

Bibliography of
Quantitative approaches to lexical comparison

Michael Cysouw <cysouw@eva.mpg.de>

- Atkinson, Quentin D., Geoff Nicholls, David Welch & Russell D. Gray (2005) 'From words to dates: water into wine, mathemagic or phylogenetic inference?' *Transactions of the Philological Society* 103(2): 193-219.
- Atkinson, Quentin D. & Russel D. Gray (2005) 'Curious parallels and curious connections: phylogenetic thinking in biology and historical linguistics.' *Systematic Biology* 54(4): 513-526.
- Atkinson, Quentin D., Andrew Meade, Chris Venditti, Simon J. Greenhill & Mark Pagel (2008) 'Languages Evolve in Punctuational Bursts' *Science* 319(5863): 588.
- Bergsma, Shane & Grzegorz Kondrak (2007a) 'Alignment-Based Discriminative String Similarity.' *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*. Prague: Association for Computational Linguistics, 656-663.
- Bergsma, Shane & Greg Kondrak (2007b) 'Multilingual Cognate Identification using Integer Linear Programming.' *RANLP Workshop on Acquisition and Management of Multilingual Lexicons*. Borovets, Bulgaria:
- Black, Paul (1976) 'Multidimensional scaling applied to linguistic relationships' in: Isidore Dyen & Guy Jucquois (eds.) *Lexicostatistics in Genetic Linguistics: Proceedings of the Montreal Conference, Centre de Recherches Mathematiques, Université de Montreal, May 19-20, 1973*. (Cahiers de l'Institute de Linguistique, 3/5-6) Louvain: Peeters, 43-92.
- 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(4): 285-308.
- Chrétien, C. Douglas (1943) *The quantitative method for determining linguistic relationships; interpretation of results and tests of significance*, (University of California publications in linguistics, 1.2) Berkeley: University of California Press, 11-19.
- (1962) 'The Mathematical Models of Glottochronology' *Language* 38(1): 11-37.

- Covington, Michael A. (1996) 'An algorithm to align words for historical comparison' *Computational Linguistics* 22(4): 481-496.
- (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.
- (2004) 'The Number of Distinct Alignments of Two Strings' *Journal of Quantitative Linguistics* 11(3): 173-182.
- Cysouw, Michael, Søren Wichmann & David Kamholz (2006) 'A critique of the separation base method for genealogical subgrouping, with data from mixe-zoquean' *Journal of Quantitative Linguistics* 13(2): 225-264.
- Davies, Paul & Alan S. C. Ross (1975) 'Close Relationship in the Uralian Languages' *Finnisch-Ugrische Forschungen* 14(1-3): 25 - 48.
- Dobson, Annette J., Joseph B. Kruskal, David Sankoff & Leonard J. Savage (1972) 'The mathematics of glottochronology revisited' *Anthropological Linguistics* 14: 93-113.
- Dobson, Annette J. (1978) 'A Stochastic Model of Lexical Change' *Journal of Applied Probability* 15(1): 38-45.
- Dolgopolsky, Aron B. (1986) 'A probabilistic hypothesis concerning the oldest relationships among the language families in northern Eurasia' in: Vitalii V. Shevoroshkin & Thomas L. Markey (eds.) *Typology, relationship and time: a collection of papers on language change and relationship by Soviet Linguists*. Ann Arbor: Karoma Publishers, 27-50.
- Dyen, Isidore, A. T. James & J. W. L. Cole (1967) 'Language Divergence and Estimated Word Retention Rate' *Language* 43(1): 150-171.
- Dyen, Isidore, Joseph B. Kruskal & Paul Black (1992) 'An Indoeuropean Classification: A Lexicostatistical Experiment' *Transactions of the American Philosophical Society* 82(5): iii-iv + 1-132.

- Embleton, Sheila (1985) 'A New Technique for Dialectometry' *LACUS forum* 12(2): 91 - 98.
- (1986) *Statistics in Historical Linguistics*, (Quantitative Linguistics, 30) Bochum: Brockmeyer.
- (1991) 'Mathematical methods of genetic classification' in: Sydney M. Lamb & E. Douglas Mitchell (eds.) *Sprung from a Common Source: Investigations into the Prehistory of Languages*. Stanford, Calif.: Stanford University Press, 365-388.
- Goebel, Hans (1984) *Dialektometrische Studien: anhand italoromanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und AFL*, (Beihefte zur Zeitschrift für Romanische Philologie, 191) Tübingen: Niemeyer.
- (2006) 'Recent Advances in Salzburg Dialectometry' *Literary and Linguistic Computing* 21(4): 411-435.
- Gray, Russell D. & Fiona M. Jordan (2000) 'Language trees support the express-train sequence of Austronesian expansion' *Nature* 405: 1052-1055.
- Gray, Russell D. & Quentin D. Atkinson (2003) 'Language-tree divergence times support the Anatolian theory of Indo-European origin' *Nature* 426: 435-439.
- Gray, Russell D. (2005) 'Pushing the time barrier in the quest for language roots' *Science* 309: 2007-2008.
- Greenberg, Joseph H. (1993) 'Observations concerning Ringe's "Calculating the Factor of Chance in Language Comparison"' *Proceedings of the American Philosophical Society* 137(1): 79-90.
- Heeringa, Wilbert (2004), 'Measuring dialect pronunciation differences using Levenshtein distance', (Rijksuniversiteit Groningen).
- Heeringa, Wilbert, Keith Johnson, & Charlotte Gooskens (2005) 'Measuring Norwegian Dialect Distances using Acoustic Features.' *UC Berkeley Phonology Lab Annual Report*. University of California at Berkeley, 312-336.
- Heeringa, Wilbert, Peter Kleiweg, Charlotte Gooskens, & John Nerbonne (2006) 'Evaluation of String Distance Algorithms for Dialectology.' *Proceedings of the Workshop on Linguistic Distances*. Sydney, Australia Association for Computational Linguistics, 51-62.

- Heggarty, Paul (2000a) 'Quantifying change over time in phonetics' in: Colin Renfrew, April McMahon, & Larry Trask (eds.) *Time Depth in Historical Linguistics*. (Papers in the Prehistory of Languages, 2) Oxford: McDonald Institute for Archaeological Research, 531-562.
- (2000b), 'Quantification and comparison in language structure: An exploration of new methodologies', PhD Thesis (University of Cambridge).
- 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(1493): 793-799.
- Holm, Hans J. (2000) 'Genealogy of the main Indo-European branches applying the Separation Base Method' *Journal of Quantitative Linguistics* 7(2): 73-95.
- (2003) 'The proportionality trap. Or: what is wrong with lexicostatistical subgrouping?' *Indogermanische Forschungen* 108: 39-47.
- (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(2): 167-214.
- Holman, Eric W., Søren Wichmann, Cecil H. Brown, Viveka Velupillai, André Müller, and Dik Bakker. 2008. Explorations in automated language classification. *Folia Linguistica* 42(3-4): 331-354.
- Justeson, John S. & Laurence D. Stephens (1979) 'Change cognation: A probabilistic model and decision procedure for historical inference' in: Elizabeth Closs Traugott, Rebecca Labrum, & Susan Shephard (eds.) *Papers from the 4th International Conference on Historical Linguistics*. (Current Issues in Linguistic Theory, 14) Amsterdam: Benjamins, 37-45.
- Kessler, Brett (2001) *The Significance of Word Lists*, (Dissertations in Linguistics, Stanford, Calif.: CSLI Publications.
- (2005) 'Phonetic comparison algorithms' *Transactions of the Philological Society* 103(2): 243-260.
- Kessler, Brett & A. Lehtonen (2006) 'Multilateral comparison and significance testing of the Indo-Uralic question' in: Peter Forster & Colin Renfrew (eds.) *Phylogenetic Methods and the Prehistory of Languages*. (McDonald Institute for Archaeological Research, Cambridge, England, 33-42.

- Kessler, Brett (2007) 'Word similarity metrics and multilateral comparison' (eds.) *Proceedings of Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology*. Stroudsburg, PA: Association for Computational Linguistics, 6-14.
- Kondrak, Gregorz (2002), 'Algorithms for Language Reconstruction',
——— (2003) 'Phonetic Alignment and Similarity' *Computers and the Humanities* 37(3): 273-291.
- Kondrak, Gregorz & Tarek Sherif (2006) 'Evaluation of Several Phonetic Similarity Algorithms on the Task of Cognate Identification.' *Proceedings of the COLING-ACL Workshop on Linguistic Distances*. Sydney: 43-50.
- Kroeber, A. L. & C. Douglas Chrétien (1937) 'Quantitative Classification of Indo-European Languages' *Language* 13(2): 83-103.
——— (1939) 'The Statistical Technique and Hittite' *Language* 15(2): 69-71.
- Kroeber, A. L. (1960) 'Statistics, Indo-European, and Taxonomy' *Language* 36(1, Part 1): 1-21.
- Lees, Robert B. (1953) 'The Basis of Glottochronology' *Language* 29(2): 113-127.
- Lohr, Marisa (1999), 'Methods for the Genetic Classification of Languages', (University of Cambridge).
- Mackay, Wesley & Gregorz Kondrak (2005) 'Computing Word Similarity and Identifying Cognates with Pair Hidden Markov Models.' *Proceedings of the Ninth Conference on Computational Natural Language Learning*. Ann Arbor, MI: 40-47.
- McMahon, April & Robert McMahon (2003) 'Finding families: quantitative methods in language classification' *Transactions of the Philological Society* 101(1): 7-55.
- McMahon, April (2005) 'Introduction' *Transactions of the Philological Society* 103(2): 113-119.
- McMahon, April, Paul Heggarty, Robert McMahon & Natalia Slaska (2005) 'Swadesh Sublists and the benefits of borrowing: an Andean case study¹' *Transactions of the Philological Society* 103(2): 147-170.
- McMahon, April M. S and Robert McMahon (2005) *Language classification by numbers*, (Oxford linguistics) Oxford : New York: Oxford University Press, xvi, 265 p.

- 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 & Randal L. Whitman (1966) 'New Mathematics for Glottochronology [and Comments and Reply]' *Current Anthropology* 7(4): 485-500.
- Nakhleh, Luay, Tandy Warnow, Don Ringe & Steven N. Evans (2005a) 'A comparison of phylogenetic reconstruction methods on an Indo-European dataset' *Transactions of the Philological Society* 103(2): 171-192.
- Nakhleh, Luay, Donald A Ringe & Tandy Warnow (2005b) 'Perfect Phylogenetic Networks: A New Methodology for Reconstructing the Evolutionary History of Natural Languages' *Language* 81: p. 382-420.
- Nerbonne, John, Wilbert Heeringa, & Peter Kleiweg (1999) 'Recent Advances in Salzburg Dialectometry' in: David Sankoff & Joseph Kruskal (eds.) *Time warps, string edits, and macromolecules*. (David Hume Series, Prague: Mouton, v-xv.
- Nerbonne, John, Mark T. Ellison, & Grzegorz Kondrak (2007) 'Computing and Historical Phonology.' *Proceedings of Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology*. Prague, Czech Republic: Association for Computational Linguistics, 1-5.
- Pagel, Mark (2000) 'Maximum-likelihood models for glottochronology and for reconstructing linguistic phylogenies' in: Colin Renfrew, April MacMahon, & Larry Trask (eds.) *Time depth in historical linguistics*. Cambridge: McDonald Institute for Archaeological Research, 189-207.
- Pagel, Mark, Quentin D. Atkinson & Andrew Meade (2007) 'Frequency of word-use predicts rates of lexical evolution throughout Indo-European history' *Nature* 449(7163): 717-720.
- Platnick, Norman I. & H. Don Cameron (1977) 'Cladistic Methods in Textual, Linguistic, and Phylogenetic Analysis' *Systematic Zoology* 26(4): 380-385.
- Renfrew, Colin, April MacMahon, and Larry Trask (eds.) (2000) *Time depth in historical linguistics*, Cambridge: McDonald Institute for Archaeological Research)

- Rexová, Kateřina, Daniel Frynta & Jan Zrzavý (2002) 'Cladistic analysis of languages: Indo-European classification based on lexicostatistical data' *Cladistics* 19: 120-127.
- Rexová, Kateřina, Yvonne Bastin & Daniel Frynta (2006) 'Cladistic analysis of Bantu languages: a new tree based on combined lexical and grammatical data' *Naturwissenschaften* 93(4): 189-194.
- Ringe, Donald (1992) *On Calculating the Factor of Chance in Language Comparison*, (Transactions of the American Philosophical Society, 82:1) Philadelphia: American Philosophical Society.
- Ringe, Donald, Tandy Warnow, Ann Taylor, & Alexander Michailov (1998) 'Computational cladistics and the position of Tocharian' in: Victor H. Mair (eds.) *The Bronze Age and Early Iron Age Peoples of Eastern Central Asia*. (Journal of Indo-European Studies Monograph, 1: Archeology, Migration and Nomadism, Linguistics 26) Washington, DC: Institute for the Study of Man, 391-414.
- Ringe, Donald, Tandy Warnow & Ann Taylor (2002) 'Indo-European and computational cladistics' *Transactions of the Philological Society* 100(1): 59-129.
- Ringe, Donald A. (1993) 'A Reply to Professor Greenberg' *Proceedings of the American Philosophical Society* 137(1): 91-109.
- Ross, Alan S. C. (1950) 'Philological Probability Problems' *Journal of the Royal Statistical Society. Series B (Methodological)* 12(1): 19-59.
- Rødseth, Øystein J. & James A. Sellers (2006) 'Improving calculations of the number of distinct alignments of two strings*' *Journal of Quantitative Linguistics* 13(1): 45-55.
- Sankoff, David (1970) 'On the Rate of Replacement of Word-Meaning Relationships' *Language* 46(3): 564-569.
- (1973) 'Mathematical developments in lexicostatistical theory' in: Henry M. Hoenigswald (eds.) *Diachronic, areal and typological linguistics*. (Current trends in linguistics, 11) The Hague: Mouton, 93-113.
- Séguy, Jean (1971) 'La dialectométrie dans l'Atlas linguistique de la Gascogne' *Revue de linguistique romane* 37: 1-24.
- Slaska, Natalia (2005) 'Lexicostatistics away from the armchair: handling people, props and problems' *Transactions of the Philological Society* 103(2): 221-242.

- 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(4): 452-463.
- Wang, William S-Y & James W. Minett (2005) 'Vertical and horizontal transmission in language evolution' *Transactions of the Philological Society* 103(2): 121-146.
- Warnow, Tandy, Donald Ringe, & Ann Taylor (1996) 'Reconstruction the evolutionary history of natural languages' (eds.) *Proceedings of the seventh annual ACM-SIAM Symposium on Discrete Algorithms*. New York: Association for Computing Machinery, 314-322.
- Warnow, Tandy (1997) 'Mathematical approaches to comparative linguistics' *Proceedings of the National Academy of Sciences of the United States of America* 94(13): 6585-6590.