DIGITALCOMMONS
@WAYNESTATE

Human Biology

Volume 81
Issue 2 *Special Issue on Demography and Cultural Macroevolution*

Article 8

2009

Population Size and Rates of Language Change

Søren Wichmann
Leiden University, wichmann@eva.mpg.de

Eric W. Holman
University of California - Los Angeles

Follow this and additional works at: http://digitalcommons.wayne.edu/humbiol

Recommended Citation
Wichmann, Søren and Holman, Eric W. (2009) "Population Size and Rates of Language Change," Human Biology: Vol. 81: Iss. 2-3, Article 8. Available at: http://digitalcommons.wayne.edu/humbiol/vol81/iss2/8

Population Size and Rates of Language Change

Abstract
Previous empirical studies of population size and language change have produced equivocal results. We therefore address the question with a new set of lexical data from nearly one-half of the world's languages. We first show that relative population sizes of modern languages can be extrapolated to ancestral languages, albeit with diminishing accuracy, up to several thousand years into the past. We then test for an effect of population against the null hypothesis that the ultrametric inequality is satisfied by lexical distance among triples of related languages. The test shows mainly negligible effects of population, the exception being an apparently faster rate of change in the larger of two closely related variants. A possible explanation for the exception may be the influence on emerging standard (or cross-regional) variants from speakers who shift from different dialects to the standard. Our results strongly indicate that the sizes of speaker populations do not in and of themselves determine rates of language change. Comparison of this empirical finding with previously published computer simulations suggests that the most plausible model for language change is one in which changes propagate on a local level in a type of network in which the individuals have different degrees of connectivity.

Keywords
language evolution, language size, Levenshtein distance, lexicostatistics, World Atlas of Language Structures (WALS), Automated Similarity Judgment Program (ASJP).

Cover Page Footnote
We thank Dietrich Stauffer, James Steele, and three anonymous referees for their comments on earlier versions of this paper. For their efforts in gathering and analyzing the lexical data described here, we would like to acknowledge the other members of the Automatic Similarity Judgment Project, in alphabetical order: Dik Bakker, Oleg Belyaev, Cecil H. Brown, Pamela Brown, Matthew Dryer, Dmitri Egorov, Patience Epps, Anthony Grant, Hagen Jung, Robert Mailhammer, André Müller, Viveka Velupillai, and Kofi Yakpo.

This open access article is available in Human Biology: http://digitalcommons.wayne.edu/humbiol/vol81/iss2/8

Population Size and Rates of Language Change

SØREN WICHMANN[1] AND ERIC W. HOLMAN[2]

Abstract Previous empirical studies of population size and language change have produced equivocal results. We therefore address the question with a new set of lexical data from nearly one-half of the world's languages. We first show that relative population sizes of modern languages can be extrapolated to ancestral languages, albeit with diminishing accuracy, up to several thousand years into the past. We then test for an effect of population against the null hypothesis that the ultrametric inequality is satisfied by lexical distances among triples of related languages. The test shows mainly negligible effects of population, the exception being an apparently faster rate of change in the larger of two closely related variants. A possible explanation for the exception may be the influence on emerging standard (or cross-regional) variants from speakers who shift from different dialects to the standard. Our results strongly indicate that the sizes of speaker populations do not in and of themselves determine rates of language change. Comparison of this empirical finding with previously published computer simulations suggests that the most plausible model for language change is one in which changes propagate on a local level in a type of network in which the individuals have different degrees of connectivity.

Scholars have addressed the issue of whether there is a relation between demography and cultural evolution and have found confirmative answers (Henrich 2004; Powell et al. 2009; Richerson and Boyd 2009; Shennan 2000). Although language is at least partly a cultural product, models for cultural innovation, where ecological adaptation and survival are crucial factors, do not directly carry over to language change, because it is highly doubtful that language structures are adaptive in any Darwinian sense. Models and simulations can be used to explore the question of a relation between language and demography, but first and foremost empirical evidence should be explored. A first step is to ask whether there is a relation between rates of language change and population sizes. Although conceptually simple, this

[1]Department of Linguistics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany; and Faculty of Archaeology, Leiden University, P.O. Box 9515, 2300 RA Leiden, The Netherlands.

[2]Department of Psychology, University of California, Los Angeles, CA 90095-1563.

Human Biology, April–June 2009, v. 81, nos. 2–3, pp. 259–274.
Copyright © 2009 Wayne State University Press, Detroit, Michigan 48201-1309

KEY WORDS: LANGUAGE EVOLUTION, LANGUAGE SIZE, LEVENSHTEIN DISTANCE, LEXICOSTATISTICS, WORLD ATLAS OF LANGUAGE STRUCTURES (WALS), AUTOMATED SIMILARITY JUDGMENT PROGRAM (ASJP).

question requires innovative approaches. Unlike archaeologists who might draw on a material record (e.g., Chamberlain 2009), linguists do not have direct access to evidence regarding population sizes for prehistoric speaker communities but must make inferences based on present populations. Similarly, rates of change must also be inferred from modern languages, through the way they relate to one another.

Nettle (1999a, 1999c) was the first to raise the question of whether there is a relationship between population size and language change. He described a computer simulation that generated a higher rate of change in features of languages spoken by smaller populations, and he also found indirect empirical evidence for the same effect of population. Subsequent research on the question has produced alternative explanations for Nettle's empirical findings along with mixed results in other empirical tests, and computer simulations have shown either the presence or the absence of an effect depending on the model adopted. In this paper we present a new kind of empirical test that shows that during the breakup of dialects into separate languages, there is a tendency for a somewhat faster rate of change in the dialect of the majority. For languages that have been temporally separated for longer periods, we do not find any relation between language change and population sizes, although the possibility remains that a more sensitive test may reveal a longer-lasting effect. We speculate that the faster rate of change in majority dialects is mainly due to the influence of speakers shifting from minority variants when a standard language is emerging.

Previous Work

Computer Simulations. Computer simulations have given mixed results on the relation between population and language change. The results are heavily dependent on the kind of social interaction model used. In the following discussion we briefly summarize the results of the implementations of different models. Because we are primarily concerned with the presentation of new empirical results and not with simulations, the summaries are brief.

Nettle (1999a: 121) argues that "spreading an innovation over a tribe of 500 people is much easier and takes much less time than spreading one over five million people." His paper mainly contains a computer simulation of language change for just two linguistic features, and he found that the rate at which the majority of the population switches between these two choices decreases to a small but nonzero limit if the population increases from 120 to 500. The model used is described by Nettle (1999b) and is based on the social impact theory of Nowak et al. (1990). In Nettle's model the impact of a linguistic variant is a function of the statuses and ages of the individuals using this variant, their social distance from the learner, and their number.

Wichmann et al. (2008) tested two versions of one and the same basic model. Their model contains parameters that are similar but not identical to those of Nettle (1999a, 1999c). The major difference between Wichmann's model and Nettle's is that Wichmann and co-workers operate with many languages, each

of which has several features. Another difference is that rather than assigning variable statuses to individuals, the Wichmann model operates with a scale-free network (Barabási and Albert 1999), in which the impact of a certain individual increases during network growth with a probability that is proportional to the impact that the individual already has had. The most connected individuals are responsible for most, if not all, linguistic change. Social distances correspond to distances among individuals in the network. In the global version of the model a speaker randomly adopts variants from the entire population. In this version the size of the population that has a given linguistic variant indirectly affects the probability that this variant will diffuse further, and different parameter settings give different results. For small diffusion rates there is no impact on language change rates from population sizes, but for large diffusion rates there is.

Wichmann et al. (2008) also tested the situation in which the donor of a diffused linguistic item must sit on a neighbor node of the network rather than on any randomly selected node, as in the global version described in the previous paragraph. For the local version there is no strong variation of change rates with population sizes (equal to the number of network nodes) regardless of the diffusion probabilities.

Ke et al. (2008) also tested the effect of population sizes in different kinds of networks: regular networks, small-world networks, random networks, and scale-free networks. A regular network is built as a ring, with each node having an equal number of connections to its nearest neighbors. A small-world network (Watts and Strogatz 1998) starts from a regular network and rewires a number of regular links randomly according to a constant probability that determines how many regular connections are changed into shortcuts. In a random network two nodes are connected based on a constant probability that is determined by the given connectivity of the network. Finally, in a scale-free network (Barabási and Albert 1999) an important feature is the existence of nodes that are extremely highly connected (so-called hubs). Ke et al. found that for regular networks, changes propagate more slowly as the population increases, whereas in the other three types of network the population size has no effect on rates of change.

Empirical Approaches. Nettle (1999a) presented two indirect lines of empirical evidence to suggest that languages with few speakers change more rapidly than languages with many speakers. First, languages in the Americas have relatively few speakers and are divided into many small unrelated families, and languages in Africa and Eurasia have relatively many speakers and are divided into a few large families. To explain this pattern, Nettle proposed that small languages, such as those in the Americas, change too rapidly for families of related languages to grow very large before evidence of their relationship is lost, whereas large languages, such as those in Africa and Eurasia, change slowly enough for their families to grow larger. Elsewhere, however, Nettle (1999b) presented an alternative explanation for the high degree of linguistic diversity in the Americas that does not involve population sizes but instead assumes a simple model of rapid

growth of diversity followed by a somewhat slower decay; this model allows for a diversity that is higher for areas settled relatively recently than for areas that were settled early in prehistory.

As a second line of evidence, Nettle (1999a) tabulated population sizes for all the known languages with the rare object-initial word orders; 13 of the 14 languages proved to have fewer than 5,000 speakers, the global median at that time. His explanation is that the rapid evolution of small languages makes them more likely to enter rare and unstable states. The 14 languages are distributed among eight different families but only three separate geographic areas: New Guinea, northern Australia, and tropical South America. Nettle therefore acknowledged that his results could represent as few as three independent cases, not enough for statistical conclusions.

Since the publication of Nettle's pioneering work, more data have become available that allow more comprehensive tests of population effects. One large database is the World Atlas of Language Structures (WALS) (Haspelmath et al. 2005). WALS contains 142 maps showing the distribution of different phonological, grammatical, and lexical features, each of which has from two to nine discrete values. These maps refer to 2,560 languages, although most features are attested for far fewer languages. WALS contains a two-level classification of languages into families and genera (Dryer 2005a). Families are defined as the most inclusive groups considered by a majority of specialists to have descended from a common ancestral language, and genera are defined as the most inclusive groups descended from an ancestral language thought to have been spoken within the last 3,500 to 4,000 years.

The WALS data are particularly ample for word order phenomena. Thus the chapter by Dryer (2005b) on the order of subject, object, and verb contains data from 1,228 languages. Dryer cites 13 languages that have object-initial order, only 5 of which are also cited by Nettle (1999a). Dryer's languages are found in ten different families and five different areas: Africa (Päri), northern Australia (Ungarinjin, Mangarrayi, Wik Ngathana), New Guinea (Tobati), the greater Amazon (Urarina, Cubeo, Nadëb, Hixkaryana, Warao, Tiriyo, Asuriní), and Tierra del Fuego (Selknam). According to the figures in *Ethnologue* (Gordon 2005), 10 of the 13 languages have populations smaller than the current global median of 6,000 speakers. In the one instance from Africa the population is 28,000. Thus the tendency for object-initial word order to correlate with small populations is still there, but it is not as strong as in Nettle (1999a).

It is clear that a more general investigation of possible explanations for rare features is needed because it will always be possible to isolate a given linguistic feature that shows a degree of correlation with some particular nonlinguistic phenomenon. The paper by Cysouw (2009) is precisely this kind of study. Cysouw uses the WALS database to produce a general study of rarity. He defines a rarity index that expresses, for each language, the degree to which the values of the features attested for that language have low relative frequencies in the entire database. Cysouw's Table 1 lists the 15 languages in WALS with the highest rarity

indexes; 12 of these languages have fewer than 6,000 speakers. Although this finding is consistent with Nettle's, Cysouw did not explain it in terms of population size. Instead, he emphasized the alternatives of genealogical inheritance and the geographic locations of languages as the factors most relevant to rarity. In particular, he found relatively high rarity indexes for the large languages of northwestern Europe, and he noted that a Eurocentric bias in the selection of features included in WALS goes some way toward explaining this phenomenon, but he acknowledged the incompleteness of this explanation.

Wichmann et al. (2008) used the WALS data for a more direct test of the relation between population and rate of language change. The test is based on the fact that change is the opposite of stability. Nichols (1995) proposed that if a feature is stable, then genealogically related languages should be more likely to share the same value of the feature than unrelated languages. Following this idea, Wichmann and Holman (2009) then estimated the stability of each feature in WALS by comparing the percentages of shared values for related and unrelated languages [see the appendixes in Holman et al. (2007) for summaries of the method and its results]. Rate of change can be inferred from the instability of a feature, which can be defined as 100% minus its stability. Wichmann and co-workers therefore estimated instability separately for sets of languages with different numbers of speakers; related languages were defined either as those in the same family or as those in the same genus, and unrelated languages were defined as those in different families. They found only a small and statistically nonsignificant effect of number of speakers on instability with either definition.

A weakness of the test is that it assumes the average time since the divergence of language pairs within families and genera to be independent of the average size of their languages. This assumption is undermined by a natural sampling bias that is derived from the tendency for different dialects to be better described in WALS for large languages than for small languages. Because of this bias, the genealogically closest, and therefore the most similar, languages also tend to be large languages, thus producing a greater representation of dialect pairs from languages with larger populations than those from languages with few speakers. Wichmann et al. (2008) counteracted the bias by excluding language pairs in the same dialect cluster, where dialect clusters were based on published information, such as lists of dialects, estimates of mutual intelligibility, inferred time depths, and percentages of shared cognates.

A completely different line of indirect evidence comes from dialectology. Trudgill (1974) introduced from geography the so-called gravity model, according to which the amount of diffusion between two dialects is proportional to the product of their populations divided by the square of their distance. To the extent that diffusion causes change, this model predicts faster rather than slower change in dialects with more speakers. Subsequent tests of the gravity model, reviewed by Nerbonne and Heeringa (2007), have found both positive and negative effects of population on diffusion between dialects, with most effects small and no clear overall trend.

In summary, most empirical tests so far have dealt with structural properties of languages or diffusion among dialects, with equivocal results. In this paper we describe another test, based on pairwise dissimilarities among languages, that is applied to lexical data.

Materials and Methods

The data come from the Automated Similarity Judgment Program (ASJP), a collaborative research program for objectively comparing languages by computer analysis of diagnostic word lists of the sort introduced by Swadesh (1950, 1955). In the first attempt to derive linguistic phylogenies from word comparisons within the ASJP, Brown et al. (2008) formulated a set of rules by which cognates (related words) were identified by matching up segments of words. They described the transcription procedure and the matching rules and presented lexicostatistical classifications of a sample of 245 languages based on automatic cognate identifications in the 100-item list of Swadesh (1955). Most of the classifications agreed well with published classifications by experts working from the standard comparative method. With the same set of 245 languages, Holman et al. (2008) then measured the relative stability of each of the 100 items. It turned out that a shorter selection of the 40 most stable items gave just as good lexicostatistical results as the 100-item list. The shorter lists and the inclusion of new project members made it possible to speed up data collecting, with the result that the present sample contains 3,161 languages for which the number of speakers is provided by *Ethnologue* (Gordon 2005), not counting pidgins, creoles, and constructed languages. The languages are classified in the same families and genera as in WALS.

Application of the cognate matching rules to the expanded database requires more memory than is available in personal computers. Fortunately, we have found that even better agreement with expert classifications can be achieved by using Levenshtein distances (Levenshtein 1966) as the basis of comparison between languages. Levenshtein distances have previously been used to compare dialects (e.g., Heeringa 2004) and, more recently, languages (e.g., Serva and Petroni 2008). The Levenshtein distance (LD) (aka edit distance) between two words is defined as the minimum number of successive changes necessary to turn one word into another, where each change is the insertion, deletion, or substitution of one symbol. To correct for the larger number of changes possible in longer words, the normalized LD (LDN) is defined as the LD divided by the number of symbols in the longer word. For meanings represented by two synonymous words, the LDN is averaged across synonyms. For a given pair of languages, the LDN between words for the same meaning in the two languages is first averaged across all the meanings on the list attested in both languages. As a baseline for phonological similarity without semantic similarity, the LDN is also averaged across all pairs of different meanings attested in the two languages. The divided normalized LD (abbreviated LDND) between the languages is then defined as the average LDN for the same meaning divided by the average LDN for different meanings. Expressed

as a percentage, the LDND is 0 between identical languages and 100 on average between unrelated languages.

The method of measuring the relative performances of cognate matching rules versus the LDND is the same one used by Holman et al. (2008) to measure the performances of different subsets of the 100-item Swadesh list. We test how well the distance measures obtained for all language pairs correlate with the distances measures in the standard classifications used by WALS and *Ethnologue*. WALS operates with three taxonomic levels—languages, genera, and families—leading to three different distances: (1) same genus, different languages; (2) same family, different genera; and (3) different families. These distances may be correlated with lexical distances, with the result expressed as Pearson's *r*. The *Ethnologue* classification has varying taxonomic levels for different families, which requires a different kind of correlation metric, the Goodman-Kruskal gamma. When we made the comparison, the database contained 1,645 languages, which is near the upper limit of computer memory for the matching rules. The Pearson correlation with taxonomic distances in the WALS classification was 0.58 for distances based on the matching rules and increased to 0.64 for the LDND. Similarly, the Goodman-Kruskal gamma with taxonomic distances in *Ethnologue* was 0.61 for distances based on the matching rules and increased to 0.73 for the LDND. The correlations for the LDND were also higher than those for either the LDN or the LD. Thus there is a clear improvement to be gained by using the LDND.

The LDND, although weighted for word length and chance resemblances as just described, is not weighted for phonetic distances among segments. Thus, for instance, substituting *p* for *b* is precisely as costly as substituting *p* for *t*. Although this may seem counterintuitive, initial probings suggest that it may be difficult to establish sound changes that are universal across different areas. Moreover, through the experience of setting up matching rules to identify cognates in the early part of the project, we learned that a criterion according to which segments should be identical to count as a match worked better than criteria by which matches of the "natural" type such as *p*:*b* were deemed less costly than "unnatural" matches such as *p*:*t*. In a fine-grained classification of dialects it may be worthwhile to weight substitutions by phonetic distances, following the practice of dialectologists (e.g., Heeringa 2004), but in large-scale classifications of languages we do not expect results to change because what feeds into difference measures among languages will mainly be replacements of entire lexical items rather than subtle replacements of single phonemes.

The correlations between the LDND and expert classifications are useful for comparing the LDND with alternatives but less useful for judging its performance in absolute terms. The fact that the Pearson correlation with WALS is 0.64 and the Goodman-Kruskal gamma with *Ethnologue* is 0.73 cannot be measured against a yardstick where the optimal performance results in a 1.0 correlation because expert classifications are clearly not always correct. When the LDND is correlated with expert classifications for individual families, it gives the best results for families that are also the best studied in terms of their historical relations, and

the apparently worst results are obtained for families that have large numbers of languages or are controversial among experts. To judge the performance of the LDND, readers are invited to look qualitatively at the actual tree for the world's languages that it produces (Müller et al. 2009).

Results

Extrapolation of Relative Population Sizes into the Past. When a language splits into two (or more) descendants, the descendants may have different numbers of speakers, particularly if one is spoken by an isolated group of migrants or colonists. An immediate question is whether languages tend to become more or less different in number of speakers with the passage of time since their divergence. We can answer this question because our lexical data provide a measure of time depth.

Swadesh (1950, 1955) used the percentage of shared cognates in his word lists to infer the time since the divergence of languages, based on the assumption that words in basic vocabulary are replaced at a stochastically constant rate. His original analogy was with radioactive decay as used in carbon-14 dating. Subsequent evidence for variation in the replacement rate suggests that a better analogy is with changes in DNA, which occur at a rate that is not constant but close enough to be used as a molecular clock in dating the divergence of biological species. Serva and Petroni (2008) recently showed that Swadesh's dating procedure could be generalized to the LDN. With the use of the LDND to correct for chance similarity, the formula of Serva and Petroni implies that time depth is proportional to $-\log(1 - \text{LDND})$.

Figure 1 shows the ratio of population size for pairs of languages in the same family as a function of $(1 - \text{LDND})$. The horizontal axis shows $(1 - \text{LDND})$ as a percentage on a reversed logarithmic scale, which is proportional to time depth. Language pairs are grouped by $(1 - \text{LDND})$; boundaries between groups are 80%, 60%, 40%, 30%, 20%, 15%, 10%, 8%, 6%, 4%, and 2%. For each pair the population of the larger language is divided by the population of the smaller; the vertical axis shows the geometric mean population ratio for each group, plotted on a logarithmic scale. Separate curves show the results for languages in Africa, Eurasia, Australia and New Guinea, and the Americas.

In all four geographic areas, population ratio increases as $(1 - \text{LDND})$ decreases; therefore languages are more similar in populations shortly after they diverge than they are later. Moreover, the increase in population ratio continues until $(1 - \text{LDND})$ has decreased to about 10% in Eurasia, 6% in the Americas, and 3% in Africa, Australia, and New Guinea. For comparison, $(1 - \text{LDND})$ is about 5% between the major living branches of the Indo-European language family, which are generally thought to have diverged about 5,500 years ago (Nichols and Warnow 2008). The shapes of the curves show the persistence of the initial population ratio. They flatten where fluctuations in population sizes have completely dampened the initial differences. As one traverses the curves from left to right

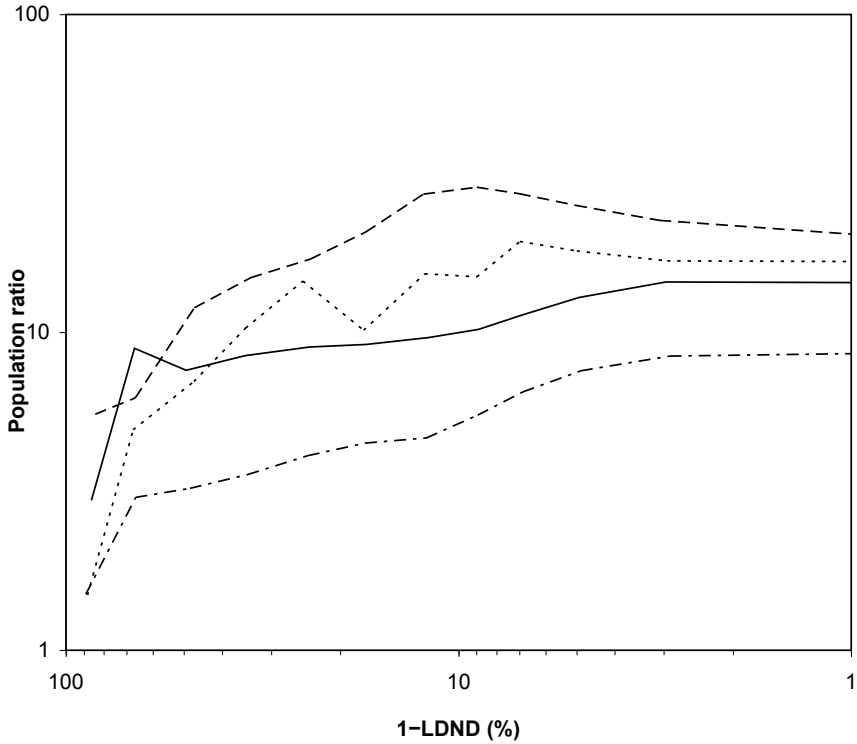

Figure 1. Population ratio as a function of percent (1 − LDND) for pairs of related languages in Africa (solid line), Eurasia (dashed line), Australia and New Guinea (dot-dashed line), and the Americas (dotted line).

until reaching the plateaus, an increasingly large number of pairs of language populations have shifted their mutual relationship such that a former minority population has become a majority population, and vice versa. The initial population ratio, however, is visible to some extent for a period corresponding to the time depth of an old language family such as Indo-European. Thus current population ratios can be extrapolated, albeit with diminishing accuracy, up to several thousand years into the past.

Language Change Rates in Relation to Population Sizes. Because the LDND is a measure of distance, it can be used to test whether languages change their basic vocabulary at the same average rate independently of their size. If languages change at a constant rate, then any two related languages are equidistant from their common ancestor. The two languages are therefore also equidistant from any third language that shares a more remote common ancestor with them. It follows that in any set of three related languages, the two closest languages should

be equidistant from the third (this implication is called the ultrametric inequality in the mathematical literature). Alternatively, if smaller languages change more rapidly, then the smaller of the two closest languages should be farther from the third on average. For example, the LDND is 63% between English and Dutch, 67% between Dutch and Danish, and 70% between English and Danish. These figures are consistent with the conventional wisdom that English and Dutch share a common ancestor that is more recent than the ancestor they both share with Danish. The figures also imply that English has diverged more than Dutch has from their common ancestor, although English has more speakers than Dutch, suggesting that other historical factors were more important than population in this case.

The test for the effect of population uses triples of languages (within the same family) in which the average LDND from the two closest languages to the third is less than 90%, because Figure 1 shows that current population ratios can be extrapolated at least as far back as the time corresponding to a (1 − LDND) of 10%. The triples are grouped by the population ratio of the two closest languages, which is sorted into the ranges 1–10, 10–100, 100–1,000, and so on up to 100,000+. For the two closest languages in each triple, the LDND from the larger language to the third is subtracted from the LDND from the smaller language to the third, producing a difference score that will be positive if the rate of change is higher for the smaller language. For English and Dutch the population ratio is in the 10–100 group and the difference score relative to Danish is −3%.

Figure 2 shows the cumulative distribution of difference scores for each group of population ratios. If smaller languages change more rapidly, then most difference scores should be positive; thus the distributions should be displaced to the right of 0, more so for higher population ratios. Most of the distributions are practically symmetric around 0 and are so close to each other as to be indistinguishable in the figure. The curve for the highest population ratios of 100,000+ is displaced slightly to the right of 0, however, suggesting that the smallest languages may change slightly more rapidly than the largest languages, but nearly all the triples contributing to this curve are confined to a single language family (Austronesian). This difference is small compared with the variability of the scores within each distribution. For instance, about 30% of each distribution is below the difference score of −3% observed for English and Dutch relative to Danish. Thus any effect of population is dwarfed by the combined effects of all the other historical factors that influence the divergence of languages.

Results of this sort are summarized with a standard measure of effect size that is the mean difference score divided by the standard deviation of the scores. The mean depends on the effect of interest (in this case, population), whereas the standard deviation depends on all the other factors that influence the scores. Cohen (1992) reported that surveys of effect sizes in various fields have found average effect sizes of about 0.5. The solid line in Figure 3 shows the effect size for population in each group as a function of the geometric mean population ratio for the group. Consistent with Figure 2, effect sizes are small, with the largest at the highest population ratio.

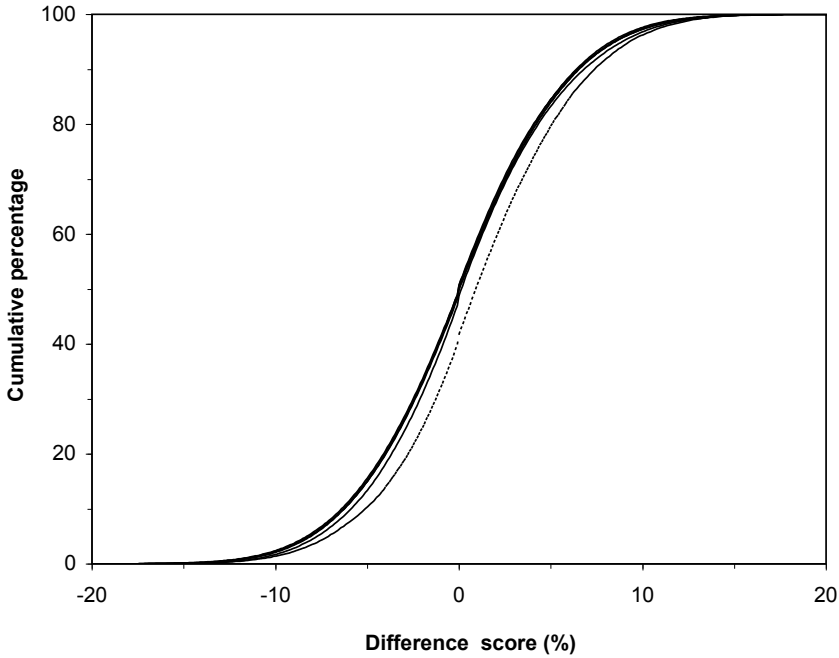

Figure 2. Cumulative distribution of difference scores (percent lexical LDND) for population ratios of 1–10, 10–100, 100–1,000, and 1,000–10,000 (superimposed solid lines), 10,000–100,000 (dashed line), and 100,000+ (dotted line).

A possible reason for the small effect sizes is that in each set of three languages, the phylogeny is inferred from the pairwise LDND percentages. Because of random errors in the percentages, the two closest languages according to the LDND will not always be the ones that diverged most recently. These errors will increase the variability of the difference scores and reduce the effect sizes. This problem can be alleviated by basing the phylogeny on the classification into genera, which are relatively noncontroversial groups inferred from information other than the ASJP data. The dashed line in Figure 3 is therefore constructed with triples in which two languages are in the same genus and the third language is in a different genus in the same family, with the first two languages considered the closest. Once again, effect sizes are small but positive if anything. So the limited effects of population size are not specific to the ASJP data.

The solid line in Figure 3 is based on triples in which the average LDND from the two closest languages to the third ranges up to 90%. In some of these triples the LDND between the two closest languages themselves can be almost as high as 90%, and thus (1 − LDND) can be almost as low as 10%. Although Figure 1 shows that current population ratios retain some trace of earlier population ratios back to time depths corresponding to a (1 − LDND) of 10%, the same figure

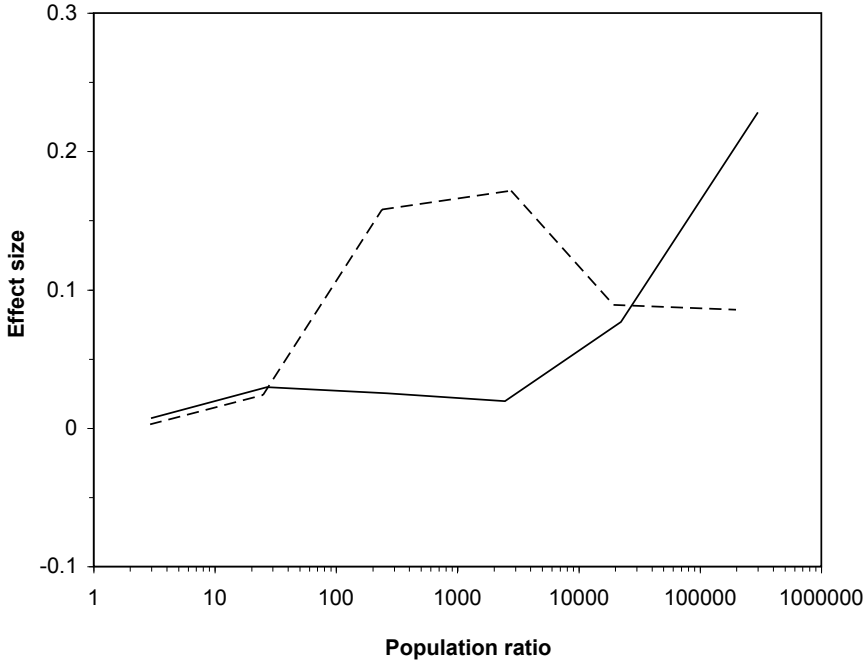

Figure 3. Effect of population on lexical LDND as a function of ratio of populations for triples of languages in the same family for which the average LDND from the two closest languages to the third is less than 90% (solid line) and for triples with two languages in the same genus and the third in a different genus in the same family (dashed line).

shows retention of much more information from more recent times, particularly those corresponding to a (1 − LDND) above 80%. This fact suggests testing the effect of population separately in subgroups of triples with a (1 − LDND) between their two closest languages of 100%–80%, 80%–60%, 60%–40%, 40%–20%, and 20%–10%. To measure the effect of population in each subgroup, we calculated the Pearson correlation coefficient between difference score and population ratio across the triples in each subgroup.

Table 1 gives the correlation for each subgroup, followed by the number of triples in the subgroup and the number of families with triples in the subgroup. The last row of the table gives the same information for all the triples taken together. The last column of the table gives the results of bootstrap significance tests on the correlations, with families as the unit of analysis. For each bootstrap sample, 188 families are chosen at random with replacement from the set of 188 families represented in the ASJP data. Each of 1,000 such samples is analyzed just like the real data. The test statistic is the proportion of the bootstrap samples in which the correlation is positive; this proportion is given in the last column of the table. The proportion is expected to be near 0 if the true correlation is negative and

Table 1. Correlation of Difference Score with Population Ratio, Numbers of Triples and Families, and Proportion of Bootstrap Samples with Negative Correlation for Triples with Different Ranges of (1 − LDND) Between Their Two Closest Languages

1 − LDND	*Correlation*	*Number of Triples*	*Number of Families*	*Proportion*
100%–80%	−0.27	12,949	10	0.017
80%–60%	−0.05	183,699	44	0.324
60%–40%	−0.05	885,758	56	0.058
40%–20%	0.02	10,489,349	65	0.637
20%–10%	0.00	7,807,360	41	0.669
100%–10%	0.01	19,379,115	77	0.649

near 1 if the true correlation is positive. For a two-tailed test at the 0.05 level, the two critical values are set at 0.025 and 0.975, such that the observed correlation is significantly different from 0 if the proportion is below 0.025 or above 0.975. Because families are the unit of analysis, the tests assume that families are independent of each other, but the tests do not assume independence among triples within families.

Only one correlation is significant: the negative correlation for the subgroup of triples in which the two closest languages are very close, with a (1 − LDND) above 80%. For comparison, (1 − LDND) is 66% between Dutch and Afrikaans, languages that have been separate for only a few hundred years. The negative correlation suggests that for a short time after their divergence, the smaller of two languages or dialects actually changes more slowly than the larger one. The fact that the correlation is significant suggests that the bootstrap test is capable of detecting an effect of population size if it actually exists in the data. The other correlations are small and not significant, despite the much larger numbers of triples and families in the other subgroups. The bottom line of the table shows that the positive trend in the solid line in Figure 3 is far from significant. This null result is consistent with the small effect sizes in Figure 3.

Several of the pairs that contribute to the effect of population sizes on language change rates for closely related dialects [(1 − LDND) above 80%] contain a standard or cross-regional variant (trade language) paired with an endangered minority variant. The standard or cross-regional variants that feed into the results are Western Panjabi (Pakistan), Plateau Malagasy (Madagascar), Ibibio (Nigeria), Amganad Ifugao (Philippines), Maori (New Zealand), and Rarotongan (Cook Islands). We speculate that as standard variants emerge, these are influenced by speakers shifting from different dialects such that a short-term increase in the rate of change is observed. The apparent effect of population sizes on language change rates is already gone by the time the dialects have been separated for a few hundred years (e.g., Afrikaans and Dutch). Therefore what matters is not the population size as such, but factors pertaining specifically to dialects. The lexical database we draw on is constructed with the intention of maximal coverage of different languages of

the world, and only occasionally do we have good coverage of different dialects. Thus it is not presently possible to present a more systematic investigation of the restricted effect observed.

Conclusions

In this paper we have reviewed different types of evidence for an effect of population sizes on the rate of language change. The preponderance of rare features in small languages and the greater genealogical diversity in areas with small languages are consistent with faster rates of change in smaller languages, as Nettle (1999a) suggested, but both findings are also amenable to alternative explanations. Nerbonne and Heeringa (2007) described mixed results in studies of diffusion among dialects as a function of the number of speakers. Wichmann et al. (2008) showed that typological features do not seem to change faster in smaller populations. The test assumed a degree of uniformity in the average time depth of language pairs within families and genera, which is a potential weakness because families and genera are not defined with consistency across the world's languages. In the present paper, therefore, we derived another test from the null hypothesis that the ultrametric inequality is satisfied by lexical distances among triples of related languages. The test was applied to lexical data from nearly one-half of the world's languages, and again the result was mainly negative, the exception being an apparently faster rate of change in the larger of two closely related variants. A possible explanation may be the influence on emerging standard (or cross-regional) variants from speakers who shift from different dialects to the standard. This question needs to be explored with more data from closely related dialects. Whether the explanation holds up or an alternative one is needed, we can exclude from consideration a purely demographic effect. Our empirical investigations strongly indicate that the sizes of speaker populations do not in and of themselves determine rates of language change.

Different computer simulation models have been applied in the investigation of the same issue. Among the models that assumed a crucial influence on changes in a language issuing from key individuals, two produced results that suggested that large languages change more slowly than small ones under some conditions at least. A third model, which assumed that linguistic changes spread more locally, did not show any relation between language change rates and population sizes. A comparison of different kinds of networks showed an effect of population only in the perhaps rather unrealistic situation in which all individuals have an equal number of connections to their nearest neighbors, whereas no effect emerged for asymmetric networks (random, small-world, and scale-free networks).

In conclusion, several different empirical data sets and methods suggest that any relation between population size and language change is negligible in comparison to the other factors involved in language change. We therefore offer the absence of such a relationship as one criterion for choosing among models of language change. Although the computer simulations cover only a small subset

of the possible models and parameter settings, they predict different results from different models and thus can help us to select the most plausible model for the propagation of linguistic changes. Given the evidence presented in this paper, the most plausible model for the greater part of human history—the one that has led to the present linguistic diversification—is one in which changes propagate at a local level in a type of network where the individuals have different degrees of connectivity.

Acknowledgments We thank Dietrich Stauffer, James Steele, and three anonymous referees for their comments on earlier versions of this paper. For their efforts in gathering and analyzing the lexical data described here, we would like to acknowledge the other members of the Automatic Similarity Judgment Project, in alphabetical order: Dik Bakker, Oleg Belyaev, Cecil H. Brown, Pamela Brown, Matthew Dryer, Dmitri Egorov, Patience Epps, Anthony Grant, Hagen Jung, Robert Mailhammer, André Müller, Viveka Velupillai, and Kofi Yakpo.

Received 13 January 2009; revision accepted for publication 16 April 2009.

Literature Cited

Barabási, L. A., and R. Albert. 1999. Emergence of scaling in random networks. *Science* 286:509–512.

Brown, C. H., E. W. Holman, S. Wichmann et al. 2008. Automated classification of the world's languages: A description of the method and preliminary results. *STUF Lang. Typol. Universals* 61:285–308.

Chamberlain, A. 2009. Archaeological demography. *Hum. Biol.* 81(2–3):275–286 (this issue).

Cohen, J. 1992. A power primer. *Psychol. Bull.* 112:155–159.

Cysouw, M. 2009. Quantitative explorations of the worldwide distribution of rare characteristics, or the exceptionality of northwestern European languages. In *Expecting the Unexpected: Exceptions in Grammar*, H. Simon and H. Wiese, eds. Berlin: Mouton de Gruyter (in press).

Dryer, M. S. 2005a. Genealogical language list. In *The World Atlas of Language Structures*, M. Haspelmath, M. Dryer, D. Gil et al., eds. Oxford, U.K.: Oxford University Press, 584–643.

Dryer, M. S. 2005b. Order of subject, object, and verb. In *The World Atlas of Language Structures*, M. Haspelmath, M. Dryer, D. Gil et al., eds. Oxford, U.K.: Oxford University Press, 334–337.

Gordon, R. G., Jr., ed. 2005. *Ethnologue*, 15th ed. Dallas: SIL International.

Haspelmath, M., M. Dryer, D. Gil et al., eds. 2005. *The World Atlas of Language Structures*. Oxford, U.K.: Oxford University Press.

Heeringa, W. 2004. *Measuring Dialect Pronunciation Differences Using Levenshtein Distance*. Ph.D. dissertation, Rijksuniversiteit Groningen, Groningen, Netherlands.

Henrich, J. 2004. Demography and cultural evolution: How adaptive cultural processes can produce maladaptive losses—The Tasmanian case. *Am. Antiq.* 69:197–214.

Holman, E. W., C. Schulze, D. Stauffer et al. 2007. On the relation between structural diversity and geographical distance among languages: Observations and computer simulations. *Linguist. Typol.* 11:395–423.

Holman, E. W., S. Wichmann, C. H. Brown et al. 2008. Explorations in automated language comparison. *Fol. Linguist.* 42:331–354.

Ke, J., T. Gong, and W. S.-Y. Wang. 2008. Language change and social networks. *Commun. Computat. Phys.* 3:935–949.

Levenshtein, V. I. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Cybernet. Control Theor.* 10:707–710.

Müller, A., V. Velupillai, S. Wichmann et al. 2009. *ASJP World Language Tree: Version 1* (April 2009). Available at http://email.eva.mpg.de/~wichmann/ASJPHomePage.htm.

Nerbonne, J., and W. Heeringa. 2007. Geographic distributions of linguistic variation reflect dynamics of differentiation. In *Roots: Linguistics in Search of Its Evidential Base*, S. Featherston and W. Sternefeld, eds. Berlin: Mouton de Gruyter, 267–297.

Nettle, D. 1999a. Is the rate of linguistic change constant? *Lingua* 108:119–136.

Nettle, D. 1999b. Linguistic diversity of the Americas can be reconciled with a recent colonization. *Proc. Natl. Acad. Sci. USA* 96:3325–3329.

Nettle, D. 1999c. Using social impact theory to simulate language change. *Lingua* 108:95–117.

Nichols, J. 1995. Diachronically stable structural features. In *Historical Linguistics 1993: Selected Papers from the 11th International Conference on Historical Linguistics*, H. Andersen, ed. Amsterdam: John Benjamins, 337–355.

Nichols, J., and T. Warnow. 2008. Tutorial on computational linguistic phylogeny. *Lang. Linguist. Compass* 2:760–820.

Nowak, A., J. Szamrej, and B. Latané. 1990. From private attitude to public opinion: A dynamical theory of social impact. *Psychol. Rev.* 97:362–376.

Powell, A., S. Shennan, and M. G. Thomas. 2009. Late Pleistocene demography and the appearance of modern human behavior. *Science* 324:1298–1301.

Richerson, P. J., R. Boyd, and R. L. Bettinger. 2009. Cultural innovations and demographic change. *Hum. Biol.* 81(2–3):211–235 (this issue).

Serva, M., and F. Petroni. 2008. Indo-European languages tree by Levenshtein distance. *Europhys. Lett.* 81(6):68005. Available at http://www.iop.org/EJ/journal/EPL.

Shennan, S. J. 2000. Population, culture history, and the dynamics of culture change. *Curr. Anthropol.* 41:811–835.

Swadesh, M. 1950. Salish internal relationships. *Intl. J. Am. Linguist.* 16:157–167.

Swadesh, M. 1955. Towards greater accuracy in lexicostatistic dating. *Intl. J. Am. Linguist.* 21:121–137.

Trudgill, P. 1974. Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. *Language Soc.* 3:215–246.

Watts, D. J., and S. H. Strogatz. 1998. Collective dynamics of small-world networks. *Nature* 393:440–442.

Wichmann, S., and E. W. Holman. 2009. *Temporal Stability for Linguistic Typological Features*. Munich: LINCOM Europa (in press). Prepublication version available at http://email.eva.mpg.de/~wichmann/WichmannHolmanIniSubmit.pdf.

Wichmann, S., D. Stauffer, C. Schulze et al. 2008. Do language change rates depend on population size? *Adv. Complex Syst.* 11(3):1–13.