Understanding Typological Distributions

Balthasar Bickel



- In the phonological words example, we saw that languages tend to have many distinct p-words
- One approach to analyze them is to take each p-word as a datapoint of its own and then analyze the distribution of these datapoints in the world seeking non-accidental skewings.
- A hypothesis: since stress patterns systematically interact with higher-level domains (phrases, utterances; information structure), they target larger domains.

- Hypothesized to be very common: Limbu (Sino-Tibetan) Stress: [prefix-'stem-suffix=clitic] [mε-'thaŋ-e=aŋ] 3ns-come.up-PST=and
- Hypothesized to be much less common: Mon (Austroasiatic) Stress: ['cl]=[pf<infix>'stem]=['cl]
 [k<ə>'lp?]
 <CAUS>cross
 ['kp]=['klp?]
 <CAUS=cross

 To test this hypothesis, we need a measure of relative size, 'coherence':

 $C(p; L) = \frac{N \text{ (morpheme types referenced by } p)}{N \text{ (morpheme types in } L)}$



- The INFERENCE problem: how to extrapolate from a sample to all human languages in the absence of random sampling? (NB: there already have been at least 500,000 languages, but we know only about 500!)
- 2. The **DIACHRONY** problem: how to tell apart the factors determing the observed cross-linguistic distribution?
 - Structural Pressure ('what works best, replicates best'; a.k.a. 'selection', 'functional pressure', 'preferred pathways of change', 'linguistic principles')
 - Language Contact ('what is most popular, replicates best')
 - Blind Inheritance ('what was best for our parents, is best for us')
 - Random fluctuation

- If we can solve the Diachrony Problem, we have solved the Inference Problem:
 - if we know that a distribution is due to Structural Pressure, independent of any other diachronic factor, this Structural Pressure determines how languages develop over time, beyond what we can observe now.

- Useful reformulation of empirical universals: instead of $p \rightarrow q$,
- $E(q) \sim p$,
 - where E represent the expected mean of q, given p
 - E(q) can be directly linked to p if q is continuous; nominal q need to be transformed first by what is called a 'link function' g(), usually the natural logarithm of odds ('logit', 'logistic regression') or counts ('loglinear analysis'). This defines the Generalized Linear Model:

$$g(E(q)) = \alpha + \beta_1 p_1 + \beta_2 p_2 \dots + \beta_k p_k$$

- bidirectional universals have an additional condition: odds(q) ≠ 1 under any level of p.
- unrestricted universal: $g(E(q)) = \alpha$, and α deviates from H₀.

Using Generalized Linear Models for testing universals

- p can be on any scale. If it is binomial, β represents the difference in q between p=0 (e.g. OV) and p=1 (e.g. VO). All multinomial p with k levels can be 'parametrized' into k-1 paramaters (a.k.a. 'dummy variables)
- Note that p can be a complex vector of predictors ('competing motivations'), including also linguistic areas!
- And it can include interactions p_ip_j, i.e. differences in effects between levels of p_i across levels of p_j.
- The goal of statistical analysis is to find those coefficients that best describe the data (e.g. via Least Squares or Maximum Likelihood Estimation) and then to test whether these coefficients are significantly different from zero, i.e. 'belong to the model'.

- Likely predictors for p-word sizes:
- Structural Pressure, [±stress pattern], as per the hypothesis
- Language Contact *within areas*, e.g. [±Southeast Asia]
- Blind Inheritance within proven families, i.e. groups with a single ancestor, e.g. [±Austroasiatic]
- Model to test: $E(c) = \mu(c)$, thus:

 $\mu(c) = \alpha + \beta_1[FAMILIES] + \beta_2[AREAS] + \beta_3[STRESS] + \beta_4[FAMILIES][STRESS] + \beta_5[AREAS][STRESS]$

• Random Fluctuation = what is left unexplained by the equation, e.g. $100-R^2$.

Plausible areal factors in phonological domains

Southeast Asia (Matisoff 2001, Enfield 2005), South-Southwest Asia (Masica 1976, 2001, Ebert 2001); Europe (Dahl 1990, Haspelmath 2001, Heine & Kuteva 2006)



Plausible family factors in phonological domains

The chosen areas allow similarly-sized family samples, with one representative per sub-branch of major branches in three families (or two if phonologies are known to be diverse and data are sufficient): Austroasiatic (11), Indo-European (12), Sino-Tibetan (17)



Model testing

NB: The sample is not random; therefore we test the model by Conditional (aka (Approximatively) Exact) Inference — specifically, Monte-Carlo permutation of the response — rather than via Random Sampling Theory (Janssen, Bickel & Zúñiga 2006), e.g.



Based on 238 sound patterns in 40 languages, we find:

- no evidence for any interactions between any factors;
- no evidence for an AREA effect (F(2)=.92, p=.51) a significant main effect of STOCK (F(2)=11.03, p<.001)
- a significant main effect of PATTERN TYPE (F(1)=20.99, p<.001)

Model testing

The best-fitting model is

 $\hat{\mu}(c) = .69 - .30[IE vs AA] - 1.4 [ST vs AA] + .26 [STRESS vs OTHER]$



Reliability Analysis

Since there are many less stress-related pw-patterns (19) than others (222), we also need a Reliability Analysis (Janssen et al 2006), replacing critical values of *c* by their grand mean:



- There is evidence for structural pressure leading to phonologies in which stress domains are larger than other domains.
 - There are no other statistical signals, i.e. no other type of phonological pattern seems to favor a specific domain size.
 - No evidence for hierarchical nesting anywhere
- The universal is independent of the effects from language contact (areas) and blind inheritance (families), but
 - blind inheritance also (independently) matters but without interaction!
 - This is in line with known diachronic preferences for structure preservation (Blevins 2004)

- But... how can extend the test to a worldwide database?
- Including the stock factor into regression models makes no sense in worldwide datasets because there are over 300 stocks...
- Need a completely new approach!

- Two observations:
 - 1. Distributions of linguistic structures change over time.
 - 2. Changes over time are described by reconstructed families (by which I exclusively mean units demonstrated by the comparative method).
- Therefore, any universal can be understood as universal pressure for families to develop a specific skewing, as they split over time (Greenberg 1978, 1995, Maslova 2000, Nichols 2003, etc.).

- Therefore, all universals are in fact diachronic in nature!
 - We can reformulate every universal as diachronic pressure, e.g. "VO \rightarrow NRel" can be reformulated as: π (VO&RelN \rightarrow VO&NRel) > π (VO&NRel \rightarrow VO&RelN), where " \rightarrow " symbolizes diachronic change
 - Given this, 'diachronic universals' are just a special case, concerning the explanation of the universal

"OV \rightarrow Np": π (OV&pN \rightarrow OV&Np) > π (OV&Np \rightarrow OV&pN) because N₀V frequently develops into Np (e.g. Nepali $\bar{a}j\bar{a}$ bhane 'today (N₀) saying (V)' > 'as for (P) today (N)')

Towards a new approach: the Skewed Family Method

- We can estimate such pressures for (non-singleton) families by assessing whether or not they show a specific skewing:
 - Expectation if there was no universal pressure at work:
 - if a variable is stable ('blind inheritance'), families end up skewed in whatever way the proto-language happened to be skewed and under whatever other conditions
 - If a variable is instable (random fluctuation, unknown factors), families tend to diversify over time
 - Expectation if there was universal pressure at work:
 - families tend to show the same skewing (the one proposed by the universal), whatever their protolanguage, and regardless of area and random fluctuation

Towards a new approach: the Skewed Family Method

- Therefore, if most (according to some statistical test) families are skewed in the same way regardless of their areal locations and regardless of any structural condition, this attests to universal pressure.
- To what extent is this a valid inference?

- Given universal family skewing in a distribution *D*, assume this to result from blind inheritance within each family.
 - → The current $D(G^0)$ reflects $D(G^{-1})$.
 - → Unless there was universal pressure before G⁻¹, all $D(G^k)$ must reflect $D(G^{k-1})$ until k spans the entire history of the human language faculty.
 - \rightarrow D must be super-stable over deep time.
 - → Changes in D are extremely unlikely within short time intervals.
- Assume that all reconstructible time intervals are short (up to about 8Ky, the age of provable families)
 - \rightarrow Expect to be able to *observe* almost no changes in D.
 - → Almost all *observable* families must be uniform.
- Conclusion: unless this is the reality, it is safe to reject blind inheritance as a cause of universal family skewing.

- What is the highest plausible probability of random change $(A \rightarrow B, B \rightarrow A)$ such that
 - given a significant skewing of a binary variable {A,B} in an observable sample size of 1000 languages (e.g. 40% vs 60%, or 10% vs. 90%),
 - the skewing is still detectable after at least 100 generations of languages (i.e. about 100Ky of history), and
 - the expected numbers of observable changes is significantly smaller than what one observes?

Estimates on plausible probabilities of random change prc

- Assume 130 non-singleton documented families: the largest available database (Dryer 2005 on word order) has 131
- We usually find more than 10 cases of change in 130 families, so $p_{rc} \leq .10$
- But at p_{rc} =.10, the proability of keeping statistical signals is below 5%. Sample simulations:

Sample simulations



Estimates on plausible probabilities of random change prc

- The simulation results suggest that as long as we observe at least a handful of non-uniform families, Blind Inheritance cannot explain family skewing: the probability of finding that many cases just must be much higher than what could keep a distribution stable over many generations.
- This validates the inference from universal family skewing to structural pressure.

Example: Greenberg #1 ("A before O preference")

Data: merged data on the ordering of A and O from Dryer 2005 (WALS) and AUTOTYP, 100% matched coding, total N = 1115, joined to AUTOTYP's critical (conservative) genealogical taxonomy

Raw data for Otomanguean

language	mbranch	ao
Lealao Chinantec	Chinantecan	OA
Chinantec (Comaltepec)	Chinantecan	AO
Chinantec (Quiotepec)	Chinantecan	AO
Chinantec (Palantla)	Chinantecan	AO
Mixtec (Chalcatongo)	Mixtecan	AO
Mixtec (Ocotepec)	Mixtecan	AO
Mixtec (Peñoles)	Mixtecan	AO
Mixtec (Jicaltepec)	Mixtecan	AO
Mixtec (Yosondúa)	Mixtecan	AO
Trique (Copala)	Mixtecan	AO
Otomi (Mezquital)	Otomian	OA
Ocuilteco	Otomian	AO
Chichimec	Pamean	AO
Pame	Pamean	AO
Isthmus Zapotec	Zapotecan	AO
Chatino (Sierra Occidental)	Zapotecan	AO
Chatino (Yaitepec)	Zapotecan	AO
Zapotec (Mitla)	Zapotecan	AO

Within-family skewing, based on a permutationbased χ^2 test (α =.05)

stock	distribution	majority value	diversity
Otomanguean	skewed(trend)	AO	10.89

Example: Greenberg #1 (A before O preference)

- Hypothesis: N (families with A<O skewing) > N (families with no O<A skewing or no skewing at all).
- One plausible areal control factor is the Circum-Pacific Macro-Area (on which 40% WALS variables show a statistical signal at an α -level of .05: Bickel & Nichols 2006):



Example: Greenberg #1 (A before O preference)

• Results:

	СР	Other
Skewed towards AO	67 (<i>e</i> =69)	39 (e=37)
Skewed towards OA	1* (e=1)	0 (e=0)
no skewing in stock	6 (e=4)	0 (e=2)

* Perhaps 0, since the only case is Chumashan, represented why what may have been dialects; i.e. this datapoint might actually represent a singleton stock (Mithun 1999:389)

- No evidence for CP factor, Fisher Exact p=.09
- Significant preference for A<O skewing: χ^2 =86.73, p_{rnd} <.001

Most hypothesized empirical universals have a structural predictor,

- e.g. $\mu(c) = \alpha + \beta_1[STRESS] + \beta_2[FAMILIES]$
- or the class of 'Implicational' or 'Restricted' Universals, e.g. 'VO \rightarrow NRel', i.e.

$$log(\frac{\pi(\mathsf{NRel})}{\pi(\mathsf{RelN})}) = \alpha + \beta_1[\mathsf{VO}] + \beta_2[\mathsf{AREAS}] + \beta_3[\mathsf{FAMILIES}]$$

 where the response is linked by the 'logit' or 'log-odds' function so that β1 represent the change in the logarithmic odds between VO=0 (i.e. OV, if OV is defined as the baseline/reference level) and VO=1 (i.e. VO)

• Therefore, the ratio between the odds themselves is $\theta = e^{\beta 1}$, and $\pi(\text{NRel}) = \frac{e^{\alpha + \beta_1 \text{VO} + \beta_i p_i \dots \beta_k p_k}}{1 + e^{\alpha + \beta_1 \text{VO} + \beta_i p_i \dots \beta_k p_k}}$

Solution: generalizing the Skewed Family Method

 Standard multiple logistic regression, with, e.g., a structural predictor L and a confounding areal factor A:

 $log(\frac{\pi(\text{proposed skewing})}{\pi(\text{opposite} \mid \text{diverse})}) = \alpha + \beta_1 L + \beta_2 A + \beta_3 L \cdot A$

- Example: the odds for families to be skewed towards NRel (vs. not) should be bigger under VO than under OV. (In fact, only Sinitic shows the opposite skewing and only Formosan is diverse. Both reflect the impact of A.)
- Two problems:
 - 1. How to determine the within-family skewing, given L and A?
 - 2. How to estimate $\beta_{1...k}$ and test their significance in the absence of random sampling?

Estimating within-family skewing under given predictors

- Assume a response ~ predictor model $F(q) \sim p$ (as before):
- Assess within-family skewing of the response under each (combination of each) predictor level, e.g.
 - Families skewed towards NRel in VO-Eurasia; in OV-Eurasia; VO-Africa etc.
- What if predictors 'split' a family? Afroasiatic in Africa vs. Eurasia; Sino-Tibetan VO vs OV etc?
 - It is irrelevant for the Skewed Family Method at what taxonomic level a universal exerts its skewing pressure;
 - Therefore, assess skewing at whatever happens to be the lowest unsplit taxonomic level; if there is none, assume predictor-defined pseudo-subgrouping

Example: VP order [±VO] as a predictor

- The predictor is split in the WALS database on Sino-Tibetan at the stock level, but not at the major branch level. Therefore estimate skewing at the major branch level:
 - Sinitic and Karenic (VO) vs. all other major branches (OV)
 - Establish skewing of the response within major branches:

unit	majority.response	distribution	VP
Sinitic (3)	RelN	skewed(absolute)	VO
Karenic (4)	NRel	skewed(absolute)	VO
Angami-Pochuri Group (2)	diverse	diverse	OV
Bodic (8)	RelN	skewed(absolute)	OV
Brahmaputran (4)	diverse	diverse	OV
Kiranti (6)	ReIN	skewed(absolute)	OV
Lolo-Burmese (6)	RelN	skewed(absolute)	OV
Newaric (2)	ReIN	skewed(absolute)	OV
Remnant Himalayish (2)	ReIN	skewed(absolute)	OV
Tani (3)	RelN	skewed(absolute)	OV
West Himalayish (2)	diverse	diverse	OV

Example: VP order [±VO] as a predictor

- Absence of unsplit taxa even at the lowest known level in the WALS database on Meso-Melanesian (< W. Oceanic < Oceanic < E. Malayo-Polynesian < CE Malayo-Polynesian < Austronesian):
 - 2 OV languages (Saliba, Tawala)
 - 1 VO language (Tolai)



 Assume two Meso-Melanesian pseudo-subgroups, but only for the purposes of testing the impact of VP order on some response (e.g. REL position).

- Hawkins 2004: Verb-final languages favor rich case "for reasons of on-line efficiency" ('rich' = distinct coding of agent and patient)
- Nichols 1992, Siewierska 2005, Dryer 1989, 2000, Bickel & Nichols 2007: the distribution of both case and word order is heavily affected by areal patterns:



Rich case

OV vs VO order

- Data on rich case from Comrie 2005 (WALS) and AUTOTYP, 1% mismatches
- Data on word order from Dryer 2005 (WALS) and AUTOTYP, 0% mismatches
- Total datapoints with information on both variables: N = 350
- Stocks with more than one member: N = 51
- Areal confounding factor chosen here for illustration: Eurasia, known to have more case 1.0 than the rest of the world. neutral 0.8 ('neutral' = no markers)0.6 distinguishing A from O)



Distribution of case across families ('trend': skewing with p < .05):

family.name	majority.response	distribution	final	EURASIA	taxonomic.level
Adamawa-Ubangi	neutral	skewed(absolute)	non_final	Other	stock
Arawakan (Maipurean)	neutral	skewed(absolute)	non_final	Other	stock pseudo-group
Arawakan (Maipurean)	neutral	skewed(absolute)	final	Other	stock pseudo-group
Atlantic	neutral	skewed(absolute)	non_final	Other	stock
Awyu-Dumut	neutral	skewed(absolute)	final	Other	stock
Balto-Slavic	non-neutral	skewed(absolute)	non_final	Eurasia	mbranch
Benue-Congo	neutral	skewed(trend)	non_final	Other	stock
Berber	non-neutral	skewed(absolute)	non_final	Other	stock
Cariban	neutral	skewed(absolute)	non_final	Other	stock pseudo-group
Cariban	neutral	skewed(absolute)	final	Other	stock pseudo-group
Central (West Semitic)	diverse	diverse	non_final	Other	sbranch
Central Indo-Aryan	non-neutral	skewed(absolute)	final	Other	ssbranch pseudo-group
Central Indo-Aryan	non-neutral	skewed(absolute)	final	Eurasia	ssbranch pseudo-group
Central Malayo-Polynesian	neutral	skewed(absolute)	non_final	Other	sbranch
Central Pacific	diverse	diverse	non_final	Other	Isbranch
Central Tungusic	non-neutral	skewed(absolute)	final	Other	mbranch
Chadic	neutral	skewed(absolute)	non_final	Other	stock
Chibchan	diverse	diverse	final	Other	stock
Chumashan	neutral	skewed(absolute)	non_final	Other	stock
Cushitic	non-neutral	skewed(absolute)	final	Other	stock
Dravidian	non-neutral	skewed(absolute)	final	Eurasia	stock
Eastern Mon-Khmer	neutral	skewed(absolute)	non_final	Eurasia	mbranch
Eskimo-Aleut	non-neutral	skewed(absolute)	final	Other	stock



• Model to analyze: $log(\frac{\pi(nonneutral)}{\pi(neutral \mid diverse)}) = \alpha + \beta_1 FINAL + \beta_2 EU + \beta_3 FINAL \cdot EU$

- Problem: how to estimate $\beta_{1...3}$ and their probability of being $\neq 0$, given that the data are not randomly sampled?
 - Like before, apply Monte-Carlo permutation tests, i.e. permute the raw response, estimate models by Standard Maximum Likelihood Estimation, and then count how often the Likelihood Ratio (Deviance) is as high as the one in models of the observed data (e.g. Good 1999).
 - Or (for smaller datasets): conditional regression by estimating β_k in a Markov Chain Monte-Carlo sample set with the same margin totals as those defined by the remaining parameters α and β_i , $k \neq i$ (Agresti 2002; Forster et al. 2003; Zamar et al. 2007)

Results:

- MLE: significant interaction term β_3 , LR = -4.15, $p_{(X^2)} = .042$
- Monte-Carlo permutation of LR: p = .054 (10,000 re-samples)
- Best-fitting model:

 $log(\frac{\hat{\pi}(nonneutral)}{\hat{\pi}(neutral \mid diverse)}) = -1.44 + 1.78FINAL + .18EU + 17.05FINAL \cdot EU$

• Factorial Analysis:

• $\beta_1([\pm FINAL]|Eurasia) = 20.81; OR > 10^9; p_{(Fisher Exact)} < .001$

- $\beta_1([\pm FINAL]|Other) = 1.78; OR = 5.88; p_{(Fisher Exact)} = .003$
- Suggesting that the effect is much weaker outside Eurasia, but still significantly different from zero.
- Therefore, a simple additive model fits just as well: $log(\frac{\hat{\pi}(nonneutral)}{\hat{\pi}(neutral \mid diverse)}) = -1.78 + 2.26$ FINAL + 1.21EU
- Comparing this to a saturated model: LR(df=76) = -88.04; p(X2)=.16, suggesting a good fit.

- Overall $OR([\pm FINAL]) = e^{2.26} = 9.59$
- Overall $OR([\pm EURASIA]) = e^{1.21} = 3.46$
- In words: the odds of families to develop a skewing towards languages with rich case marking are almost 10 times higher if the family has V-final order than if it has non-V-final order; and 3.5 times higher in Eurasia than elsewhere.



An additional benefit of the method

- Since area is modelled as a regression factor, factorial analysis is needed only if there is an interaction.
- If there is no interaction, factorial analysis can be missleading.
- Example: take macro-continents as an areal control, in the spirit of Dryer 1989, 2000:



Macrocontinents as areal control factor

Traditional method: test proportions within each area (Fisher Exact Test)



Macrocontinents as areal control factor

A better graph (Meyer et al. 2006, package {vcd} in R)



Macrocontinents as areal control factor

• Results from regression analysis:

Best-fitting model:

 $log(\frac{\hat{\pi}(\text{nonneutral})}{\hat{\pi}(\text{neutral} \mid \text{diverse})}) = -2.33 + 1.94 \text{FINAL} + 1.08 \text{AM} + 1.67 \text{EUR} - .12 \text{NGA}$

• FINAL: LR(1)=13.20, *p*<.001

• MACROCONTINENTS: LR(3)=7.32, p=.07



Extensibility

Do traditionally-sized areas have an impact?



 S Africa N Savannah N Africa Ethiopian Plateau Caucasus-Mesopotamia Ethiopian Plateau Southeast Asia S Africa S	 N Australia Guinea California California NE South Am Alaska-Oregon SE South America
--	---

Controlling for traditionally-sized areas



• Best-fitting model:

 $log(\frac{\hat{\pi}(nonneutral)}{\hat{\pi}(neutral \mid diverse)}) = -19.56 + 3.57FINAL + \hat{\beta}_{A}A + ... + \hat{\beta}_{Y}Y$

- FINAL: LR(1)=17.83, *p* < .001
- AREAS: LR(23)=52.92, p<.003

- If the coefficient of some structural predictor is significantly different from zero, regardless of the levels of A, there is evidence for a universal.
- Such a universal exerts pressure within the time depth of the eldest taxon in the sample.
- If the coefficient is (close to) zero (OR=1) in such a way that N(proposed skewing) ≈ N(opposite skewing) (i.e. very few 'diverse' families): languages 'blindly' inherit the protolanguage structure, i.e. the variable is superstable.
- Then, we can extrapolate this to isolates and construct a sample stripped off all within-family replicas (= Dryer-style 'genus' sample; 'g-sample': Bickel, in press), where each datapoint represents the majority response of an entire family or an isolate.

stock	Language	VP	REL
Adamawa-Ubangi	Day	VO	NRel
Adamawa-Ubangi	Doyayo	VO	NRel
Adamawa-Ubangi	Gbeya Bossangoa	VO	NRel
Adamawa-Ubangi	Linda	VO	NRel
Adamawa-Ubangi	Mbodomo	VO	NRel
Adamawa-Ubangi	Mbum	VO	NRel
Adamawa-Ubangi	Mondunga	VO	NRel
Adamawa-Ubangi	Mumuye	VO	NRel
Adamawa-Ubangi	Nzakara	VO	NRel
Adamawa-Ubangi	Samba Leko	VO	NRel
Adamawa-Ubangi	Sango	VO	NRel

$$\rightarrow$$
 1 g-unit = 1 datapoint

G-sampling

stock	Language	VP	REL
Sino-Tibetan	Bai	VO	RelN
Sino-Tibetan	Cantonese	VO	RelN
Sino-Tibetan	Hakka	VO	RelN
Sino-Tibetan	Mandarin	VO	RelN
Sino-Tibetan	Karen (Bwe)	VO	NRel
Sino-Tibetan	Karen (Pwo)	VO	NRel
Sino-Tibetan	Karen (Sgaw)	VO	NRel
Sino-Tibetan	Kayah Li (Eastern)	VO	NRel
Sino-Tibetan	Achang	OV	RelN
Sino-Tibetan	Akha	OV	RelN
Sino-Tibetan	Apatani	OV	RelN
Sino-Tibetan	Athpare	OV	RelN
Sino-Tibetan	Balti	OV	RelN
Sino-Tibetan	Burmese	OV	RelN
Sino-Tibetan	Byangsi	OV	RelN
Sino-Tibetan	Camling	OV	RelN
Sino-Tibetan	Chantyal	OV	RelN
Sino-Tibetan	Chepang	OV	RelN
Sino-Tibetan	Chin (Siyin)	OV	RelN
Sino-Tibetan	Mishmi (Digaro)	OV	RelN
Sino-Tibetan	Dimasa	OV	RelN
Sino-Tibetan	Gallong	OV	RelN
Sino-Tibetan	Gurung	OV	RelN
Sino-Tibetan	Hani	OV	RelN
Sino-Tibetan	Hayu	OV	RelN
Sino-Tibetan	Jinghpo	OV	RelN
Sino-Tibetan	Khaling	OV	RelN
Sino-Tibetan	Kham	OV	RelN
Sino-Tibetan	Lahu	OV	RelN
Sino-Tibetan	Limbu	OV	RelN
Sino-Tibetan	Maru	OV	RelN

Sino-Tibetan	Meithei (Manipuri)	OV	RelN
Sino-Tibetan	Mising	OV	RelN
Sino-Tibetan	Mao Naga	OV	RelN
Sino-Tibetan	Nar-Phu	OV	RelN
Sino-Tibetan	Newar (Dolakha)	OV	RelN
Sino-Tibetan	Newar (Kathmandu)	OV	RelN
Sino-Tibetan	Nocte	OV	RelN
Sino-Tibetan	Purki	OV	RelN
Sino-Tibetan	Rawang	OV	RelN
Sino-Tibetan	Sikkimese	OV	RelN
Sino-Tibetan	Tamang	OV	RelN
Sino-Tibetan	Thulung	OV	RelN
Sino-Tibetan	Tibetan (Modern Literary)	OV	RelN
Sino-Tibetan	Angami Naga	OV	NRel
Sino-Tibetan	Garo	OV	NRel
Sino-Tibetan	Pattani	OV	NRel

Dryer/WALS: 4 g-units

		Bickel, in press:		
OV	RelN	36	1	(trend in ST)
OV	NRel	3	3	(deviant within OV)
VO	RelN	4	1	(trend in ST)
VO	NRel	4	4	(deviant within VO)

Bickel, in press in Sprachtypologie und Universalienforschung 2008

- If such a g-sample shows an effect, there may be a universal at work at a time depth that is too large to leave a signal in reconstructible taxonomies.
- But, at this time depth, we can't determine whether the distribution is independent of very ancient skewings, precisely because we know that the variables are super-stable (Maslova 2000).

Alternative methods that have been proposed

- Dryer 1989, 2000: use g-samples throughout
 - Problem: no guarantee that the g-sample picks up a stationary distribution rather than some unreconstructible earlier areality or other incidents (Maslova 2000).
- Maslova 2000, 2007: compute universally constant rates of change ('transition probabilities', akin to biological clocks)
 - Problems:
 - there are no constant rates of change in language!
 - unclear how the method works for multifactorial designs
 - computation is based on language pairs, but families are often larger (a sampling problem)

Alternative methods that have been proposed

- However, for all its methodological problems, Dryer-style genus sampling has the distinct advantage that it can include isolates; both Maslova's and my method can be only be applied to families with more than one member.
- And, g-sampling has a very useful, practical cousin: predefined standard samples, e.g. the WALS 200-languages sample.

 Reconciliation: for many hypotheses considered sofar, the methods lead to converging results

Skewed Family Method:



G-sampling Method:



 Results of statistical analysis of the g-sampled data: no evidence for an interaction term, but also no evidence for an area effect, i.e. the best-fitting model is:

 $log(\frac{\hat{\pi}(nonneutral)}{\hat{\pi}(neutral)}) = -.79 + 1.02FINAL; \ \widehat{OR}(FINAL) = 2.77$

- Linguistics can move towards standard methods shared with other sciences by
 - measuring instead of reducing variation
 - interpreting empirical universals in regression models
 - which allow statistical estimation of 'competing factors' from structure, language contact and other domains
 - and can be tested by Monte-Carlo permutation

- However, the most challenging part in this is to control for blind inheritance effects. This can be done through
 - the Skewed Family Method (capturing the dynamics of universals but leaving out isolates) or
 - the G-Sampling Method (presupposing stationary distributions but including isolates)

- What is a 'universal' then?
 - Empirical universal = sign. structural factor in a model
 - Absolute universal = the terms of our metalanguage descriptively needed in every language = that which necessarily follows from the Descriptive A Priori
- What do we urgently need?
 - More fine-grained and more precise systems of analytical variables (inventories of types) that accomodate all sources of variation.
 - More research on families and subgrouping
 - Detailed research on each factor to be entered into a model