

Bibliography of
Quantitative Historical Linguistics

Michael Cysouw <cysouw@lmu.de>

- Atkinson, Quentin D., and Russel D. Gray. 2005. Curious parallels and curious connections: phylogenetic thinking in biology and historical linguistics. *Systematic Biology* 54.513-526.
- Atkinson, Quentin D., Andrew Meade, Chris Venditti, Simon J. Greenhill, and Mark Pagel. 2008. Languages Evolve in Punctuational Bursts. *Science* 319.588.
- Atkinson, Quentin D., Geoff Nicholls, David Welch, and Russell D. Gray. 2005. From words to dates: water into wine, mathemagic or phylogenetic inference? *Transactions of the Philological Society* 103.193-219.
- Bergsma, Shane, and Greg Kondrak. 2007. Multilingual Cognate Identification using Integer Linear Programming. *RANLP Workshop on Acquisition and Management of Multilingual Lexicons*, Borovets, Bulgaria:
- Bergsma, Shane, and Grzegorz Kondrak. 2007. Alignment-Based Discriminative String Similarity. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, 656-663. Prague: Association for Computational Linguistics.
- Bhargava, Aditya, and Grzegorz Kondrak. 2009. Multiple Word Alignment with Profile Hidden Markov Models. *Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Student Research Workshop and Doctoral Consortium (NAACL-HLT 2009)*: 43-48. Boulder, Colorado: Association for Computational Linguistics.
- Black, Paul. 1976. Multidimensional scaling applied to linguistic relationships. *Lexicostatistics in Genetic Linguistics: Proceedings of the Montreal Conference, Centre de Recherches Mathematiques, Université de Montreal, May 19-20, 1973*, ed. by Isidore Dyen, and Guy Jucquois, *Cahiers de l'Institute de Linguistique*: 3/5-6, 43-92. Louvain: Peeters.
- Brown, Cecil H., Eric W. Holman, Søren Wichmann, and Viveka Velupillai. 2008. Automated classification of the world's languages: a description of the method and preliminary results. *Sprachtypologie und Universalienforschung* 61.285-308.
- Bryant, David, Flavia Filimon, and Russell D. Gray. 2005. Untangling our past: Languages, trees, splits and networks. *The Evolution of Cultural Diversity: A Phylogenetic Approach*, ed. by Ruth Mace, Clare J. Holden, and Stephan Shennan, 67-84. London: UCL.
- Chrétien, C. Douglas. 1943. *The quantitative method for determining linguistic relationships; interpretation of results and tests of significance*. University of California publications in linguistics. Berkeley: University of California Press.

- Chrétien, C. Douglas. 1962. The Mathematical Models of Glottochronology. *Language* 38.11-37.
- Chrétien, C. Douglas. 1966. Genetic Linguistics and the Probability Model. *Language* 42.518-530.
- Covington, Michael A. 1996. An algorithm to align words for historical comparison. *Computational Linguistics* 22.481-496.
- Covington, Michael A. 1998. Alignment of multiple languages for historical comparison. *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, 275-280.
- Covington, Michael A. 2004. The Number of Distinct Alignments of Two Strings. *Journal of Quantitative Linguistics* 11.173-182.
- Cysouw, Michael, and Hagen Jung. 2007. Cognate identification and alignment using practical orthographies. *Proceedings of Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology* 109-116.
- Cysouw, Michael, Søren Wichmann, and David Kamholz. 2006. A critique of the separation base method for genealogical subgrouping, with data from mixe-zoquean. *Journal of Quantitative Linguistics* 13.225-264.
- Cysouw, Michael & Diana Forker. 2009. Reconstruction of morphosyntactic function: Nonspatial usage of spatial case marking in Tsezic. *Language* 85(3): 588-617.
- Davies, Paul, and Alan S. C. Ross. 1975. Close Relationship in the Uralian Languages. *Finnisch-Ugrische Forschungen* 14.25 - 48.
- Dediu, Dan. 2010. A Bayesian phylogenetic approach to estimating the stability of linguistic features and the genetic biasing of tone. *Proceedings of the Royal Society B: Biological Sciences*
- Dobson, Annette J. 1978. A Stochastic Model of Lexical Change. *Journal of Applied Probability* 15.38-45.
- Dobson, Annette J., Joseph B. Kruskal, David Sankoff, and Leonard J. Savage. 1972. The mathematics of glottochronology revisited. *Anthropological Linguistics* 14.93-113.
- Dolgopolsky, Aron B. 1986. A probabilistic hypothesis concerning the oldest relationships among the language families in northern Eurasia. *Typology, relationship and time: a collection of papers on language change and relationship by Soviet Linguists*, ed. by Vitalii V. Shevoroshkin, and Thomas L. Markey, 27-50. Ann Arbor: Karoma Publishers.
- Dunn, Michael, Angela Terrill, Ger Reesink, RA Foley, and Steve C. Levinson. 2005. Structural phylogenetics and the reconstruction of ancient language history. *Science* 309.2072-2075.
- Dyen, Isidore. 1990. Homomeric lexical classification. *Linguistic Change and Reconstruction Methodology*, ed. by Philip Baldi, Trends in Linguistics: Studies and Monographs: 45, 211-230. Berlin: Mouton de Gruyter.

- Dyen, Isidore. 1990. The homomeric argument for a Slavo-Germanic subgroup of Indo-European. *Linguistic Change and Reconstruction Methodology*, ed. by Philip Baldi, Trends in Linguistics: Studies and Monographs: 45, 385-392. Berlin: Mouton de Gruyter.
- Dyen, Isidore, A. T. James, and J. W. L. Cole. 1967. Language Divergence and Estimated Word Retention Rate. *Language* 43.150-171.
- Dyen, Isidore, Joseph B. Kruskal, and Paul Black. 1992. An Indoeuropean Classification: A Lexicostatistical Experiment. *Transactions of the American Philosophical Society* 82.iii-iv + 1-132.
- Embleton, Sheila. 1985. A New Technique for Dialectometry. *LACUS forum* 12.91 - 98.
- Embleton, Sheila. 1986. *Statistics in Historical Linguistics*. Quantitative Linguistics. Bochum: Brockmeyer.
- Embleton, Sheila. 1991. Mathematical methods of genetic classification. *Sprung from a Common Source: Investigations into the Prehistory of Languages*, ed. by Sydney M. Lamb, and E. Douglas Mitchell, 365-388. Stanford, Calif.: Stanford University Press.
- Goebel, Hans. 1984. *Dialektometrische Studien: anhand italo-romanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und AFL*. Beihefte zur Zeitschrift für Romanische Philologie. Tübingen: Niemeyer.
- Goebel, Hans. 2006. Recent Advances in Salzburg Dialectometry. *Literary and Linguistic Computing* 21.411-435.
- Gray, R. 2005. Evolution. Pushing the time barrier in the quest for language roots. *Science* 309.2007-2008.
- Gray, RD, AJ Drummond, and SJ Greenhill. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323.479-483.
- Gray, Russell D. 2005. Pushing the time barrier in the quest for language roots. *Science* 309.2007-2008.
- Gray, Russell D., and Quentin D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426.435-439.
- Gray, Russell D., and Fiona M. Jordan. 2000. Language trees support the express-train sequence of Austronesian expansion. *Nature* 405.1052-1055.
- Greenberg, Joseph H. 1993. Observations concerning Ringe's "Calculating the Factor of Chance in Language Comparison". *Proceedings of the American Philosophical Society* 137.79-90.
- Greenhill, Simon J, Robert Blust, and Russell D. Gray. 2008. The austronesian basic vocabulary database: from bioinformatics to lexomics. *Evol Bioinform Online* 4.271-283.
- Hartman, Steven Lee. 1981. A universal alphabet for experiments in comparative phonology. *Computers and the Humanities* 15.75-82.
- Hartman, Steven Lee. 1994. Commentary on Lowe and Mazaudon. *Comput. Linguist.* 20.418-418.

- Heeringa, Wilbert. 2004. Measuring dialect pronunciation differences using Levenshtein distance. Ph.D. Thesis Rijksuniversiteit Groningen.
- Heeringa, Wilbert, Keith Johnson, and Charlotte Gooskens. 2005. Measuring Norwegian Dialect Distances using Acoustic Features. *UC Berkeley Phonology Lab Annual Report*, 312-336. University of California at Berkeley.
- Heeringa, Wilbert, Peter Kleiweg, Charlotte Gooskens, and John Nerbonne. 2006. Evaluation of String Distance Algorithms for Dialectology. *Proceedings of the Workshop on Linguistic Distances*, 51-62. Sydney, Australia Association for Computational Linguistics.
- Heggarty, Paul. 2000. Quantification and comparison in language structure: An exploration of new methodologies. University of Cambridge.
- Heggarty, Paul. 2000. Quantifying change over time in phonetics. *Time Depth in Historical Linguistics*, ed. by Colin Renfrew, April McMahon, and Larry Trask, Papers in the Prehistory of Languages: 2, 531-562. Oxford: McDonald Institute for Archaeological Research.
- Hewson, John. 1973. Reconstructing prehistoric languages on the computer: The triumph of the electronic neogrammarian. *Proceedings of the 5th conference on Computational linguistics-Volume 1*: 263-273. Association for Computational Linguistics.
- Hewson, John. 1989. Computer-aided research in comparative and historical linguistics. *Computational Linguistics*, ed. by Istvan S. Batori, Winfried Lenders, and Wolfgang Putschke, 576-580. Berlin: Walter de Gruyter.
- Hewson, John. 1993. *A computer-generated dictionary of proto-Algonquian*. Quebec: Canadian Museum of Civilization.
- Hewson, John. 1994. Commentary on Lowe and Mazaudon. *Comput. Linguist.* 20.419-419.
- Holden, Clare Janaki. 2002. Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum-parsimony analysis. *Proceedings of the Royal Society B: Biological Sciences* 269.793-799.
- Holm, Hans J. 2000. Genealogy of the main Indo-European branches applying the Separation Base Method. *Journal of Quantitative Linguistics* 7.73-95.
- Holm, Hans J. 2003. The proportionality trap. Or: what is wrong with lexicostatistical subgrouping? *Indogermanische Forschungen* 108.39-47.
- Holm, Hans J. 2007. The new arboretum of Indo-European "trees"; Can new algorithms reveal the phylogeny and even prehistory of Indo-European?*. *Journal of Quantitative Linguistics* 14.167-214.
- Holman, Eric W. 1996. Quantitative properties of the evolution and classification of languages. *Journal of Classification* 13.27-56.
- Holman, Eric W., Søren Wichmann, Cecil H. Brown, Viveka Velupillai, André Müller, and Dik Bakker. 2008. Explorations in automated language classification. *Folia Linguistica*

42.331-354.

- Justeson, John S., and Laurence D. Stephens. 1979. Change cognation: A probabilistic model and decision procedure for historical inference. *Papers from the 4th International Conference on Historical Linguistics*, ed. by Elizabeth Closs Traugott, Rebecca Labrum, and Susan Shephard, *Current Issues in Linguistic Theory*: 14, 37-45. Amsterdam: Benjamins.
- Kessler, Brett. 2001. *The Significance of Word Lists*. Dissertations in Linguistics. Stanford, Calif.: CSLI Publications.
- Kessler, Brett. 2005. Phonetic comparison algorithms. *Transactions of the Philological Society* 103.243-260.
- Kessler, Brett. 2007. Word similarity metrics and multilateral comparison. *Proceedings of Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology*, 6-14. Stroudsburg, PA: Association for Computational Linguistics.
- Kessler, Brett, and A. Lehtonen. 2006. Multilateral comparison and significance testing of the Indo-Uralic question. *Phylogenetic Methods and the Prehistory of Languages*, ed. by Peter Forster, and Colin Renfrew, McDonald Institute for Archaeological Research, Cambridge, England: 33-42.
- Kitchen, A, C Ehret, S Assefa, and CJ Mulligan. 2009. Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East. *Proc Biol Sci* 276.2703-2710.
- Kondrak, Grzegorz. 2000. A new algorithm for the alignment of phonetic sequences. Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference: 288-295. Morgan Kaufmann Publishers Inc.
- Kondrak, Grzegorz. 2002. Algorithms for Language Reconstruction. Toronto: Ph.D Thesis University of Toronto.
- Kondrak, Grzegorz. 2003. Phonetic Alignment and Similarity. *Computers and the Humanities* 37.273-291.
- Kondrak, Grzegorz. 2009. Identification of Cognates and Recurrent Sound Correspondences in Word Lists. *Traitement automatique des langues et langues anciennes* 50.201-235.
- Kondrak, Grzegorz, and Tarek Sherif. 2006. Evaluation of Several Phonetic Similarity Algorithms on the Task of Cognate Identification. *Proceedings of the COLING-ACL Workshop on Linguistic Distances*, 43-50. Sydney:
- Kroeber, A. L. 1960. Statistics, Indo-European, and Taxonomy. *Language* 36.1-21.
- Kroeber, A. L., and C. Douglas Chrétien. 1937. Quantitative Classification of Indo-European Languages. *Language* 13.83-103.
- Kroeber, A. L., and C. Douglas Chrétien. 1939. The Statistical Technique and Hittite. *Language* 15.69-71.
- Lees, Robert B. 1953. The Basis of Glottochronology. *Language* 29.113-127.

- Lees, Robert B. 1980. Language and the Genetic Code. *The Signifying Animal*, ed. by Irmengard Rauch, and Gerald F. Carr, 218-226. Bloomington, IN: Indiana University Press.
- Lohr, Marisa. 1999. Methods for the Genetic Classification of Languages. Ph.D. Thesis University of Cambridge.
- Lowe, John Brandon. 1995. Cross-linguistic lexicographic databases for etymological research, with examples from Sino-Tibetan and Bantu languages. United States -- California: University of California, Berkeley.
- Lowe, John Brandon, and Martine Mazaudon. 1994. The reconstruction engine: a computer implementation of the comparative method. *Computational Linguistics* 20.381-417.
- Mackay, Wesley, and Gregorz Kondrak. 2005. Computing Word Similarity and Identifying Cognates with Pair Hidden Markov Models. *Proceedings of the Ninth Conference on Computational Natural Language Learning*, 40-47. Ann Arbor, MI:
- McMahon, April. 2005. Introduction. *Transactions of the Philological Society* 103.113-119.
- McMahon, April, Paul Heggarty, Robert McMahon, and Natalia Slaska. 2005. Swadesh Sublists and the benefits of borrowing: an Andean case study¹. *Transactions of the Philological Society* 103.147-170.
- McMahon, April, and Robert McMahon. 2003. Finding families: quantitative methods in language classification. *Transactions of the Philological Society* 101.7-55.
- McMahon, April M. S, and Robert McMahon. 2005. *Language classification by numbers*. Oxford : New York: Oxford University Press.
- Merwe, Nikolaas J. van der, Olga Akhmanova, A. Capell, C. Douglas Chrétien, Harold E. Driver, Isidore Dyen, Roberto Escalante, Sarah C. Gudschinsky, Jr, Robert A. Hall, Shiro Hattori, Andre G. Haudricourt, Dell Hymes, Hisanosuke Izui, Jr, William H. Jacobsen, Wilhelm Milke, Alfred G. Smith, Morris Swadesh, George L. Trager, and Randal L. Whitman. 1966. New Mathematics for Glottochronology [and Comments and Reply]. *Current Anthropology* 7.485-500.
- Muzaffar, Towhid Bin. 1997. Computer Simulation of Shawnee Historical Phonology. Corner Brook: Master Thesis Memorial University of Newfoundland.
- Nakhleh, Luay, Donald A. Ringe, and Tandy Warnow. 2005. Perfect Phylogenetic Networks: A New Methodology for Reconstructing the Evolutionary History of Natural Languages. *Language* 81.382-420.
- Nakhleh, Luay, Tandy Warnow, Don Ringe, and Steven N. Evans. 2005. A comparison of phylogenetic reconstruction methods on an Indo-European dataset. *Transactions of the Philological Society* 103.171-192.
- Nerbonne, John, Mark T. Ellison, and Grzegorz Kondrak. 2007. Computing and Historical Phonology. *Proceedings of Ninth Meeting of the ACL Special Interest Group in*

- Computational Morphology and Phonology*, 1-5. Prague, Czech Republic: Association for Computational Linguistics.
- Nerbonne, John, Wilbert Heeringa, and Peter Kleiweg. 1999. Recent Advances in Salzburg Dialectometry. *Time warps, string edits, and macromolecules*, ed. by David Sankoff, and Joseph Kruskal, David Hume Series: v-xv. Prague: Mouton.
- Pagel, Mark. 2000. Maximum-likelihood models for glottochronology and for reconstructing linguistic phylogenies. *Time depth in historical linguistics*, ed. by Colin Renfrew, April MacMahon, and Larry Trask, 189-207. Cambridge: McDonald Institute for Archaeological Research.
- Pagel, Mark, Quentin D. Atkinson, and Andrew Meade. 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449.717-720.
- Petroni, Filippo, and Maurizio Serva. 2008. Language distance and tree reconstruction. *Journal of Statistical Mechanics: theory and experiment* 2008.P08012.
- Platnick, Norman I., and H. Don Cameron. 1977. Cladistic Methods in Textual, Linguistic, and Phylogenetic Analysis. *Systematic Zoology* 26.380-385.
- Prokić, Jelena. 2010. Families and Resemblances. Groningen: Ph.D. Thesis Rijksuniversiteit Groningen.
- Prokić, Jelena, Martijn Wieling, and John Nerbonne. 2009. Multiple sequence alignments in linguistics. Proceedings of the EACL 2009 Workshop on Language Technology and Resources for Cultural Heritage, Social Sciences, Humanities, and Education (LaTeCH-SHELT&R 2009): 18-25. Association for Computational Linguistics.
- Renfrew, Colin. 2002. The emerging synthesis: the archeogenetics of farming/language dispersals and other spread zones. *Examining the Farming/Language Dispersal Hypothesis*, ed. by Peter Bellwood, and Colin Renfrew, McDonald Institute Monographs: 3-16. Cambridge: McDonald Institute for Archaeological Research.
- Renfrew, Colin. 2009. Where Bacteria and Languages Concur. *Science Science* 323.467-468.
- Renfrew, Colin, April MacMahon, and Larry Trask, eds. 2000. *Time depth in historical linguistics*. Cambridge: McDonald Institute for Archaeological Research.
- Rexová, Kateřina, Yvonne Bastin, and Daniel Frynta. 2006. Cladistic analysis of Bantu languages: a new tree based on combined lexical and grammatical data. *Naturwissenschaften* 93.189-194.
- Rexová, Kateřina, Daniel Frynta, and Jan Zrzavý. 2002. Cladistic analysis of languages: Indo-European classification based on lexicostatistical data. *Cladistics* 19.120-127.
- Ringe, Donald. 1992. *On Calculating the Factor of Chance in Language Comparison*. Transactions of the American Philosophical Society. Philadelphia: American Philosophical Society.

- Ringe, Donald, Tandy Warnow, and Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100.59-129.
- Ringe, Donald, Tandy Warnow, Ann Taylor, and Alexander Michailov. 1998. Computational cladistics and the position of Tocharian. *The Bronze Age and Early Iron Age Peoples of Eastern Central Asia*, ed. by Victor H. Mair, Journal of Indo-European Studies Monograph: 1: Archeology, Migration and Nomadism, Linguistics 26, 391-414. Washington, DC: Institute for the Study of Man.
- Ringe, Donald A. 1993. A Reply to Professor Greenberg. *Proceedings of the American Philosophical Society* 137.91-109.
- Rødseth, Øystein J., and James A. Sellers. 2006. Improving calculations of the number of distinct alignments of two strings*. *Journal of Quantitative Linguistics* 13.45-55.
- Ross, Alan S. C. 1950. Philological Probability Problems. *Journal of the Royal Statistical Society. Series B (Methodological)* 12.19-59.
- Sankoff, David. 1970. On the Rate of Replacement of Word-Meaning Relationships. *Language* 46.564-569.
- Sankoff, David. 1973. Mathematical developments in lexicostatistical theory. *Diachronic, areal and typological linguistics*, ed. by Henry M. Hoenigswald, Current trends in linguistics: 11, 93-113. The Hague: Mouton.
- Séguy, Jean. 1971. La dialectométrie dans l'Atlas linguistique de la Gascogne. *Revue de linguistique romane* 37.1-24.
- Serva, Maurizio, and Fillippo Petroni. 2008. Indo-European languages tree by Levenshtein distance. *EPL (Europhysics Letters)* 81.68005.
- Slaska, Natalia. 2005. Lexicostatistics away from the armchair: handling people, props and problems. *Transactions of the Philological Society* 103.221-242.
- Sokal, Robert R. 1988. Genetic, geographic, and linguistic distances in Europe. *Proceedings of the National Academy of Sciences of the United States of America* 85.1722-1726.
- Stevick, Robert D. 1963. The Biological Model and Historical Linguistics. *Language* 39.159-169.
- Swadesh, Morris. 1950. Salish internal relationships. *International Journal of American Linguistics* 16.157-167.
- Swadesh, Morris. 1952. Lexico-Statistic Dating of Prehistoric Ethnic Contacts: With Special Reference to North American Indians and Eskimos. *Proceedings of the American Philosophical Society* 96.452-463.
- Wang, William S-Y., and James W. Minett. 2005. The Invasion of Language: emergence, change and death. *Trends in Ecology and Evolution* 20
- Wang, William S-Y., and James W. Minett. 2005. Vertical and horizontal transmission in language evolution. *Transactions of the Philological Society* 103.121-146.
- Wang, William S-Y., and Feng Wang. 2004. Basic words in language evolution. 1 - 21.

- Warnow, Tandy. 1997. Mathematical approaches to comparative linguistics. *Proceedings of the National Academy of Sciences of the United States of America* 94.6585-6590.
- Warnow, Tandy, Donald Ringe, and Ann Taylor. 1996. Reconstruction the evolutionary history of natural languages. *Proceedings of the seventh annual ACM-SIAM Symposium on Discrete Algorithms*, 314-322. New York: Association for Computing Machinery.
- Weng, Zili, and Robert R. Sokal. 1995. Origins of Indo-Europeans and the Spread of Agriculture in Europe: Comparison of Lexicostatistical and Genetic Evidence. *Human Biology* 67.577 - 594.
- Whitfield, John. 2008. Across the Curious Parallel of Language and Species Evolution. *PLoS Biol PLoS Biol* 6.e186.