FINAL PROJECT REPORT

2009 - 2011

Project Investigator: **Prof. Dr. Balthasar Bickel**, University of Leipzig
(Since April 1, 2011: University of Zürich)

http://www.spw.uzh.ch/bickel
mailto:balthasar.bickel@uzh.ch

Summary

The goal of this project was (i) to deliver theoretical foundations for what I call Distributional Typology, i.e. for statistical research on the dynamics of linguistic diversity, and (ii) to develop methods for exploring this dynamics.

   The key aspects of the theoretical foundations that were developed in the project build in part on earlier insights in the literature (e.g. work by Matthew Dryer): The classical implication-based approach to typology is first shown to be incompatible with the approach that is the standard in many other sciences, where descriptive generalizations are statistical and distinct from the causes that are hypothesized to explain the generalizations. This standard-science approach is found to be superior for exploring the distribution of linguistic structures in the world, for the following reasons. Typical causes that shape cross-linguistic distributions relate to processing preferences, language contact, pragmatic principles etc. Causes of this kind cannot be strictly implicational because they always compete with other factors, most importantly imitation of arbitrary inherited norms — sometimes to the extent that 'non-fitting' languages can survive for generations. The effects of most causes must therefore be probabilistic, and the relevant probabilities must be diachronic in nature, i.e. preferences for bringing and keeping languages in line with the causes operating on them.

   What is needed then is a suitable metalanguage for capturing statistical generalizations and a method for exploring diachronic probabilities. These two issues have been put center stage in the methodological part of the project:

1. Received analytical concepts (e.g. 'subordination') systematically reduce substantial cross-linguistic variation before this variation can be mined for distributional patterns. In response to this problem, Distributional Typology relies instead on large sets of fine-grained descriptive variables. The resulting multivariate typologies, and the multivariate analyses based on them, need no deductive justification because description is distinct from explanation; it is sufficient for variables and analyses to be observationally and psycholinguistically adequate and to be universally applicable.

2. Using empirical studies of large multivariate typological databases and also computer simulations I showed that a distributional bias towards a structure S within a language family (including the limiting case where all languages of a family have the same feature) cannot be plausibly attributed to extreme historical inertia but attest to an inequality stating that the changes towards S are more probable than changes away from S, i.e. the daughter languages of the family develop S if they don't have it, and they keep S stable if they have it. Given this, it is possible to estimate the relevant inequality and therefore the probabilities of change and stability within each family. Applying standard techniques from probability theory, such estimates can also be extended to small families and isolates, within a quantifiable but limited margin of error.

Equipped with the tools of multivariate typology and analysis and a method for estimating probabilities of keeping and changing structures, Distributional Typology makes it possible to formulate statistical models (e.g. regression models) assessing the effects of the causes that one expects to drive diachronic developments, from idiosyncratic drift or inertia within specific families to areal diffusion and universal principles grounded in language processing.

1 Scientific results

The goal of this project was twofold: (i) to deliver and discuss the theoretical foundations needed for a probabilistic approach to linguistic typology; and (ii) to develop statistical methods for evaluating hypotheses on typological distributions. In the following I summarize the main results, based on chapters and articles that are published or accepted for publication. I first discuss two main aspects of the theoretical work and then focus on the methodological work.

1.1 Theoretical foundations: basic approach

Following up on Dryer (1997), I criticize in Bickel (in press-b) the classical approach to typological generalizations which is centered on finding exceptionless laws. This classical approach is deeply grounded in what one may call the *Pāṇinian* technique: seek maximally general statements, try to explain away as many counterexamples as possible by reanalyzing them as falling outside the scope of the statement, and then account for the remaining exceptions by specifying their conditions in terms that are again as general as possible. The paragon of the Pāṇinian technique in typology is the absolute universal, expressed as a material implication such as 'VO order → Prepositions'. Applied to infinite data sets, as is the case for typology, the Pāṇinian technique comes with the same justification problem that ?? identified for linguistic analysis in general and that can arguably only be resolved by assuming deductive models of explanation.

However, given the fact that all linguistic structures arise from diachronic processes and that all these processes are subject to probabilistic effects, it is essential to develop a non-deductive, statistically-based alternative to the Pāṇinian technique. A promising alternative comes from what is the standard methodology that most other disciplines (from particle physics and genetics to sociology and psychology) apply to complex phenomena: formulate statistical predictions from a complex causal theory and test these predictions against a sample of data, under certain assumption of random noise. Under such an approach, causes and their theories can be grounded in any structure that can plausibly be expected to have an impact on how languages develop, from the nature of the human brain and principles of communication to diffusion processes in language contact and to local cultural traditions. In order to determine the effects of any such cause, what is needed is a precise statistical model that estimates the probabilities of diachronic processes and that can be fitted to and tested against empirical samples.

Such models require a linguistic metalanguage in which they can be formulated and techniques of estimating the probabilities of diachronic processes. I take these two issues in the following.

1.2 Theoretical foundations: capturing variation

A linguistic metalanguage for modeling probabilities needs to be descriptively precise and well-operationalized, psycholinguistically and neurolinguistically adequate, and universally applicable. What is not needed in the approach of Distributional Typology is any explanatory

power, because explanations are located in causal theories that predict diachronic probabilities, and not in deductive derivations.

While descriptive precision and psycholinguistic adequacy are tall orders, universal applicability is particularly challenging for typological work because linguistic structures tend to be similar to each other but hardly ever identical in the languages of world. This leads to a basic problem of comparability. In Bickel (2010, 2011) I propose a programmatic solution to the problem: because similarities mean that structures are identical in some but different in other regards, metalanguages are universally fully applicable once they keep all these regards separate, i.e. once all similarities are fully resolved into vectors of individual identity vs. difference statements. What we need, therefore, is metalanguages that decompose analytical terms (e.g. 'subordination') to such an extent that each difference between structures (within and across languages) can be captured precisely. The result is a large sets of very fine-grained variables (e.g. on the scope of specific markers, their positional and extraction possibilities etc.): a **multivariate typology**.

Such typologies can then be mined for associations between variables as well as for cross-linguistically recurrent patterns or areal signals (Bickel et al. 2009, Bickel 2010, Stoll & Bickel in press). At the same time, the individual variables provide a useful tool for analyzing individual structures and languages. Once they are applied to sufficiently large datasets, they form suitable testing grounds for statistical models of causal theories.

1.3   Statistical methods

One of the traditional challenges to statistical modeling in typology has always been the fact that languages are not independent 'trials' like subjects in an experiment but that they form historical chains of dependencies (Maslova 2000). These chains can be successfully captured by demonstrating 'language families', relying on the Comparative Method. The question is how the dependencies can be accounted for in statistical modeling.

To answer this question, I first considered models that treat language family as a factor on a par with area membership and structural predictors. For example, Bickel et al. (2009) explore regression models with phonological domain size as the dependent variable and, as independent variables, families and areas along with the structural type of phonological processes (stress rules vs. other processes). This 'families-as-factors' approach runs into several problems, however. If families are treated as fixed factors, models with families always fit the data better than models without families just because there are more families in the world than structural types: detailed individual-level information like family membership fits data necessarily better than abstract types. As a result, such models tend to underestimate structural factors that still may have a small effect across families. This particular problem can be solved in principle by modeling families not as fixed but as random factors (as recently proposed by Jaeger et al. 2011), but this may not be a valid alternative because like languages, families cannot be randomly sampled, especially not if we also want test area effects (Janssen et al. 2006). Also, in standard random-and-fixed factor models, the fixed factors can freely vary over the random factors (e.g. experimental conditions over subjects). But this condition can often not be met if families are modeled as random factors because some structural types of interest may be limited to some families. Finally, and this

is the biggest problem, the 'family-as-factor' approach treats small and large families alike. However, large families provide more information about diachronic developments than small families; for example, a distributional pattern in a family with over hundred members tells us more about how languages developed over time in this family than what we can possibly learn from a small family with only a handful of members or less.

The general problem behind the 'families-as-factors' approach is the same problem that has also always haunted the many attempts to control for diachronic dependencies by selecting balanced samples, with one representative per family (Bakker 2011): the idea that the only family-based effect that matters for typology is shared inheritance. For example, a dominance of VO structures in a family is taken to result from faithful replication of VO over generations. Under this idea, one clearly must remove this replication effect from the model because otherwise, large families can inflate counts by replication of their proto-structures. However, as Nichols (2003) emphasizes, replication can result not only from 'blind' inheritance ('do what the parents do, whatever it is') but also from the very factors that are of typological interest, e.g. the chances for keeping VO order may be higher if a language is surrounded by VO languages rather than OV languages, or if there is no case marking in the same language.

In response to this, I developed an alternative approach to families, the **Family Bias Method** (starting with a first sketch in Bickel 2008 and then gradually developing it in Bickel 2008, 2011, in press-b; with applications in Bornkessel-Schlesewsky et al. 2008 and Bickel et al. in press). Families are taken not as a confounding factor but as natural laboratories for studying typological change over time. As long as a family is demonstrable by the Comparative Method, it is (by definition) a set of structures that must have developed from a single proto-language. Therefore we can take the observed modern structures within each family as a dataset for estimating developmental trends.

In Bickel (2011, in press-a) I propose that such estimates can rely on a relatively simple method: if there is a synchronic bias towards some structure S as opposed to non-S within a family, this can be interpreted as reflecting an inequality of probabilities of change, $Pr(\text{non-S} \succ \text{S}) > Pr(\text{S} \succ \text{non-S})$ in this family (where the succession symbol stands for diachronic change). In other words, a bias towards S means that if the proto-language had S, there was a high probability of keeping it and so it kept it often; if the proto-language did not have S, there was a high probability of acquiring it and so it was acquired quickly early on (perhaps at the first split of the proto-language), and stayed, or it was acquired later but often (in many branches). Given this, we can estimate probabilities of diachronic processes from synchronic biases, separately for each family, and independently of any assumptions of the nature of the proto-language. These probability estimates can then be used to test and explore specific causes that drove change and non-change across many families: are the probabilities idiosyncratic to individual families, are they influenced by areal diffusion processes, or are they affected by universal preferences?

The key assumption of the Family Bias Method is that a synchronic bias can indeed be taken to reflect an inequality and not an equality of very low probabilities of change, i.e. extreme and functionally 'blind' stability that would go back to the accidental history of the first human languages (Maslova 2000). To substantiate this assumption, I explored

in Bickel (in press-a) for 386 typological variables how small probabilities of change can be assumed to be, given the known number of changes, i.e. the number of observed divergences within families. It turned out that the data can be fitted for all variables only with assumed probabilities of change greater than $Pr(\text{change}) \geq .58$; for some (25%) variables sufficient fits can be reached by $Pr(\text{change}) \geq .10$, but lower values only allow fits of variables that suggest an extreme functional pressure, e.g. variables registering whether or not a language makes an interrogative vs. declarative distinction. It is more plausible to assume that such distinctions result from functionally motivated inequalities favoring the maintenance and development of the distinction rather than from just extreme and functionally 'blind' stability. This is further confirmed in Bickel (in press-a) by a computer simulation study which shows that already after 50 generations of language change, even an extremely strong distributional biases (99% vs. 1%) is no longer detectable already for probabilities of change of $Pr(\text{change}) \leq .01$. Together with Taras Zakharko I am currently preparing further simulation studies that provide additional evidence for the key assumption that synchronic biases can be interpreted as the result of diachronically biased probabilities of change (Zakharko & Bickel 2012).

With this support, the Family Bias Method turns out to be a solid method for estimating diachronic probabilities, at least for sufficiently large families. In Bickel (in press-a) I apply probability-theoretical reasoning to also arrive at estimates for smaller families (including one-member families, i.e. isolates): first, the estimated biases (in any direction) in large families allow estimating the probability of a small family to come from a larger but no longer attested family with vs. without a bias (in any direction). Also, from the large families we can estimate the probability of a given member to reflect the bias in its family as opposed to deviate from this bias as an exception. Together with the actual observations in the small families, we can use these two probabilities to estimate the probable biases in small families. Like in the case of large families, these synchronic biases can be interpreted as reflecting probabilities of diachronic processes.

The Family Bias Method, which is computationally implemented in an R package (Zakharko & Bickel 2011ff), has several advantages over alternative approaches that have recently been proposed: it is the only method that generalizes to data from small families and isolates. Furthermore, unlike the method proposed by Maslova (2004), Maslova & Nikitina (2007), the Family Bias Method allows models for causal theories of any complexity, with any number and kind of variables, including scalar and multinomial variables. (For a sample application to scalar variables, see Bickel et al.'s (in press) evaluation of referential hierarchy effects on the alignment of case markers.) Finally, unlike the method recently adopted by Dunn et al. (2011), the Family Bias Method is also sensitive to inertia/persistence, and not only to change, as possible signals of universal or areal preferences

Once diachronic probability estimates are obtained, they can be statistically modeled in order to evaluate the relative effects of various causal factors of interest. For example, one can formulate statistical models that evaluate the relative explanatory weight of areal diffusion factors as opposed to processing factors leading to harmonic word order patterns. In Bickel (in press-a) I assess the goodness of fit of such models by using exact binomial tests and, for two-factor models, Fisher exact tests. However, a more promising line of research is in terms of Poisson regressions because these generalize to models of variable complexity

and because there are good reasons to assume that linguistic diachrony is best modeled as a Poisson process (Cysouw 2010). Poisson regression models are explored and used in Bickel (2011) and Bickel (in press-b).

1.4 Conclusions

The three pieces of development summarized here give a unified shape to the Distributional Typology framework: Distributional Typology starts with a causal theory on factors that are expected to influence how languages develop over time. Such a theory predicts models of cross-linguistic distributions that can be formulated in terms of descriptively precise variables that are based in a multivariate typology. The resulting models are statistical, not categorical, and they are diachronic, not synchronic. The parameters of these models can be determined on the basis of probabilities of diachronic processes. In turn, these probabilities can be estimated from synchronic distributions using the Family Bias Method.

References (not listed in the tabular report below, Section 2.1):

Bakker, D., 2011. Language sampling. In Song, J. J. (ed.) *The Oxford Handbook of Language Typology*, 90 − 127. Oxford: Oxford University Press.

Chomsky, N., 1986. *Knowledge of Language: Its Nature, Origin, and Use*. Westport, CT: Praeger.

Cysouw, M., 2010. On the probability distribution of typological frequencies. In Ebert, C., G. Jäger, & J. Michaelis (eds.) *The Mathematics of Language*, 29 − 35. Berlin: Springer.

Dryer, M. S., 1997. Why statistical universals are better than absolute universals. *Papers from the 33rd Annual Meeting of the Chicago Linguistic Society* 123 − 145.

Dunn, M. J., S. J. Greenhill, S. C. Levinson, & R. D. Gray, 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473, 79 − 82.

Jaeger, T. F., P. Graff, W. Croft, & D. Pontillo, 2011. Mixed effect models for genetic and aral dependencies in linguistic typology. *Linguistic Typology* 15, 281–320.

Janssen, D., B. Bickel, & F. Zúñiga, 2006. Randomization tests in language typology. *Linguistic Typology* 10, 419 − 440.

Maslova, E., 2000. A dynamic approach to the verification of distributional universals. *Linguistic Typology* 4, 307 − 333.

Maslova, E., 2004. Dinamika tipologičeskix raspredelenij i stabil'nost' jazykovyx tipov. *Vorprosy Jazykoznanija* 2004, 3–16.

Maslova, E. & T. Nikitina, 2007. Stochastic universals and dynamics of cross-linguistic distributions: the case of alignment types. Ms. Stanford University, `http://anothersumma.net/Publications/Ergativity.pdf`.

Nichols, J., 2003. Diversity and stability in language. In Janda, R. D. & B. D. Joseph (eds.) *Handbook of Historical Linguistics*, 283 − 310. London: Blackwell.

2 Publications and presentations

2.1 Project-related publications

In reverse chronological order:

Bickel, B., A. Witzlack-Makarevich, & T. Zakharko, in press. Typological evidence against universal effects of referential scales on case alignment. In Richards, M., A. Malchukov, & I. Bornkessel-Schlesewsky (eds.) *Scales*. Berlin: Mouton de Gruyter.

Bickel, B., in press-b. Distributional typology: statistical inquiries into the dynamics of linguistic diversity. In Heine, B. & H. Narrog (eds.) *The Oxford Handbook of Linguistic Analysis, 2nd edition*. Oxford: Oxford University Press [pre-print available at `http://www.spw.uzh.ch/bickel-files/papers/statistical_typology_Bickel2012.pdf`].

Bickel, B., in press-a. Distributional biases in language families. In Bickel, B., L. A. Grenoble, D. A. Peterson, & A. Timberlake (eds.) *Language typology and historical contingency: studies in honor of Johanna Nichols*. Amsterdam: Benjamins [pre-print available at `http://www.spw.uzh.ch/bickel-files/papers/stability.fsjn.2011bickelrevised.pdf`].

Stoll, S. & B. Bickel, in press. Capturing diversity in language acquisition research. In Bickel, B., L. A. Grenoble, D. A. Peterson, & A. Timberlake (eds.) *Language typology and historical contingency: a festschrift to honor Johanna Nichols*. Amsterdam: Benjamins [pre-print available at `http://www.psycholinguistics.uzh.ch/stoll/publications/stollbickel.sampling2012rev.pdf`].

Bickel, B., 2011. Statistical modeling of language universals. *Linguistic Typology* 15, 401 – 414.

Bickel, B., 2011. Multivariate typology and field linguistics: a case study on detransitivization in Kiranti (Sino-Tibetan). In Austin, P. K., O. Bond, D. Nathan, & L. Marten (eds.) *Proc. Conf. Language Documentation and Linguistic Theory 3*, 3–13. London: SOAS.

Bickel, B., 2010. Capturing particulars and universals in clause linkage: a multivariate analysis. In Bril, I. (ed.) *Clause-hierarchy and clause-linking: the syntax and pragmatics interface*, 51 – 101. Amsterdam: Benjamins.

Bickel, B., 2010. Absolute and statistical universals. In Hogan, P. C. (ed.) *The Cambridge Encyclopedia of the Language Sciences*, 77–79. Cambridge: Cambridge University Press.

Bickel, B. & J. Nichols, 2009. The geography of case. In Malchukov, A. & A. Spencer (eds.) *The Oxford Handbook of Case*, 479 – 493. Oxford: Oxford University Press.

Bickel, B., K. Hildebrandt, & R. Schiering, 2009. The distribution of phonological word domains: a probabilistic typology. In Grijzenhout, J. & B. Kabak (eds.) *Phonological domains: universals and deviations*, 47–75. Berlin: Mouton de Gruyter.

Bickel, B., 2008. A refined sampling procedure for genealogical control. *Language Typology and Universals* 61, 221–233.

Bickel, B., 2008. A general method for the statistical evaluation of typological

distributions. Working paper, University of Leipzig, [http://www.spw.uzh.ch/bickel-files/papers/testing_universals_bickel2008.pdf].

Bornkessel-Schlesewsky, I., K. K. Choudhary, A. Witzlack-Makarevich, & B. Bickel, 2008. Bridging the gap between processing preferences and typological distributions: initial evidence from the online comprehension of control constructions in Hindi. In Malchukov, A. & M. Richards (eds.) *Scales (Linguistische ArbeitsBerichte 86)*, 397 − 436. Leipzig: Institut für Linguistik [http://www.uni-leipzig.de/~asw/lab/lab86/LAB86_Bornkessel_et_al.pdf].

2.2 Project-related software

Zakharko, T. & B. Bickel, 2011ff. `familybias`: Family bias estimation. R package, http://www.spw.uzh.ch/software.html.

2.3 Project-related conference presentations (peer-reviewed or plenary [*])

Zakharko, T. & B. Bickel, 2012. Synchronic and diachronic biases in language families. 45th Annual Meeting of the Societas Linguistica Europaea, Stockholm, August 29 - September 1.

Bierkandt, L., A. Witzlack-Makarevich, T. Zakharko, & B. Bickel, 2011. Synchronic variation and diachronic trends in the alignment of Kiranti agreement. International Workshop on Gradualness in Change and its Relation to Synchronic Variation and Use, Pavia, 30-31 May [slides available at http://www.spw.uzh.ch/witzlack/2011_Pavia_gradualness.pdf].

Bickel, B., A. Witzlack-Makarevich, L. Bierkandt, & T. Zakharko, 2011. Semantic role clustering: an empirical assessments of semantic role types. Workshop on Role Complexes, Zurich, 4-5 April 2011 [slides available at http://www.spw.uzh.ch/witzlack/2011_Zurich_SemanticRoles.pdf].

Bickel, B., 2011b. Putting variation center stage: beyond 'language' (or 'dialect') as the basic data unit in typology (and elsewhere). Keynote lecture, SKY Conference on Variation and Typology, Helsinki, August 25 [slides available at http://www.spw.uzh.ch/bickel/presentations/variation-centered-typology-helsinki2011.pdf].

Bickel, B., 2011a. Multivariate typology and field linguistics: a case study on detransitivization in Kiranti (Sino-Tibetan). Keynote lecture, Conference on Language Documentation and Linguistic Theory 3, London, November 19.

Bickel, B., 2011. Government and agreement: what's where why? Keynote lecture, Conference on Syntactic Government and Subcategorization, Cambridge, England, September 2 [slides available at http://www.spw.uzh.ch/bickel/presentations/cambridge2011.pdf].

Bierkandt, L., A. Witzlack-Makarevich, Z. Taras, & B. Bickel, 2011. Diachronic trends in the alignment of the Kiranti agreement system. 20th International Conference on Historical Linguistics, Osaka, 25-30 July.

Bickel, B., 2010. Hidden diversity. Invited presentation, Internationales Nachwuchskolloquium EMUS, Gesellschaft für Angewandte Linguistik, Leipzig, September 13.

Bickel, B., 2010b. The bias against ergativity: typology, diachrony, and processing. Keynote lecture, Conference on Variation and Change in Argument Realization, Capri, May 28.

Bickel, B., 2010a. Areas and universals. Plenary talk, 43rd Ann. Meeting of the Soc. Ling. Europ., Vilnius, 3-9-2010 [`http://www.spw.uzh.ch/bickel/presentations/areas.and.universals.vilnius2010bickel.pdf`].

Bickel, B., 2009. Typological patterns and hidden diversity. Plenary talk, 8th Bi-Annual Meeting of the Asssociation for Linguistic Typology, July 24 [`http://www.uzh.ch/spw/bickel/presentations/alt2009bickel-plenary.pdf`].

Bickel, B., A. Witzlack-Makarevich, & T. Zakharko, 2009. Is differential case marking caused by a universal referential scale? 8th Biannual Meeting of the Association for Linguistic Typology, Berkeley, July 26 [`http://www.spw.uzh.ch/witzlack/scales@alt8berkeley.pdf`].

Bickel, B., 2009. Affixes and clitics: towards a multivariate typology. The Morphology of the World's Languages, Leipzig, June 26, 2009.

In addition, various aspects of the theory and methods of Distributional Typology were presented in invited workshops colloquia and guest lecture:

2012 Patterns of alignment in verb agreement, *Anna Siewierska Memorial Workshop*, Max-Planck-Institute for Evolutionary Anthropology, April 23 (with Alena Witzlack-Makarevich and Giorgio Iemmolo)

2011 Multivariate Typologie, Guest lecture, University of Bern, December 7

2010a Multivariate Typologie, Guest lecture, University of Cologne, November 9

2010b A preference against ergativity? Evidence from real-time language processing and quantitative typology, *Workshop on the Acquisition of Ergativity*, Leipzig, November 1 (with Ina Bornkessel-Schlesewsky)

2010c Warum sind Ergative unbeliebt? Neurolinguistische und sprachtypologische Perspektiven, *Ehrenkolloquium anlässlich des 80. Geburtstages von Manfred Bierwisch*, Sächsische Akademie der Wissenschaften zu Leipzig, October 1

2010d Linguistic universals and hidden diversity. Guest lecture, University of Marburg, May 7

2009a Typological distributions over time. *Workshop "Visions for Linguistics"*, Konstanz, November 20

2009b Linguistic universals and hidden diversity. Guest lecture, *University of Leiden*, November 3

2009c Explaining the S/A preference in incremental comprehension and diachronic transmission. *Workshop on Polyfunctionality and Underspecification*, Leucorea, Wittenberg, Sepember 1 (with Ina Bornkessel-Schlesewsky and Alena Witzlack-Makarevich)