Problems testing typological correlations with the online *WALS*

MATTHEW S. DRYER

Abstract

The ease with which WALS *allows users to combine features from two maps and determine numbers of languages of the resulting types means that there is a danger of misusing the data from* WALS *to arrive at unsupported conclusions regarding typological correlations. I examine two instances where the overall numbers suggest a correlation and show that in only one of the two instances is there any reason to believe that there is in fact a correlation. In the case where the apparent correlation turns out to be an illusion, namely between tone and the order of object and verb, the illusion arises because most of the tone languages in* WALS *are in two areas which happen to be primarily VO. This illustrates the need to examine how the languages are distributed geographically. But this is information that* WALS *also provides, on the maps.*

Keywords: *consonant inventory, linguistic areas, linguistic atlas, methodology, phonology, sampling, tone, word order*

1. Introduction

The online *WALS* (Haspelmath et al. (eds.) 2008) allows users to combine the features for any two maps and in that way appears to provide users with a means of testing possible typological correlations or associations between typological variables. However, as I have argued in a number of previous publications (Dryer 1989, 2000, 2003), one cannot use raw language numbers to test for typological correlations because of problems of genealogical and areal biases. While this article largely repeats discussion of these problems, the ease with which *WALS* allows users to combine features from two maps and determine numbers of languages of the resulting types means that there is now a particularly great danger of misusing the data from *WALS* to arrive at unsupported conclusions regarding typological correlations.

Linguistic Typology 13 (2009), 121–135
DOI 10.1515/LITY.2009.007
1430–0532/2009/013-0121

In this article, I illustrate the problem with two possible typological correlations using features from *WALS* based on Maddieson's (2005a, 2008a; 2005b, 2008b; 2005c, 2008c) and Dryer's chapters (2005b, 2008b) and show how, although the raw language numbers suggest a correlation in both cases, careful examination shows that there is evidence of a correlation in only one of these two cases.[1] The method that I will use for testing for a correlation is a variation of one that I have used in a number of previous publications, one that crucially tests for whether the apparent correlation is found in different parts of the world and cannot be attributed to a small number of regions. The variation involves counting languages rather than genera and because of the hesitation that some linguists have had with counting genera, this variation may be more attractive to many linguists. I argue that most of the problems associated with raw language numbers, i.e., with counting total numbers of different types of languages in the world as a whole, arise due to failing to examine the distribution across geographical areas rather than due to counting languages rather than genera.

2. Uvular consonants and glottalized consonants

The first possible correlation that I will test involves two phonological features, based on Maddieson (2005a, 2008a; 2005b, 2008b), the presence of uvular consonants on the one hand and the presence of ejective consonants or glottalized resonants. Maddieson (2005b, 2008b) also includes a third type of glottalized consonant, namely implosives, which I exclude here. For ease of presentation I will henceforth use the term glottalized consonant to refer only to ejectives and glottalized resonants. Using the online *WALS* (Maddieson 2008a, 2008b), we find the raw totals for the number of languages with and without consonants of the two sorts given in Table 1.

Table 1. *Glottalized consonants and uvular consonants*

	Uvulars	No Uvulars
Glottalized	47	52
No Glottalized	51	416

1. I cite here the versions of these chapters in Haspelmath et al. (eds.) 2005, since I cite data in terms of numbers of genera or numbers of families and/or numbers of languages or genera divided by area that is only easily obtainable from the CD-ROM version of Haspelmath et al. (eds.) 2005. Because this data is only easily obtainable from Haspelmath et al. (eds.) 2005, the problems discussed in this article are particularly problematic for the online *WALS* (Haspelmath et al. (eds.) 2008).

Table 1 suggests a correlation between these two features, since although languages with uvular consonants are approximately as frequent as languages lacking uvular consonants among languages with glottalized consonants (47 vs. 52), languages lacking uvular consonants are much more common among languages without glottalized consonants (by 416 to 51).[2]

In various previous publications (Dryer 1989, 1992, 2003), I have used a method in which I divided the languages into six large continental areas and then counted the number of genera within each area that contain languages of that type.[3] Applying this method to the question at hand, we get the data in Table 2. The numbers represent the number of genera in the area specified that contain at least one language of the type specified. For example, the "4" in the first row, first column means that there are four genera in Africa containing languages with both glottalized consonants and uvular consonants and the "5" below it means that there are five genera in Africa containing languages with glottalized consonants but not uvular consonants. For each area, the larger number on the first and second lines is in boldface and the same is done for the third and fourth lines.

Table 2. *Genera containing languages with vs. without glottalized consonants and with vs. without uvular consonants*

	Africa	Eurasia	Southeast Asia & Oceania	Australia & New Guinea	North America	South America	Total
Glottalized & Uvular	4	8	1	1	**19**	**4**	37
Glottalized & No Uvular	**5**	**15**	**8**	1	6	1	36
No Glottalized & Uvular	14	1	5	0	15	7	42
No Glottalized & No Uvular	**43**	**20**	**35**	**59**	**25**	**39**	221

2. It is tempting to try to apply a Chi-Square or Fisher Exact Test to the data in Table 1, but, as discussed in Dryer 1989 and 2003, the data in Table 1 fails to meet a fundamental condition for the application of these tests, namely that the elements within each cell be independent. In other words, a significant Chi-Square or Fisher Exact Test result could simply reflect genealogical or areal bias in the sample rather than a typological correlation.
3. In Dryer 1989, I used five areas, but in this article I use the six areas I have used in subsequent publications, such as Dryer 1992 and 2003: Africa, Eurasia (which excludes Sino-Tibetan and other languages of southeast Asia), Southeast Asia & Oceania, Australia & New Guinea, North America, and South America.

The first two lines of Table 2 show that among languages with glottalized consonants, languages with uvular consonants and languages without uvular consonants are about equally common, the former being more common in two areas, the latter more common in three areas, and the two equally common in one area. In other words, we cannot say that languages with glottalized consonants tend to have uvular consonants as well. But the question of whether there is a correlation is a question of whether languages with glottalized consonants are more likely than languages without glottalized consonants to have uvular consonants. The last two lines of Table 2 suggest that this is the case: in all six areas, it is more common (in fact considerably more common) for languages without glottalized consonants to lack uvular consonants as well.

However, in order to rigourously test for a correlation, what we need to do is compare proportions over areas, as I have done in Dryer 1989, 1992, and 2003. The relevant figures are given in Table 3. Each of the figures in Table 3 is computed from a corresponding pair of figures in Table 2. The first line of Table 3 gives the number on the first line of Table 2 as a proportion of the sum of the numbers on the first and second lines. For example, the ".44" under Africa on the first line of Table 3 is computed by dividing 4 by 4+5. This gives the proportion of genera in Africa that contain languages with uvular consonants among those that contain languages with glottalized consonants.[4] The second line of Table 3 gives analogous proportions for the third and fourth lines of Table 2, i.e., for genera containing languages that lack glottalized consonants. The larger proportion for each area is in boldface.

Table 3. *Proportions of genera containing languages with uvular consonants among languages with vs. without glottalized consonants*

	Africa	Eurasia	Southeast Asia & Oceania	Australia & New Guinea	North America	South America	Mean
Glottalized	**.44**	**.35**	.11	**.50**	**.76**	**.80**	.49
No Glottalized	.25	.05	**.13**	.00	.38	.15	.16

4. Because it is possible for a genus to contain languages of two types and thus be represented in both cells, the number of genera containing languages with glottalized consonants might be less than the sum of the two numbers. Hence, more accurately, the proportion .44 in the upper leftmost cell in Table 3 is really the number of genera in Africa containing languages with glottalized consonants and uvular consonants as a proportion of the sum of the number of genera in Africa containing languages with glottalized consonants and uvular consonants and the number of genera containing languages with glottalized consonants but not uvular consonants.

As discussed in Dryer 1992 and 2003, in order for a correlation to be considered statistically significant, the proportion of one type must be higher in all six areas. If the proportion were higher on the first line of Table 3 for all six areas, then we could conclude that languages with glottalized consonants are indeed more likely than languages without glottalized consonants to have uvular consonants. However, the data in Table 3 falls short of this, since the proportion of genera containing languages with glottalized consonants is higher among languages with uvular consonants in only five of the six areas. However, the one proportion in Table 3 that does not conform, that for Southeast Asia & Oceania, is only slightly higher (.13 vs. .11) for languages without glottalized consonants, while the proportion for languages with glottalized consonants is greater by a larger amount in the other five areas (by an amount that varies from .19 in Africa to .65 in South America), so we are probably justified in concluding that there is a correlation, that is, that a language with glottalized consonants is more likely to have uvular consonants than one without glottalized consonants.

However, there is a variation on the method just applied that does show a preference for glottalized consonants among languages with uvular consonants in all six areas. The method illustrated in the preceding paragraphs differs from simply counting raw numbers of languages in two respects. First, it counts genera rather than languages. And second it counts them within six areas. Now it turns out that it is the second of these two differences that is by far the most crucial for controlling for genealogical and areal bias. The alternative method to the one used above is to simply count languages rather than genera within each of the six areas. Let me apply this method and then discuss the methodological issues surrounding the two methods.

Table 4. *Languages with vs. without glottalized consonants and with vs. without uvular consonants*

	Africa	Eurasia	Southeast Asia & Oceania	Australia & New Guinea	North America	South America	Total
Glottalized & Uvular	4	12	1	1	**23**	**6**	47
Glottalized & No Uvular	**8**	**23**	**8**	1	10	1	51
No Glottalized & Uvular	19	1	5	0	20	7	52
No Glottalized & No Uvular	**100**	**53**	**82**	**81**	**34**	**65**	415

Table 4 gives data comparable to that given above in Table 2, except that Table 4 gives numbers of languages rather than numbers of genera. Although the numbers in Table 4 are greater than those in Table 2, the overall pattern is the same: among languages with glottalized consonants, languages with uvular consonants are more common in two areas, languages without uvular consonants are more common in three areas and there is one area in which the two types have the same frequency. Furthermore, the particular areas that are of one sort in Table 2 are of the same sort in Table 4. For example, in both tables, it is North and South America where languages with uvular consonants are more common than languages without uvulars among languages with glottalized consonants.

Table 5 gives data comparable to Table 3 above, this time, however, giving proportions of languages (rather than proportions of genera) within each area.

Table 5. *Proportions of languages with uvular consonants among languages with vs. without glottalized consonants*

	Africa	Eurasia	Southeast Asia & Oceania	Australia & New Guinea	North America	South America	Mean
Glottalized	**.33**	**.34**	**.11**	**.50**	**.70**	**.86**	.47
No Glottalized	.16	.02	.06	.00	.37	.10	.12

When we compare proportions of languages within each area, we find that the proportion of languages which have uvular consonants is higher among languages with glottalized consonants in all six areas. The chance of this happening in each area is 1 in 2 and since there are six areas, the chance of this happening simultaneously in all six areas is 1 in 2^6, or 1 in 64. On the basis of this, we can conclude that languages with glottalized consonants are more likely to have uvular consonants than languages without glottalized consonants.

3. Counting genera vs. counting languages

Let us consider the methodological issues surrounding counting genera within each area versus counting languages within each area. Is it acceptable to count languages or is there some reason to count genera rather than languages? In Dryer 2000, I defended the value of counting genera in the face of arguments by Maslova 2000; however, while I believe that the particular arguments of Maslova were problematic and while there are domains where it does make a difference whether one counts genera or languages (see Section 5 below), there are reasons to believe that it actually does not make much difference when testing hypotheses of typological correlations (or associations) as long as one tests for whether the pattern is found in different areas.

There are four points to consider with respect to the issue of whether to count genera or languages. The first consideration is that, whether one counts genera or languages, the only assumption that underlies the method is that the six areas are independent of each other. If one makes that assumption, then the chances of all six areas patterning in a particular way due to chance is 1 in 64.[5] No assumptions are required about what is counted INSIDE each area. It is clear that even if one counts genera, there are strong areal and genealogical biases within each area. For example, the majority of languages in Southeast Asia & Oceania are Austronesian and many of the other languages in southeast Asia show strong similarities due to contact (Sinitic, Hmong-Mien, Tai-Kadai, and Mon-Khmer languages). So there is genealogical and areal bias within each area, whether one counts languages or genera. But this does not matter as far as the logic of the test is concerned; all the test assumes is that the six areas are independent of each other.

Why then have I counted genera rather than languages in my previous publications? The answer is that in selecting languages for my database, I have made no attempt to avoid genealogical bias within the sample of languages I include. The main reason for this is that I have attempted to include all languages for which there is relevant data and attempting to avoid bias in my sample would have meant that I was deliberately excluding languages. I have included many languages from the Oceanic branch of Austronesian, for example, even though these all belong to a single genus. The only consideration that genealogical classification has played in selecting languages for my database has been that I have made a greater effort to include languages in genera that were previously unrepresented in my database or that were poorly represented (for example I have attempted to get data from more languages from a particular genus, if I have data for only one language from that genus, but very little data for that one language). However, the samples of languages that most other authors of chapters in Haspelmath et al. (eds.) 2005 and 2008 use are in fact less biased genealogically. Many authors use a 200-language sample of languages that was constructed to cover a wide range of languages both genealogically and areally and was only slightly biased in two respects. First, it deliberately included a number of major languages, like English, German, French, and Spanish, and this introduced a slight bias towards Indo-European languages of Europe. And second, in order that the maps would not have large empty spaces in Africa and the Pacific, the sample included multiple languages from two large genera, Bantoid and Oceanic. However, overall the 200-language sample is relatively unbiased genealogically. Some authors, such as Maddieson, use larger lan-

5. Note that the chance of all six areas patterning the same way is 1 in 32, since there are two ways in which they might pattern the same way. But the chance of their patterning in one of these two ways in 1 in 64.

guage samples that they have chosen on their own, but these samples were also chosen with an aim at minimizing areal and genealogical bias. What this means is that in testing correlations using *WALS* chapters other than my own, such as the one under discussion using two of Maddieson's chapters, there is less of a need to count genera. For these, counting languages works almost as well. This also applies to testing correlations between features which involve one of my chapters and one of someone else's chapters (as in the next possible correlation to be discussed below) since the sample for such chapters will be the intersection of the samples for the two chapters and the intersection will generally be as unbiased as the less biased of the two samples involved. In other words, the second consideration is that in so far as there is a point to counting genera rather than languages, there is less of a need of this when at least one of the *WALS* chapters is one authored by someone other than myself.

A third consideration bearing on the choice between counting genera within areas and counting languages within areas is that in practice the results are usually the same. The example above, where all six areas pattern the same way when we count languages but only five areas pattern the same way when we count genera, is actually rather exceptional. In the vast majority of cases I have examined, either all six areas pattern the same way, regardless of whether one counts languages or genera, or they do not. And this is because the proportions of languages of a particular type within a particular area tend to be similar to the proportions of genera. This is illustrated by the fact that the proportions of languages in Table 5 are actually fairly similar to the proportions of genera in Table 3, as can be seen by examining Table 6, which gives the differences between the proportions in the two tables, where the number given is the difference between the proportion of genera in Table 3 and the proportion of languages in the corresponding cell in Table 5. We see that the largest difference in Table 6 is .11 and the other eleven figures are less than .10. In fact the mean difference in Table 6 is only .04. Hence in practice, there is usually little difference between counting languages and counting genera.

Table 6. *Difference between proportions of genera in Table 3 and proportions of languages in Table 5*

	Africa	Eurasia	Southeast Asia & Oceania	Australia & New Guinea	North America	South America	Mean
Glottalized	.11	.01	.00	.00	.06	.06	.04
No Glottalized	.09	.03	.07	.00	.01	.05	.04

A final consideration is the status of genera. The groups that are treated as genera in *WALS* (and in various publications of mine) are at best educated

guesses as to which groups are comparable to each other, which satisfy the criteria of being groups that are separated at a distance not more than 3,500 or 4,000 years, analogous to the subfamilies of Indo-European. However, considerable guesswork has gone into determining which groups should be considered genera and for many families there are many alternative ways in which one could assign genera. In other words, although counting genera rather than languages controls for a certain amount of genealogical bias, it also introduces some indeterminacy in the sense that what constitutes a genus is often unclear. Since by far the largest problem with counting raw numbers of languages in the world as a whole is the problem that differences in numbers may be due to a small number of geographical areas, issues about genera run the risk of distracting attention from the areal issue as the primary issue. Linguists who have expressed reluctance about counting genera should be encouraged to count languages, as long as they make some effort to determine whether a particular pattern is found throughout the world. In other words, most of the problems associated with counting raw numbers of languages is addressed if one counts proportions of languages within the six areas. However, I will discuss below other situations where it is still more important to count genera rather than to count languages.

4. Tone and order of object and verb

The second possible correlation that I will test is one involving a correlation between whether a language is a tone language and the order of object and verb, based on Maddieson (2005c, 2008c) and Dryer (2005b, 2008b). Maddieson (2005c, 2008c) identifies three types of languages as far as the presence of tone is concerned, those without tone, those with simple tone systems (either languages with pitch accent systems, like Japanese, or languages with only two tones, like Tibetan), and those with complex tone systems (with three or more tones). In my discussion here I will ignore the second of these three types and restrict attention to the first and third, those without tone and those with complex tone. Using the online *WALS* (Maddieson 2008c, Dryer 2008b), one can calculate the overall frequency of four types of languages, defined in terms of whether the language has complex tone or no tone and whether the language is VO or OV, given in Table 7. The raw numbers here immediately suggest a correlation: among languages with complex tone systems, VO languages outnumber OV languages by a considerable margin of 40 to 15, while among languages without tone, OV languages are slightly more common, by 120 to 89.

Table 7. *Tone and order of object and verb: raw numbers of languages*

	VO	OV
Complex Tone	40	15
No Tone	89	120

Table 8. *Tone and the order of object and verb: numbers of languages*

	Africa	Eurasia	Southeast Asia & Oceania	Australia & New Guinea	North America	South America	Total
Complex Tone & VO	**21**	0	**13**	0	**6**	0	40
Complex Tone & OV	6	0	5	**2**	1	**1**	15
No Tone & VO	**8**	14	**35**	8	**17**	7	89
No Tone & OV	2	**45**	4	**31**	14	**24**	120

Table 9. *Proportions of languages that are VO among languages with complex tone vs. languages without tone*

	Africa	Eurasia	Southeast Asia & Oceania	Australia & New Guinea	North America	South America	Mean
Complex Tone	.71	–	.72	.00	**.86**	.00	.46
No Tone	**.80**	.24	**.90**	**.21**	.55	**.23**	.59

Table 8 gives a breakdown of the language numbers from Table 7 into the six areas, analogous to Table 4 above. Once again, in order to determine whether there is a correlation, we need to compare the proportions within each of the six areas. The relevant data is given in Table 9, computed from Table 8. Each figure on the first line of Table 9 gives the proportion of languages in that area that are VO among those with complex tone, while each figure on the second line gives the proportion of languages that are VO among those that lack tone. The proportion on the first line of Table 9 for Eurasia is undefined because there are no languages (at least in this sample) with complex tone so that the

proportion is 0/0.[6] As before, the larger of the two figures within each area is in boldface. Table 9 shows clearly that there is no tendency for languages with complex tone to be more likely than languages without tone to be VO; if there were, the proportions would be higher on the first line of Table 9. But to the contrary, in only one area (North America) is the proportion of languages that are VO higher among languages with complex tone than among languages without tone. And in four areas the proportion of languages that are VO is higher among languages that do not have tone, so that in so far as there is a trend, it is in the opposite direction from that which the raw language numbers in Table 7 suggested.

Why are the results so different when we compare proportions of languages within areas from what the raw language numbers suggest? If we go back and examine Table 8, we see that although there are considerably more VO languages than OV languages with complex tone, by 40 to 15, these 40 languages with complex tone in the sample are found in only three of the six areas, Africa, Southeast Asia & Oceania, and North America, and in fact 34 of them are found in the first two of these areas. But these two areas are also two areas in which VO languages outnumber OV languages (in terms of numbers of languages in *WALS*, by 268 to 65 in Africa and by 169 to 112 in Southeast Asia & Oceania). What appears to be the case is that these two areas are two areas in which VO word order and complex tone are independently features that are common. It is precisely because these two features are common in these two areas that we find so many VO languages in these areas with complex tone. However, what shows that these two features are independent is the fact that in these two areas, the incidence of tone is just as common among the OV languages as it is among the VO languages. We find more VO languages than OV languages with complex tone in these areas only because there are more VO languages than OV languages in these areas. But the proportion of languages with tone is actually higher among the OV languages than among VO languages in these two areas. This can be shown more directly by rearranging the data in Table 8, by word order rather than by tone type, as shown in Table 10.

6. Recall that Eurasia is defined to exclude southeast Asia. Tone is in fact quite common in southeast Asia.

Table 10. *Order of object and verb vs. tone: numbers of languages*

	Africa	Eurasia	Southeast Asia & Oceania	Australia & New Guinea	North America	South America	Total
OV & Complex Tone	**6**	0	**5**	2	1	1	15
OV & No Tone	2	**45**	4	**31**	**14**	**24**	120
VO & Complex Tone	**21**	0	13	0	6	0	40
VO & No Tone	8	**14**	**35**	**8**	**17**	**7**	89

Table 11. *Proportions of languages that have complex tone among OV vs. VO languages*

	Africa	Eurasia	Southeast Asia & Oceania	Australia & New Guinea	North America	South America	Mean
OV	**.75**	.00	**.56**	**.06**	.07	**.04**	.25
VO	.72	.00	.27	.00	**.26**	.00	.21

Table 11 gives the proportions of languages that have complex tone, among OV languages and among VO languages, based on Table 10, and shows how the proportion of languages with complex tone is higher among OV languages than among VO languages in both Africa and Southeast Asia & Oceania (though only slightly in the former). But, as noted above, because these two areas contain more VO languages than OV languages, there are more VO languages than OV languages with complex tone in these two areas: there are 34 VO languages in these two areas with complex tone and 11 OV languages with complex tone. But these are most of the languages in the sample with complex tone; outside this area there are only 6 VO languages with complex tone and only 4 OV languages with complex tone. This example shows particularly clearly the danger of using data from raw totals of languages without examining their distribution over areas.

5. Conclusion

I have used proportions of languages within areas rather than proportions of genera in this article, in order to emphasize the fact that the primary problem with counting raw language numbers is not in counting languages rather than genera but in failing to look at their distribution over areas. Does this mean that there is no value in counting genera at all? Is my defence of counting genera in Dryer 2000 now irrelevant? As mentioned earlier, for samples of languages that are already relatively unbiased, it makes less difference in practice whether we count languages or genera. However, when looking at

possible correlations that involve only my own database, I intend to continue to examine proportions of genera rather than proportions of languages, given that my own database has not been constructed with an aim towards having an unbiased sample. Furthermore, although the difference between comparing proportions of languages over areas and comparing proportions of genera over areas is relatively small, there are other domains in which it makes a greater difference whether one counts genera or languages. Consider the basic question of whether there is a typological preference for SOV over SVO word order (see also Dryer 2000). If we simply compare the total language numbers of these two types, using the data from Dryer 2008a, we find that SOV outnumbers SVO by a relatively small difference of 497 languages to 436, suggesting that there is no typological preference for SOV over SVO. However, when we compare the number of genera containing these two types of languages, we find 205 genera containing SOV languages and only 108 genera containing SVO languages, in other words approximately twice as many genera containing SOV languages.[7] While this difference itself may well be due to historical accident, it at least shows that there may indeed be a typological preference for SOV over SVO that is obscured when we consider numbers of languages, showing that at least in cases like this one, counting genera does yield different results from counting languages.[8]

The current version of the online *WALS* does not provide any means for counting either languages or genera within specific areas, nor in fact does it provide any means for counting genera at all; it only provides raw language numbers. The CD-ROM version of *WALS* which came with Haspelmath et al. (eds.) 2005 does provide a way to count both languages and genera by the six areas, though the process is slightly cumbersome because one must define filters for the six areas and then select each filter individually to get the data for each area. This is how I derived the numbers cited in Tables 2, 4, and 8. The most that one can do with the current version of the online *WALS* to address the problem discussed here is to look at the maps in addition to the raw language numbers to see whether a pattern is geographically widespread. If one uses the online *WALS* to combine the features for Tone and the Order of Object

7. The data cited for numbers of languages comes from Dryer 2008a rather than Dryer 2005a, since it includes a correction of an error in Dryer 2005b. The data for numbers of genera comes from Dryer 2005a (since data on numbers of genera are not available in Haspelmath et al. (eds.) 2008), but the error does not affect the number of genera.
8. Dryer 2005a also provides data in terms of numbers of families, though I consider such numbers of questionable value since it is quite unclear in many cases whether different groups should be considered as belonging to the same family, and different assumptions as to what the families are can lead to very different results. But ignoring this, we find 106 families containing SOV languages and only 41 containing SVO, showing an even stronger preference for SOV over SVO.

and Verb, using Maddieson 2008c and Dryer 2008b, and suppresses the dots (by choosing "no icon") for all feature value combinations other than OV & Complex Tone and VO & Complex Tone, one sees that although there are far fewer dots for OV languages with complex tone, these dots are as spread out as if not more spread out than the dots for VO languages with complex tone, that the dots for VO languages with complex tone are very areally concentrated, with a clear cluster in southeast Asia, a small cluster in southern Mexico, and a somewhat more diffuse distribution in Africa, with a belt of languages across central Africa from Senegal to Ethiopia plus two other languages in southern Africa. Significantly, not only are there also OV languages with complex tone in all these areas, but there are also OV languages with complex tone in New Guinea, Brazil, and the southeastern United States. In other words, even the map just referred to shows a broader geographical distribution of OV languages with complex tone. What this map does not show, however, is the fact that the two areas where one finds most of the VO languages with tone are also two areas where VO languages greatly outnumber OV languages so that it is not obvious from the online *WALS* that in these two areas the incidence of tone is as high among OV languages as it is among VO languages.

The general moral should be clear. Examining raw numbers of languages from the online *WALS* can be very misleading. There is little question that one of the unfortunate results from the online *WALS* will be erroneous conclusions. Hopefully, in the long run, these confusions will be cleared up.

Received: 25 June 2008
Revised: 25 December 2008

University at Buffalo

Correspondence address: Linguistics Department, University at Buffalo, 609 Baldy Hall, Buffalo, NY 14260, U.S.A.; e-mail: dryer@buffalo.edu

Acknowledgements: Part of the research for this article was made possible by Social Sciences and Humanities Research Council of Canada Grants 410-810949, 410-830354, and 410-850540, by National Science Foundation Research Grant BNS-9011190, and by support from the Max-Planck-Institut für evolutionäre Anthropologie in Leipzig, Germany.

References

Dryer, Matthew S. 1989. Large linguistic areas and language sampling. *Studies in Language* 13. 257–292.

Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68. 81–138.

Dryer, Matthew S. 2000. Counting genera vs. counting languages: A reply to Maslova. *Linguistic Typology* 4. 334–350.

Dryer, Matthew S. 2003. Significant and non-significant implicational universals. *Linguistic Typology* 7. 108–128.

Dryer, Matthew S. 2005a. Order of subject, object and verb. In Haspelmath et al. (eds.) 2005, 330–333.

Dryer, Matthew S. 2005b. Order of object and verb. In Haspelmath et al. (eds.) 2005, 338–341.
Dryer, Matthew S. 2008a. Order of subject, object and verb. In Haspelmath et al. (eds.) 2008, Chapter 81. http://wals.info/feature/81 (21 June 2008)
Dryer, Matthew S. 2008b. Order of object and verb. In Haspelmath et al. (eds.) 2008, Chapter 83. http://wals.info/feature/83 (21 June 2008)
Haspelmath, Martin, Matthew S. Dryer, David Gil & Bernard Comrie (eds.). 2005. *World atlas of language structures*. Oxford: Oxford University Press.
Haspelmath, Martin, Matthew S. Dryer, David Gil & Bernard Comrie (eds.). 2008. The world atlas of language structures online. München: Max Planck Digital Library. http://wals.info/
Maddieson, Ian. 2005a. Uvular consonants. In Haspelmath et al. (eds.) 2005, 30–33.
Maddieson, Ian. 2005b. Glottalized consonants. In Haspelmath et al. (eds.) 2005, 34–37.
Maddieson, Ian. 2005c. Tone. In Haspelmath et al. (eds.) 2005, 58–61.
Maddieson, Ian. 2008a. Uvular consonants. In Haspelmath et al. (eds.) 2008, Chapter 6. http://wals.info/feature/6 (21 June 2008)
Maddieson, Ian. 2008b. Glottalized consonants. In Haspelmath et al. (eds.) 2008, Chapter 7. http://wals.info/feature/7 (21 June 2008)
Maddieson, Ian. 2008c. Tone. In Haspelmath et al. (eds.) 2008, Chapter 13. http://wals.info/feature/13 (21 June 2008)
Maslova, Elena. 2000. A dynamic approach fo the verification of distributional universals. *Linguistic Typology* 4. 307–333.