CAES Vol. 2, № 4 (December 2016)

An examination of the calibration of linguistic distance. II. Covariates

Simon Brown

Deviot Institute, Deviot, Tasmania, Australia;
College of Public Health, Medical and Veterinary Sciences, James Cook University, Queensland, Australia;
e-mail: Simon.Brown@deviotinstitute.org

Abstract

The assessment of linguistic relatedness is often based on the analysis of word lists to estimate the distance between languages. To relate the distance to a divergence time requires established calibration data and a model representing the underlying processes of language change. However, the available calibration data may be based on different sorts of evidence, come from different geographical regions and represent different language families, for example. Covariates such as these can impact significantly on the estimated divergence time. This reinforces concerns about the selection of calibration data, especially when just a small number of values are chosen.

Key words: dating evidence; geographical region; linguistic distance; model dependence

Introduction

Despite the difficulties associated with the use of lexical distance in the measurement of relatedness (Akulov 2015b, Brown 2016b, Hoijer 1956, Matisoff 1990, Rea 1958, Teeter 1963), two common measures are based on the analysis of word lists. The first is the normalised Levenshtein distance (LDN) from which the second, the LDN divided (LDND), is derived (Bakker *et al.* 2009). For two languages (α and β) LDN is the average Levenshtein distance (d_L, (Levenshtein 1966)) between n pairs of words normalised to the length (L_{ii}) of the longer of the two words

$$D_o(\alpha,\beta)=\frac{1}{n}\sum_{i=1}^{n}\frac{d_L(\alpha_i,\beta_i)}{L_{ii}}, \tag{1}$$

where α_i and β_i represent word i in languages α and β, respectively. The LDND is

$$D_s(\alpha,\beta)=\frac{D(\alpha,\beta)}{\Gamma(\alpha,\beta)}, \tag{2}$$

where LDN is normalised to the 'global distance'

$$\Gamma(\alpha,\beta)=\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\frac{d_L(\alpha_i,\beta_j)}{L_{ij}} \tag{3}$$

(Petroni & Serva 2010) between the $n(n-1)$ pairs of words of different meanings in the word lists of α and β (Brown 2016c).

Part 1 was published in CAES vol. 2 no. 2

To relate either of these measures of distance to the divergence time (t) between languages requires established calibration values. While this process is model-dependent (Brown 2016b), one simple expression for this relationship is

$$D_x = \phi_x\left(1 - \exp\left(-k_x t\right)\right), \tag{4}$$

where D_x may be D_o (1) or D_s (2) depending on whether the distance measure is LDN or LDND, respectively, and that the values of ϕ_x and k_x also depend on whether D_x is D_o (in which case $x = o$) or D_s ($x = s$). However, (4) is also based on values of the rate constant (k_x) for the divergence process and the upper limit (ϕ_x) of D_x that are estimated from appropriate calibration data (Brown 2016b). Given (4), it is unsurprising that the calibration of D_o is relatively insensitive to differences between large divergence times. For example, this is the case for t in excess of about 4500 years for Indo-European languages, although this estimate was based on values of k_o and ϕ_o derived from only a few calibration values (Brown 2016b).

It has been shown that estimates of D_o for Indo-European languages are intrinsically uncertain (Brown 2016a). Moreover, it is mathematically inevitable that LDND is greater than LDN, as has been confirmed for Indo-European languages (Brown 2016c), and, as LDND includes the global distance (3), it is likely that it has at least the same uncertainty as LDN with which it is highly correlated (Brown 2016c). This means that the uncertainty in the calibration of D_x has several sources. Of these, the distribution of $D_o(\alpha_i, \beta_i)$ has been considered (Brown 2016a) and some thought has been given to the estimation of k and ϕ using LDN (Brown 2016b). However, the calibration data on which this is based have not been considered. Here, some of the factors intrinsic to the data (the covariates) that contribute to the calibration and its uncertainty and their impact on the estimated divergence time are considered.

Calibration data

Some recent work has been based on just a small number of calibration values (Brown 2016b). Specifically, the LDN values given by Gray and Atkinson (2003) and Petroni and Serva (2008) provided a convenient basis for analysing some of the issues associated with the sensitivity of the relationship between D_o and t. However, having shown that LDND is highly correlated with LDN and that $D_s \geq D_o$ when the same Indo-European word lists are used (Brown 2016c), it is reasonable to make use of the much larger set of LDND-based calibration values given by Holman *et al.*(2011). These data are expressed in terms of similarity (S), expressed as a percentage, which is related to D_s by

$$S = 100\left(1 - D_s\right). \tag{5}$$

While the implicit assumption that $D_s \leq 1$ is theoretically incorrect (Brown 2016c), it is a reasonable approximation and, using (5), the original D_s values can be calculated from S (Table 1). Holman *et al.* (2011) analysed their calibration data by considering log(S) as a function of t

$$\log(S) = 2\log(r)t + \log(S_0), \tag{6}$$

where S_0 is the initial value of S and r is the average proportion of lexical similarity retained after 1000 y (Swadesh 1950). Using (5), (6) is equivalent to

$$D_s = 1 - 0.01S_0 r^{2t} = \left(1 - 0.01S_0\right) + 0.01S_0\left(1 - \exp\left[2\ln(r)t\right]\right), \tag{7}$$

which is related to (4), but differs from it in both the initial ($D_x(t = 0) = 0$, $D_s(t = 0) > 0$) and final values ($D_x(t \rightarrow \infty) = \phi_x$, $D_s(t \rightarrow \infty) = 1$) because (4) and (6) are based different models (Brown 2016b).

CAES Vol. 2, № 4 (December 2016)

Table 1. The 52 calibration values given by Holman *et al.* (2011) grouped according to the dating evidence. Values of D_s were calculated from the reported similarity (S) using (5).

	Time (y)	D_s (site^{-1})		Time (y)	D_s (site^{-1})
Archaeological					
Benue-Congo	6500	0.964	Malayo-Chamic	2400	0.753
Central S. African Khoisan	2000	0.883	Malayo-Polynesian	4250	0.874
Dardic	3550	0.730	Mississippi Valley Siouan	2475	0.718
E. Polynesian	1050	0.517	Ongamo-Maa	1150	0.548
E. Malayo-Polynesian	3350	0.924	Pama-Nyungan	4500	0.945
Ga-Dangme*	600	0.509	Saami	1750	0.664
Indo-Aryan	3900	0.752	Southern Nilotic	2500	0.866
Indo-European	5500	0.947	Swahili*	1200	0.492
Indo-Iranian	4400	0.917	Temotu	4300	0.926
Inuit	800	0.396	Tupi-Guarani (coastal)*	1750	0.760
Iranian	3900	0.859	Turkic*	2500	0.902
Ma'anyan-Malagasy	1350	0.697	Wakashan	2500	0.852
Epigraphic					
Cholan	1600	0.567	Hmong-Mien	2500	0.943
Czech-Slovak	1050	0.328	Oromo	460	0.365
Ethiopian Semitic	2450	0.811	Scandinavian	1100	0.672
Goidelic	1050	0.725	Sorbian (Lusatian)	450	0.312
Historical					
Breton-Welsh	1450	0.574	Maa	600	0.325
Cham	529	0.449	Maltese-Maghreb Arabic	910	0.664
Chamic	1550	0.841	Mongolic	750	0.792
Chinese	2000	0.870	Northern Roglai-Tsat	1000	0.742
Common Turkic	1419	0.621	Romance	1729	0.710
E. Slavic	760	0.605	Romani	650	0.381
English-Frisian	1550	0.694	Slavic	1450	0.570
Germanic	2100	0.708	Southern Songhai	550	0.371
Italo-Western Romance	1524	0.678	Southwest Tungusic	236	0.467
Ket-Yugh	1300	0.513	Western Turkic	900	0.546

* Dates for which both archaeological and historical evidence were available.

In addition to the larger number of values, these data have at least three other properties that were recognised by Holman *et al.* (2011). First, they represent estimates based on three different sorts of dating evidence: archaeological, epigraphic and historical (although four of the 52 values are based on both historical and archaeological data and, therefore, they are included in each of these categories in the analysis). In contrast, the LDN values given by Gray and Atkinson (2003) were predominantly historical and epigraphic, whereas those used by Petroni and Serva (2008) were not well defined. Second, they include representatives of 17 different language families (Holman *et al.* 2011), rather than being restricted to Indo-European languages. However, of these, only the Indo-European and Austronesian families have more than five[1] values among the data. Third, they include values relating to four geographical regions (Holman *et al.* 2011), although there are only five values from the Americas.

For at least two reasons these calibration data are not directly comparable to those used previously (Brown 2016b). First, the words in the lists are represented phonetically using the 41 symbol ASJPcode rather than the 26 letter Latin alphabet (Holman *et al.* 2011). Second, even if this were not the case, Holman *et al.* (2011) report similarity based on LDND rather than on LDN. While it might be expected that these two measures are correlated when the same word lists are used (Brown 2016c), there is no reason to expect that this would necessarily extend to a situation in which both the lists and the representation of the words in those lists differ.

Nevertheless, even if the data themselves are not strictly comparable, it does not preclude the possibility that the estimates of k_x and ϕ_x obtained from the two sets of calibration data might be compared. Here it is assumed that ϕ_s tends to be greater than ϕ_o, because $D_s \geq D_o$ when the same Indo-European word lists are used (Brown 2016c), and that comparison is of limited value.

Influence of data selection

Holman *et al.* (2011) reported that the type of dating evidence, the language family and the geographical region were not significant factors in the discrepancy (expressed as a percentage) between the calibration dates and the Automated Similarity Judgment Program dates. This does not necessarily mean that these factors are insignificant in the relationship between t and D_s (or, equivalently, S (5)).

In fact, the type of dating evidence, the language family and the geographical region do have some impact on the estimates of the parameters of (4). Moreover, the relationship between these and the estimation of divergence time using (4) means that even statistically insignificant differences among these estimates can have a considerable effect on the estimated t (Brown 2016b). One of the inferences to be drawn from this is that more and improved calibration data are desirable.

Considering the data by dating evidence, and remembering that ϕ_s is not bounded from above (Brown 2016c), the estimated upper limit of LDND (ϕ_s) ranges from about 1.0 for the epigraphic data (Figure 1A(i)) to 0.77 for the historical data (Figure 1A(ii)), whereas the archaeological data and the entire set of calibration data yield values of about 0.9 (Figure 1A, (iii) and (iv)). As might be anticipated, the range of dates over which the data are available is a significant consideration. For example, the lack for data for longer divergence times precludes reliable estimates using epigraphic or historical data alone, although these data are useful for smaller t (Figure 1A, (i) and (ii)). The archaeological data are sufficient to yield a reasonable estimate of ϕ_s, which is not significantly different from that obtained using all 52 values (Table 2). The rate constants range from 0.8×10^{-3} y^{-1} to 1.5×10^{-3} y^{-1} and the estimates obtained from epigraphic or historical data alone have considerably greater

[1] Five is the minimum number of data points required to obtain even a very poor estimate of the parameters of (4) by nonlinear regression. This is apparent from the fit and the parameter estimates obtained using the calibration data from the Americas (Figure 2A(iv) and Table 2).

uncertainty than that (0.9×10^{-3} y^{-1}) obtained using archaeological data (Table 2). The considerable differences in the uncertainty of these parameter estimates and their interdependence are illustrated by the 95% confidence regions in (ϕ_s, k_s)-space (Figure 1B). The confidence regions estimated for the epigraphic and historical data are considerably larger than that obtained using the archaeological data, consistent with their much greater uncertainty, but the latter lies predominantly within those of the others (Figure 1B).

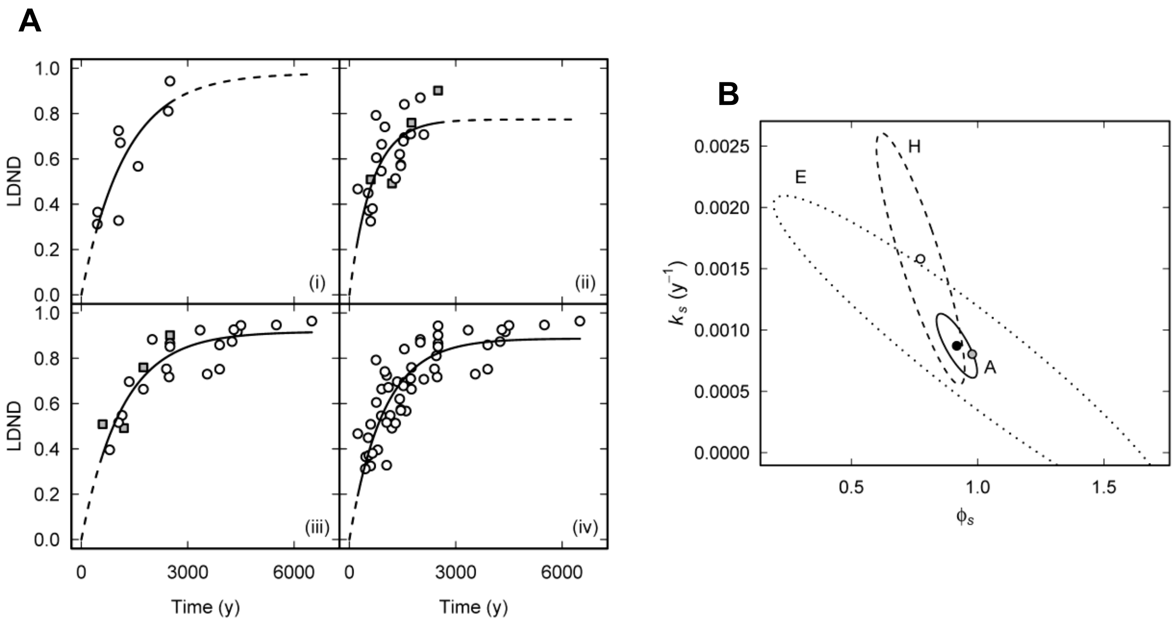

Figure 1. The calibration data based on the nature of the dating evidence. In (A) the epigraphic (i), historical (ii) or archaeological (iii) evidence or the complete set of values given in Table 1 (iv) are shown. In panels (ii) and (iii) the grey squares represent those four values for which there is both archaeological and historical evidence (E. W. Holman *et al.* 2011). In each panel, the solid curves represent least squares fits of (4) to the data and the dashed curves represent extrapolations of these. In (B) the 95% confidence regions of the parameter estimates (Beale 1960) for the epigraphic (E, ⋯⋯), historical (H, – – –) and archaeological (A, ——) calibration data are shown. The grey, white and black circles are the coordinates of the parameter estimates for epigraphic, historical and archaeological evidence, respectively, used in the curves in panel (A) and these are given in Table 2.

Only three of the four geographical regions are represented by more than five values, which means that the parameter estimates for the fourth (the Americas) have limited reliability. However, for these subsets of the data ϕ_s ranges from 0.84 to about 1 and k_s ranges from about 0.6×10^{-3} y to 1.2×10^{-3} y (Figure 2A). Despite the uncertainty associated with these estimates (Table 2), the range of dates for which values are available are reasonable for the African, Eurasian and Pacific data (Figures 2A, (i), (ii) and (iii)), consistent with the confidence regions shown in Figure 2B, but this is not the case for the data from the Americas (Figure 2A(iv)) for which no reliable estimates of the parameter uncertainty could be obtained (Table 2).

Of the 17 language families represented in Table 1 (Holman *et al.* 2011), only two contribute more than five values. The data for Indo-European languages yield $\phi_s = 0.85$ and $k_s = 1 \times$

10^{-3} y and for the Austronesian languages $\phi_s = 0.91$ and $k_s = 1.2 \times 10^{-3}$ y (Table 2). It is notable that the estimate of k_s for Indo-European languages is larger than the values of k_o reported previously (Brown 2016b, Serva & Petroni 2008).

Table 2. Properties of the curves (4) for the calibration data classified by dating evidence (Figure 1), geographical region (Figure 2) or selected language families. Parameter estimates (± SE) and the mean square of the residuals (MSR) were obtained using R (Ihaka & Gentleman 1996).

		ϕ_s	k_s ($\times 10^{-3}$ y^{-1})	MSR
All data		0.89 ± 0.04	1.0 ± 0.1	0.012
Evidence	Archaeological	0.92 ± 0.03	0.9 ± 0.1	0.006
	Epigraphic	1.0 ± 0.3	0.8 ± 0.5	0.021
	Historical	0.77 ± 0.08	1.5 ± 0.5	0.015
Region	Africa	0.97 ± 0.07	0.9 ± 0.1	0.005
	Americas	1	0.6	0.007
	Eurasia	0.84 ± 0.06	1.2 ± 0.3	0.017
	Pacific	0.91 ± 0.04	1.2 ± 0.2	0.007
Family[*]	Austronesian	0.91 ± 0.04	1.2 ± 0.2	0.006
	Indo-European	0.85 ± 0.06	1.0 ± 0.2	0.011

[*] Only those language families with more than five values are included.

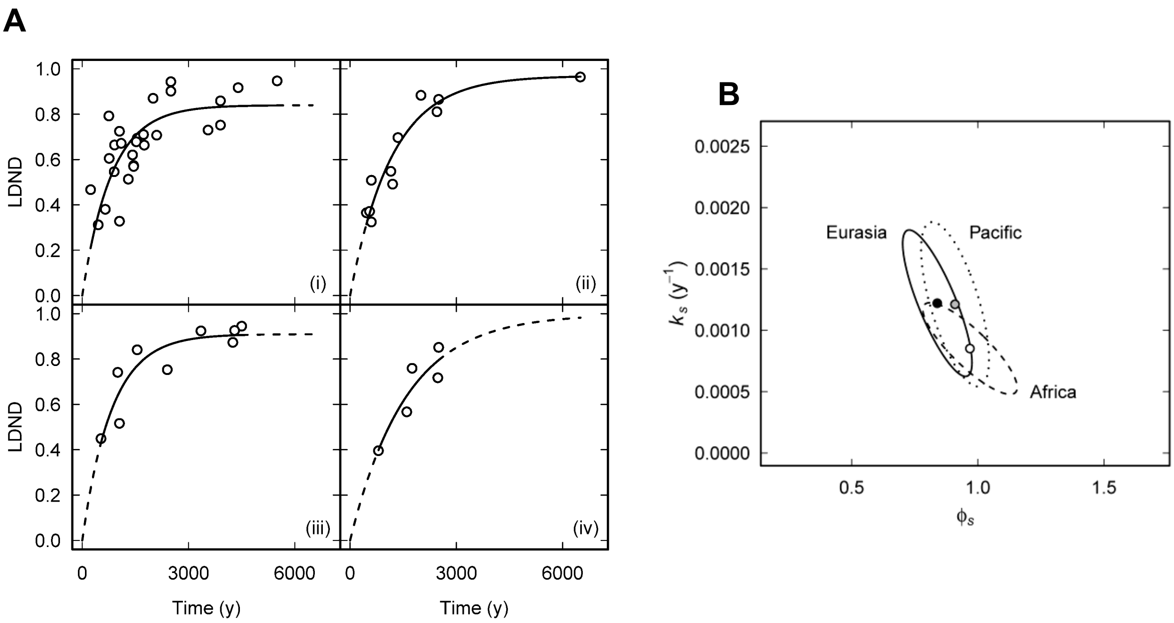

Figure 2. The calibration data based on the geographical region. In (A) the Eurasian (i), African (ii), Pacific (iii) or American (iv) calibration data given in Table 1 are shown. In each panel, the solid curves represent least squares fits of (4) to the data and the dashed curves represent extrapolations of these. In (B) the 95% confidence regions of the parameter estimates (Beale 1960) for the Pacific (······), African (– – –) and Eurasian (——) calibration data are shown. The grey, white and black circles are the coordinates of the parameter estimates for the Pacific, African and Eurasian regions, respectively, used in the curves in panel (A) and these are given in Table 2.

Model-dependence

The model used to implement the calibration is a significant factor in determining the relationship between t and D_s (Brown 2016b). Holman *et al.* (2011) used ordinary least squares regression to fit (6) to the calibration data. This yielded a correlation between t and log(S) of –0.84 and the coefficients of (6) were $r = 0.72$ and $S_0 = 92\%$ for the complete set of data. Acknowledging that both t and S are inherently uncertain, they also fitted the inverse function for which they reported the equivalent of $r = 0.79$ and $S_0 = 62\%$. If estimates of the "analytical" variances of both t and log(S) were available, a regression line, based on (6), obtained by taking into account the uncertainties of both variables would lie between these two extremes (Francq & Govaerts 2014).

The fits of (6) to the calibration values obtained by Holman *et al.* (2011) have a mean square of the residuals (MSR) of 0.013 for (r, S_0) = (0.72, 92%) and 0.012 for (0.79, 62%), and that shown in Figure 1A(iv), obtained by nonlinear regression using (4), has an MSR of 0.012. There is, therefore, no significant difference in the quality of the fits, but they are clearly different (Figure 3). Specifically, (4) passes though the origin ($t = 0$, $D_x = 0$), whereas (7) passes through ($t = 0$, $D_s = 1 - 0.01S_0$), which implies that LDND can never be smaller than 0.08 or 0.38. The latter lower limit, corresponding to (r, S_0) = (0.79, 62%), is clearly inconsistent with the five calibration values that are smaller, rendering this fit unusable. A second difference between (4) and (7) is that the latter asymptotically approaches 1, whereas the former approaches ϕ_s, as t increases. This difference reflects the consideration of the possibility of convergent evolution in the model on which (4) is based, but not in that underlying (7) (Brown 2016b). Which of these fits best represents the data presumably depends largely on which model is preferred.

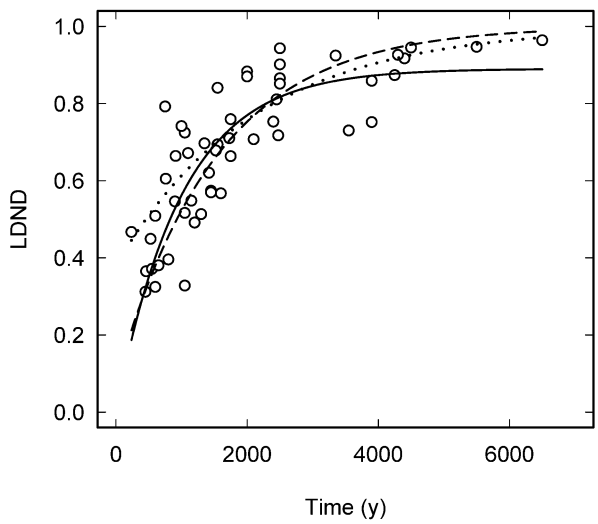

Figure 3. Comparison of the curves fitted to all the calibration data (○) given in Table 1. The solid line (——) is the fit of (4) shown in Figure 1A(iv), the dashed (– – –; (r, S_0) = (0.72, 92%)) and dotted lines (·····; (r, S_0) = (0.79, 62%)) are the fits of (6) reported by Holman *et al.* (2011) expressed in the form of (7).

While (4) and (7) are different because the latter has an extra constant term, parameter values of (7) can be used to estimate those corresponding to a parallel curve described by (4), and *vice versa* (Appendix). In the case of the curves shown in Figure 3, (r, S_0) = (0.72, 92%) and (0.79, 62%) correspond to (k_s, ϕ_s) = (0.66×10^{-3} y^{-1}, 0.92) and (0.47×10^{-3} y^{-1}, 0.62), respectively. For the first of these, the estimated ϕ_s is not significantly different from that in Table 2, but the estimated rate constant (k_s) is about 34% smaller than the value of 1×10^{-3} y^{-1} given in Table 2, similar to the values of k_o for Indo-European languages reported previously (Brown 2016b, Serva & Petroni 2008), and is smaller than any of the other rate constants in Table 2.

The impact on the estimated divergence time

The parameter estimates summarised in Table 2 do not differ greatly from one another. However, as has been shown previously (Brown 2016b), small differences in ϕ_x or k_x can be associated with significant differences in the estimated divergence time. In order to examine this effect for the subsets of the calibration data considered the t values for as many of the entries in Table 1 as possible were calculated using each of the parameter sets in Table 2. This required the inverse of (4)

$$t = -\frac{1}{k_x}\ln\left(1 - \frac{D_x}{\phi_x}\right), \tag{8}$$

which is undefined unless $D_x < \phi_x$. For consistency, this constraint was interpreted as requiring that D_s was smaller than any of the ϕ_s in the subsets being considered. This meant that the number of values varied because, for example, it required that $D_s < 0.77$ for the comparison of the dating evidence (so $n = 35$ in Figure 4A), but $D_s < 0.84$ for the comparison by geographic region (so $n = 37$ in Figure 4B).

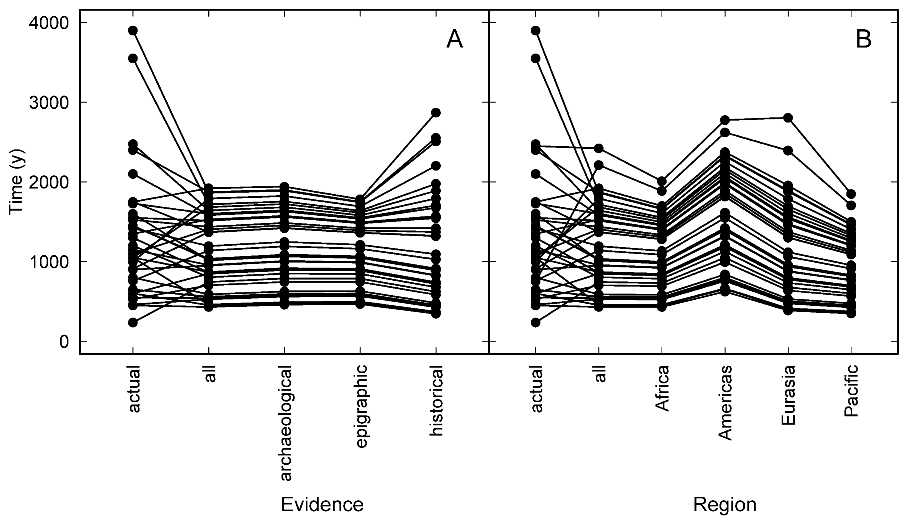

Figure 4. The impact of the selection of calibration data by dating evidence (A) or geographic region (B) on the estimated divergence time. Values of D_s in Table 1 were used to calculate t (8) using the parameter estimates obtained from the indicated subsets of the data (Table 2). In each panel only those values of $D_s < \min(\phi_s)$ were used and the actual t for these values are also shown.

Consistent with the regression approach taken (Figures 1A(iv) and 2A(iv)), the range of t values estimated using the parameters derived from the entire set of calibration data (the 'full model') was smaller than that of the actual values used in the dating evidence (Figure 4A) and the regional (Figure 4B) comparisons. Those obtained using the archaeological or epigraphic parameters were very similar to that obtained from the full model, although the epigraphic estimates were slightly decreased at larger D_s. In contrast, the historical estimates were almost 1000 y larger at high D_s and were smaller at low D_s (Figure 4A). The effect of the parameters for different geographical regions is more complex (Figure 4B). At small Ds the estimates obtained using the African, Eurasian and Pacific data were similar to that of the full model, but the Eurasian model increased the estimates at large D_s compared with the African and Pacific models and the full model. The estimates derived from the parameters based on data representing the Americas gave consistently larger estimates than the other models (Figure 4B), but this should be treated with caution because the parameter estimates were based on only a very small number of values (Figure 2A(iv)).

Some general observations

Two issues arise from the analyses described here. First, that the selection of calibration data is a critical step. Second, that the choice of the model on which to base the calibration is also important. The divergence time estimate depends on both of these irrespective of whether LDN, LDND or some other measure is used.

The data used to calibrate the estimation of t certainly matter. The values of t calculated using parameters based on calibrations founded on historical dating evidence can be quite different from those based on archaeological or epigraphic evidence (Figure 4A). It may be that this reflects the relatively restricted range of dates (236-2100 y) of the historical data (Figure 1A(ii)), but the epigraphic data in Table 1 are similarly restricted (450-2500 y, Figure 1A(i)) and they yield estimates comparable with those based on the archaeological evidence (Figure 4A). Similarly, calibrations based on data from different geographical regions (even discounting the few data representing the Americas) can also yield different estimates of t (Figure 4B). Whether this reflects real regional variation or underlying differences between language families is not clear. One of the weaknesses of LDN or LDND as measures of relatedness is that k_x is too large for them to be useful at longer t (Brown 2016b), especially if it is not constant in time (Atkinson *et al.* 2008, Nettle 1999) or space (Atkinson 2011). Other measures, that change more slowly, are necessary to overcome this difficulty (Brown 2016b), which requires the further development of different approaches (Akulov 2015a, Grimes & Agard 1959, Kroeber & Chrétien 1937). However, it is not inevitable that more sophisticated measures will yield a significant improvement unless great care is taken (Brown 2015). Nevertheless, the choice of calibration data is a critical step irrespective of whether lexical distance or some better measure is employed.

The model-dependence of the calibration of linguistic distance illustrated in Figure 3 has been considered previously (Brown 2016b). The choice of model influences the size of the estimated divergence time whether the data are based on LDN, LDND or some other measure. This is because the model encodes a particular view of the processes underlying language change and so the choice deserves careful consideration.

Conclusions

The calibration data of Holman *et al.* (2011) provides a means of assessing the intrinsic variation due to the dating evidence, geographic region and, to a lesser extent, language family. The selection of the data used in calibration affects the relationship between D_s and t (Figures 1 and 2, Table 2) and can strongly influence the estimates of divergence time derived from it (Figure 4). This reinforces the concern previously expressed (Brown 2016b) about the use of a small number of apparently arbitrarily chosen calibration values. This issue merits further investigation. Similarly, the choice of the model that underpins the analysis is an important consideration (Brown 2016b) because it also strongly influences the apparent relationship between D_s and t (Figure 3). While these observations were drawn from data based on LDND, there is no justification for thinking that data based on LDN would have been very different.

Appendix

The relationship between the parameters of (4) and (7) can be clarified by writing the latter as

$$D_s = 1 - 0.01S_0 + 0.01S_0\left(1 - \exp\left(-2\times10^{-3}\ln\left(\frac{1}{r}\right)t\right)\right), \qquad \text{(A1)}$$

in which it is recognised that $D_s(t = 0) = 1 - 0.01S_0$ and that r is dimensionless, but is based on t expressed in thousands of years. While (A1) is not identical to (4), the last term on the right of (A1) does have the same form as (4) so it is possible to match the shapes of the curves by matching these terms. Comparing the last term on the right of (A1) with (4) gives $k_x = -2\times10^{-3}\ln(r)$ and $\phi_x = 0.01S_0$. The parameter value estimates of Holman *et al.* (2011) are (r, S_0) = (0.72, 92%) and (0.79, 62%), which correspond to $(k_s, \phi_s) = (0.66 \times 10^{-3}\ \mathrm{y}^{-1}, 0.92)$ and $(0.47 \times 10^{-3}\ \mathrm{y}^{-1}, 0.62)$, respectively. The curves described by (r, S_0)-(k_s, ϕ_s) pairs are parallel and offset by the constant in (A1).

References

Akulov A. 2015a. Verbal grammar correlation index (VGCI) method: a detailed description. *Cultural Anthropology and Ethnosemiotics*, vol. 1 no. 4; pp.: 19 – 42

Akulov A. 2015b. Why conclusions about genetic affiliation of certain language should be based on comparison of grammar but not on comparison of lexis? *Cultural Anthropology and Ethnosemiotics*, vol. 1 no. 3; pp.: 5 – 9

Atkinson Q. D. 2011. Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science*, vol. 332; pp.: 346 – 349

Atkinson Q. D., Meade A., Venditti C., Greenhill S. J., Pagel M. 2008. Languages evolve in punctuational bursts. *Science*, vol. 319; pp.: 588

Bakker D., Müller A., Velupillai V., Wichmann S., Brown C. H., Brown P., Egorov D., Mailhammer R., Grant A., Holman E. W. 2009. Adding typology to lexicostatistics: a combined approach to language classification. *Linguistic Typology*, vol. 13; pp.: 169 – 181

Beale E. M. L. 1960. Confidence regions in non-linear estimation. *Journal of the Royal Statistical Society*, vol. 22B no. 1; pp.: 41 – 88

Brown S. 2015. A bioinformatic perspective on linguistic relatedness. *Cultural Anthropology and Ethnosemiotics*, vol. 1 no. 4; pp.: 43 – 52

Brown S. 2016a. An analysis of the distributions of linguistic distances. *Cultural Anthropology and Ethnosemiotics*, vol. 2 no. 1; pp.: 2 – 12

Brown S. 2016b. An examination of the calibration of linguistic distance. I. Sensitivity. *Cultural Anthropology and Ethnosemiotics*, vol. 2 no. 2; pp.: 1 – 9

Brown S. 2016c. Two measures of linguistic distance. *Cultural Anthropology and Ethnosemiotics*, vol. 2 no. 3; pp.: 2 – 12

Francq B. G., Govaerts B. B. 2014. Hyperbolic confidence bands of errors-in-variables regression lines applied to method comparison. *Journal de la Société Française de Statistique*, vol. 155 no. 1; pp.: 23 – 45

Gray R. D., Atkinson Q. D. 2003. Language tree divergence times support the Anatolian theory of Indo-European origin. *Nature*, vol. 426; pp.: 435 – 439

Grimes J. E., Agard F. B. 1959. Linguistic divergence in Romance. *Language*, vol. 35 no. 4; pp.: 598 – 604

Hoijer H. 1956. Lexicostatistics: a critique. *Language*, vol. 32 no. 1; pp.: 49 – 60

Holman E. W., Brown C. H., Wichmann S., Müller A., Velupillai V. V., Hammarström H., Sauppe S., Jung H., Bakker D., Brown P., Belyaev O., Urban M., Mailhammer R., List J.-M., Egorov D. 2011. Automated dating of the world's language families based on lexical similarity. *Current Anthropology*, vol. 52 no. 6; pp.: 841 – 875

Ihaka R., Gentleman R. 1996. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, vol. 5 no. 299 – 314

Kroeber A. L., Chrétien C. D. 1937. Quantitative classification of Indo-European languages. *Language*, vol. 13 no. 2; pp.: 83 – 103

Levenshtein V. I. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics - Doklady*, vol. 10 no. 8; pp.: 707 – 710

Matisoff J. A. 1990. On megalocomparison. *Language*, vol. 66 no. 1; pp.: 106 – 120

Nettle D. 1999. Is the rate of linguistic change constant? *Lingua*, vol. 108 no. 119 – 136

Petroni F., Serva M. 2008. Language distance and tree reconstruction. *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008 no. 8; pp.: P08012

Petroni F., Serva M. 2010. Measures of lexical distance between languages. *Physica A*, vol. 389; pp.: 2280 – 2283

Rea J. A. 1958. Concerning the validity of lexicostatistics. *International Journal of American Linguistics*, vol. 24 no. 2; pp.: 145 – 150

Serva M., Petroni F. 2008. Indo-European languages tree by Levenshtein distance. *Europhysics Letters*, vol. 81; pp.: 68005

Swadesh M. 1950. Salish internal relationships. *International Journal of American Linguistics*, vol. 16 no. 4; pp.: 157 – 167

Teeter K. V. 1963. Lexicostatistics and genetic relationships. *Language*, vol. 39 no. 4; pp.: 638 – 648