

A full-scale test of the language farming dispersal hypothesis

Harald Hammarström

Radboud Universiteit Nijmegen and Max Planck Institute for Evolutionary Anthropology

One attempt at explaining why some language families are large (while others are small) is the hypothesis that the families that are now large became large because their ancestral speakers had a technological advantage, most often agriculture. Variants of this idea are referred to as the Language Farming Dispersal Hypothesis. Previously, detailed language family studies have uncovered various supporting examples and counterexamples to this idea. In the present paper I weigh the evidence from ALL attested language families. For each family, I use the number of member languages as a measure of cardinal size, member language coordinates to measure geospatial size and ethnographic evidence to assess subsistence status. This data shows that, although agricultural families tend to be larger in cardinal size, their size is hardly due to the simple presence of farming. If farming were responsible for language family expansions, we would expect a greater east-west geospatial spread of large families than is actually observed. The data, however, is compatible with weaker versions of the farming dispersal hypothesis as well with models where large families acquire farming because of their size, rather than the other way around.

Keywords: language farming dispersal hypothesis, language families, large families, small families, isolates, agriculture, hunting and gathering

1. Introduction

Some language families are ‘large’, like the Austronesian family, which is both geographically widespread (from Madagascar to Easter Island) and have a large number of member languages. Yet other families are minimal in both geographic size and in the number of member languages, e.g. isolates like Burushaski. A natural question is: why are some small when others are large?

One attempt at explaining why some language families are big (while others are small) is the hypothesis that the families that are now large became large because their ancestral speakers had a technological advantage, most often that of farming. In this paper, I focus on this hypothesis, to see how well it accounts for surface properties of the language families in the world. I have developed a database of *all* language families and approximations of their cardinal size (number of member languages), geospatial size and subsistence type (i.e., whether speakers have a predominantly hunting-gathering or a farming subsistence). This database enables us to perform a number of statistical tests of hypotheses involving size and subsistence type of language families. In this way, some merits of farming-dispersal hypotheses that were previously opaque on a worldwide scale are elucidated.

2. Language families and data

2.1. Language families

There are some 7,000 attested languages in the world (see Lewis 2009 for a fair catalogue of the living ones). A language family is defined as follows:

- a **set of languages** (possibly a one-member set)
- with at least one **sufficiently attested** member language
- that has been **demonstrated in publication**
- to **stem from a common ancestor**
- by **orthodox comparative methodology** (Campbell & Poser 2008)
- for which there are **no** convincing published attempts to demonstrate a **wider affiliation**.

I know of no dedicated effort at a systematic application of this definition that spans the whole world. Therefore, we have used our own attempt in the present paper, yielding ca. 400 families for the 7,000 languages. A list with sources and brief commentary is given as an online appendix. Discussion of the most accurate assessment of language families is beyond the scope of this study.

A more fine-grained test could be done with a classification into both families and lower-level subfamilies within families. However, much more work would be needed for systematic subgrouping of the world's language families than for mere family membership. At this time, reliable information of this kind is not within reach for most of the world's families.

The concept of size has different senses, and there are different measures of the size of a language family. In this study, I use two measures, cardinal size and geospatial size.

2.2 Cardinal size

I define the **cardinal size** of a family as *the number of languages belonging to it*. For example, Austronesian is said to encompass 1261 languages, so its cardinal size is 1261. Burushaski is a language isolate, a language family with only one member, so it has a cardinal size of 1. Since language differentiation mainly comes about through human migration, the cardinal size of a family is a measure of the expansional spread of a family. In the tests in this study, I have used the readily available language counts of Lewis (2009). For language families not listed in Lewis (2009) — mainly extinct families — I have added counts from the individual sources listed in the online appendix. Note that only the language *counts* are from Lewis (2009), the actual family membership is as defined above.

2.3 Geospatial size

The geospatial size of a family may be approximated using the coordinates of the languages belonging to the family. This study uses a database of language coordinates, also derived from Lewis (2009). For language families not listed — mainly extinct families — I have supplemented with coordinates derived from the individual sources listed in the online appendix. Unfortunately, in the available data, each language is represented as a point at the geographic centre of where its speakers reside, rather than as a polygon (or the like). This affects the above approximation of family size somewhat, but I do not expect that any of the conclusions in this study would require major revisions with more sophisticated approaches to identifying their locations.

A small number of families have outlier languages or speaker populations as a result of recent, modern migrations. For example, a Hindi variety (labelled Caribbean Hindustani) is spoken in Guyana in South America. Conversely, some families have been reduced to almost nothing of their historically attested size as a result of much the same modern migrations, e.g. Chumashan. Since these enlargements and reductions are well documented and have nothing to do with prehistoric family-size patterns, for the present study, the sizes of these families have been restored to that of the first eyewitness documentation drawing on individual sources. As a result, the sizes of the language families in the Americas are restored to that of witnessed locations in 16th–19th century sources, and Indo-European is reduced to Eurasia. Prehistoric family size information (e.g., from archaeological or other considerations) has not been taken into account, as information is not available for all language families.

2.4 Subsistence type

For the purposes of the present study, I have designed the database on family subsistence type as follows. Every family is judged either **AGRICultural** (AGR) or

Hunter–Gatherer (HG). This categorisation is based on an aggregation (see below) of the subsistence type of the speakers of its individual member languages. Since speakers of one and the same language typically have the same subsistence type, I have categorised the subsistence type of a single language accordingly (a language which is not HG is categorised as AGR):

- A language is a Hunter–Gatherer (HG) language if and only if its speakers subsist more than 50% on
- **hunted/gathered** food (= reproduction of species not controlled)
- based on **ethnographic evidence** at
- **time of first eyewitness documentation.**

A few comments are in order. First, the line between foraging and practices which help certain foods to succeed is contentious. Here, the line has been drawn by whether the speaker population controls the reproduction of the species in question. In particular, sago is assessed to be a predominantly non-cultivated foraged food type. Second, I have used ethnographic evidence only (as opposed to, e.g. archaeological or linguistic evidence), since systematic ethnographic evidence is available non-controversially for nearly all language families. I use the earliest possible evidence, as the goal is to elucidate the earlier patterns for the language families in question. Third, given that most subsistence types attested are a mix of foraged and non-foraged foods, I have rather arbitrarily set the threshold at 50%. This choice is more or less the only possibility if we want to classify every attested language family using ethnographic evidence. Many ethnographic sources are content with saying that one of the kinds of food is more or less important (or something similar corresponding to 50%) — only rarely are there measurements of a granularity sufficient for a different threshold. I freely admit that, for questions in a narrower domain than in the present paper, more interesting patterns may emerge with a different threshold.

If a family consists only of languages of the same subsistence type, i.e., only HG or only AGR families, the classification of the whole family is simply that of its member languages. However, some, if not many, families contain a mix of HG and AGR languages. It turns out, empirically, that in all these families, one subsistence type is peripheral and the other is predominant (as of first eyewitness documentation) in the family. For example, all of the over 400 Sino-Tibetan languages are agricultural except the Chepang, Raute and Raji, which are dwindling groups of nomadic hunter–gatherers. Similarly, all Algic languages are hunter–gatherer except a few southeastern ones which took part in the agricultural complex of their southern neighbours. Since the database is designed to test whether one subsistence type can account for the size of the family, it makes sense to classify the family by its dominant subsistence type.

It is worth noting that the subsistence type of a family as defined here is logically independent of the subsistence type of the proto-language of the family. Information on subsistence type of a proto-language is very powerful information for testing

causalities of family sizes and subsistence. Unfortunately, systematic information about the subsistence type of (all) proto-languages is not available.

3. Explananda on language family sizes

The first questions about language families have been formulated in terms of cardinal size. As already noted, the ca. 7,000 languages fall into ca. 400 families, and more importantly, the families are of very **unequal** size. Figure 1 shows a plot of rank versus family size. Language family sizes are not normally distributed, i.e., it is not the case that there are equal numbers of families (at any level) above and below the mean family size (≈ 17.87). In essence a few families are very big, while almost all others are tiny in comparison. This kind of distribution is commonly referred to as Zipfian (or power-law or log-linear) and has been observed with a multitude of other phenomena (Li 2002).

The distribution of cardinal sizes follows a Zipfian rather than normal distribution also when broken down over continents and also for HG and AGR families separately, though the slope varies.

The question arises why some are big while most are small? As will be discussed, one hypothesis is that subsistence type has a role to play, in that larger families are the result of an agricultural subsistence type.

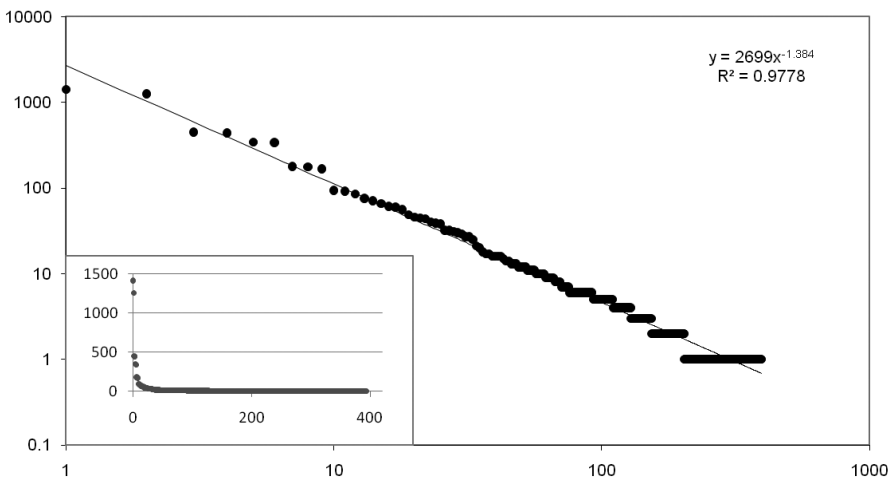


Figure 1. A log-log rank-size plot of language families and their cardinal sizes. The inset x-axis has the largest (rank one) family leftmost, the second largest (rank two), and so on, while the y-axis has the size of the corresponding family. The main plot has log on both axes. The slope of the line is -1.38 , with Pearson's $r \approx 0.99$.

But first, it is not clear that an explanation in terms of a logically independent variable, such as subsistence type, is necessary at all. A Zipfian distribution is the expectation of various kinds of population models with random selection (Mitzenmacher 2004). Indeed, it has been pointed out by Wichmann (2005) that a Galton-Watson process may well be sufficient to account for cardinal sizes, and fitting each continent individually gives parameters, which may be interpreted as time estimates (Hammarström 2008).

If simple language-split models can also produce the family sizes observed, why explore more complicated hypotheses involving subsistence type? As I shall show, there are good reasons for involving subsistence type as (part of) the explanation for cardinal language family sizes.

If the sole reason for the distribution of language family sizes is splitting and merging of randomly selected families, then which families end up large should not be correlated with any independent variable, such as subsistence type. In essence, whether a language is agricultural, or, say, has a present name that begins with a consonant, should have no influence on whether the language family ends up large or not. However, as I show with farming-size correlations, the large families are not a random selection, as one would expect from a simple language split model, but are biased towards agriculture.

4. The language farming dispersal hypothesis

4.1 Previous investigations of farming expansions

The modern idea that farming may have played a role in some or all cases of large families observed today seems to have extended from the idea that one particular family, Indo-European, spread via farming — a hypothesis famously defended by Renfrew (1987). Renfrew has since promoted the analogue hypothesis for other language families, primarily in Eurasia (Renfrew 1997, 1999, 2002). Further popularisation has come from the archaeologist Bellwood's research on Oceania (Diamond & Bellwood 2003, Bellwood 1997, Bellwood & Renfrew 2002a, Bellwood 2001) and not least by the popular Diamond (1997).

Experts on individual (sub-)families have taken up the question, sometimes within a larger framework, with detailed case studies, often including the question of agricultural vocabulary in the proto-language. For Africa, we have a recent book-length perspective (Blench 2006), earlier broad papers (Ehret 2002, MacDonald 1998) and special attention to Bantu (Holden 2002, Phillipson 2002) and Afroasiatic (Militarev 2002, Ehret et al. 2004). For Eurasia, aside from the large literature on Indo-European, the families in South and Southeast Asia have been the subject of case studies (Fuller 2002, Higham 2002, 1998, van Driem 2002, Hudson 2002, Southworth 2006)

and wider perspectives from Blench (2005, 2008b). For Oceania, Austronesian is well studied in many respects, including farming (Pawley 2002, Bellwood & Dizon 2008, Blench et al. 2008) and the potential of subsistence type to explain the pattern of Papuan families, notably Trans-New Guinea has been noted (Ross 2006, Pawley 2005, 2007). For one large Australian family, Pama-Nyungan, agriculture could not have been the cause of expansion, but other technological advantages could have played similar roles (Evans & Jones 1997, Evans & McConnell 1998). For the Americas, apart from a rather superficial study (Blench 2008a), interest and optimism have been restricted, and counterexamples to the expansional role of farming have been discussed. One of the best studied non-farming expansions is the Numic spread (Hill 2001, 2002, 2006), where a HG subfamily expands from an otherwise farming language family. One sceptical paper in particular, Campbell (2002), brings up a number of counterexamples, but lacks explicit definitions and sources and does not have the fullest possible scope. Similarly, Blench (2008c) summarises the positive and negative positions for a number of families, but cuts too many corners and lacks definitions, sufficient sources and does not take all known families into scope. Finally, Wichmann (to appear) investigates whether farming correlates with family *density*, defined as the ratio of numbers of languages in a family and the internal linguistic differentiation of the family, where linguistic differentiation is measured in differences in vocabulary. He concludes that a high density is predictive of farming and a low density of hunting and gathering, with a zone of indeterminacy in between high and low densities. This paper takes a fairly wide scope and uses a combination of ethnographic and archaeological evidence for assessing the subsistence of a family. However, as with other studies, it lacks explicit sources and definitions, e.g. of what counts as agriculture and with respect to language families, and it does not take all known families into account.

So far we have many case studies of individual families which support the association between farming and their dispersal, but also many counterexamples, both of individual widespread families with no association to farming, e.g. Na-Dene, and of farming families without traceable wide spread, e.g. Kamsá. The present paper will not review the status of individual families. Instead, I continue the line of investigation of Campbell (2002), Wichmann (to appear) and Blench (2008c) by aggregating the pros and cons, this time from the largest possible set of families. I further add the previously unchecked matter of east-west versus north-south family expansions.

4.2 Definition of language farming dispersal hypothesis

There is a certain vagueness in the formulation of the LFDH by adherents and sceptics alike. For instance, Bellwood & Renfrew (2002b:xiii) use this formulation:

The farming/language dispersal hypothesis makes the ... proposal that the present-day distributions of many of the world's ... language families can be traced back to the early developments and dispersals of farming ...

The fate of LFDH fares differently depending on what we take it to entail. In particular, major choices are the following:

- a. *All* or only *some* (more than random?) dispersals are due to farming?
- b. *All* or only *some* (more than random?) instances of farming give rise to a dispersal?
- c. Is the connection between farming and dispersals (if any) causal (farming causes dispersal) or simply co-temporal?

I do not wish to commit any previous author to all or any of these claims. However, in general, scientists seek to find the strongest possible generalisation. Therefore, I first test the strongest formulation of the LFDH. It can be simply stated as follows:

LFDH-S-a: Farming is the cause for every large language family.

LFDH-S-b: Every instance of farming causes a family expansion.

If we find reason to reject it, the counterexamples for such a rejection will guide a weaker reformulation to be tested next.

More precisely, I will put LFDH-S to the following tests, which are within the power of the database at hand, and which are individually able to refute the LFDH-S (but if LFDH-S passes them, it does not follow that LFDH-S is true):

- Does farming have explanatory power in predicting which families are large and which are not?
- Does the geospatial distribution of the observed farming language families show an east-west spread (rather than a north-south spread) as predicted if the cause of their spread is farming (Diamond 1997)?

5. Farming and cardinal size

Statistics on AGR versus HG families are shown in Table 1 and Table 2 lists the biggest families of either category.

As can be seen, the largest families are agricultural, but there are a few fairly large hunter-gatherer families as well. The median sizes testify to the numerical dominance of very small families among both farming and non-farming families. The latter fact means that LFDH-S-b must be rejected (or at least weakened), as the counterexamples far outnumber the confirming examples.

On the other hand, pertaining to LFDH-S-a, is the correlation of AGR and mean size statistically significant? I use the following test:

Sample 1000 subsets of size 176, from the universe of 394 families, to check how many of the subsets have a sum.

Table 1. Statistics on cardinal size of AGR versus HG families.

	AGR	HG	ALL
# families	176	218	394
Σ -size	6139	956	7095
Mean size	34.88	4.39	18.01
Median size	2	1	2

Table 2. The largest farming and non-farming families.

AGR		HG	
Atlantic-Congo	1415	Pama-Nyungan	176
Austronesian	1261	Sepik	49
Sino-Tibetan	449	Eyak-Athapaskan-Tlingit	45
Indo-European	440	Algic	44
Afro-Asiatic	347	Lower Sepik-Ramu	32
Trans New Guinea	340	Salishan	27
Otomanguean	178	Lakes Plain	20
Austroasiatic	170	Tor-Orya	17
East Sudanic	94	Siouan	17
Tai-Kadai	92	Western Daly	11
Dravidian	85	Eskimo-Aleut	11
Tupí	76	Cenderawasih Bay	11
Mande	71	Miwok-Costanoan	10
Central Sudanic	66	Bosavi	10
Uto-Aztecan	61	Great Andamanese	10
Arawak	60	Gunwinyguan	9
...		Cochimi-Yuman	9
		

The correlation between AGR and (average) size turns out to be highly significant ($p < 0.001$). Importantly, this means that pure random birth-death processes are very unlikely to be historically true models of language family size dynamics (cf. Holman, this volume). It also means that LFDH-S-a was not immediately refuted, i.e., farming families are significantly larger than non-farming ones, despite a few counterexamples.

6. Farming and east-west spreads

The database allows very crude estimates of geospatial areal size by simply multiplying the horizontal and vertical geospatial size. But areal size turns out to be correlated in general — both for AGR and HG families — with cardinal size, so this offers no essentially new hurdle for the LFDH-S. (Actual computations are omitted due to space restrictions.)

A tangent test for the LFDH-S is whether farming families are more extended in the east-west direction than in the north-south direction. This is because, as explained eloquently by Diamond (1997), agriculture spreads more easily horizontally, where the climate is similar. An agricultural dependence on a smaller number of crops and animals is much more sensitive to climate surroundings than a multi-faceted hunter-gatherer subsistence. Therefore, if the cause of expansion of language families is really farming, farming families should show a significant east-west spatial alignment.

I quantify the horizontality (HOR) of a family as follows, using the coordinates of its member languages.

- East-West (EW) expansion is the difference (in degrees) between the eastern and western endpoint languages of the family.
- North-South (NS) expansion is the difference (in degrees) between the northern and southern endpoint languages of the family.
- Horizontality is the ratio between east-west distance denoted (in km, at the NS-midpoint) and north-south distance denoted (in km).

Table 3. Computation of horizontalness for the Saharan family. $EW = 32.72 - 10.85 = 21.88$ and $NS = 20.61 - 11.01 = 9.60$, so the NS-midpoint is at 15.81. The width of one longitudinal degree on latitude 15.81 is $\cos 15.81 \cdot 111.325 \approx 107.113$ kilometres. Thus, the east-west km-distance is $107.113 \cdot 21.88 \approx 2343.65$. The NS km-distance is straightforwardly $10.60 \cdot 111.325 \approx 1180.05$ km. The horizontality is thus $2343.65 / 1180.05 \approx 1.986$.

W Endpoint	Kanuri, Manga	kby	Niger	10.85 E
E Endpoint	Berti	byt	Sudan	32.72 E
S Endpoint	Kanuri, Central	knc	Nigeria	11.01 N
N Endpoint	Berti	byt	Sudan	20.61 N

Table 4. Observed horizontality for (non-isolate) AGR-families and HG-families.

	AGR	HG	ALL
# families	94	104	198
Mean HOR	2.04	1.84	1.94
Median HOR	1.24	1.12	1.18

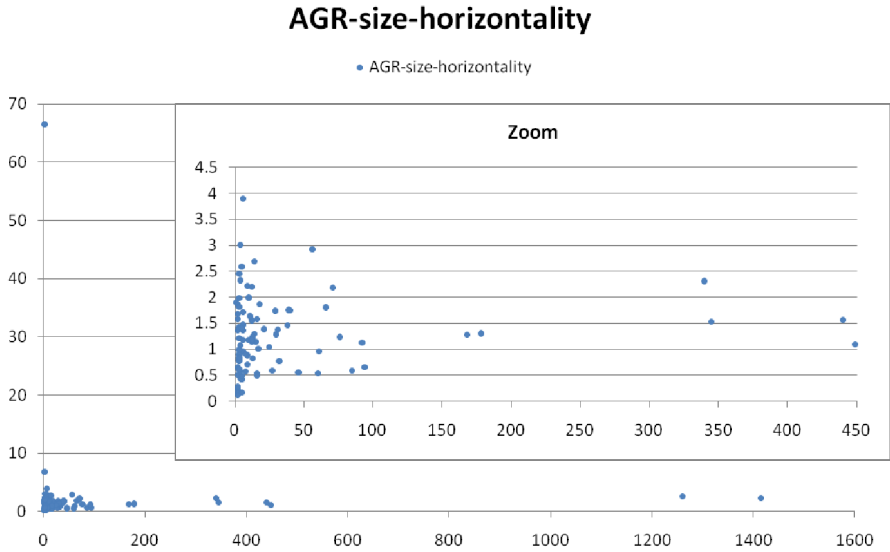


Figure 2. A size-horizontality plot for agricultural language families.

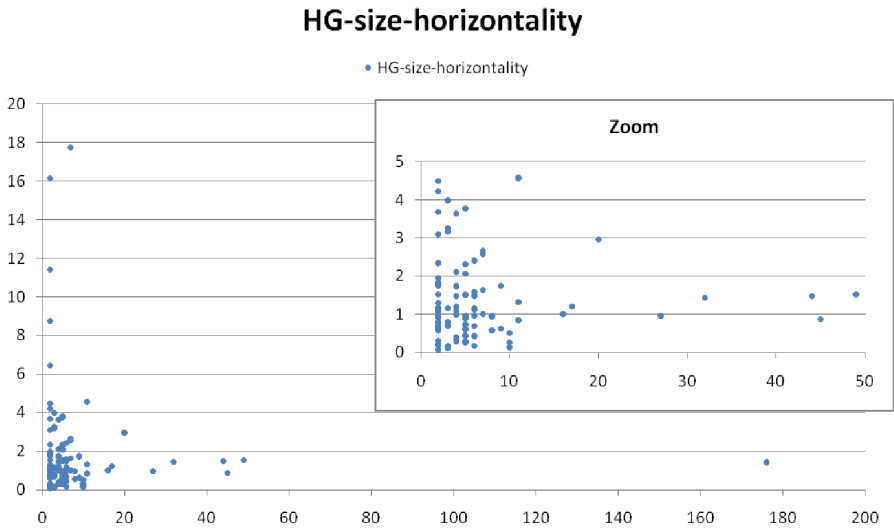


Figure 3. A size-horizontality plot for hunter-gatherer language families.

An important note is that isolates are excluded since we cannot approximate their horizontalness with only one language centrepont. 198 data points remain, more than enough for statistical tests. Table 3 exemplifies the quantification for one language family.

I will plot size and horizontality with AGR and HG families separately. If LFDH-S-a is true, then the AGR-plot should show a size-horizontality correlation, because then farming is the only way in which a family can grow large, and thanks to this farming, the spread should have been easier in the horizontal direction. Figures 2, 3 show the size-horizontality plot for AGR-families and HG-families respectively.

It appears from the plots that there is not significantly more horizontality in AGR families than in HG (not even for the largest families). Table 4 summarises the actual figures. There are no statistically significant differences between AGR and HG families. It follows that LFDH-S-a must be rejected or that other factors, e.g. mountains or bodies of water, significantly influence the picture.

While both AGR and HG families have larger than neutral horizontality, the plots show that cardinal size and horizontality do not correlate for either AGR or HG families (nor jointly), which means that horizontality is not involved as a cause or corollary in cardinal size expansion. I leave the question open as to what the real reason is that horizontality is larger than neutral.

7. Discussion and conclusions

Based on a worldwide test, the following statements are true:

- Most families are small, whether agricultural or hunter–gatherer.
- Agricultural families are significantly larger than hunter–gatherer families (on average).
- Small agricultural families more often than not have (only) agricultural neighbours.
- If agriculture was the cause of the larger agricultural families, one would expect them to show horizontality (at least more so than the corresponding hunter–gatherer families), but this is **not** the case.

It is always possible that the tests reported here fall short when more factors are taken into account. However, apparently relevant factors — mountains, forests, seas, extinct unattested branches, simplifying assumptions and erroneous/outlier coordinates — affect agricultural just as much as hunter–gatherer families. The bottom line is that the two kinds of families contrast when it comes to cardinal size, but are indistinguishable as regards horizontality. Thus, on the present evidence, both LFDH-S-a and LFDH-S-b need to be revised to fit our observed language family data.

Here is one tentative idea for revision: Before farming, an ecology of families exhibited small and large units randomly. When farming swept in, the larger families were naturally more likely to adopt it. (Imagine geographically coherent families of different sizes, and a stroke of a large agricultural brush.) Families with farming members slowly but steadily shifted wholesale to farming (due to population size advantages of

farming), but farming itself spread (via contact) much faster than farming families could extinguish non-farming neighbours. A small family which had adopted farming could not expand on its neighbours through farming if the neighbours were already farming as well. The advantages of this scenario are that larger average size of farming families is predicted without the necessity of horizontality, and the presence of small agricultural families is predicted if and only if they neighbour other agricultural families. If 'small' is defined as smaller than or equal to 10 members then there are some 164 small AGR-families; A majority (ca. 100) of these are found surrounded by other AGR families in East Papua, Sahel, Mexico, Andes and Eurasia, and the rest in HG-surroundings in the Amazon and New Guinea. Possibly some remaining AGR-families in HG-surroundings can be singled out in terms of geographical factors, i.e., the spread of farming is hindered by jungles, seas and mountains.

Future work should include more rigorous testing of weaker/refined LFDH-hypotheses and the inclusion of geographical factors preventing farming from spreading. Also, for many families, information on subsistence status via the reconstruction of the proto-language is to be expected in the future, which will allow for better tests of causality.

References

- Bellwood, Peter. 1997. "Prehistoric cultural explanations for widespread language families". *Archaeology and Linguistics: Aboriginal Australia in global perspective*, ed. by Patrick McConnell & Nicholas Evans, 123–134. Melbourne: Oxford University Press.
- Bellwood, Peter. 2001. "Archaeology and the historical determinants of punctuation in language-family origins". *Areal Diffusion and Genetic Inheritance*, ed. by Alexandra Aikhenvald & R. M. W. Dixon, 27–43. Oxford: Oxford University Press.
- Bellwood, Peter & Eusebio Dizon. 2008. "Austronesian cultural origins: Out of Taiwan, via the Batanes Islands, and onwards to Western Polynesia". *Past Human Migrations in East Asia: Matching archaeology, linguistics and genetics*, ed. by Alicia Sanchez-Mazas, Roger Blench, Malcolm D. Ross, Ilia Peiros & Marie Lin, 23–39. (= *Routledge Studies in the Early History of Asia* 5.) London: Routledge.
- Bellwood, Peter & Colin Renfrew, eds. 2002a. *Examining the Farming/Language Dispersal Hypothesis*. Oxford: McDonald Institute for Archaeological Research.
- Bellwood, Peter & Colin Renfrew. 2002b. "Foreword". *Examining the Farming/Language Dispersal Hypothesis*, ed. by Peter Bellwood & Colin Renfrew, xiii–xiv. Oxford: McDonald Institute for Archaeological Research.
- Blench, Roger M. 2005. "From the mountains to the valleys: Understanding ethnolinguistic geography in Southeast Asia". *The Peopling of East Asia*, ed. by Laurent Sagart, Roger M. Blench & Alicia Sanchez-Mazas, 31–50. London: Routledge.
- Blench, Roger M. 2006. *Language, Archaeology and the African Past*. (= *African Archaeology Series* 10.) Lanham, MD: Altamira Press.
- Blench, Roger M. 2008a. "Accounting for the pattern of the pattern of Amerindian Languages: What role did agriculture play in phylac dispersals?" Manuscript.

- Blench, Roger M. 2008b. "Stratification in the peopling of China: How far does the linguistic evidence match genetics and archaeology?" *Past Human Migrations in East Asia: Matching archaeology, linguistics and genetics*, ed. by Alicia Sanchez-Mazas, Roger Blench, Malcolm D. Ross, Ilia Peiros & Marie Lin, 105–132. (= *Routledge Studies in the Early History of Asia*, 5.) London: Routledge.
- Blench, Roger M. 2008c. "Agriculture and phylic dispersals: Reevaluating the evidence". Presented at *Us and Them: Modelling past genetic, linguistic, and cultural boundaries*, Bordeaux, 15–17 May 2008.
- Blench, Roger M., Malcolm D. Ross & Alicia Sanchez-Mazas. 2008. "Methodological issues: Linking genetic, linguistic and archaeological evidence". *Past Human Migrations in East Asia: Matching archaeology, linguistics and genetics*, ed. by Alicia Sanchez-Mazas, Roger Blench, Malcolm D. Ross, Ilia Peiros & Marie Lin, 3–19. London: Routledge.
- Campbell, Lyle. 2002. "What drives linguistic diversification and language spread?" *Examining the Farming/Language Dispersal Hypothesis*, ed. by Peter Bellwood & Colin Renfrew, 49–63. Oxford: McDonald Institute for Archaeological Research.
- Campbell, Lyle & William J. Poser. 2008. *Language Classification: History and method*. Cambridge University Press.
- Diamond, Jared. 1997. *Guns, Germs and Steel: The fates of human societies*. London: Cape.
- Diamond, Jared & Peter Bellwood. 2003. "Farmers and their languages: The first expansions". *Science* 300.596–603.
- Ehret, Christopher. 2002. "Language family expansions: Broadening our understandings of cause from an African perspective". *Examining the Farming/Language Dispersal Hypothesis*, ed. by Peter Bellwood & Colin Renfrew, 163–176. Oxford: McDonald Institute for Archaeological Research.
- Ehret, Christopher, S. O. Y. Keita & Paul Newman. 2004. "The origins of Afroasiatic". *Science* 306.5702.1680–1680.
- Evans, Nicholas & Rhys Jones. 1997. "The cradle of the Pama-Nyungans: Archaeological and linguistic speculations". *Archaeology and Linguistics: Aboriginal Australia in global perspective*, ed. by Patrick McConvell & Nicholas Evans, 385–417. Melbourne: Oxford University Press.
- Evans, Nicholas & Patrick McConvell. 1998. "The enigma of Pama-Nyungan expansion in Australia". *Archaeology and Language, II*, ed. by Roger M. Blench & Matthew Spriggs, 174–191. (= *One World Archaeology*, 29.) London: Routledge.
- Fuller, Dorian. 2002. "An agricultural perspective on Dravidian historical linguistics: Archaeological crop packages, livestock and Dravidian crop vocabulary". *Examining the Farming/Language dispersal hypothesis*, ed. by Peter Bellwood & Colin Renfrew, 191–213. Oxford: McDonald Institute for Archaeological Research.
- Hammarström, Harald. 2008. "Computations on language family dynamics". *Proceedings of the Second Swedish Language Technology Conference (SLTC-08), November 20–21, 2008, Stockholm* ed. by Rolf Carlson, Björn Granström, Anna Hjalmarsson, Viggo Kann, Beata Megyesi, Joakim Nivre, & Anna Sågvald Hein, 33–34. Stockholm: KTH.
- Higham, Charles F. W. 2002. "Language and farming dispersals: Austroasiatic languages and rice cultivation". *Examining the Farming/Language Dispersal Hypothesis*, ed. by Peter Bellwood & Colin Renfrew, 223–232. Oxford: McDonald Institute for Archaeological Research.
- Higham, Charles F. W. 1998. "Archaeology, linguistics and the expansion of the East and South-east Asian Neolithic". *Archaeology and Language, II*, ed. by Roger M. Blench & Matthew Spriggs, 103–114. (= *One World Archaeology*, 29.) London: Routledge.

- Hill, Jane H. 2001. "Proto-Uto-Aztecan: A community of cultivators in Central Mexico?" *American Anthropologist* 103:4.913–934.
- Hill, Jane H. 2002. "Proto-Uto-Aztecan cultivation and the northern devolution". *Examining the Farming/Language Dispersal Hypothesis*, ed. by Peter Bellwood & Colin Renfrew, 331–340. Oxford: McDonald Institute for Archaeological Research.
- Hill, Jane H. 2006. "Uto-Aztecan Hunter–Gatherers: Language diversity in the Takic spread and the Numic spread compared". Paper presented at *Historical Linguistics and Hunter–gatherer Populations in Global Perspective* Leipzig, 10–12 August 2006.
- Holden, Clare Janaki. 2002. "Bantu language trees reflect the spread of farming across sub-Saharan Africa: A maximum-parsimony analysis". *Proceedings of the Royal Society of London, Series B* 269.793–799.
- Hudson, Mark. 2002. "Agriculture and language change in the Japanese Islands". *Examining the Farming/Language Dispersal Hypothesis*, ed. by Peter Bellwood & Colin Renfrew, 311–318. Oxford: McDonald Institute for Archaeological Research.
- Lewis, Paul M., ed. 2009. *Ethnologue: Languages of the world*. 16th edn. Dallas: SIL International.
- Li, Wentian. 2002. "Zipf's Law everywhere". *Glottometrics* 5.14–21.
- MacDonald, Kevin C. 1998. "Archaeology, language and the peopling of West Africa: A consideration of the evidence". *Archaeology and Language*, II, ed. by Roger M. Blench & Matthew Spriggs, 33–66. (= *One World Archaeology* 29.) London: Routledge.
- Militarev, Alexander. 2002. "The Prehistory of a dispersal: The Proto-Afrasian (Afroasiatic) farming lexicon". *Examining the either Dispersal Hypothesis*, ed. by Peter Bellwood & Colin Renfrew, 135–150. Oxford: McDonald Institute for Archaeological Research.
- Mitzenmacher, Michael. 2004. "A brief history of generative models for power law and lognormal distributions". *Internet Mathematics* 1:2.226–251.
- Pawley, Andrew. 2002. "The Austronesian dispersal". *Examining the Farming/Language Dispersal Hypothesis*, ed. by Peter Bellwood & Colin Renfrew, 251–273. Oxford: McDonald Institute for Archaeological Research.
- Pawley, Andrew. 2005. "The chequered career of the Trans New Guinea hypothesis: Recent research and its implications". *Papuan Pasts: Studies in the cultural, linguistic and biological history of the Papuan-speaking peoples*, ed. by Andrew Pawley, Robert Attenborough, Jack Golson & Robin Hide, 67–108. (= *Pacific Linguistics*, 572.) Canberra: Research School of Pacific and Asian Studies, Australian National University.
- Pawley, Andrew. 2007. "Recent research on the historical relationships of the Papuan languages, or, what does linguistics say about the prehistory of Melanesia?" *Population Genetics, Linguistics and Culture History in the Southwest Pacific*, ed. by Jonathan S. Friedlaender, 36–59. Oxford: Oxford University Press.
- Phillipson, David W. 2002. "Language and farming dispersals in Sub-Saharan Africa, with particular reference to the Bantu-speaking peoples". *Examining the Farming/Language Dispersal Hypothesis*, ed. by Peter Bellwood & Colin Renfrew, 177–187. Oxford: McDonald Institute for Archaeological Research.
- Renfrew, Colin. 1987. *Archaeology and Language: The puzzle of Indo-European origins*. London: Cape.
- Renfrew, Colin. 1997. "World linguistic diversity and farming dispersals". *Archaeology and Language*, I, ed. by Roger M. Blench & Matthew Spriggs, 82–90. (= *One World Archaeology*, 27.) London: Routledge.

- Renfrew, Colin. 1999. "Reflections on the archaeology of linguistic diversity". *The Human Inheritance: Genes, language, and evolution* ed. by Bryan Sykes, 1–32. Oxford: Oxford University Press.
- Renfrew, Colin. 2002. "'The emerging synthesis': The archaeogenetics of farming/language dispersals and other spread zones". *Examining the Farming/Language Dispersal Hypothesis*. ed. by Peter Bellwood & Colin Renfrew, 3–16. Oxford: McDonald Institute for Archaeological Research.
- Ross, Malcolm. 2006. "Clues to the linguistic situation in Near Oceania before agriculture". Paper for presentation at the symposium Historical linguistics and hunter-gatherer populations in global perspective, MPI-EVA Leipzig, 10–12 August 2006.
- Southworth, Franklin C. 2006. "New light on three South Asian language families". *Mother Tongue* 11.124–159.
- van Driem, George. 2002. "Tibeto-Burman phylogeny and prehistory". *Examining the Farming/Language Dispersal Hypothesis*, ed. by Peter Bellwood & Colin Renfrew, 233–249. Oxford: McDonald Institute for Archaeological Research.
- Wichmann, Søren. 2005. "On the power-law distribution of language family sizes". *Journal of Linguistics* 41.117–131.
- Wichmann, Søren. To appear. "Neolithic linguistics". *Language and Prehistory of the Indo-European peoples: A cross-disciplinary perspective*, ed. by Gojko Barjamovic, Irene Elmerot, Adam Hyllested, Benedicte Nielsen & Bjørn Okholm Skaarup. Budapest: Archaeolingua.

Appendix. Language families: Listing, sizes, sources

This appendix can be found at <http://dx.doi.org/10.1075/dia.27.2.02ham.app>

Résumé

Il existe une hypothèse permettant d'expliquer la taille importante de certaines familles de langues (par opposition à des familles de taille plus restreinte) : On soutient en effet que ces familles sont importantes du fait que leurs locuteurs ont eu accès à des avantages technologiques, le plus souvent dans le domaine de l'agriculture. On appelle cette idée *Language/Farming Dispersal Hypothesis*. Auparavant, des études détaillées de certaines familles ont abouti à des résultats contradictoires d'une famille de langues à l'autre. Dans cette présentation nous examinons divers aspects de toutes les familles de langues connues. Pour chaque famille de langues nous nous servons du nombre de langues en tant que mesure cardinale. Par ailleurs, nous nous servons de données géographiques de ces langues afin de mesurer leur importance spatiale et nous nous appuyons sur des données ethnographiques afin de déterminer le type d'économie en question. On en conclut que, bien que les familles de langues tendent à être plus importantes dans les sociétés agricoles, leur taille n'est presque jamais liée à la simple présence d'un système basé sur l'agriculture. Si les techniques agricoles étaient responsables de l'expansion des familles de langues, la diffusion d'est en ouest de diverses grandes familles de langues serait plus importante que ce qui s'observe. Les résultats sont compatibles avec : 1- une version plus faible de la Lan-

guage/Farming Dispersal Hypothesis, ou 2-un modèle dans lequel des familles déjà importantes évoluent vers des sociétés agricoles justement en raison de leur taille, et non pas le contraire.

Zusammenfassung

Eine mögliche Erklärung der Tatsache, dass manche Sprachfamilien groß sind (während andere es nicht sind), ist die Hypothese, dass diese Sprachfamilien ihre Größe einem technologischen Vorteil verdanken, meistens dem Ackerbau. Diese Theorie wird als *Language/Farming Dispersal Hypothesis* (Sprache-Ackerbau-Verbreitungshypothese) bezeichnet. Bisherige Studien zu verschiedenen Sprachfamilien haben sowohl bestätigende als auch gegenteilige Beispiele zu dieser Theorie ergeben. In diesem Artikel sollen die vorliegenden Aussagen zu allen dokumentierten Sprachfamilien betrachtet und abgewogen werden. Für jede Sprachfamilie verwenden wir die Anzahl der Sprachen als Kardinalgröße, außerdem benutzen wir die geographischen Koordinaten der Einzelsprachen, um die geographische Ausdehnung zu bestimmen und verwenden ethnographische Informationen zur Feststellung der Wirtschaftsweise. und zeigen, dass, obwohl Ackerbau betreibende Sprachfamilien normalerweise größer sind, ihre Größe nicht einfach mit der Präsenz des Ackerbaus zu erklären ist. Wenn die Expansion von Sprachfamilien durch den Ackerbau werden könnte, wäre eine deutlichere Ost-West-Ausdehnung entlang der Breitengrade zu erwarten als zu beobachten ist. Die Daten sind jedoch vereinbar mit schwächeren Versionen der Sprache-Ackerbau-Verbreitungshypothese sowie mit Modellen, bei denen große Sprachfamilien aufgrund ihrer Größe zum Ackerbau übergehen statt umgekehrt.

Author's address

Harald Hammarström
Radboud Universiteit Nijmegen
Faculteit der Letteren
Postbus 9103
6500 HD Nijmegen, The Netherlands

h.hammarstrom@let.ru.nl