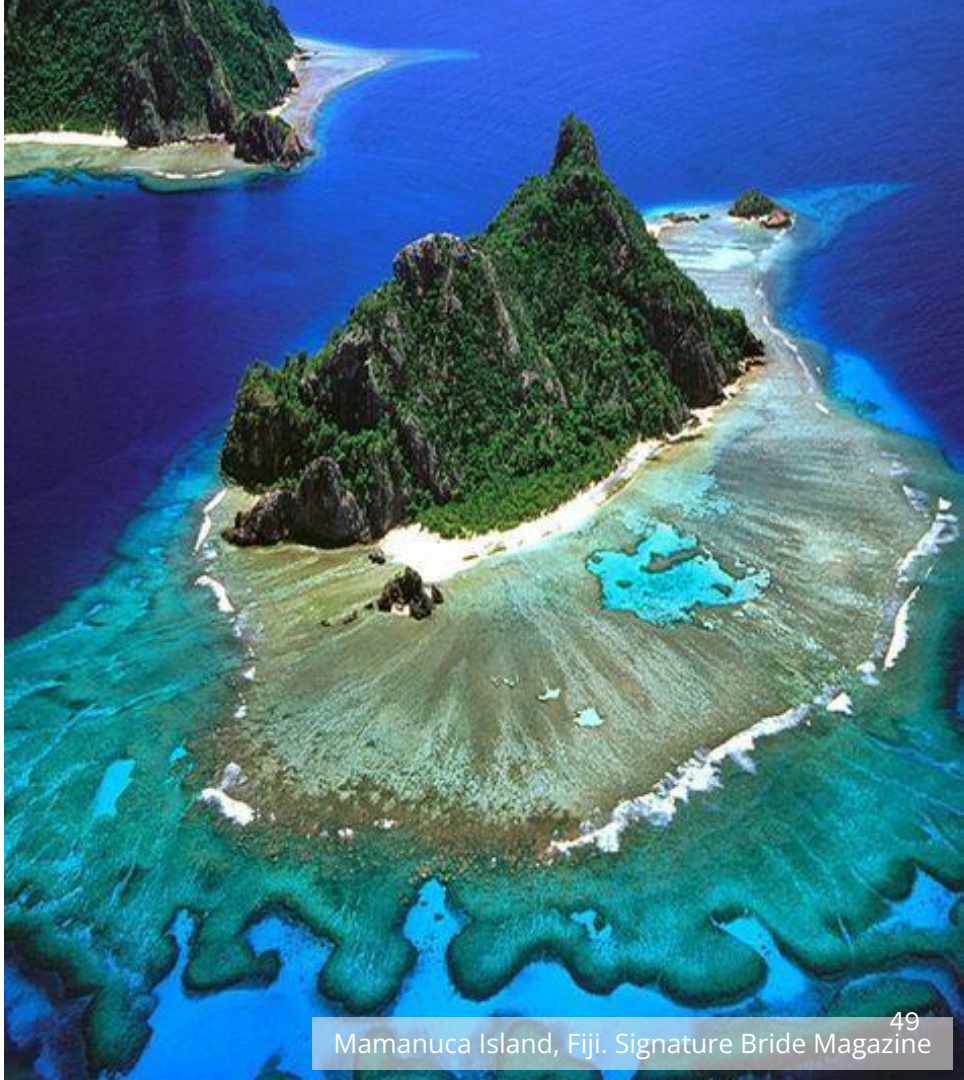
Table 8: Summary of conceptual pros and cons of the ASR-methods

ASR-Method	**Pros**	**Cons**
Conventional HL	widely used and attested; human-friendly ; takes into account complexities regarding item- and language-specific nuance and context	may ignore branch lengths; plausibility/rates of changes and plausibility of combined states are under-specified which leads to hard-to-resolve conflicts; possible: assumes slowest rate = most plausible rate
Maximum Parsimony	easy to understand; consistent; explicit	ignores branch lengths; assumes slowest rate = most plausible rate; does not allow asymmetric transition rates
Maximum Likelihood	consistent; explicit; takes into account branch lengths; dynamically estimates rates; can take further input such as priors on root state, rates etc	requires more knowledge of computational mathematics
Most Common	easy to understand	ignores the tree altogether; estimates no rates

Table 9: Summary of conceptual pros and cons of the trees.

Tree		Pros	Cons
Glottolog 4.5		includes all Oceanic languages	has no branch lengths; possibly inconsistent sub-grouping; many polytomies (10%); lowest proportion of D-estimates similar to 0
Gray et al. (2009) - MCCT		has branch lengths; is based on explicit lexical data; transparent methodology at each step; fewer polytomies (3%)	includes fewer languages
Gray et al. (2009) - random sample of 100 from posterior		has branch lengths; is based on explicit lexical data; transparent methodology at each step; much fewer polytomies (0.15%); encompasses more variation than MCCT; highest proportion of D-estimates similar to 0	includes fewer languages; takes longer time to calculate over

Conclusions



Mamanuca Island, Fiji. Signature Bride Magazine

Conclusions: case study

- ★ Several of the computational methods perform very similar to historical linguists
- ★ Historical Linguists may be mainly relying on Max Parsimony. It is conceivable that ML's way of using branch lengths estimates some HL plausibility knowledge
- ★ ML + posterior are conceptually best
- ★ The agreement lends support to computational methods, which can then make predictions that the comparative method haven't yet or struggle to make due to the amount of data involved.
- ★ D-estimates didn't correlate with HL-concurrence, but prop did. Maybe D-estimates don't measure the right thing?
- ★ might do this with IE too, if I can work up the courage

Conceptual thoughts

- ➔ evaluating the input data: phylogenetic signal wasn't decisive in determining how much the methods agree
  - ◆ does that also suggest it's an inadequate metric for determining if the data is appropriate to do ASR on?
  - ◆ or is it just a product of there being little variation to go by, most traditional historical linguists don't make risky predictions?

The open question

➔ how to integrate what we know is different about grammar (dependencies, re-inventing, different evolutionary pressures etc) into historical modelling of grammar?

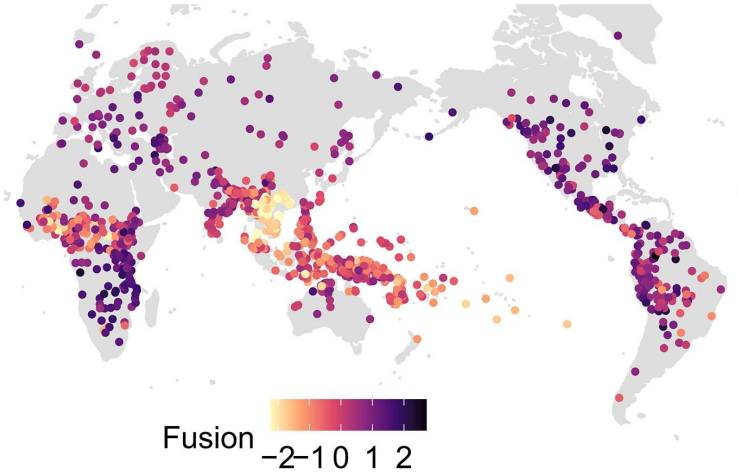
One of our recent studies

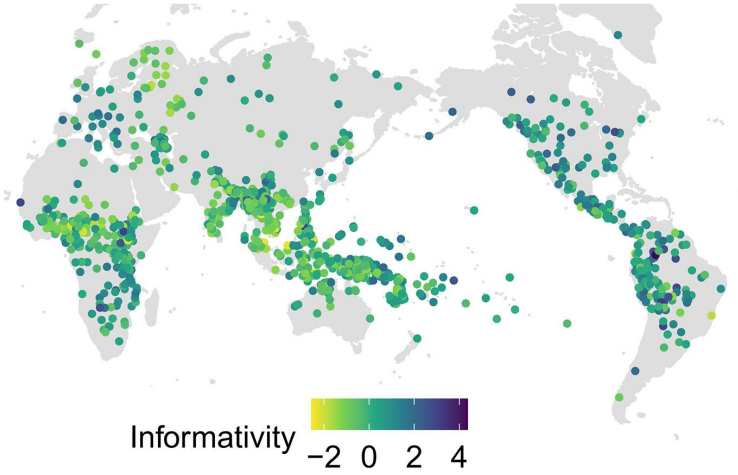
Shcherbakova, O., Michaelis, S. M., Haynie, H. J., Passmore, S., Gast, V., Gray, R. D., Greenhill, S., Blasi, D. & Skirgård, H. (2023). Societies of strangers do not speak less complex languages. Science Advances, 9(33), eadf7704.

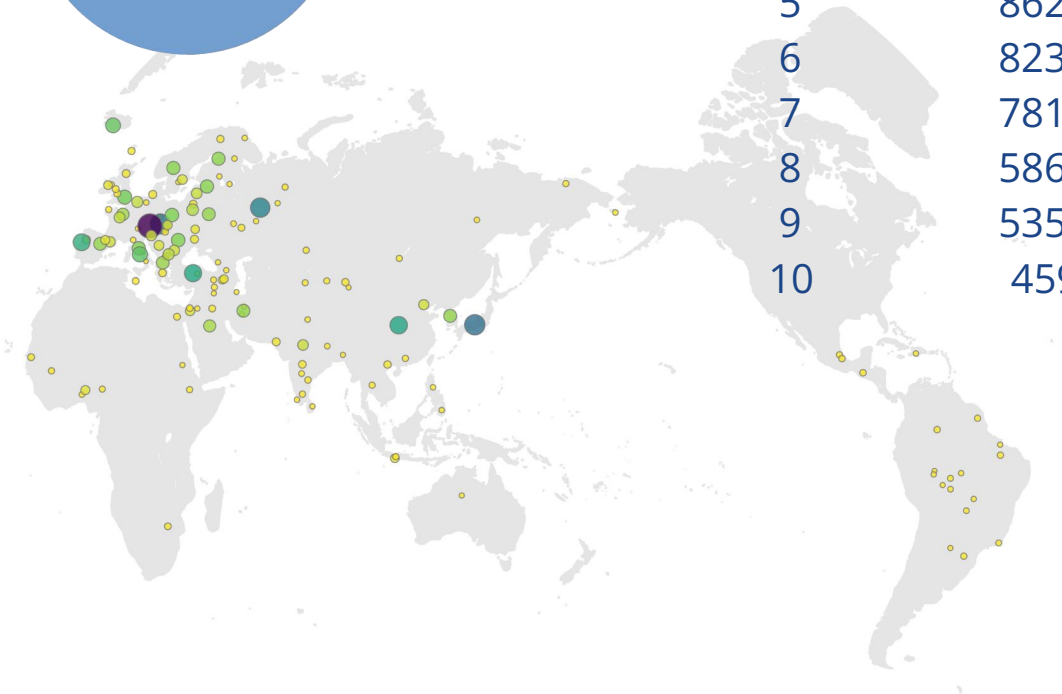
two dimensions of complexity using Grambank (questionnaire-based typology)

- ➔ boundedness (fusion)
- ➔ informativity

these were not affected by population size, a contrary result to Lupyan & Dale (2010)







1	208438 stan1295	German
2	132418 nucl1643	Japanese
3	127794 czec1258	Czech
4	116324 russ1263	Russian
5	86239 lite1248	Literary Chinese
6	82319 nucl1301	Turkish
7	78141 port1283	Portuguese
8	58683 lati1261	Latin
9	53564 icel1247	Icelandic
10	45982 stan1293	English

Other corpora to consider

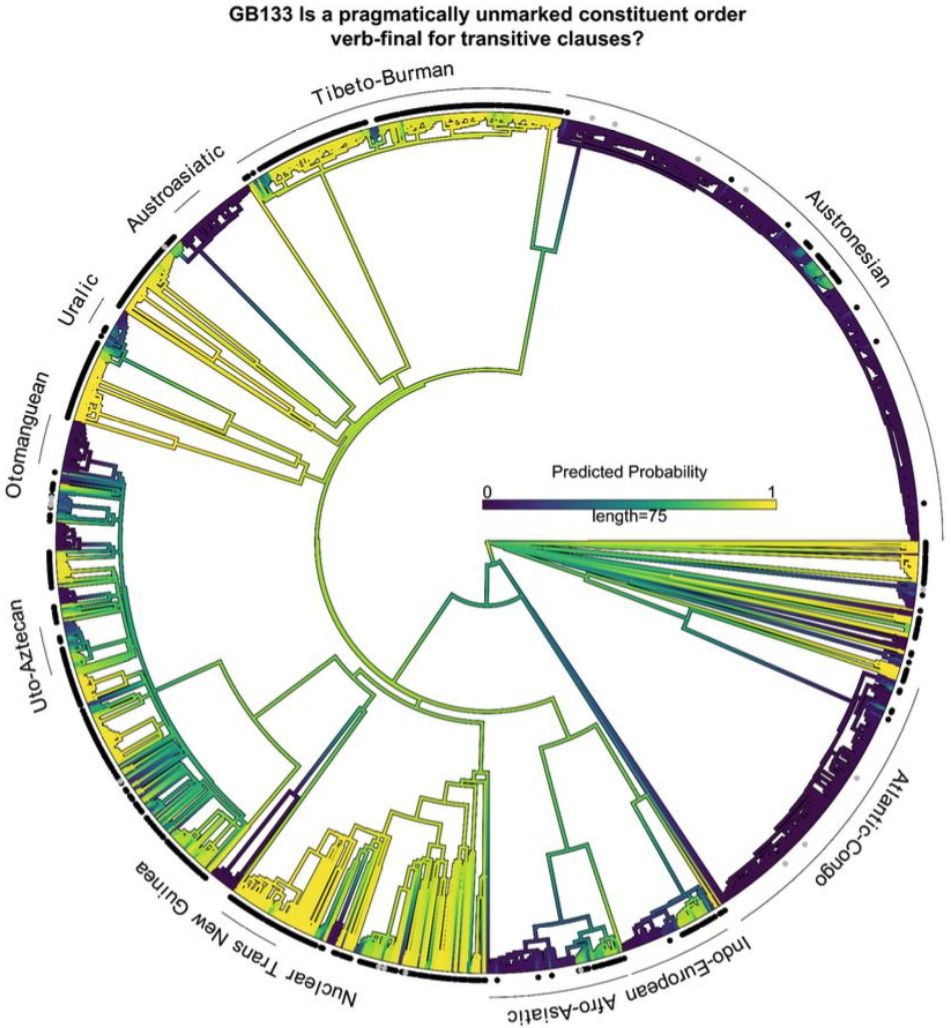
MultiCast
Universal Dependencies
DoReCo
ELAR
PARADISEC
CHILDES & TalkBank generally
INESS
Pangloss
The Language Archive (TLA)

Leipzig Corpora Collection
NLTK Corpora
Open Subtitles
Open parallel corpus
Europarl Parallel Corpus (EPC)
MULTEXT (Multilingual Text Tools and Corpora)

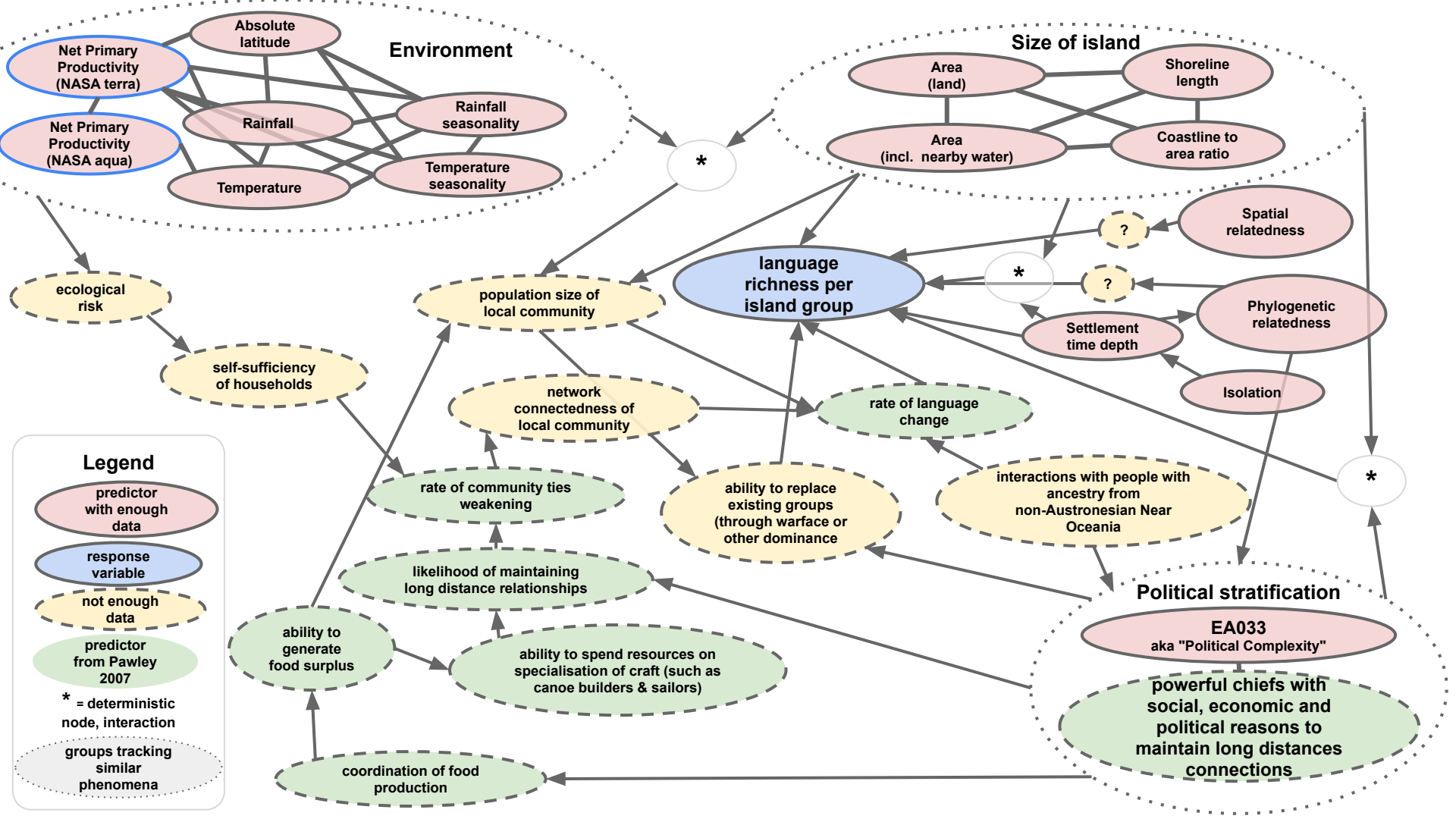
ASR with space?

regression models can impute missing values for ancestral nodes using information from the tips, tree and space

include horizontal effects better?



speaking of causality


Environment
Absolute latitude
Net Primary Productivity (NASA terra)
Net Primary Productivity (NASA aqua)
Rainfall
Rainfall seasonality
Temperature
Temperature seasonality
Size of island
Area (land)
Shoreline length
Area (incl. nearby water)
Coastline to area ratio
*
?
Spatial relatedness
language richness per island group
ecological risk
population size of local community
Phylogenetic relatedness
Settlement time depth
Isolation
self-sufficiency of households
network connectedness of local community
rate of language change
rate of community ties weakening
ability to replace existing groups (through warface or other dominance
interactions with people with ancestry from non-Austronesian Near Oceania
likelihood of maintaining long distance relationships
Political stratification
EA033
aka "Political Complexity"
ability to generate food surplus
ability to spend resources on specialisation of craft (such as canoe builders & sailors)
powerful chiefs with social, economic and political reasons to maintain long distances connections
coordination of food production
Legend
predictor with enough data
response variable
not enough data
predictor from Pawley 2007
* = deterministic node, interaction
groups tracking similar phenomena

the end

- ➔ thank you to all of my collaborators in the Grambank team, especially Russell Gray, Simon Greenhill, Olena Shcherbakova and Hannah Haynie
- ➔ thank you to all grambank coders, grammar writers and language communities who have made this all possible
- ➔ thank you to the workshop organisers in Edinburgh

References

- Chung, Sandra. 1978. Case Marking and Grammatical Relations in Polynesian Languages. Austin: University of Texas.
- Crowley, Terry. 1985. Common Noun Phrase Marking in Proto-Oceanic. Oceanic Linguistics 24(1/2). 135–193.
- Clark, D Ross. 1976. Aspects of Proto-Polynesian Syntax, vol. 6. Linguistic Society of New Zealand.
- Crowley, T. and Bowern, C. 2010. *An introduction to historical linguistics*, 4th edn. Oxford University Press.
- Evans, Bethwyn. 2001. A Study of Valency-Changing Devices in Proto Oceanic. Research School of Pacific and Asian Studies, The Australian National University. PhD Thesis.
- Pawley, Andrew. 1970. Grammatical Reconstruction and Change in Polynesia and Fiji. In SA Wurm & DC Laycock (eds.), Studies in Honour of Arthur Capell, 301–368. Canberra: Pacific Linguistics.
- Pawley, Andrew. 1973. Some Problems in Proto-Oceanic Grammar. Oceanic Linguistics 12(1/2). 103–188.
- Jonsson, Niklas. 1998. Det Polynesiska Verbmorfemet - Cia; Om Dess Funktion i Samoanska.
- Jäger, Gerhard & Johann-Mattis List. 2018. Using Ancestral State Reconstruction Methods for Onomasiological Reconstruction in Multilingual Word Lists. Language Dynamics and Change 8(1). 22–54.
- Kikusawa, Ritsuko. 2001. Rotuman and Fijian Case-Marking Strategies and Their Historical Development. Oceanic Linguistics 40(1). 85–111.
- Kikusawa, Ritsuko. 2002. Proto Central Pacific Ergativity: Its Reconstruction and Development in the Fijian, Rotuman and Polynesian Languages. Canberra: Pacific Linguistics.
- Lynch, John, Malcolm Ross & Terry Crowley. 2011. Proto Oceanic. In John Lynch, Malcolm Ross & Terry Crowley (eds.), The Oceanic Languages Curzon Language Family Series, 54–91. Richmond: Curzon 2nd edn.
- Marck, Jeffrey C. 2000a. Polynesian Languages. In J. Garry & C. Rubino (eds.), Facts About the World's Languages: An Encyclopaedia of the World's Major Languages, Past and Present, New York: H.W. Wilson.
- Marck, Jeffrey C. 2000b. Topics in Polynesian Language and Culture History. Canberra: Pacific Linguistics.
- Ross, Malcolm D. 2004. The Morphosyntactic Typology of Oceanic Languages. Language and Linguistics 5(2). 491–541.

References

Skirgård, Hedvig, Hannah J. Haynie, Damián E. Blasi, Harald Hammarström, Jeremy Collins, Jay J. Latarche, Jakob Lesage, Tobias Weber, Alena Witzlack-Makarevich, Sam Passmore, Angela Chira, Luke Maurits, Russell Dinnage, Michael Dunn, Ger Reesink, Ruth Singer, Claire Bowern, Patience Epps, Jane Hill, Outi Vesakoski, Martine Robbeets, Noor Karolin Abbas, Daniel Auer, Nancy A. Bakker, Giulia Barbos, Robert D. Borges, Swintha Danielsen, Luise Dorenbusch, Ella Dorn, John Elliott, Giada Falcone, Jana Fischer, Yustinus Ghanggo Ate, Hannah Gibson, Hans-Philipp Göbel, Jemima A. Goodall, Victoria Gruner, Andrew Harvey, Rebekah Hayes, Leonard Heer, Roberto E. Herrera Miranda, Nataliia Hübler, Biu Huntington-Rainey, Jessica K. Ivani, Marilen Johns, Erika Just, Eri Kashima, Carolina Kipf, Janina V. Klingenberg, Nikita König, Aikaterina Koti, Richard G. A. Kowalik, Olga Krasnoukhova, Nora L.M. Lindvall, Mandy Lorenzen, Hannah Lutzenberger, Tônia R.A. Martins, Celia Mata German, Suzanne van der Meer, Jaime Montoya Samamé, Michael Müller, Saliha Muradoglu, Kelsey Neely, Johanna Nickel, Miina Norvik, Cheryl Akinyi Oluoch, Jesse Peacock, India O.C. Pearey, Naomi Peck, Stephanie Petit, Sören Pieper, Mariana Poblete, Daniel Prestipino, Linda Raabe, Amna Raja, Janis Reimringer, Sydney C. Rey, Julia Rizaew, Eloisa Ruppert, Kim K. Salmon, Jill Sammet, Rhiannon Schembri, Lars Schlabbach, Frederick W.P. Schmidt, Amalia Skilton, Wikaliler Daniel Smith, Hilário de Sousa, Kristin Sverredal, Daniel Valle, Javier Vera, Judith Voß, Tim Witte, Henry Wu, Stephanie Yam, Jingting Ye 葉婧婷, Maisie Yong, Tessa Yuditha, Roberto Zariquiey, Robert Forkel, Nicholas Evans, Stephen C. Levinson, Martin Haspelmath, Simon J. Greenhill, Quentin D. Atkinson & Russell D. Gray (2023) Grambank reveals global patterns in the structural diversity of the world's languages. Science Advances 9. doi:10.1126/sciadv.adg6

References

Bouckaert, R., Redding, D., Sheehan, O., Kyritsis, T., Gray, R., Jones, K. E., & Atkinson, Q. (2022, July 20). Global language diversification is linked to socio-ecology and threat status. https://doi.org/10.31235/osf.io/f8tr6