Two ways in which linguistic typology is not butterfly collection

Hedvig Skirgård
Emergent Topics in Typology
Stockholm 2024

1

Me

➔ Born and raised in Uppsala
➔ Bachelor and masters degrees here in linguistics from Stockholm University
➔ Worked as a research assistant for Dr Hammarström in Nijmegen for 1.5 years
➔ PhD from Australian National University (2015 - 2020)
➔ 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

This talk

➔ I have never given a keynote before
➔ I was given very free reins
➔ I want to talk about two emergent topics I think about a lot
➔ I've included things I wish I knew earlier and that I think may be useful for some of you
➔ Some things may be very obvious to you

Two ways in which linguistic typology is _not_ like collecting butterflies

1 linguistic typology is not like collecting butterflies because in addition to describing the extent of the design space and finding patterns we are also looking to explain them. The explanations likely involve disciplines outside of linguistics (ecology, psychology, phylogenetics, philosophy etc).

2 linguistic typology is not like collecting butterflies because we are studying fellow humans, their communities, history etc. Language matters to people in a way that butterflies (mostly) don't, and most languages are under threat of dormancy

emergent topic 1

explanations and multidisciplinary

Linguistic typology

Large and diverse research field, broadly interested in

➔ systematic cross-linguistic comparison: finding similarities and differences, documenting where languages are in the possible **design space** (c.f. *typ*-ology)
➔ explaining said cross-linguistic variation, usually through cognitive, areal, cultural, pragmatic, historical and/or evolutionary **constraints**

More specifically

★ tends to be focussed on grammar, but not necessarily. There's lexical/semantic typology, phonological typology and pragmatic typology (as we have seen at this conference)
★ includes the testing of universals (different from Universal Grammar), finding of sprachbunds etc
★ linguistic typology is often identified in opposition to the generative linguistics enterprise, however there is also generative-based cross-linguistic comparison (e.g. SSWL, Longobardi et al. etc)
★ not about classifying languages into families - but can have a diachronic perspective
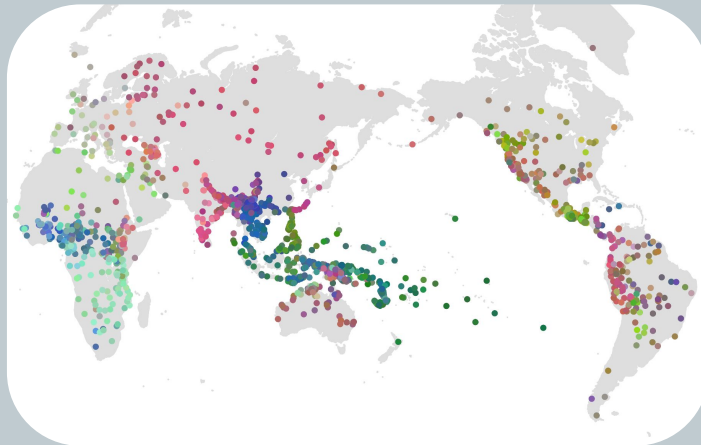
[*the aim of linguistic typology/theory*] *must be to show which structures are possible, in general, and why it is just those structures, and not others, that are possible*
(Hjelmslev)

What is where and why?
(Bickel)

Looking for explanations

➔ if cross-linguistic patterns are possibly impacted by X
   - then researchers in linguistic typology might also engage with Y & Z

predictor	linguistic field	non-linguistic field
social dynamics	sociolinguistics	sociology, anthropology, population genetics
cognitive constraints (e.g. working memory)	psycholinguistics	psychology, neurology, cognitive sciences
history (inheritance/space)	historical linguistics, areal linguistics	history, evolutionary biology, ecology
acquisition bottleneck (child and adult)	psycholinguistics, language acquisition	psychology
analogy / co-occurrence	Natural Language Processing, corpus ling	computer science

an example of looking for explanations multidisciplinarity

estimating the effect of genealogy and contact on grammar

Example 1: estimating spatial and genealogical effects

Grambank release paper

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

Detour: Grambank basics
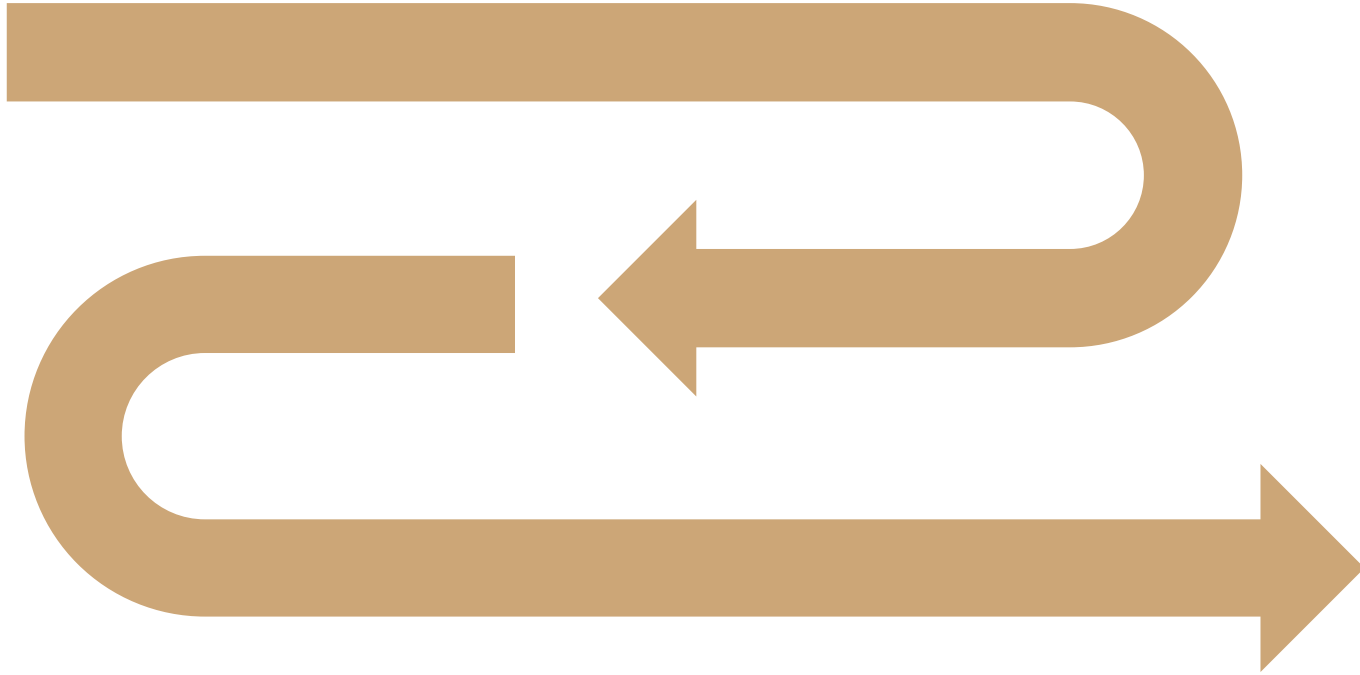
Grambank: overview

★ Part of the Glottobank consortium

★ Funded by the Department of Linguistic and Cultural Evolution at the Max Planck Institute for Evolutionary Anthropology

★ 195 features (including multistate) - 201 binarised

  ○ GB020 Are there definite or specific articles?

  ○ GB111 Are there conjugation classes?

  ○ GB159 Are nouns reduplicated?

★ based on NTS, Sahul, Pioneers and WALS-questionnaire
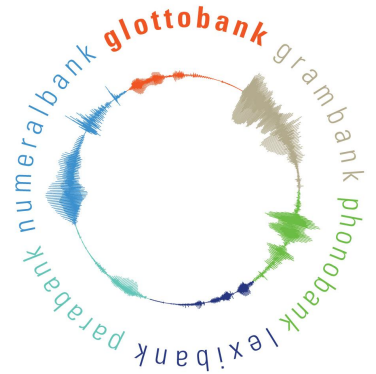
Grambank: coverage
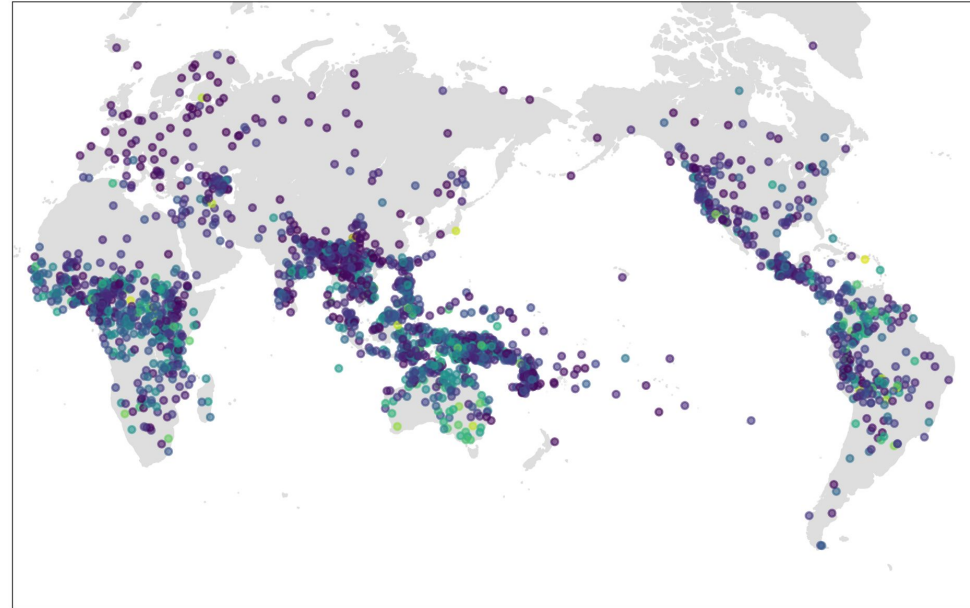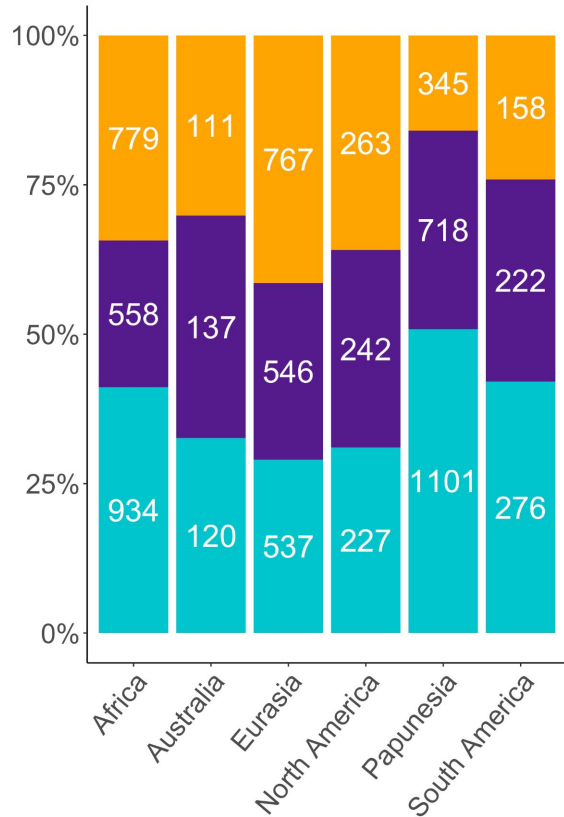
Grambank: a pragmatic approach

➔ definitions in comparative work needn't match language-specific description definitions
➔ sophisticated theoretical concepts are often hard to apply
➔ science relies on generalisations, and there won't be a satisfying perfect questionnaire
➔ finding the saturation point of challenges
➔ consistency is great
➔ we cannot be everything to everyone

verb

adjektiv

verbpartiklar

prepositions
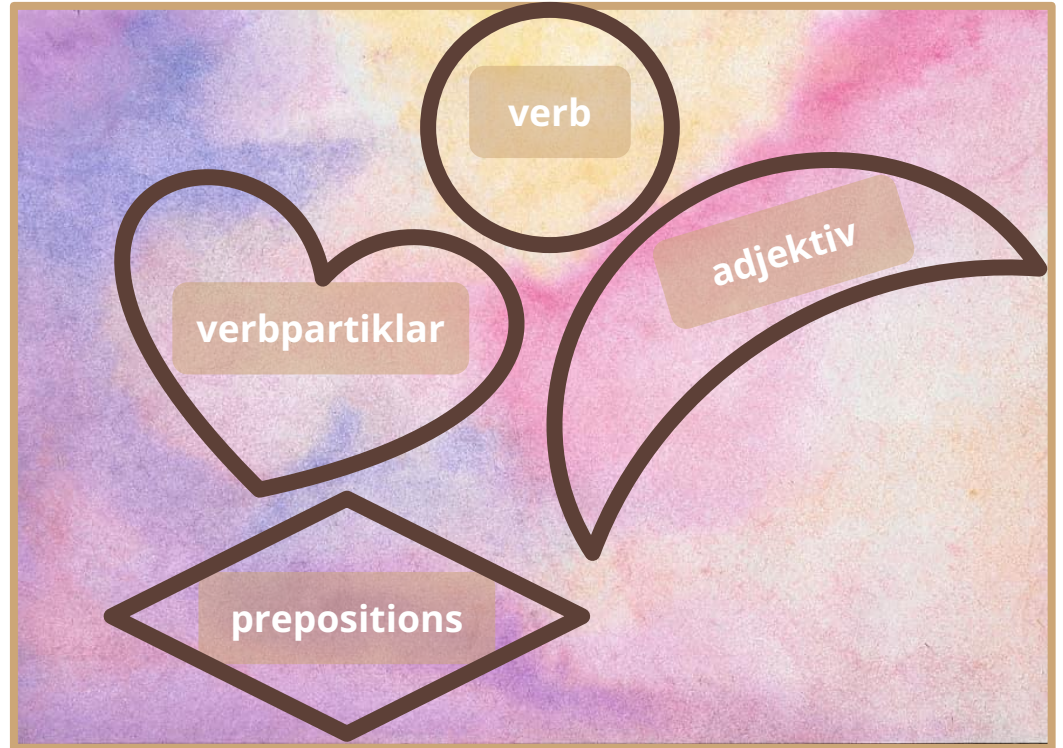
Grambank: acknowledgements

the Grambank project would not work without

➔ language communities sharing their knowledge (very large unknown number)
➔ grammar writers (over 5,000 unique authors are cited in Grambank)
➔ language experts (123)
➔ grambank coders (60+)
➔ grambank patrons and coordinators (8)
   ◆ Jeremy Collins, Hannah Haynie, Jay Latarche, Jakob Lesage, Hedvig Skirgård, Tobias Weber, Harald Hammarström and Alena Witzlack-Makarevich

Want to know more about Grambank inner workings?

Recommended Grambank wiki articles for anyone wanting
to know how to run a project like this

➜ Advice for typological database construction
➜ Encountering coding problems and solving them
➜ Language internal variation and different strategies for one
   function
➜ Automatic quality checks
➜ Absence of evidence
➜ Background of the Grambank questionnaire
➜ Contribute
➜ Dedicated marking
➜ Obligatoriness
➜ Productivity

grambank

Interactive break!
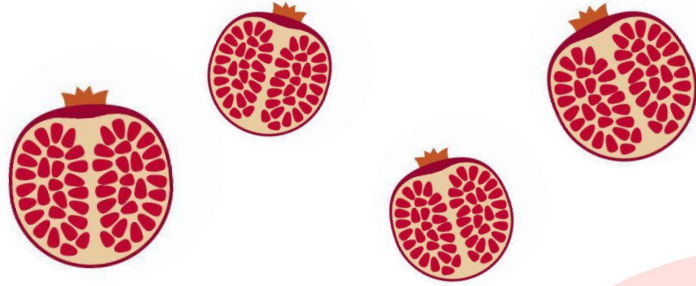
Please stand up

➔ Pick apples!
➔ Are you making a typology questionnaire-type dataset of your own?
   ◆ if yes: pick apples forward rather than up

➔ Would a template of a CLDF structure dataset (Grambank-flavoured) be useful to you?
   ◆ if yes: pick apples forward rather than up

➔ Thank you, you can sit down now!

back to spatial and genealogical effects

Example: estimating spatial and genealogical effects

➔ **question: among our Grambank questionnaire features, which ones are more predictable based on genealogy or geography**

➔ **inspiration from evolutionary biology:**
   ◆ R. Dinnage, A. Skeels, M. Cardillo, Spatiophylogenetic modelling of extinction risk reveals evolutionary distinctiveness and brief flowering period as threats in a hotspot plant genus. *Proc. R. Soc. B* **287**, 20192817 (2020).

➔ **specifics**
   ◆ regression model (Integrated Nested Laplace Approximations - INLA - which is similar to Bayesian Regression Modelling)
   ◆ calculate co-variation of languages in a global tree and geographic points
   ◆ run models with only tree, only geo and both (comparing with WAIC)
   ◆ brought on two model-specific experts as co-authors: Sam Passmore and Russell Dinnage
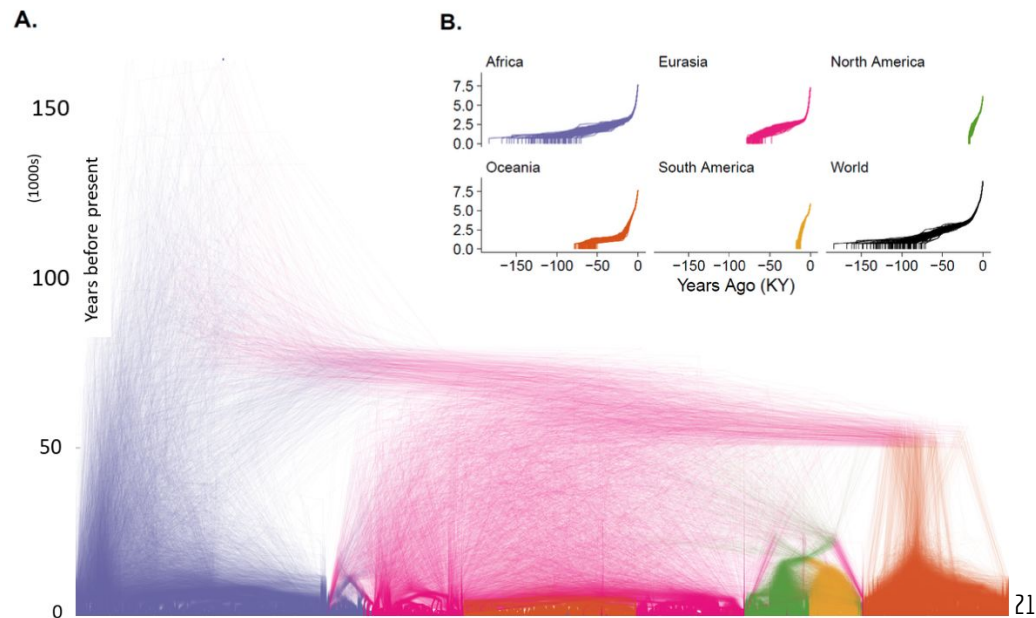
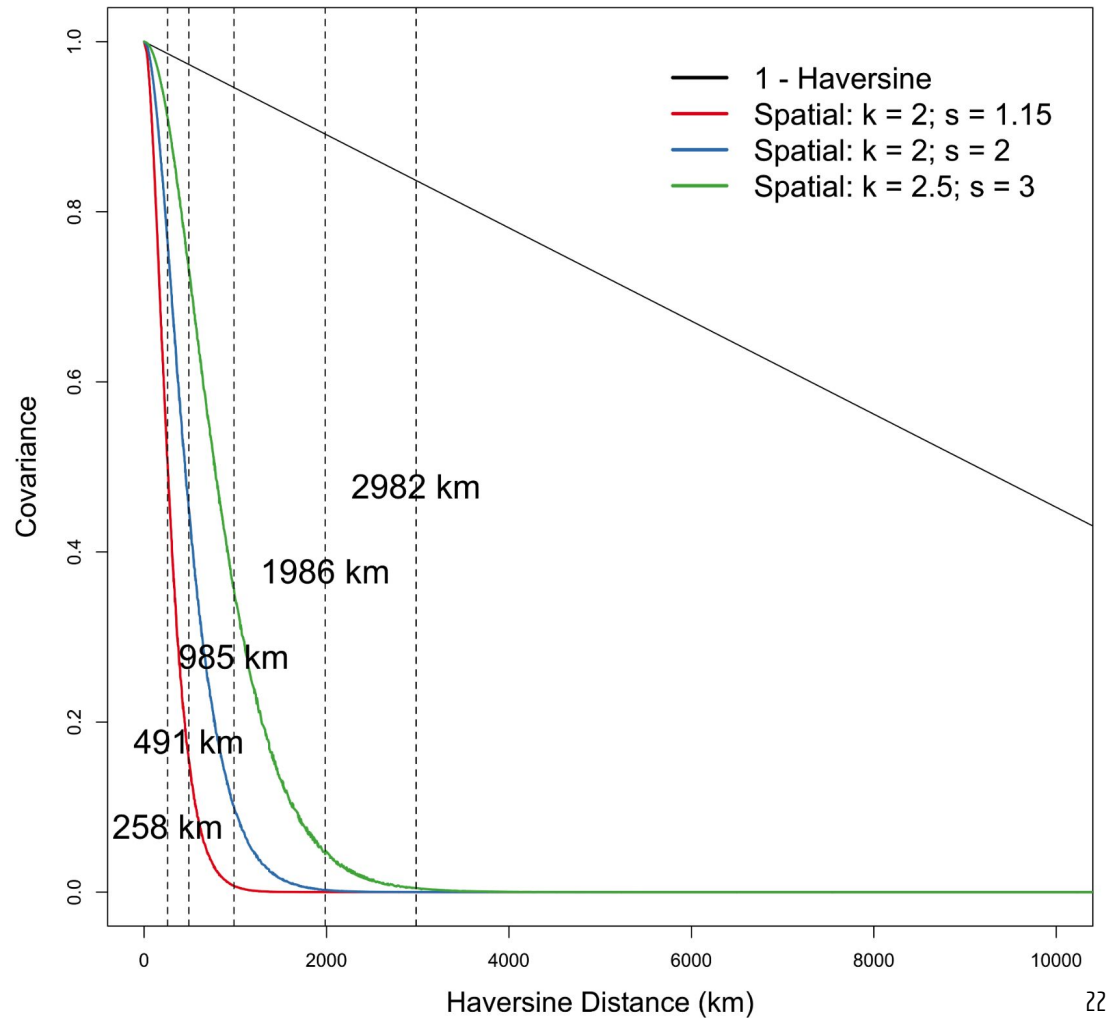# Example: estimating spatial and genealogical effects

global tree from:

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

➔ INLA uses a Brownian model of change

➔ language locations from Glottolog 4.5
➔ decay model of covariance using a Matérn function
➔ we varied sigma and kappa in the covariance, but it made little difference
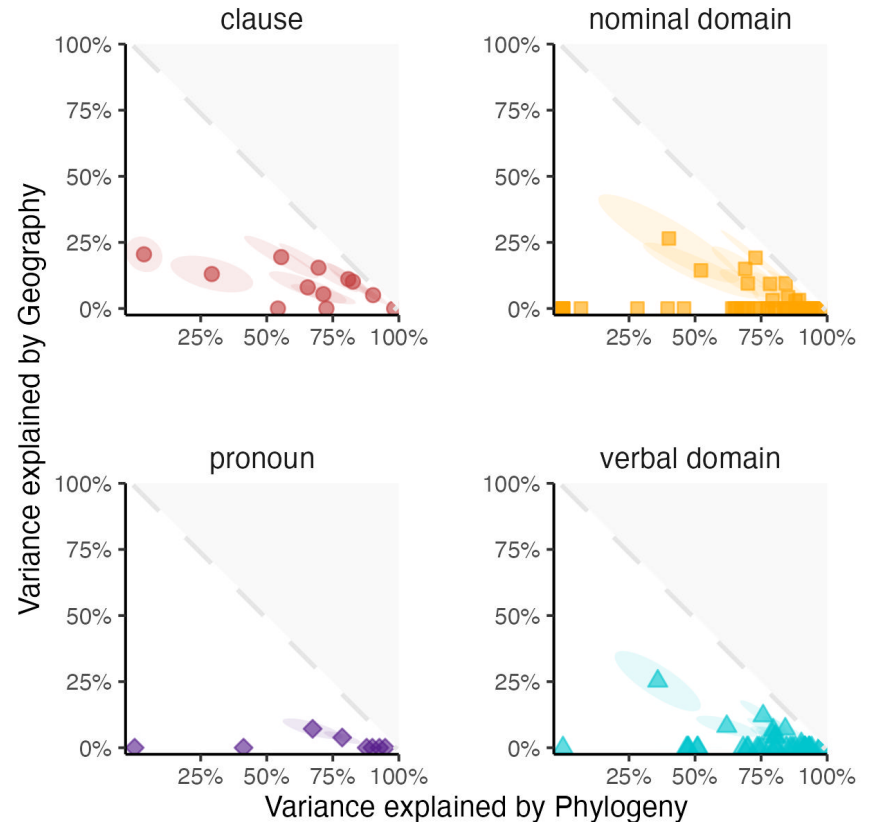
Example: estimating spatial and genealogical effects

➔ **result:**

- ◆ the vast majority of features were predictable mainly from phylogeny

➔ keep in mind, phylogenetic relations and spatial are correlated

other examples of multidisciplinary tools in LT

→ distance/co-variation
  ◆ Gower/relative Hamming, patristic distances, geographic cost-surfaces, cultural fixation scores
→ phylogenetic signal
  ◆ D-estimates, Bloomberg's K, Pagel's Lambda, neigbournets (visualisation and Q-residuals)
→ areality
  ◆ sBayes (Ranacher, Peter, et al. 2021), two-dimensional Gaussian Process (Guzmán Naranjo & Becker 2022), Matérn covariance in regression models (Skirgård et al 2023)
→ complexity/efficiency
  ◆ Shannon entropy , EEG (N400 spike), compression
→ testing strength of association by co-evolutionary modelling
  ◆ BayesTraits, regression modelling (brms, inla etc), phylogenetic path-analysis
→ causality
  ◆ Directed Acyclic Graphs

Remember…

You can borrow and adapt individual methods from other disciplines without assuming that the study objects and mechanisms are identical. The assumptions of the specific method may still hold.

For example, there are methods from evolutionary biology that can be applied to linguistics data *without* assuming that biological and cultural evolution are identical. Furthermore, methods always make reductive assumptions, even *within* disciplines researchers should carefully investigate if the data violates the assumptions of a given method in catastrophic ways.
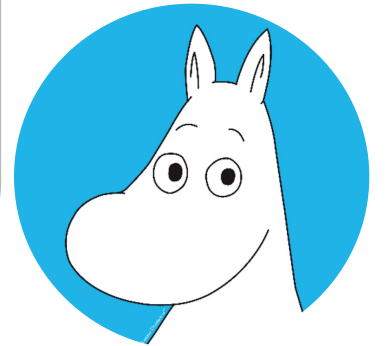
Paraphrasing Leino, U., Syrjänen, K., & Vesakoski, O. (2020). Linguistic change and biological evolution. *The Philosophy and Science of Language: Interdisciplinary Perspectives*, 179-193.

Things people have said I cannot understand/
would take me (specifically) considerable time to learn

✅ Classical Multidimensional scaling/Principal Coordinates Analysis
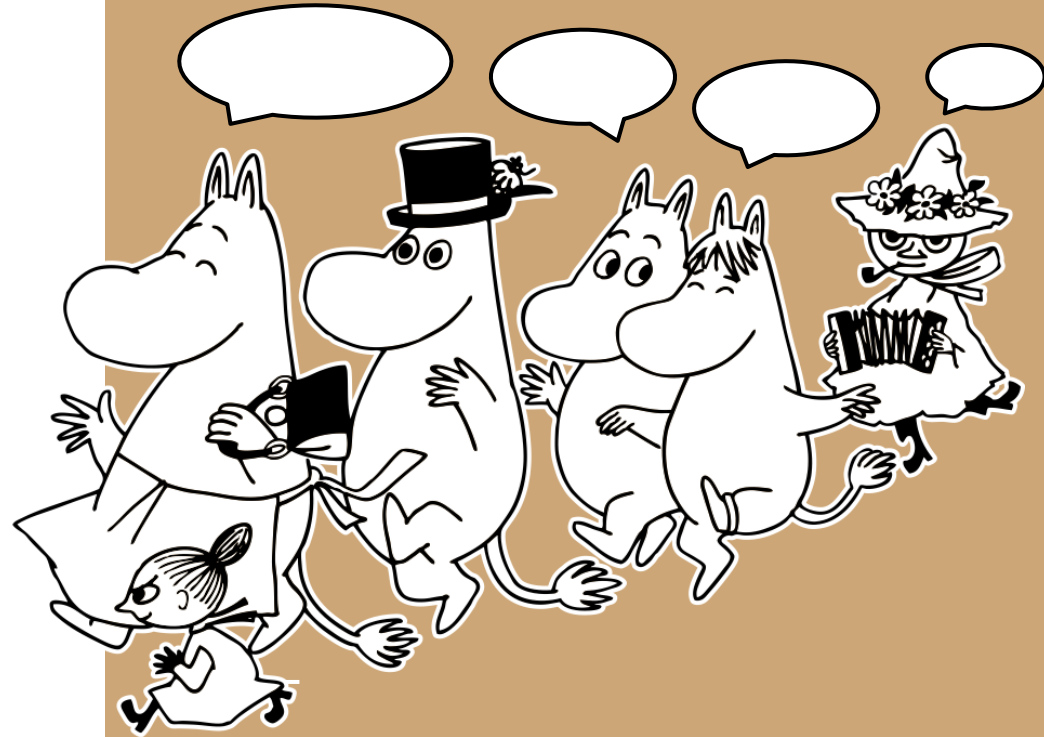
✅ Regression Models

❓ Co-evolutionary modelling (phylopath / BayesTraits)

emergent topic 2

languages cannot
be removed from people
and people matter

Languages are not butterflies, because people are not butterflies

≠

Languages are not butterflies

➔ languages exists in social landscapes
  ◆ contact, efficiency of production and reception, societal norms & stereotypes, convergence and divergence motivations
➔ people care about languages in a way (most people) don't care about butterflies
➔ most languages are endangered and endangerment is connected to other harm as well

➔ linguistics exists in a social context as well, and can be abused
  ◆ SIL International having to discourage US immigration from using Ethnologue to make decisions about asylum
  ◆ prescriptivism broadly
  ◆ political agendas regarding urheimat and historical linguistics broadly (e.g. Iran & Turkey)
  ◆ small languages or contact languages being described as useless, primitive etc

United Nations Educational, Scientific and Cultural Organization's

2022-2032 | INTERNATIONAL DECADE OF

Indigenous Languages

Status quo

★ most languages of the world are endangered

★ language endangerment is associated with health problems (Auger 2016; Durie et al 2009; Oster et al (2014)

★ Language is often used as a metonym for losing culture, identity, cohesion etc.

★ Education tip: SBS Australia's My Grandmother's Lingo

★ linguists are interested in language diversity as a way to understand the possibilities of being human, less diversity → less knowledge

★ circa 4,000 languages have at least a grammar sketch (see Glottolog's glottoscope)

What can linguists do?

➔ a lot of the same
➔ explore possibilities of explicitly connecting research to large questions
➔ be mindful about biases in research
➔ when possible, consider supporting indigenous-driven research
➔ corpus based linguistic typology can be done on smaller languages too (to a certain extent) , e.g. DoReCo, ELAR, PARADISEC etc.
➔ conferences and institutions: ELDP, LREC, langdoc, etc.
➔ do public outreach to talk about value of language diversity
 ◆ 3 minute thesis, high-school/secondary school visits, your university's outreach/media department, talk to your mum
➔ More ideas?

General linguistic wisdoms for the public which may seem trivially obvious to linguists

➔ all languages are communicatively equivalent
➔ aesthetic judgments are just opinions, they are not objectively valuable
➔ multilingualism does not harm children & is the norm
➔ all languages are valuable
➔ language change and variation is not bad

Lest we forget

➔ Naturally, losing language is just one of many ways minority groups suffer
➔ Languages are not only valuable in terms of their unique record of humanity (if so, we could just document them all once and be done)
➔ Language endangerment is connected with colonialist and nation state ideologies broadly
   ◆ More roads, more education → more language endangerment (Bromham et al 2022)
➔ Let's not lose sight of people's agency and subjectivity
   ◆ *When a linguist or a language activist asks : "Why do speakers of a language stop speaking it?" what they usually really mean is: "Why have the speakers of that language failed us?"*

Kulick (2019) *A death in the rainforest - how a language and a way of life came to an end in Papua new Guinea.* Pages 24-25

Research in light of large-scale language dormancy

two examples

Example 2: what happens to language disparity when endangered languages are gone?

Grambank release paper

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

# Example 2: what happens to language disparity when languages are lost?

Functional richness - the amount of niche space occupied by a set of observations

Another method borrowed from ecology

The Functional Richness metrics are based on multidimensional scaling of Gower-distances of Grambank data, computed with all and then with only those that are not labelled as endangered

Captained by co-author Simon Greenhill

Example 2: Grambank functional richness by area
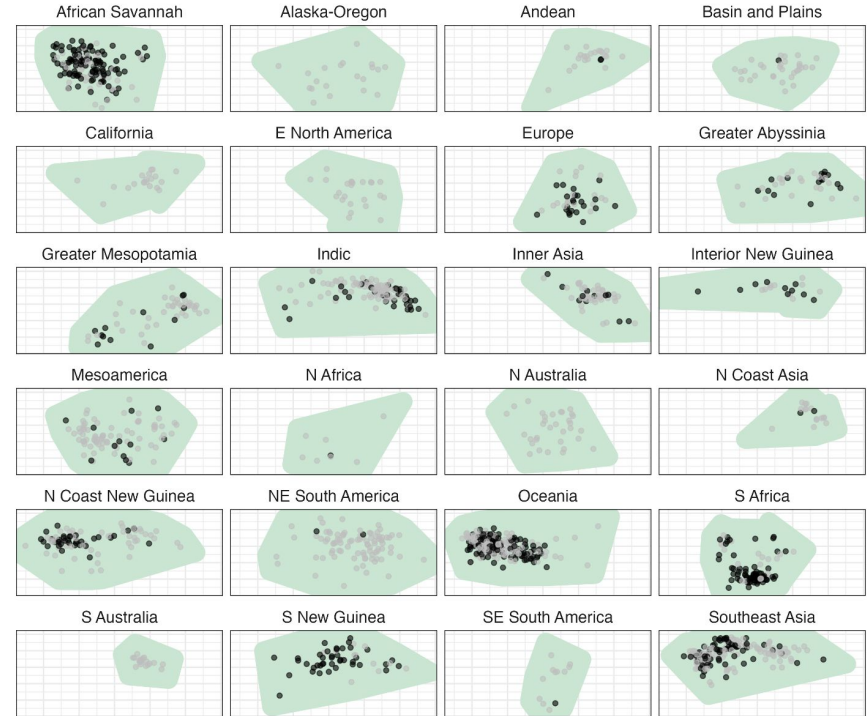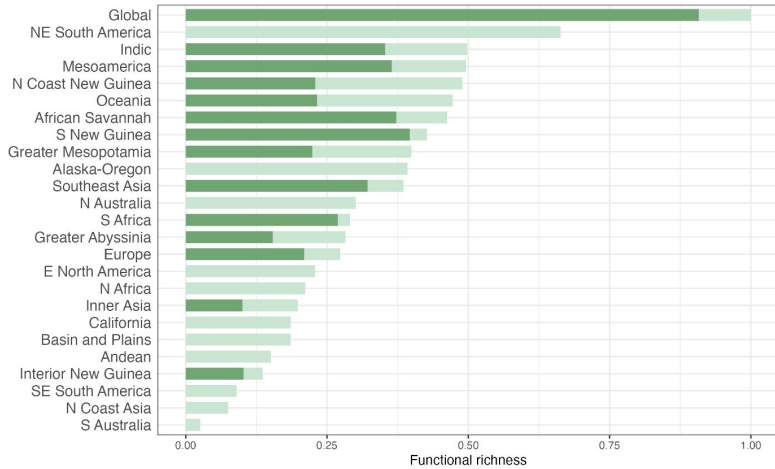
The pronounced reduction of nearly half the functional space occupied by languages, even in regions with many non-threatened languages (e.g., Oceania, North Coast New Guinea, Greater Abyssinia, and Greater Mesopotamia), will undermine our ability to investigate the basic structures of language and the diverse expressions used to encode them.

Example 3: gramgaps

Lesage, J., Haynie, H. J., Skirgård, H., Weber, T., & Witzlack-Makarevich, A. (2022). Overlooked data in typological databases: What grambank teaches us about gaps in grammars. In *13th Conference on Language Resources and Evaluation (LREC 2022)* (pp. 2884-2890). European Language Resources Association (ELRA)

- using grambank to measure the amount of description and straightforwardness of features
- using datapoints + comments of coders
- give indications of where more description is needed

Example 3: Gramgaps

Level of description = **presence of evidence vs. absence of evidence**

Straightforwardness = **how easy is it to translate a described topic into a value** in Grambank?
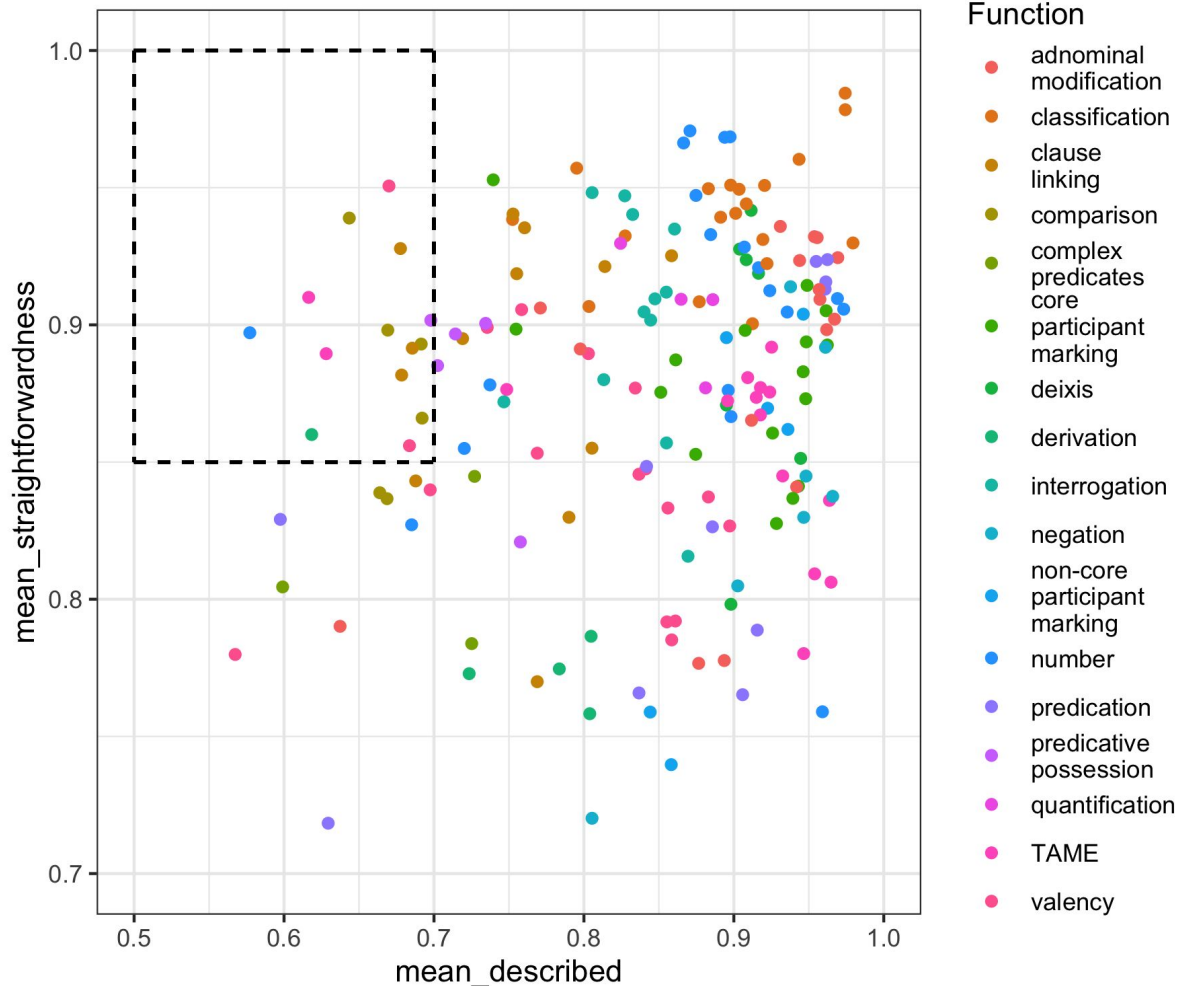
Indigenous perspectives on research

➜ "Something's gotta change" by Lesley Woods (2023):
https://press.anu.edu.au/publications/series/asia-pacific-linguistics/somethings-gotta-change (Ngiyampaa Wangaaypuwa, Australia)
   ◆ podcast Because Language episode 35

➜ "Decolonizing Methodologies" by Linda Tuhiwai Smith (1999):
https://en.wikipedia.org/wiki/Decolonizing_Methodologies (Māori, New Zealand)

➜ "Studying the vernacular in the vernacular – Luqa literacy in the Solomon Islands" by Alpheaus Graham Zobule. Talk at the Australian Research Council's Centre of Excellence for the Dynamics of Language, Canberra. 2020
https://youtu.be/InBDa6uq9zA?si=zb3KJNdR-CHTgqqq (Luqa, Solomon Islands)

This is the end, thank you

References

Bromham et al 2022), L., Dinnage, R., Skirgård, H., Ritchie, A., Cardillo, M., Meakins, F., ... & Hua, X. (2022). Global predictors of language endangerment and the future of linguistic diversity. *Nature ecology & evolution*, *6*(2), 163-173.)

Lesage, J., Haynie, H. J., Skirgård, H., Weber, T., & Witzlack-Makarevich, A. (2022). Overlooked data in typological databases: What grambank teaches us about gaps in grammars. In *13th Conference on Language Resources and Evaluation (LREC 2022)* (pp. 2884-2890). European Language Resources Association (ELRA)

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

M. D. Auger, Cultural Continuity as a Determinant of Indigenous Peoples' Health: A Metasynthesis of Qualitative Research in Canada and the United States. Int. Indig. Policy J. 7, 3 (2016).

M. Durie, H. Milroy, E. Hunter (2009) "Mental Health and the Indigenous Peoples of Australia and New Zealand" in Healing Traditions: The Mental Health of Aboriginal Peoples in Canada, L. J. Kirmayer, G. G. Valaskakis, Eds. (UBC Press, Vancouver, 2009), pp. 36–55.

Oster, R. T., Grier, A., Lightning, R., Mayan, M. J., & Toth, E. L. (2014). Cultural continuity, traditional Indigenous language, and diabetes in Alberta First Nations: a mixed methods study. International journal for equity in health, 13, 1-11.

Three tips

➜ CrossGram (Haspelmath, Englisch and Forkel)

◆ https://crossgram.clld.org/

➜ CLDF for dummies
◆ https://github.com/dlce-eva/dlce-eva/blob/main/doc/cldf_for_dummies.md

➜ R-scripting with Grambank
◆ https://github.com/HedvigS/R_grambank_cookbook
◆ https://github.com/HedvigS/Oceanic_computational_ASR
◆ https://github.com/grambank/grambank-analysed

questions in linguistics often overlap with biology ones specifically, in particular ecology and evolutionary biology