Analyzing Correlated Evolution of Multiple Features Using Latent Representations: Supplementary Materials

Yugo Murawaki
Graduate School of Informatics, Kyoto University
Yoshida-honmachi, Sakyo-ku, Kyoto, 606-8501, Japan
murawaki@i.kyoto-u.ac.jp

S.1 Calibration Points

Table S.1 lists calibration points used in our experiments. We collected these points from secondary literature and mapped them to Glottolog 3.2 (Hammarström et al., 2018). Not all parameters are present in the sources cited, and they are educated guesses at best.

S.2 Preliminary Analysis of the Inferred Trees

It remains a challenging task to quantitatively evaluate the time-trees we inferred from binary latent parameters because what they should look like is largely an open question. Like the present study, Dediu (2010) also inferred the (relative) dates and states of the internal nodes, but he did not assess their quality either.

We need gold standard data for past languages that have been documented or reconstructed from descendants. Unfortunately, WALS (Haspelmath et al., 2005), the database of typological features we used in the experiments, only covers modern languages. To facilitate research on diachronic typology, we urge typologists to collect what they know about typological profiles of past languages. Such a dataset can be used not only for evaluation but for more accurate phylogenetic inference (Chang et al., 2015).[1]

We acknowledge that creating a typological database of past languages is much more challenging than creating a phonological database. There are at least two reasons for this. First, the success of traditional historical-comparative linguistics is mostly limited to phonology and morphology; syntactic reconstruction remains highly controversial (Barðdal and Eythórsson, 2012). Second, even if we give up reconstructed languages and focus on attested languages, determining their feature values is not easy. While historical-comparative linguists share the common goal of identifying phonological inventories, typological features are highly theory-dependent. Even the order of subject, object and verb, which appears to be theory-free at first glance, turns out to be non-trivial because it is not necessarily clear how to determine the dominant order for flexible order languages (Dryer, 2013).

Nevertheless, we investigated several proto-languages of which we have limited knowledge. We confirmed that the model did not go in the wrong direction. When there was a single dominant value among its children, the parent usually picked it up as expected. Of course, we are more interested in cases where children disagreed with each other, but we ourselves simply did not have the answer. In the following, we discuss a data point for which we expected the model to fail and it did to a large degree.

A frequent criticism against phylogenetic approaches to diachronic typology is directed at its failure to take contact into account (Croft et al., 2011). A common counter-argument is that because phylogenetic methods are agnostic to the source of a change, contact-induced changes are only an indication of the trait's low phylogenetic stability (Dediu and Cysouw, 2013). We are, however, more pessimistic about phylogenetic reconstruction based only on a snapshot of dynamic pro-

[1] Marsico et al. (2018) recently published phonological inventory data of ancient and reconstructed languages that are comparable with PHOIBLE (Moran et al., 2014), a database of modern languages. What we need is a typological version of the phonological database.

Autotyp 0.1.0 (Bickel et al., 2017), a smaller, more focused database of linguistic typology, contains some ancient languages, mostly of the Middle East. DiACL (Carling, 2017) explicitly aims at covering historical and reconstructed languages but its geographical coverage is limited to Eurasia, Pacific and the Amazon. Nevertheless, Cathcart et al. (2018) made effective use of these languages to identify change events on branches of a time-tree.

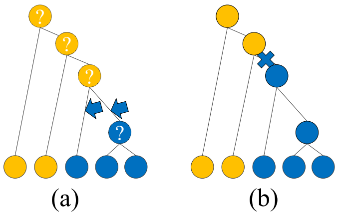

Figure S.1: Contact-induced change. (a) A tree with actual node states. The arrows denote contacts. (b) A tree that phylogenetic models are likely to reconstruct from the leaf nodes.

cesses. Areal neighbors in contact often happen to be genetic relatives, as illustrated in Figure S.1. Take the well-known areal feature, tone, as an example. Even though it is almost certain that Old Chinese was atonal (Baxter and Sagart, 2014), all modern Sinitic languages are tonal. Not surprisingly, Proto-Sinitic had a complex tone system with a probability of 98.3%, according to our analysis. The phylogenetic stability of tone must have been overestimated. We believe that if we have data on ancestral languages, using them as constraints (Chang et al., 2015; Cathcart et al., 2018) would mitigate the problem.

References

Jóhanna Barðdal and Thórhallur Eythórsson. 2012. Reconstructing syntax: Construction grammar and the comparative method. In Hans Christian Boas and Ivan A. Sag, editors, *Sign-based Construction Grammar*, pages 257–308. Center for the Study of Language and Information.

William H. Baxter and Laurent Sagart. 2014. *Old Chinese: A New Reconstruction*. Oxford University Press.

Balthasar Bickel, Johanna Nichols, Taras Zakharko, Alena Witzlack-Makarevich, Kristine Hildebrandt, Michael Rießler, Lennart Bierkandt, Fernando Zúñiga, and John B. Lowe. 2017. The AUTOTYP typological databases. version 0.1.0.

Remco Bouckaert, Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. Suchard, and Quentin D. Atkinson. 2012. Mapping the origins and expansion of the Indo-European language family. *Science*, 337(6097):957–960.

Gerd Carling. 2017. Diachronic atlas of comparative linguistics online. Accessed: 2018-08-27.

Chundra Cathcart, Gerd Carling, Filip Larsson, Niklas Johansson, and Erich Round. 2018. Areal pressure in grammatical evolution. *Diachronica*, 35(1):1–34.

Will Chang, Chundra Cathcart, David Hall, and Andrew Garrett. 2015. Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language*, 91(1):194–244.

William Croft, Tanmoy Bhattacharya, Dave Kleinschmidt, D. Eric Smith, and T. Florian Jaeger. 2011. Greenbergian universals, diachrony, and statistical analyses. *Linguistic Typology*, 15(2):433–453.

Dan Dediu. 2010. A Bayesian phylogenetic approach to estimating the stability of linguistic features and the genetic biasing of tone. *Proceedings of the Royal Society of London B: Biological Sciences*, 278(1704):474–479.

Dan Dediu and Michael Cysouw. 2013. Some structural aspects of language are more stable than others: A comparison of seven methods. *PLoS ONE*, 8(1):1–20.

Matthew S. Dryer. 2013. Order of subject, object and verb. In Matthew S. Dryer and Martin Haspelmath, editors, *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology.

Russell D. Gray, Alexei J. Drummond, and Simon J. Greenhill. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *science*, 323(5913):479–483.

Rebecca Grollemund, Simon Branford, Koen Bostoen, Andrew Meade, Chris Venditti, and Mark Pagel. 2015. Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences*, 112(43):13296–13301.

Harald Hammarström, Robert Forkel, Martin Haspelmath, and Sebastian Bank, editors. 2018. *Glottolog 3.2*. Max Planck Institute for the Science of Human History.

Martin Haspelmath, Matthew Dryer, David Gil, and Bernard Comrie, editors. 2005. *The World Atlas of Language Structures*. Oxford University Press.

Eric W. Holman, Cecil H. Brown, Søren Wichmann, André Müller, Viveka Velupillai, Harald Hammarström, Sebastian Sauppe, Hagen Jung, Dik Bakker, Pamela Brown, Oleg Belyaev, Matthias Urban, Robert Mailhammer, Johann-Mattis List, and Dmitry Egorov. 2011. Automated dating of the world's language families based on lexical similarity. *Current Anthropology*, 52(6):841–875.

Egidio Marsico, Sebastien Flavier, Annemarie Verkerk, and Steven Moran. 2018. BDPROTO: A database of phonological inventories from ancient and reconstructed languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, pages 1654–1658.

Luke Maurits and Thomas L. Griffiths. 2014. Tracing the roots of syntax with Bayesian phylogenetics. *Proceedings of the National Academy of Sciences*, 111(37):13576–13581.

Steven Moran, Daniel McCloy, and Richard Wright, editors. 2014. *PHOIBLE Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Name	Glottocode	Prior	Src.
Aceh-Chamic	cham1327	Uniform$(1{,}800, 2{,}500)$	G
Afro-Asiatic	afro1255	Uniform$(9{,}500, 25{,}000)$	M
Anglo-Frisian	angl1264	Uniform$(1{,}490, 1{,}590)$	H
Austronesian	aust1307	$\mathcal{N}(\mu = 7{,}000, \sigma = 1{,}000)$	M
Balto-Slavic	balt1263	$\mathcal{N}(\mu = 3{,}100, \sigma = 600)$	B
Benue-Congo	benu1247	Uniform$(6{,}000, 7{,}000)$	H
Brythonic	bryt1239	$\mathcal{N}(\mu = 1{,}550, \sigma = 25)$	B
Celtic	celt1248	Lognormal$(\mu = 2{,}000, \sigma = 0.6) + 1{,}200$	B
Cham	cham1328	$\mathcal{N}(\mu = 529, \sigma = 25)$	H
Chamic	cham1330	Uniform$(1{,}500, 1{,}600)$	H
Cholan	chol1287	$\mathcal{N}(\mu = 1{,}600, \sigma = 250)$	H
Common Turkic	comm1245	$\mathcal{N}(\mu = 1{,}419, \sigma = 250)$	H
Czech-Slovak	czec1260	Uniform$(1{,}000, 1{,}100)$	H
East Bantu	east2731	$\mathcal{N}(\mu = 2{,}500, \sigma = 25)$	R
East Polynesian	east2449	Uniform$(1{,}150, 1{,}800)$	G
East Slavic	east1426	$\mathcal{N}(\mu = 760, \sigma = 25)$	H
Eastern Baltic	east2280	$\mathcal{N}(\mu = 1{,}350, \sigma = 25)$	B
Ellicean	elli1239	Uniform$(1{,}000, 2{,}000)$	G
Ethiosemitic	ethi1244	Uniform$(2{,}400, 2{,}500)$	H
Goidelic	goid1240	Uniform$(1{,}000, 1{,}100)$	H
Hmong-Mien	hmon1336	$\mathcal{N}(\mu = 2{,}500, \sigma = 500)$	H
Indo-Aryan	indo1321	Lognormal$(\mu = 1{,}000, \sigma = 1.0) + 2{,}150$	B
Indo-European	indo1319	$0.7\mathcal{N}(\mu = 6{,}000, \sigma = 750) + 0.3\mathcal{N}(\mu = 8{,}750, \sigma = 750)$	M
Indo-Iranian	indo1320	Uniform$(4{,}000, 4{,}800)$	H
Inuit	inui1246	$\mathcal{N}(\mu = 800, \sigma = 50)$	H
Iranian	iran1269	Lognormal$(\mu = 400, \sigma = 0.8) + 2{,}600$	B
Kipchak	kipc1239	$\mathcal{N}(\mu = 900, \sigma = 100)$	H
Malayo-Polynesian	mala1545	Uniform$(3{,}600, 4{,}500)$	G
Micronesian	micr1243	Uniform$(1{,}900, 2{,}200)$	G
Mississippi Valley	miss1254	Uniform$(2{,}250, 2{,}700)$	H
Mongolic	mong1329	$\mathcal{N}(\mu = 750, \sigma = 100)$	H
Narrow Bantu	narr1281	Uniform$(4{,}000, 5{,}000)$	R
North and East Malayo-Sumbawan	nort3170	Uniform$(2{,}000, 3{,}000)$	G
North Germanic	nort3160	Uniform$(950, 1{,}250)$	H
Northwest Germanic	nort3152	$\mathcal{N}(\mu = 1{,}875, \sigma = 67)$	B
Nuclear Oromo	nucl1736	$\mathcal{N}(\mu = 460, \sigma = 50)$	H
Oceanic	ocea1241	Uniform$(3{,}200, 3{,}600)$	G
Pama-Nyungan	pama1250	Uniform$(4{,}000, 5{,}000)$	H
Saami	saam1281	Uniform$(1{,}500, 2{,}000)$	H
Sinitic	sini1245	$\mathcal{N}(\mu = 2{,}500, \sigma = 250)$	H
Sino-Tibetan	sino1245	$\mathcal{N}(\mu = 7{,}000, \sigma = 1{,}000)$	M
Slavic	slav1255	Lognormal$(\mu = 300, \sigma = 0.6) + 1{,}200$	B
Southeast Barito	sout2919	Uniform$(1{,}300, 1{,}400)$	H
Southern Nilotic	sout2830	Uniform$(2{,}000, 3{,}000)$	H
Romance	roma1334	$\mathcal{N}(\mu = 1{,}729, \sigma = 100)$	H
Romani	roma1329	Uniform$(600, 700)$	H
Tupi-Guarani	tupi1276	Uniform$(1{,}500, 2{,}000)$	H
Turkic	turk1311	$\mathcal{N}(\mu = 2{,}500, \sigma = 500)$	H
Wakashan	waka1280	$\mathcal{N}(\mu = 2{,}500, \sigma = 500)$	H
West Germanic	west2793	$\mathcal{N}(\mu = 1{,}550, \sigma = 25)$	B

Table S.1: List of calibration points. The sources are as follows. B: Bouckaert et al. (2012). G: Gray et al. (2009). H: Holman et al. (2011). M: Maurits and Griffiths (2014). R: Grollemund et al. (2015).