Role of Demographic Dynamics and Conflict in the Population-Area Relationship for Human Languages

Susanna C. Manrubia[1]*, **Jacob B. Axelsen**[1¤], **Damián H. Zanette**[2]

1 Centro de Astrobiologa, Instituto Nacional de Técnica Aeroespacial and Consejo Superior de Investigaciones Científicas, Torrejón de Ardoz, Madrid, Spain, 2 Consejo Nacional de Investigaciones Cientficas y Técnicas, Centro Atómico Bariloche and Instituto Balseiro, San Carlos de Bariloche, Argentina

Abstract

Many patterns displayed by the distribution of human linguistic groups are similar to the ecological organization described for biological species. It remains a challenge to identify simple and meaningful processes that describe these patterns. The population size distribution of human linguistic groups, for example, is well fitted by a log-normal distribution that may arise from stochastic demographic processes. As we show in this contribution, the distribution of the area size of home ranges of those groups also agrees with a log-normal function. Further, size and area are significantly correlated: the number of speakers p and the area a spanned by linguistic groups follow the allometric relation $a \propto p^z$, with an exponent z varying accross different world regions. The empirical evidence presented leads to the hypothesis that the distributions of p and a, and their mutual dependence, rely on demographic dynamics and on the result of conflicts over territory due to group growth. To substantiate this point, we introduce a two-variable stochastic multiplicative model whose analytical solution recovers the empirical observations. Applied to different world regions, the model reveals that the retreat in home range is sublinear with respect to the decrease in population size, and that the population-area exponent z grows with the typical strength of conflicts. While the shape of the population size and area distributions, and their allometric relation, seem unavoidable outcomes of demography and inter-group contact, the precise value of z could give insight on the cultural organization of those human groups in the last thousand years.

Introduction

Despite the extraordinary sociocultural diversity of human communities [1], a number of statistical features characterizing human populations follow remarkably simple patterns. Often, those patterns arise from basic mechanisms with a relatively small contribution of cultural traits. As examples we find the distribution of city sizes across the world –grounded in the fluctuating growth and decline of their populations [2]– or the use of space in hunter-gatherer groups –which relies on metabolic and environmental constraints and group dynamics [3]. A macroecological approach to social organization focusing on the exchange of energy and resources among individuals and societies, and with their environment, is a promising avenue toward a quantitative understanding of human space use, of population structure, and of cultural and linguistic diversity patterns [4].

One of the most basic traits that define a group of human beings is sharing a common language: the evolution of languages can be closely mapped onto the evolution of cultural groups despite the fact that, at present, the number of the latter is almost certainly larger than that of languages [1]. Actually, studies of linguistic phylogenies can be as informative as studies on gene evolution [5,6], and linguistic taxonomy presents scaling properties [7] analogous to those of biological taxonomy [8,9]. The similarities

between biological and linguistic patterns go beyond a mere analogy [10–12]. As with biological species richness [13], linguistic diversity attains its maximum value close to the equator [14] and decreases with latitude. Ecological processes, historical factors, and home ranges for biodiversity [15], as well as environmental variables and political complexity for ethnolinguistic groups [16], have been put forward as the main drivers behind tropical richness. Still, there is no agreement on the origin of the latitudinal gradient in species richness [17], probably indicating that, in this case, several different mechanisms contribute in similar amounts.

A quantity related to linguistic diversity is the average area (or home range) covered by a human group. Data on the geographic ranges of languages [18] show that the area covered by a group increases with latitude [14] and thus follows Rapoport's rule, which describes this same phenomenon for home ranges of biological species [19]. Beyond environmental variables, the home range of ethnnolinguistic groups has been found to depend on their subsistence strategy and on their political complexity [16]. In the case of hunter-gatherers, the home range as a function of group size has been taken as an indicator of its environmental needs and of its social behavior [3].

Though the home range is an important feature characterizing biological species or human groups, there is a quantity that seems

to be more directly affected by environmental characteristics and cultural practices: the size of the group, measured by the number of members. An analysis of the correlation between the sizes p of linguistic groups and the areas a over which they spread reveals that both variables are significantly correlated and fulfill an allometric relationship $a \propto p^z$, with a variable exponent z. The distribution of abundance and range of biological species fulfils an analogous relationship, though there is no agreement on the processes originating it [20,21]. In this paper, we present evidence on the universality of the relation between population and area for human groups and suggest, on the basis of a well-motivated mathematical model, that this relationship, as well as the functional form of the distribution of areas, stems from demographic processes and the interaction between neighboring groups.

Distribution of Sizes of Linguistic Groups

According to linguists' classifications, there are about 6900 languages currently spoken all over the world [22]. Among them, 516 are classified as nearly extinct (with less than 100 speakers each), while the ten most abundant languages are spoken by a total of 2.6×10^9 people. The uneven distribution of language sizes [23,24], measured as the number of speakers per language (see Fig. 1), is well described by a log-normal frequency distribution [11]. Only rare languages deviate significantly from the lognormal shape and are more frequent than expected. Early analyses of the distribution of the number of individuals in a given biological species pointed out to a log-normal distribution as well [25], though, subsequently, other distributions for species abundance have been also observed in different ecosystems [26].

It is estimated that the population of the world in year 1000 was $P_0 = 3.1 \times 10^8$ [27]. This number has grown at least eighteen-fold in ten centuries to reach the $P_T = 5.7 \times 10^9$ people accounted for by the Ethnologue in year 2000 [22]. The log-normal distribution of language sizes can be explained by means of a model that considers the stochastic multiplicative growth of independent populations in the last thousand years as the main mechanism behind the evolution in size of linguistic groups [28]. The only parameters of the model are those retrieved by assuming an exponential growth of the world population from P_0 to P_T, and the current distribution of language sizes, yielding for the annual growth rate a mean value $\bar{\alpha} \approx 1.0029$ and a mean square dispersion $\sigma_\alpha \approx 0.096$ [28]. Simple models for the distribution of biological species abundance also rely on demographic dynamics, though they typically keep the total population bounded [26,29].

Results

Home Range Distribution

Data on the geographical distribution of languages have been obtained from the fifteenth edition of the Ethnologue [22]. In order to avoid boundary effects due to small islands and coasts, only languages whose centroids are located more than 5 kilometers inland and belong to the 100 largest terrestrial landmasses have been considered (see Materials and Methods). An example of how territory is fragmented into different linguistic groups is shown in Fig. 2, which also illustrates the quality of the data analyzed.

The areas covered by linguistic groups also follow a broad distribution compatible with a log-normal function (Fig. 1). This is a remarkable observation considering the plethora of different mechanisms that underlie the spread of languages and the existence of patterns such as described by Rapoport's rule [19]. In analogy with the distribution of language sizes, one may hypothesize that the variation in the home range of a group should be described as well by a multiplicative process. However, the justification of such a mechanism is not as natural as it appears for the size of linguistic groups, where population growth and decline are chiefly driven by the multiplicative stochastic effect of birth and death events. Should demographic dynamics be also behind
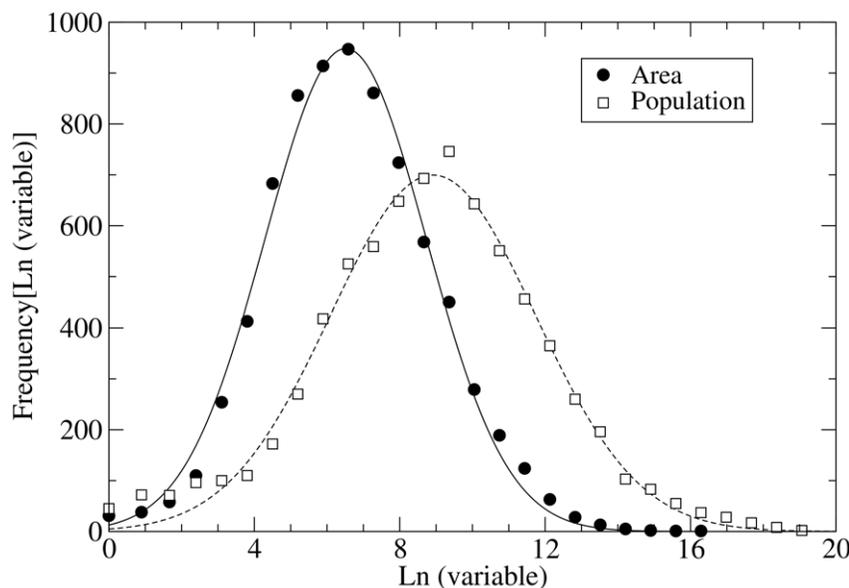
Figure 1. Normalized frequency distribution of language sizes and language areas (home ranges) over the world. When a language is spoken in two or more disconnected domains, the total area (sum of all disconnected polygons) is represented. The data set includes all living languages plus about 700 recently extinct languages listed in the Ethnologue [22]. Curves stand for Gaussian fittings on the logarithmic variables. Both fittings are statistically highly significant, yielding correlation coefficients $r^2 = 0.93$ for the distribution of population sizes and $r^2 = 0.94$ for the distribution of areas.
doi:10.1371/journal.pone.0040137.g001

Figure 2. Linguistic diversity in Papua New Guinea and surrounding islands. Each polygon corresponds to a domain where a single language is spoken. The matrix where polygons ar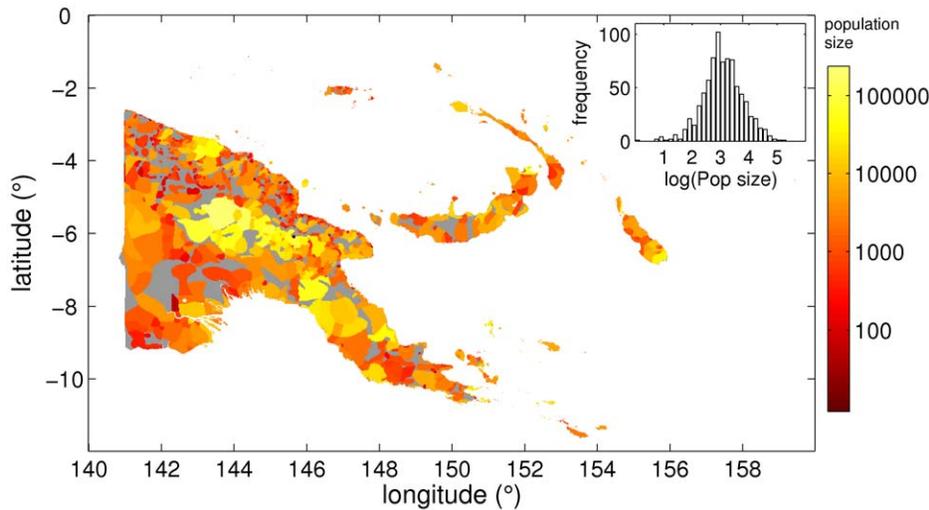e embedded (see Materials and Methods) has been mapped to an indexed image using log_{10} transformation of the population sizes for each language, as catalogued by SIL in the Ethnologue DB ver.15. The index values are colored according to the scale in the right, which indicates the absolute size of the group speaking each language. Unhabited landmass of Papua New Guinea is colored grey. The inset contains the population size histogram, with six bins per decade from 1 to 10^6.
doi:10.1371/journal.pone.0040137.g002

the observed distribution of language areas, population sizes and home ranges should be significantly correlated.

Relationship between Linguistic Group Size and Area

Following the observation that the size p_i of the population speaking language i and the area a_i it covers are log-normally distributed –and likely correlated– we analyzed the dependence between both variables, as shown in Fig. 3. Six different representative regions are shown: Africa, Asia, Europe, Papua New Guinea, and North and South America. In all cases, the correlation between the two variables is high, though it is also affected by a substantial dispersion.

The log-normal distributions of language sizes and areas imply that the transformed variables $q_i = \ln p_i$ and $s_i = \ln a_i$ are well fitted by Gaussian distributions. Their joint distribution can be approximated by a bivariate normal distribution (BND, see Materials and Methods). The two main quantities characterizing this joint distribution are the exponent z, which yields the slope of the major ellipse axis of the scatter plot (and is represented as a straight line in all plots of Fig. 3), and the coefficient ρ, which quantifies the degree of correlation between the two variables: the larger $|\rho|$, the narrower the ellipse and the more correlated q_i and s_i are.

A summary of the quantities characterizing each of the regions analyzed is shown in Table 1. The first four columns contain the averages $\bar{q}$ and $\bar{s}$, the slope z, and the area-population correlation ρ for human languages in each of the six regions mentioned, for the whole world, and for modern populations of hunter-gatherers (see below). As can be seen, the variation in the exponent z characterizing the allometric relationship is broad, while the correlation between population sizes and areas is similar in all cases, as shown by the narrow range spanned by ρ.

Model

The occurrence of a BND for the distribution of the logarithms of population sizes and areas for human languages plausibly suggests that their joint evolution can be accounted for by a stochastic multiplicative process correlating the two variables.

Specifically, the demographic pressure associated with population growth of a given human group should promote its geographical expansion. This tendency, however, will generally clash with similar trends in neighboring populations and a conflict may ensue due to the limited area available. On the other hand, a decrease in population numbers leads to a contraction of the group, with the consequent decrease in the occupied area.

Let us assume that the (logarithmic) number of individuals in a human group, characterized by their common language, obeys the recurrence equation.

$$q_i(t+1) = q_i(t) + \beta_i(t), \qquad (1)$$

where the index i identifies the group, and t is measured in years. The additive process (1) is the direct translation of the multiplicative process on the original variable p_i. At each step, the growth rate $\beta_i(t)$ is randomly drawn from a uniform distribution $\Pi(\beta)$ in the interval $(\epsilon - \eta, \epsilon + \eta)$ (see Materials and Methods). The half-width $\eta = 0.169$ and the mean value $\epsilon = -0.00186$ are taken from the average evolution of linguistic groups in the last thousand years [28].

Similarly, the evolution of the corresponding (logarithmic) area is assumed to obey.

$$s_i(t+1) = s_i(t) + \xi_i(t). \qquad (2)$$

The key ingredient of the model is how the dynamics of q_i and s_i are correlated or, in other words, how the corresponding growth rates depend on each other. We assume that $\xi_i(t)$ depends on the growth rate of the population size, $\beta_i(t)$, according to the following rules:

1. If $\beta_i(t) \leq 0$, $\xi_i(t)$ is drawn at random from a uniform distribution in the interval $[-r|\beta_i(t)|, 0]$. This represents a shrinking of the area when the population decreases.

2. If $\beta_i(t) > 0$, the group i enters into conflict with other groups. To represent a typical conflict event, a second group j is randomly
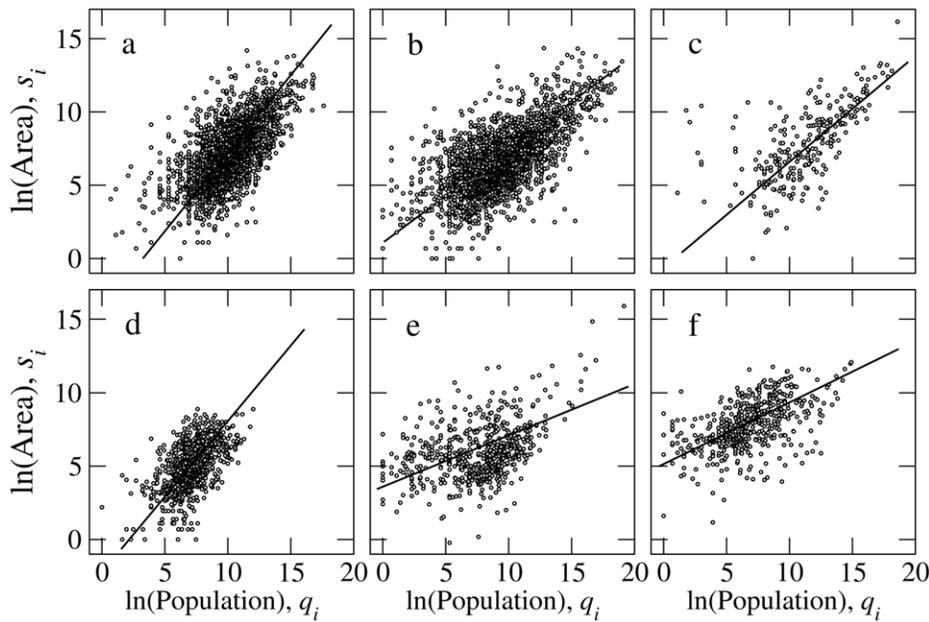
Figure 3. Scatter plot of the logarithm of the area ($s_i = \ln a_i$) and the logarithm of the population size ($q_i = \ln p_i$) for all languages. Six different representative regions are plotted: (a) Africa (2314 languages), (b) Asia (2333 languages), (c) Europe (260 languages), (d) Papua New Guinea (813 languages), (e) North America (585 languages), and (f) South America (517 languages). North and South American languages have been separated according to whether their centroids laid to the north or to the south of the 12°27′N parallel, which stands for the northernmost point of South America. The corresponding values of ρ and z are compiled in Table 1. The value of z for each case coincides with the slope of the line drawn in each panel. Axes labels and sizes are the same in all cases.
doi:10.1371/journal.pone.0040137.g003

chosen and the growth rates of i and j in the last year are compared:

(a) If $\beta_j(t) < \beta_i(t)$, $\xi_i(t)$ is drawn from the interval $[0, w\beta_i(t)]$. Since group i presents the largest growth rate, it prevails and its area grows.

(b) If $\beta_j(t) \geq \beta_i(t)$, $\xi_i(t) \in [-w\beta_i(t), 0]$. Group j prevails, and i's area shrinks.

According to these rules, the stochastic evolution of $s_i(t)$, which becomes correlated to that of $q_i(t)$, turns out to be fully determined by the two parameters r and w, respectively quantifying the spontaneous retreat and the outcome of conflicts, and by the distribution $\Pi(\beta)$. The exponent z and the correlation ρ, which quantify the statistical interdependence of language sizes and

areas, can be written as functions of the model parameters in analytical form (see Materials and Methods). For each choice of r and w, the model yields a single pair z, ρ.

In the present model, the number of human groups –and, therefore, of different languages– is preserved along the evolution, so that such events as language birth, fragmentation and death are not explicitly taken into account. Language extinction, which arguably was the most important among these processes in the last thousand years [11,22], can nevertheless be introduced by eliminating those groups whose population falls below a prescribed level, for instance, of one individual. However, we have verified by means of numerical simulations that this additional ingredient has no significant effect on the resulting values of z and ρ. On the

Table 1. Parameters characterizing the distribution of population size and area covered by human languages in several representative world regions.

	$\bar{q}$	$\bar{s}$	z	ρ	r	w
Africa	9.63	7.13	0.94 ± 0.02	0.64 ± 0.01	0.75 ± 0.02	2.30 ± 0.04
North America	7.61	5.97	0.35 ± 0.03	0.40 ± 0.03	–	–
South America	6.76	7.82	0.42 ± 0.03	0.53 ± 0.03	0.11 ± 0.01	1.59 ± 0.10
Asia	9.07	6.66	0.64 ± 0.01	0.66 ± 0.01	0.67 ± 0.01	1.73 ± 0.03
Europe	10.31	7.85	0.71 ± 0.05	0.64 ± 0.04	0.63 ± 0.04	1.92 ± 0.12
Papua N G	7.10	5.00	1.03 ± 0.05	0.56 ± 0.02	0.34 ± 0.02	2.61 ± 0.16
World	**8.91**	**6.51**	**0.57 ± 0.01**	**0.52 ± 0.01**	**0.100 ± 0.003**	**1.92 ± 0.04**
Hunter-gatherers	6.73	4.57	2.02 ± 0.10	0.48 ± 0.03	0.020 ± 0.002	3.84 ± 0.18

Columns represent the following data: $\bar{q}$ and $\bar{s}$ are the (logarithmic) averages of population size and area; z corresponds to the slope of the major ellipse axis relating q_i and s_i, and ρ measures their degree of correlation. Errors in both variables are shown. The values of r and w correspond to model parameters yielding the measured values of z and ρ within the estimated interval.
doi:10.1371/journal.pone.0040137.t001

other hand, the model does not exclude that a language evolves along time, as long as it preserves its identity through association with a given human population.

Note that the above rules impose no restriction on the growth of the geographical region occupied by each population. Thus, the total area covered by the system can increase indefinitely. This feature, which is manifestly unrealistic for recent stages in the evolution of human populations, can be corrected by assuming any prescribed time dependence for the total area –for instance, that it remains constant or that it grows monotonically approaching a limit value– and, accordingly, rescaling individual areas at each evolution step. This rescaling does not affect the correlation between areas and populations and, therefore, the values of z and ρ are not modified.

Expected Values of z and ρ

The two surfaces represented in Fig. 4 correspond to the values of z (Fig. 4(a)) and ρ (Fig. 4(b)) obtained by numerically solving the exact equations for the two variables in the intervals $r \in [0,1]$ and $w \in [0,4]$. Except for North America, whose area-population correlation is sensibly lower than for other regions, the model can quantitatively recover all empirical values for the groups included in Table 1.

Discussion

The size of linguistic groups and the area they cover are well fitted by log-normal probability distributions. The correlation between those two quantities supports that the demographic process defined by natural population growth is the main mechanism behind those distributions. The correlation between group size and area, described by our index ρ, is close to 0.52 (world average) for almost all cases analyzed. A multiplicative model relating the evolution of group sizes and areas is able to explain and quantitatively recover the observed patterns. In our analysis of the model, the exact equations for z and ρ were actually solved in a range of values of parameters $r \in [0,4]$ and $w \in [0,5]$ significantly broader than those shown in Fig. 4. Nevertheless, those pairs recovering empirical data fulfill $r < 1$ and $w > r$ (see Table 1 and Fig. 4). Values of r smaller than one imply that the decrease in area when a population shrinks is sublinear with respect to the decrease in population size. Though demographic change may be a primary cause behind areal variations, other causes such as the use of primary resources or the cultural value of the territory likely act against reducing the home range. On the other hand, $w > r$ indicates that the loss or gain in area experienced by two conflicting populations is larger than the change they would undergo when no conflict arises. This is also a plausible result of group clash in comparison to spontaneous retreat.

Additional support to the plausibility of the dynamical rules implemented in our model comes from independent data of 339 present-day traditional hunter-gatherer groups [18]. Despite the cultural control of natality they practice, the distributions of sizes and home ranges of those traditional societies are remarkably well fit by log-normal distributions. A covariance analysis of the relationship between the home range and the group size yields $z = 2.02$ and $\rho = 0.48$.

The special case of native North American languages, whose value of ρ cannot be recovered by the model, can be however qualitatively understood. Most of the linguistic territories reported by the Ethnologue database in the United States and in Canada correspond to very small areas, assimilable to reduced aboriginal communities and reservations (compare $\bar{s}$ in North America to

other continents). In these artificially created ethnic domains, it is expected that the (assigned) areas are much less correlated to the population size than in groups whose spatial distribution is governed by demographic dynamics. This explains the small value of ρ. Artificial territorial allowance may also explain the relatively small slope z, because of the limited possibilities of expansion. The population in such communities, in any case, is systematically depleted by emigration toward and assimilation by urban settlements of European origin. Consistently, other analyses of the mechanisms underlying the spread of ethnolinguistic groups have explicitly removed languages in the Americas and in Australia (which show restricted ranges) due to their colonial history and population replacement [16].

Though we cannot interpret the relation between the model parameters r and w, and the empirical quantities z and ρ in a straightforward way, there is an eye-catching correlation between the values of the slope z and those of w: the stronger the effect of a conflict, the larger the population-area exponent z (see Fig. 4(a)). Current data support that the relative number of casualties in violent conflicts has declined with the advent of modern states: *pre-state societies were far more violent than our own. (. . .) In tribal violence, the clashes are more frequent (. . .) and rates of death per battle are higher* [30]. In zones of anarchy, violence is also higher than in regions ruled by stable governments [31]. It seems reasonable that the fraction of population that perishes in violent conflicts correlates positively to changes in owned territory. If this is so, larger values of w should better describe those regions which, along the historical period relevant to the demographic processes discussed here, were socially less organized, and *vice versa*.

Though, admittedly, our model does not consider all the processes affecting the population-area relationship in human groups, the conclusions drawn from its solution are robust under changes such as group extinction (when the population falls below a single individual) or under quantitative changes in η and ϵ. We have used average values for η and   (which come from data corresponding to world population) and have applied them to all regions. Those two quantities and other relevant parameters as r and w almost certainly vary in time, while we have considered them to be constant. Analyses aimed at predicting the fate of specific (small) regions should take into account region- and time-dependent parameters, and explicitly model the spatial allocation of population density. Also, since the location of different populations with respect to each other was disregarded, we have considered that interaction between any two groups is equally probable, while in reality conflict is much more likely to arise between neighboring groups.

Certainly, environmental variables, subsistence strategies, cultural behaviour or political complexity play a role in determining the spread of linguistic groups. These facts do not contradict our main hypothesis, namely, that demographic dynamics and conflict explain the allometric relationship between group size and area, and their correlation. Specific environmental and cultural variables may be instrumental in determining which particular groups grow and which others shrink, and very likely do condition the outcome of conflicts. In this sense, their predictive power in particular cases is far beyond that of our statistical model. Still, we maintain that simple processes as the ones here implemented might be the main drivers to ascertain how human groups are distributed in space, and to what extent their statistical patterns resemble those observed in non-cultural species.
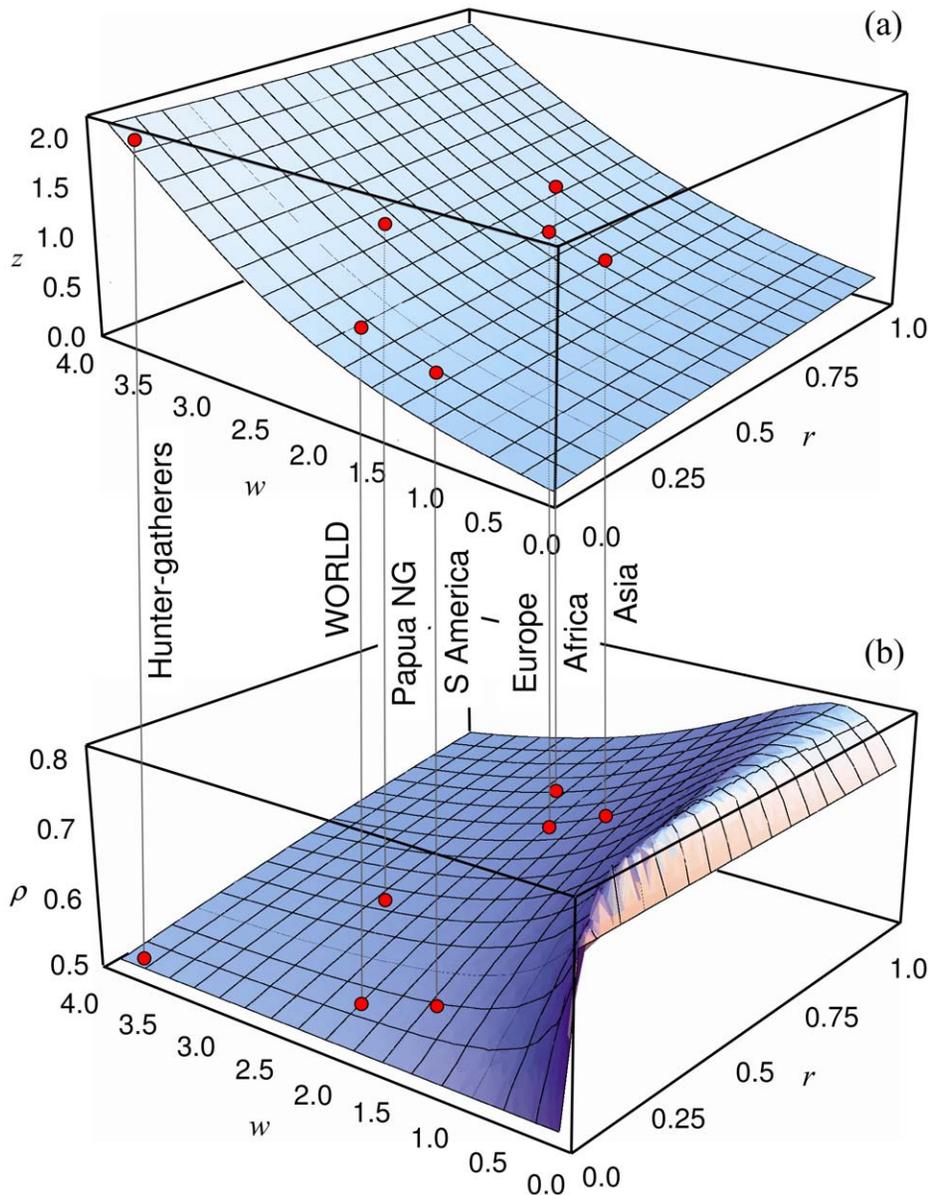
Figure 4. Model results. (a) Exponent z and (b) correlation ρ yielded by the model as a function of the parameters respectively characterizing spontaneous retreat and conflict outcome, r and w. The population-area relationship is recovered with different parameters for all world regions, except for North America (not shown in the plots). Data for the world considers all points represented in Fig. 3. Though z and ρ are, in principle independent variables (in terms of an arbitrary BND), the model establishes a quantitative relationship between them. The allowed values of ρ when z remains fixed can be read from this figure, and *vice versa*. As an example, note that, if $z \simeq 2$, ρ is to be found in the range $(0.5, 0.6)$, approximately.
doi:10.1371/journal.pone.0040137.g004

Materials and Methods

Data Origin

Data on the size of linguistic groups and their geographical properties have been obtained from the World Language Mapping System (WLMS 3.2) of the Ethnologue, version 15 by The Global Mapping Inc. (http://www.worldgeodatasets.com/). Permission to publish the processed data was received. Considered regions correspond to polygons with centroids located more than 5 kilometers inland of the shoreline. The shoreline is defined in the Global Self-consistent, Hierarchical, High-resolution Shoreline Database (GSHHS) by The National Oceanic and Atmospheric Administration (NOAA, http://www.ngdc.noaa.gov/mgg/

shorelines/gshhs.html). Polygons were embedded in a matrix at a resolution of 48 pixels per degree using the 'poly2mask' function in MATLAB. Areas were calculated using the 'areaint' function of the Mapping Toolbox in MATLAB version 7.

Bivariate Normal Distribution

The joint distribution of two variables q and s that independently follow Gaussian distributions can be approximated by the bivariate normal distribution (BND),

$$F(q,s) = N \exp\left[-\frac{1}{2}(\mathbf{r} - \bar{\mathbf{r}})\Sigma^{-1}(\mathbf{r} - \bar{\mathbf{r}})^{\dagger} \right] \qquad (3)$$

with $\mathbf{r}=(q,s)$, $\bar{\mathbf{r}}=(\bar{q},\bar{s})$, $\bar{q}=M^{-1}\sum_i q_i$, $\bar{s}=M^{-1}\sum_i s_i$, and $N=\left[2\pi\sigma_q\sigma_s\sqrt{1-\rho^2}\right]^{-1}$. M is the number of data points, and the index i runs over these data. The covariance matrix Σ is given by

$$\Sigma=\begin{pmatrix} \sigma_q^2 & \rho\sigma_q\sigma_s \\ \rho\sigma_q\sigma_s & \sigma_s^2 \end{pmatrix} \qquad (4)$$

with $\sigma_q^2=M^{-2}\sum_i(q_i-\bar{q})^2$, $\sigma_s^2=M^{-2}\sum_i(s_i-\bar{s})^2$, and $\rho\sigma_q\sigma_s=M^{-2}\sum_i(q_i-\bar{q})(s_i-\bar{s})$. The eigenvectors of Σ can be written as $(1,z)$ and $(-z,1)$, where z corresponds to the slope of the major ellipse axis of the scatter plot. The value of ρ quantifies the degree of correlation between the two variables: the larger $|\rho|$, the narrower the ellipse.

Analytical Solution of the Model

The marginal distribution for the population growth rate, $\Pi(\beta)$, which is uniform over the interval $(\epsilon-\eta,\epsilon+\eta)$, can be written as

$$\Pi(\beta)=\frac{1}{2\eta}u(\beta;\epsilon-\eta,\epsilon+\eta), \qquad (5)$$

where $u(x;x_1,x_2)=\theta(x-x_1)-\theta(x_2-x)$ and $\theta(x)$ is the Heaviside step function. Then, the joint distribution function for the growth rates of populations and areas is

$$f(\beta,\xi)=\Pi(\beta)[\theta(-\beta)u(\xi;0,r\beta)+\theta(\beta)u(\xi;-w\beta,0)\pi_+(\beta)$$

$$+\theta(\beta)u(\xi;0,w\beta)\pi_-(\beta)], \qquad (6)$$

with $\pi_\pm(\beta)=\pm\int_\beta^{\pm\infty}\Pi(\beta')d\beta'$.

By virtue of the central limit theorem, for sufficiently long times (i.e. when the contribution of the initial conditions can be neglected), the mean values and the (co)variances of $q_i(t)$ and $s_i(t)$ will be given by t multiplied by the respective quantities for β and ξ over their joint distribution $f(\beta,\xi)$. Consequently, the correlation ρ and the slope z of the (logarithmic) populations and areas will be the same as those corresponding to β and ξ. In other words, in the long-time limit, ρ and z can be expressed in terms of r, w, and the parameters ϵ and η in $\Pi(\beta)$. We recall that, as stated above, the latter can be fixed from real data on the evolution of human populations. The expressions for ρ and z in terms of r and w are algebraically very involved but, nonetheless, can be found analytically and dealt with by standard computational means.

Estimation of Errors

Errors in z and ρ, as reported in Table 1, were obtained by standard bootstrapping techniques. Given the values of z and ρ obtained from the best fit of the bivariate normal distribution for each world region, we extracted a set of data points at random from the predicted distribution. For the random set, whose number of data equaled that of the corresponding world region, the values of z and ρ were recalculated. The assigned errors were the dispersions in the distributions of z and ρ obtained from 1000 surrogate data sets. The errors in the model parameters r and w have been obtained from the analytical relationship provided by the analytical solution of the model.

Author Contributions

Conceived and designed the experiments: SCM JBA DHZ. Performed the experiments: SCM JBA DHZ. Analyzed the data: SCM JBA DHZ. Contributed reagents/materials/analysis tools: SCM JBA DHZ. Wrote the paper: SCM JBA DHZ.

References

1. Pagel M, Mace R (2004) The cultural wealth of nations. Nature 428: 275–278.
2. Zanette DH, Manrubia SC (1997) Role of intermittency in urban development: A model of large-scale city formation. Phys Rev Lett 79: 523–526.
3. Hamilton MJ, Milne BT, Walker RS, Brown JH (2007) Nonlinear scaling of space use in human hunter-gatherers. Proc Natl Acad Sci U S A 104: 4765–4769.
4. Burnside WR, Brown JH, Burger O, Hamilton MJ, Moses M, et al. (2011) Human macroecology: linking pattern and process in big-picture human ecology. Biol Rev 87: 194–208.
5. Cavalli-Sforza LL, Piazza A, Menozzi P, Mountain J (1988) Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. Proc Natl Acad Sci USA 85: 6002–6006.
6. Gray RD, Drummond AJ, Greenhill SJ (2009) Language phylogenies reveal expansion pulses and pauses in Pacific settlement. Science 323: 479–483.
7. Zanette DH (2001) Self-similarity in the taxonomic classification of human languages. Adv Compl Sys 4: 281–286.
8. Yule GU (1924) A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis. Proc R Soc Lond B 213: 21–87.
9. Burlando B (1990) The fractal dimension of taxonomic systems. J Theor Biol 146: 99–114.
10. Nettle D (1998) Explaining global patterns of language diversity. J Antrhopol Archaeol 17: 354–374.
11. Sutherland WJ (2003) Parallel extinction risk and global distribution of languages and species. Nature 423: 276–279.
12. Pagel M (2009) Human language as a culturally transmitted replicator. Nat Rev Genet 10: 405–415.
13. Kathleen Lyons S, Willig MR (2002) Species richness, latitude, and scale-sensitivity. Ecology 83: 47–58.
14. Mace R, Pagel M (1995) A latitudinal gradient in the density of human languages in North America. Proc R Soc Lond B 261: 117–121.
15. Pimm SL, Brown JH (2004) Domains of diversity. Science 304: 831–833.
16. Currie TE, Mace R (2009) Political complexity predicts the spread of ethnolinguistic groups. Proc Natl Acad Sci U S A 106: 7339–7344.
17. Willig MR, Kaufman DM, Stevens RD (2003) Latitudinal gradients of biodiversity: Pattern, process, scale, and synthesis. Ann Rev Ecol Sys 34: 273–309.
18. Moseley C, Asher RE (1994) Atlas of the World's languages. Routledge, London.
19. Rapoport EH (1982) Aerography: Geographic strategies of species. Pergamon Press, New York.
20. Harte J, Blackburn T, Ostling A (2001) Self-similarity and the relationship between abundance and range size. Am Nat 157: 374–386.
21. Gaston KJ, He F (2002) The distribution of species range size: a stochastic process. Proc R Soc Lond B 269: 1079–1086.
22. Gordon RG, ed. (2005) Ethnologue: Languages of the World, 15th edition. SIL International, Dallas.
23. Gomes MAF, Vasconcelos GL, Tsang IJ, Tsang IR (1999). Scaling relations for diversity of languages. Phys A 271: 489–495.
24. Schwämmle V, Queirós SMD, Brigatti E, Tchumatchenko T (2009). Competition and fragmentation: a simple model generating lognormal-like distributions. New J Phys 11: 093006.
25. Preston FW (1948). The commonness and rarity of species. Ecology 29: 254–283.
26. McKane A, Alonso D, Solé RV (2000) Mean-field stochastic theory for species-rich assembled communities. Phys Rev E 62: 8466–8484.
27. United Nations, Department of Economic and Social Affairs. Available: www.un.org/esa/population/publications/sixbillion. Accessed 2012 Jun 15.
28. Zanette DH (2008) Demographic growth and the distribution of language sizes. Int J Mod Phys C 19: 237–247.
29. Volkov I, Banavar JR, Hubbell SP, Maritan A (2003) Neutral theory and relative species abundance in ecology. Nature 424: 1035–1037.
30. Pinker S (2007) A history of violence. The New Republic, 19th March.
31. Pinker S (2011) The better angels of our nature – Why violence has declined. Viking Adult.