Inheritance and Diffusion of Language and Culture: A Comparative Perspective

Eric W. Holman

University of California, Los Angeles

Søren Wichmann

Max Planck Institute for Evolutionary Anthropology and Leiden University

Cecil H. Brown

Northern Illinois University

ABSTRACT Data in the *World Atlas of Language Structures (WALS)* and the *Ethnographic Atlas* are used to express the similarity between pairs of languages and cultures, respectively, as the percentage of traits shared by members of the pair. For both languages and cultures, similarity is plotted as a joint function of geographic distance and recency of common ancestry, the latter estimated from the language classification in *WALS*. Even for languages and cultures with no known common ancestor, similarity decreases as distance increases out to at least 5000 kilometers, suggesting widespread diffusion or possibly local adaptation. At any given geographic distance, similarity between languages decreases with increasing remoteness of common ancestry up to the possibly 10,000-year time depth of the oldest language families, but similarity between cultures decreases only up to a time depth of about 4000 years, suggesting that cultural evidence of common ancestry is lost more rapidly than linguistic evidence.

In general, the further apart languages or cultures are located geographically, the fewer traits they have in common and, hence, the less similar they are to each other. This empirical relationship has been confirmed in a wide range of contexts: cultural similarity among indigenous communities in Brazil, Melanesia, and California (Milke 1949); cultural and linguistic (lexical) similarity among Salishan tribes in western North America (Jorgensen 1969); lexical similarity among dialects in France (Séguy 1971), dialects in the Alps (Goebl 2001), dialects and closely related languages in Micronesia (Cavalli-Sforza and Wang 1986), and languages worldwide (Holman et al. in press); phonological similarity among dialects in the Netherlands (Nerbonne and Heeringa n.d.); and grammatical similarity among languages worldwide (Holman et al. 2007).

There are three principal processes that may contribute to the observed relation between distance and similarity. One is borrowing or diffusion between languages or cultures in contact, through a process of horizontal transmission in which people of different languages or cultures acquire information from each other. Since people are most likely to interact with others who live nearby, diffusion and thus similarity will be negatively related to distance. Another process is inheritance of attributes from a common ancestral language or culture through a process of vertical transmission in which people of earlier generations pass on information to people of later ones. To show how inheritance is related to distance, Cavalli-Sforza and Wang (1986) pointed out that after two linguistic or cultural groups diverge from a common ancestor, they become less similar to each other with the passage of time and also more likely to have migrated away from each other. A third process, also indirectly related to distance, is adaptation to the local environment, which causes similarity between cultures located in similar environments, which in turn are likely to be near each other geographically. Although adaptation has traditionally been considered relevant to cultural but not linguistic features, evidence for an effect of adaptation on linguistic phonology has recently been accumulated by Munroe et al. (1996), Fought et al. (2004), Ember and Ember (2007a, 2007b), and Munroe and Fought (2007). Moreover, as Borgerhoff Mulder et al. (2006) mentioned, adaptation can influence both diffusion and inheritance, because well-adapted traits are more likely to be borrowed or inherited.

The next step is to estimate the relative contributions of these three causal factors to the relation between distance and similarity. Most research on this question has applied correlational methods to cultural data. A contentious series of studies analyzed the material culture of villages located on the north coast of New Guinea that are inhabited by speakers of languages in several unrelated families; the data were originally analyzed by Welsch et al. (1992) and subsequently reanalyzed by Moore and Romney (1994, 1996), Roberts et al. (1995), Welsch (1996), and Shennan and Collard (2005). All the studies showed an effect of geographic proximity with language held constant, which was attributed to diffusion. Depending on how cultural similarity was defined, most of the studies also found an effect of language with geographic proximity held constant, which was attributed to inheritance. Other studies by Guglielmino et al. (1995) and Hewlett et al. (2002) analyzed the cultural data from sub-Saharan Africa in Murdock's (1967a) *Ethnographic Atlas*. As in the New Guinea studies, spatial proximity was used as a measure of diffusion and phylogenetic proximity in language classifications was used as a measure of inheritance; in addition, environmental similarity as inferred from vegetation maps was used as a measure of local adaptation. Each of these measures was correlated with the similarity between pairs of African societies with respect to each cultural trait. Differences among the correlations suggested differences among cultural traits in the relative importance of inheritance, diffusion, and adaptation.

Collard et al. (2006) took a different approach, which can be traced back more than a century to Schleicher's (1863) application of Darwinian ideas to linguistics. Using methods originally developed in biology, Collard et al. constructed phylogenetic trees for sets of animal species and human cultures. The trees fit the data about as well for cultures as for species, suggesting that inheritance plays about as strong a role in cultural evolution as in organic evolution. Moreover, since diffusion is negligible in the evolution of animals, it follows that the combined effect of diffusion and adaptation on culture is about equal to the effect of adaptation on animals.

The present study combines aspects of these previous lines of research. Like Milke and others, we construct empirical functions relating similarity in linguistic and cultural traits to spatial distance; like the correlational studies, we infer inheritance from linguistic classifications; and like Collard et al., we conduct parallel analyses in two different domains, in this case language and culture. To distinguish between causal factors, separate functions relating similarity to distance are constructed at different taxonomic levels of linguistic relationship. A difference between taxonomic levels at the same spatial distance can then be attributed to inheritance or adaptation but not diffusion. Since the highest taxonomic level corresponds to languages with no known inherited relationship, an effect of spatial distance at this level can be attributed to diffusion or adaptation but not inheritance. Given that adaptation can influence both inheritance and diffusion, our methods do not distinguish adaptation from either alternative, and we will therefore not attempt to draw inferences about adaptation.

DATA AND METHODS

The linguistic data are from the *World Atlas of Language Structures* (Haspelmath et al. 2005, henceforth *WALS*). *WALS* contains 138 maps showing the distribution of different phonological, lexical, and grammatical features. Each feature has anywhere from two to nine discrete values or traits. For example, Map 75 refers to the feature 'epistemic possibility,' which has three values, 'verbal constructions,' 'affixes on verbs,' and 'other.' The present study draws on 134 of the 138 features, excluding those with redundant data. Of the 2560 languages treated in *WALS*, this study excludes pidgins, creoles, and sign languages, leaving 2488 languages. Although *WALS* is by far the largest compilation of its kind, its data are still incomplete. With 2488 languages and 134 features, there are potentially 2488×134 or 333,392 data points, but only 54,428 of these or about 16% are actually attested in *WALS*. Languages are extremely variable in their degree of attestation: 218 languages are attested for only a single feature, one (English) is attested for all 134, and the rest are scattered across the interval between these extremes. Features are less variable, ranging from 111 to 1369 in the number of languages attested.

The cultural data are from the *Ethnographic Atlas* (henceforth *EA*), originally published by Murdock (1967a) and subsequently augmented by Murdock (1967b, 1968a, 1968b, 1968c, 1971), Barry (1980a, 1980b), and Korotayev et al. (2004). *EA* summarizes the published literature on 80 cultural features related mainly to subsistence economy, family and kinship, social stratification, division of labor, and house construction. Each feature has from two to fourteen discrete values. For example, the sixth feature is 'prevailing mode of obtaining a wife,' which has seven values, 'bride-price or bride-wealth,' 'bride-service,' 'token bride-price,' 'gift exchange,' 'exchange,' 'absence of any significant consideration,' and 'dowry.' *EA* includes societies from the historical past as well as the ethnographic present; in order to reduce variation across societies in the definition of time depth, the present study excludes the seven societies dated before 1500 A.D., leaving 1269 societies. The data in *EA* are more complete than in *WALS*. Of the 1269×80 or 101,520 potential data points, 80,069 or about 79% are actually attested. Degree of attestation ranges from 17 to 80 features across societies, and from 367 to 1269 societies across features.

Both *WALS* and *EA* include language classifications; the present study uses the more recent classification of *WALS*. This classification, described by Dryer (2005), groups languages into genera and genera into families. Genera are defined as the most inclusive groups descended from a common ancestral language spoken within the last 3500 to 4000 years. Families are defined as the most inclusive groups widely considered to be descended from a common ancestor; few if any families are estimated to be more than 10,000 years old. As examples of a family and genera, Dryer offers respectively Indo-European and its major subgroups such as Germanic and Celtic.

To parallel the literature on dialects and closely related languages, in the present analysis an additional level, called dialect clusters, is introduced into the classification below the level of genera. This level approximates the uppermost boundary between languages and dialects based on information in Bender (1997), Campbell (1997), Gordon (2005), Moseley and Asher (1994), Voegelin and Voegelin (1977), Wurm and Hattori (1981), the lexicostatistical studies cited by Holman (2004), and additional lexicostatistical data from Aikhenvald (2001), Bastin et al. (1983), Boyd (1974), Nurse (1979), and Schoenbrun (1994). Languages are grouped together in the same dialect cluster if they satisfy at least one of the four following conditions as inferred from these sources: (1) they are described as dialects, possible dialects, or emergent languages within the same language; (2) they are described as mutually intelligible or partly intelligible; (3) they are estimated to have diverged from each other no more than 1000 years ago; or (4) their percentage of shared cognates in a lexicostatistical word list is at least 74% for a 100-word list or 66% for a 200-word list, percentages that correspond to about 1000 years of separation according to Swadesh (1955).

The same classification was used both for the languages in *WALS* and for those spoken by peoples of the societies in *EA*. The 2488 languages from *WALS* are classified into 1962 dialect clusters, 458 genera, and 205 families. The 1269 societies from *EA* are classified into 866 dialect clusters, 276 genera, and 117 families.

Geographic distances between languages and between societies are calculated from latitudes and longitudes provided in *WALS* and *EA*. These distances follow the shortest path on the surface of a sphere, ignoring obstacles such as mountains, deserts, and oceans.

The data were analyzed by calculating the degree of similarity in traits as a joint function of taxonomic level and spatial distance. In principle, such data could be presented in scatter plots with each point denoting a pair of languages or societies. The 2488 languages from *WALS* would then produce 2488×2487/2 or 3,093,828 pairs, of which 1,578,175 share at least one attested feature and could produce data points; and the 1269 societies from *EA* would produce 1269×1268/2 or 804,546 pairs, all of which share attested features and could produce data points. The usual solution to this excess, which we adopted from earlier studies such as those of Milke (1949) and Séguy (1971), is to divide distance into intervals and average the points within each interval. Unless otherwise mentioned, we used intervals of 500 kilometers, such as 0–500 kilometers, 500–1000 kilometers, and so on. In order to avoid unreliable results from unrepresentative samples, only intervals represented by at least five families and 50 features were analyzed further.

The procedure for averaging within intervals was more complicated than in the earlier studies for two reasons. The first complication is the different taxonomic levels:

in each interval with enough families, pairs were sorted into one of four taxonomic levels, members of the pair being either (1) in the same dialect cluster, (2) in different dialect clusters but the same genus, (3) in different genera but the same family, or (4) in different families. The other complication is the variability of attestation in *WALS*; for this reason, the units of analysis were features, which are less variable in attestation than languages. For each feature, the similarity between the pairs in a taxonomic group at a given level was defined as the number of pairs with the same value of the feature, divided by the number of pairs for which the feature is attested in both members. Similarity was then averaged across taxonomic groups at the given level, weighted by the square root of the number of pairs in each group. The square-root weighting compensated for the fact that the number of pairs in a group increases approximately as the square of the size of the group, while the amount of data in a group increases only as the size of the group. These weighted average similarities were finally averaged across all 134 features and expressed as a percentage in order to represent the overall similarity of the pairs at a taxonomic level in a distance interval.

RESULTS

Figures 1 and 2 plot similarity in language and culture, respectively, as a function of spatial distance at each taxonomic level. The curves for members of the same family (and lower levels) stop before 7500 kilometers (or sooner) because fewer than five families contain members that far apart.

Figure 1 here

Figure 2 here

Similarity generally decreases with distance at each taxonomic level. The effect of distance is clearest in lowest curve in each figure, which refers to the highest taxonomic level. Since this curve describes the similarity between members of different language families, which are generally considered to be unrelated by inheritance, the shape of the curve is not complicated by effects of inheritance. The curve extends with adequate samples out to the longest distances, and similarity tends to decline with distance out to at least 5000 kilometers.

Similarity also tends to be lower in the curves for higher taxonomic levels. The effect of taxonomic level is clearest with distance held constant, particularly at the shortest distances, where all taxonomic levels are represented with adequate samples. In Figure 1, all the curves are separate at all distances including the shortest. In Figure 2, however, the curves for the two highest levels are practically superimposed at short distances, suggesting a more limited effect of inheritance on culture.

Standard statistical tests for these geographic and taxonomic effects require independent units of analysis. The present units of analysis are the linguistic and cultural features, which are not mutually independent. Lack of independence has been widely discussed in cross-cultural research under the name of Galton's problem (Tylor 1889), which applies to societies rather than features as units of analysis. The usual solution to Galton's problem has been to select a reduced sample of societies that can more plausibly be considered independent, but the methods proposed for doing so are still controversial (Strauss and Orans 1975, Korotayev and de Munck 2003, Dow 2007). We used reduced samples of approximately independent features, which were identified empirically with a method that is specific to features and different from the methods previously used to generate reduced samples of societies. Holman (2008) described this method in detail and used it to find 47 approximately independent linguistic features in the 134 *WALS* features. For culture, the same method discovered 16 approximately independent features among the 80 *EA* features; this process is summarized in the Appendix.

As described above, each of the curves in Figures 1 and 2 is itself an average of curves constructed for individual features. Statistical tests were therefore based on the individual curves for the independent features. Tests for effects of geographic distance are described first, followed by tests for effects of taxonomic level. A significance criterion of $p < .05$ is used in all tests.

Tests for effects of distance were conducted at the highest taxonomic level, languages and societies in different families. It is sufficient to show that similarity is significantly greater at short distances, below 1000 kilometers, than at long distances, above 10,000 kilometers. The difference is indeed significant for the 47 independent linguistic features, $t(46) = 6.14$, and for the 16 independent cultural features, $t(15) = 4.11$.

Tests for effects of taxonomic level were conducted at distances below 1000 kilometers. Adjacent levels were compared in order to approximate the duration of inherited effects. Only 43 of the independent linguistic features are attested at the lowest level, dialect clusters, but all 47 are attested at higher levels. For the 43 attested features, similarity is significantly higher within dialect clusters than within genera, $t(42) = 2.59$. For all 47 features, similarity is significantly higher within genera than within families, $t(46) = 3.23$, and also significantly higher within families than between families, $t(46) = 4.25$. All 16 of the independent cultural features are attested at all levels. For these features, similarity is significantly higher within dialect clusters than within genera, $t(15)$

= 3.14, and also significantly higher within genera than within families, $t(15) = 3.29$, but not significantly higher within families than between families, $t(15) = 0.72$.

DISCUSSION

The evidence in Figure 1 for linguistic diffusion out to distances of several thousand kilometers supports Dryer's (1992) suggestion of language areas as large as continents. According to the lowest curve in Figure 1, diffusion produces a change of about 10% in typological similarity between languages in different families, from about 50% between the closest languages to about 40% between the most distant ones. In contrast, the lowest curve in figure 3 of Holman et al. (in press) shows the effect of distance on the similarity between languages in different families with respect to a 40-item subset of Swadesh's (1955) basic vocabulary list. This curve drops by less than 1% from the closest languages to the most distant ones. The difference between the curves suggests that diffusion is much stronger for typological features than for basic lexical items. This sort of large-scale comparison of diffusion in different aspects of language was not possible until the recent publication of *WALS*.

The present finding of statistically significant linguistic inheritance from the time depth of currently accepted families is consistent with the fact that language families are themselves established in part on the basis of features like those tabulated in *WALS*. The present data neither confirm nor deny the possibility of linguistic inheritance from greater time depths, which could underlie more controversial families such as those proposed by Greenberg (1987, 2000). On the one hand, the substantial effects of diffusion apparent in Figure 1 may produce spurious evidence suggesting remote inherited relationships, and may also overshadow any real evidence for such relationships. On the other hand, the lowest curve, which depicts the effect of diffusion in the absence of acknowledged inheritance, may prove to be useful as a baseline for disentangling inheritance from diffusion.

For cultures, the results in Figure 2 confirm, with completely different methods of analysis, the finding of most correlational studies that cultural similarity depends on both geographic proximity and linguistic affiliation. In particular, there is statistically significant inheritance of cultural traits for about 4000 years within dialect clusters and genera, but none thereafter even within families. It follows that the features in *EA* are useless for exploring inherited relationships more than about 4000 years old. Inheritance can preserve ancestral features from time depths at most half as great for culture as for language. This finding suggests diffusion or adaptation rather than inheritance as explanations for the large culture areas inferred from *EA* data by Burton et al. (1996) and further discussed by Korotayev and Kazankov (2000) and Jones (2003).

The temporally limited effect of cultural inheritance in Figure 2 stands in contrast to the strong effect suggested by the comparison of cultural and biological inheritance in Collard et al. (2006). Two reasons can be offered for the discrepancy. First, most of the cultural data sets analyzed by Collard et al. include representatives of only one or a few genera; cultures this closely related can be expected to show a substantial effect of inheritance according to Figure 2. Second, the conventional identification of cultural inheritance with language classifications may be an oversimplification; the cultural phylogenies inferred by Collard et al. may include cases where language and culture are inherited along separate pathways.

In summary, for both languages and cultures similarity decreases as distance increases. This is so even for languages and cultures without known common ancestors, suggesting the influence of widespread diffusion or possibly local adaptation. At any given geographic distance, similarity between languages with known common ancestors decreases with increasing remoteness of common ancestry up to the possibly10,000-year time depth of the oldest language families. However, similarity between cultures with known ancestors in common decreases only up to a time depth of about 4000 years, suggesting that cultural evidence of common ancestry is lost more rapidly than linguistic evidence. The present findings, then, suggest that languages are more conservative than cultures.

Other than random fluctuations, the main sources of linguistic change are the development of new words to accommodate new concepts as they originate in culture, and the occasional need for groups to assert different linguistic identities. Languages almost certainly are influenced less by environmental adaptations than are cultures, and to an even lesser extent by conscious manipulations. For these reasons, it is not surprising that languages are more conservative than cultures, and consequently more likely to retain similarity as they become removed from one another, for whatever historical reasons, through time and space.

REFERENCES CITED

Aikhenvald, Alexandra Y.

2001 Areal Diffusion, Genetic Inheritance, and Problems of Subgrouping: A North Arawak Case Study. *In* Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics. Alexandra Y. Aikhenvald and R. M. W. Dixon, eds. Pp. 167-194. Oxford: Oxford University Press.

Barry, Herbert, III

1980a Ethnographic Atlas XXVIII. Ethnology 19:245-263.

1980b Ethnographic Atlas XXVIII: Part 2. Ethnology 19:367-385.

Bastin, Yvonne, Athéna Coupez, and Bernard de Halleux

1983 Classification Lexicostatistique des Langues Bantous. Bulletin des Séances de l'Académie Royale des Sciences d'Outre-Mer 27:173-199.

Bender, Lionel M.

1997 The Nilo-Saharan Languages: A Comparative Essay. Munich: Lincom Europa.

Borgerhoff Mulder, Monique, Charles L. Nunn, and Mary C. Towner

2006 Cultural Macroevolution and the Transmission of Traits. Evolutionary Anthropology 15:52-64.

Boyd, Raymond

1974 Étude Comparative dans le Groupe Adamawa. Paris: SELAF.

Burton, Michael L., Carmella C. Moore, John J. M. Whiting, and A. Kimball Romney

1996 Regions Based on Social Structure. Current Anthropology 37:87-123.

Campbell, Lyle

1997 American Indian Languages: The Historical Linguistics of Native America. New York: Oxford University Press.

Cavalli-Sforza, Luigi Luca, and William S.-Y. Wang

1986 Spatial Distance and Lexical Replacement. Language 62:38-55.

Collard, Mark, Stephen J. Shennan, and Jamshid J. Tehrani

2006 Branching, Blending, and the Evolution of Cultural Similarities and Differences among Human Populations. Evolution and Human Behavior 27:169-184.

Dow, Malcolm M.

2007 Galton's Problem as Multiple Network Autocorrelation Effects: Cultural Trait Transmission and Ecological Constraint. Cross-Cultural Research 41:336-363.

Dryer, Matthew S.

1992 The Greenbergian Word Order Correlations. Language 68:81-138.

2005 Genealogical Language List. *In* World Atlas of Language Structures. Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie, eds. Pp. 584-643. London: Oxford University Press.

Ember Carol R., and Melvin Ember

2007a Climate, Econiche, and Sexuality: Influences on Sonority in Language. American Anthropologist 109:180-185.

2007b Rejoinder to Munroe and Fought's commentary. American Anthropologist 109:785.

Fought, John G., Robert L. Munroe, Carmen R. Fought, and Erin M. Good

2004 Sonority and Climate in a World Sample of Languages: Findings and Prospects. Cross-Cultural Research 38:27-51.

Goebl, Hans

2001 Arealtypologie und Dialektologie. *In* Language Typology and Universals: An International Handbook, Vol. 2. Martin Haspelmath, Ekkehard König, Wulf Oesterreicher, and Wolfgang Raible, eds. Pp. 1471-1491. Berlin: Walter de Gruyter.

Gordon, Raymond G., Jr.

2005 Ethnologue: Languages of the World, 15th edition. Dallas: SIL International. Web version: www.ethnologue.com

Greenberg, Joseph H.

1987 Language in the Americas. Stanford: Stanford University Press.

2000 Indo-European and its Closest Relatives: the Eurasiatic Language Family. Stanford: Stanford University Press.

Guglielmino, Carmela R., C. Viganotti, Barry Hewlett, and Luigi Luca Cavalli-Sforza

1995 Cultural Variation in Africa: Role of Mechanisms of Transmission and Adaptation. Proceedings of the National Academy of Sciences of the United States of America 92:7585-7589.

Haspelmath, Martin, Matthew S. Dryer, David Gil, and Bernard Comrie, eds.

2005 World Atlas of Language Structures. London: Oxford University Press.

Hewlett, Barry S., Annalisa de Silvestri, and Rosalba Guglielmino

2002 Semes and Genes in Africa. Current Anthropology 43:313-321.

Holman, Eric W.

2004 Why Are Language Families Larger in Some Regions than in Others? Diachronica 21:57-84.

2008 Approximately Independent Typological Features of Languages. International Journal of Modern Physics C 19:215-220.

Holman, Eric, Søren Wichmann, Cecil H. Brown, Viveka Velupillai, André Müller, and Dik Bakker

In press Explorations in Automated Language Classification. Folia Linguistica.

Holman, Eric W., Christian Schulze, Dietrich Stauffer, and Søren Wichmann

2007 On the Relation between Structural Diversity and Geographical Distance among Languages: Observations and Computer Simulations. Linguistic Typology 11: 395-423.

Hubert, Lawrence, and Phipps Arabie

1985 Comparing Partitions. Journal of Classification 2:193–218.

Jones, Doug

2003 Kinship and Deep History: Exploring Connections between Culture Areas, Genes, and Languages. American Anthropologist 105:501-514.

Jorgensen, Joseph G.

1969 Salish Language and Culture. Bloomington: Indiana University.

Korotayev, Andrey, and Alexander Kazankov

2000 Regions Based on Social Structure: A Reconsideration (or Apologia for Diffusionism). Current Anthropology 41:668-690.

Korotayev, Andrey, Alexander Kazankov, Sventlana Borinskaya, Daria Khaltourina, and Dmitri Bondarenko

2004 Ethnographic Atlas XXX: Peoples of Siberia (1). Ethnology 43:83-92.

Korotayev, Andrey, and Victor de Munck

2003 "Galton's Asset" and "Flower's Problem": Cultural Networks and Cultural Units in Cross-Cultural Research (Or, Male Genital Mutilations and Polygyny in Cross-Cultural Perspective). American Anthropologist 105:353-358.

Milke, Wilhelm

1949 The Quantitative Distribution of Cultural Similarities and Their Cartographic Representation. American Anthropologist 51:237-252.

Moore, Carmela C., and A. Kimball Romney

1994 Material Culture, Geographic Propinquity, and Linguistic Affiliation on the North Coast of New Guinea: A Reanalysis of Welsch, Terrell, and Nadolski (1992). American Anthropologist 96:370-396.

1996 Will the "Real" Data Please Stand Up: Reply to Welsch (1996). Journal of Quantitative Anthropology 196:235-261.

Morey, Leslie C., and Alan Agresti

1984. The Measurement of Classification Agreement: an Adjustment to the Rand Statistic for Chance Agreement. Educational and Psychological Measurement 44:33-37.

Moseley, Christopher, and R. E. Asher, eds.

1994 Atlas of the World's Languages. London: Routledge.

Munroe, Robert L., and Fought, John G.

2007 Response to Ember and Ember's "Climate, Econiche, and Sexuality: Influences on Sonority in Language". American Anthropologist 109:784-785.

Munroe, Robert L., Ruth H. Munroe, and Stephen Winters

1996 Cross-Cultural Correlates of the Consonant-Vowel (CV) Syllable. Cross-Cultural Research 38:27-51.

Murdock, George Peter

1967a Ethnographic Atlas. Pittsburgh: University of Pittsburgh Press

1967b Ethnographic Atlas XXIII. Ethnology 6:481-487.

1968a Ethnographic Atlas XXIV. Ethnology 7:106-111.

1968b Ethnographic Atlas XXV. Ethnology 7:218-224.

1968c Ethnographic Atlas XXVI. Ethnology 7:327-329.

1971 Ethnographic Atlas XXVII. Ethnology 10:122-127.

Nerbonne, John, and Wilbert Heeringa

N.d. Geographic Distributions of Linguistic Variation Reflect Dynamics of Differentiation. Submitted to Roots: Linguistics in Search of its Evidential Base. S. Featherstone and W. Sternefeld, eds. (available at www.let.rug.nl/~nerbonne/papers/geography-2006.pdf)

Nurse, Derek

1979 Classification of the Chaga Dialects: Language and History on Kilamanjaro, the Taita Hills, and the Pare Mountains. Hamburg: Helmut Buske Verlag.

Rand, William M.

1971 Objective Criteria for the Evaluation of Clustering Methods. Journal of the American Statistical Association 66:846-850.

Roberts, John M., Jr., Carmela C. Moore, and A. Kimball Romney

1995 Predicting Similarity in Material Culture among New Guinea Villages from Propinquity and Language: A Log-Linear Approach. Current Anthropology 36:769-788.

Schleicher, August

1863 Die Darwinsche Theorie und die Sprachwissenschaft: Offenes Sendschreiben an Herrn Ernst Häckel. Weimar: Böhlau.

Schoenbrun, David L.

1994 Great Lakes Bantu: Classification and Settlement Chronology. Sprache und Geschichte in Afrika 15:91-152.

Séguy, Jean

1971 La Relation entre la Distance Spatiale et la Distance Lexicale. Revue de Linguistique Romane 35:335-357.

Shennan, Stephen, and Mark Collard

2005 Investigating Processes of Cultural Evolution on the North Coast of New Guinea with Multivariate and Cladistic Analysis. *In* The Evolution of Cultural Diversity: A Phylogenetic Approach. Ruth Mace, Clare J. Holden, and Stephen Shennan, eds. Pp. 133-164. London: UCL Press.

Strauss, David J., and Martin Orans

1975 Mighty Sifts: A Critical Appraisal of Solutions to Galton's Problem and a Partial Solution. Current Anthropology 16:573-594.

Swadesh, Morris

1955 Towards Greater Accuracy in Lexicostatistic Dating. International Journal of American Linguistics 21:121-137.

Tylor, Edward B.

1889 On a Method of Investigating the Development of Institutions. Journal of the Royal Anthropological Institute 18:245-269.

Voegelin, Charles F., and Florence M. Voegelin

1977 Classification and Index of the World's Languages. New York: Elsevier.

Welsch, Robert L.

1996 Language, Culture and Data on the North Coast of New Guinea. Journal of Quantitative Anthropology 196:209-234.

Welsch, Robert L., John Terrell, and John A. Nadolski

1992 Language and Culture on the North Coast of New Guinea. American Anthropologist 94:568-600.

Wichmann, Søren, and Eric W. Holman.

N. d. Assessing Temporal Stability for Linguistic Typological Features. Manuscript under review. Prepublication version: http://email.eva.mpg.de/~wichmann/WichmannHolmanIniSubmit.pdf.

Wurm, Stephen A., and Shirô Hattori

1981 Language Atlas of the Pacific Area. Canberra: Australian Academy of the Humanities in collaboration with the Japan Academy.

APPENDIX

The identification of independent features requires a measure of relationship between features. Among the many available measures, the adjusted Rand index has the advantage of a correction for chance that gives the index an expected value of 0 if the features are independent. To define the adjusted Rand index in the present context, let any two features be given. The features agree on the similarity of two cultures if the cultures have the same value on both features, or if the cultures have different values on both features. The original Rand index is denoted R and defined as the number of pairs of cultures for which the two features agree, divided by the number of pairs of cultures for which the two features are attested. To correct for chance agreement, the adjusted Rand index is denoted R□ and defined as [R – E(R)]/[1 – E(R)], where E(R) is the expected value of R if the features are independent of each other. Rand (1971) introduced R, Morey and Agresti (1984) introduced R□, and Hubert and Arabie (1985) derived the general formula for E(R). R□ is equal to 1 if the features always agree, and it has expected value 0 if the features are independent. R□ can take negative values by chance, and positive values either by chance or because the features are not independent.

The procedure for finding independent features is the same as described by Holman (2008) except applied to the features of *EA* rather than *WALS*. The 80 features in *EA* give rise to 80×79/2 or 3160 pairs. R□ was calculated for each pair. The average R□ across all pairs proved to be 0.0436, higher than expected if all the features are independent. The feature with the highest average R□ with the others was then excluded and the average R□ across the remaining pairs was recalculated. This process was repeated until the average R□ dropped below 0. At this point, 16 features remained with an average R□ of -0.0005, essentially as expected for a set of independent features. Table 1 lists the 16 features with their column numbers and names in *EA*. These are the features that we used as units of analysis in statistical tests.

Table 1 here

Table 1 contains a number of secondary and peripheral traits, and lacks others generally considered more important and central. This pattern can be attributed to the fact that some traits are considered important or central because they are correlated with many other traits. Such traits are thus likely to be excluded from a set of independent traits, but they are nevertheless indirectly represented in the statistical tests through their correlations with the traits on the list.

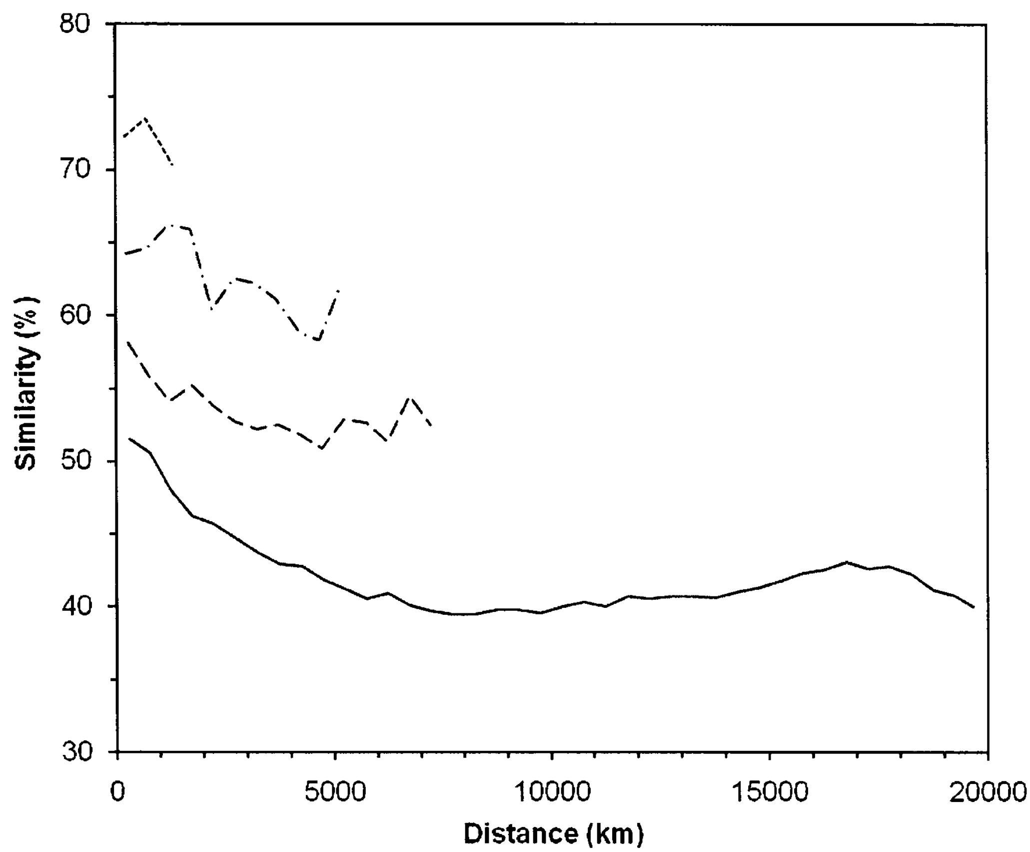

FIGURE 1. Percentage similarity between languages as a function of the spatial distance between them. Solid line: languages in different families. Dashed line: languages in the same family but different genera. Dot-dashed line: languages in the same genus but different dialect clusters. Dotted line: languages in the same dialect cluster.

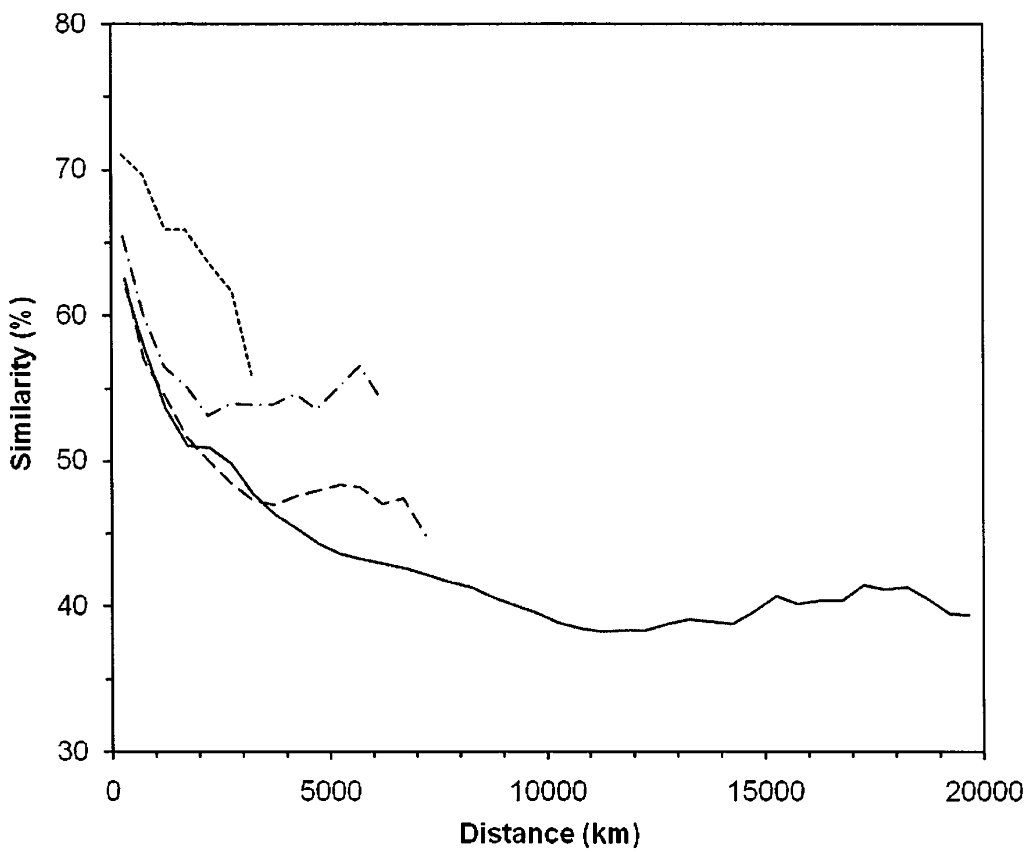

FIGURE 2. Percentage similarity between societies as a function of the spatial distance between them. Solid line: societies in different language families. Dashed line: societies in the same family but different genera. Dot-dashed line: societies in the same genus but different dialect clusters. Dotted line: societies in the same dialect cluster.

TABLE 1. Independent cultural features

Column	Name
23	Largest matrilineal exogamous kin group
24	Secondary cognatic kin groups
26	Cousin marriage: preferential or prescriptive unions
32	Jurisdictional hierarchy: local community
38	Segregation of adolescent boys
44	Sex difference in weaving
51	Age difference in boat building
52	Sex difference in house construction
54	Sex difference in gathering
55	Age difference in gathering
57	Age difference in hunting
58	Sex difference in fishing
61	Age difference in animal husbandry
72	Former presence of slavery
86	Secondary floor level of house
89	Secondary roofing material of house

NOTES

Acknowledgments. We thank Hans-Jörg Bibiko, creator of the electronic interactive version of *WALS*, for providing us with the total dataset in electronic form. We also thank Tom Boellstorff, Bernard Comrie, Dietrich Stauffer, and four anonymous referees for helpful comments on an earlier version of this paper.