

SØREN WICHMANN (Leipzig & Leiden)  
DAVID KAMHOLZ (Leipzig)

## A stability metric for typological features

In this paper we propose a method to measure the stability of typological features given a data set such as HASPELMATH *et al.* (2005). We determine the frequency with which one and the same value of a feature occurs in shallow genealogical groupings ('genus' level). If the languages within such a group tend to show the same value, there is reason to believe that the feature is inherently more stable than if much variation is found within such groups. The proposed metric may be used to evaluate the utility of linguistic features when employing typological data for inferring genealogical relationships among languages.

### 1. Introduction<sup>1</sup>

The use of typological features for making historical linguistic inferences is a subject of recent interest (DUNN *et al.* 2005, WICHMANN & SAUNDERS 2006) and it is worth investigating to what extent it can provide a valuable supplement to existing methods. If one is interested in using typological features for such purposes (or typological features in addition to some other data set), one should choose the features which tend to be most stable diachronically and which are least amenable to change, whether this change be due to internal factors or areal convergence.

In this paper, propose a measure to evaluate the stability of typological features. The approach that we suggest in this paper is to study the distribution of values for a given feature at the level of uncontroversial genetic units of the same approximate time depth in order to see to what degree the feature tends to have one and the same realization at this level. A convenient list of uncontroversial genetic units is provided by the 'genera' in DRYER's (2005a) contribution to HASPELMATH *et al.* (2005), henceforth WALS. A genus is a group of languages which are clearly related and whose time depth, in all cases where this is known, does not exceed four thousand years. The level of classification corresponding to a genus is intended to be comparable across the world. It is not an entirely objective notion, but relies to some extent on the informed intuitions of MATTHEW DRYER (see DRYER 1989 for discussion).

Ideally, also areal groups should be considered in such a measure of stability. However, it is difficult to deal with issues of areal convergence since areal distributions are not easily quantified, and since it is difficult to single out a type of areal unit by which global distributions might be measured. Most features in WALS exhibit some sort of areal patterning—not surprisingly, as they were generally selected with the goal of producing telling maps—but the size of the areas involved differ tremendously. Thus, some features have an areal distribution involving more than one continent (e.g. the Pacific Rim or Sahul), others involve most of a single

---

<sup>1</sup> We would like to thank Bernard Comrie for comments and Michael Cysouw for extensive discussion and suggestions for improvements going far beyond normal editorial duties. Eric W. Holman provided help with some of the statistical issues and input on many issues pertaining to this paper. None of these scholars, of course, necessarily shares any of our views.

continent (e.g., Africa), and yet others involve increasingly smaller areas. These factors impose the difficulty of selecting one particular size of area that would be suitable for comparing the areal behavior of different features in the same way that the genus is a suitable diachronic unit for looking at the diachronic stability of features. In the present report, then, we leave the issue of areal convergence unresolved.

An example illustrating that a feature may be relatively stable diachronically and at the same time amenable to areal influence is that of front rounded vowels. Front rounded vowels are widespread in Eurasia but exceedingly rare outside of this macro-area (MADDIESON 2005a = WALS 11). This means that for some Eurasian genera the presence of front rounded vowels are the norm and for any non-Eurasian genus the norm is for them to be absent. This distribution is very probably the result of some kind of diffusion. In the measures of genetic stability, which we will present shortly, this feature will look as if it is genealogically stable since one particular value—either presence or absence—tends to dominate throughout the world’s genera. In fact, the feature *is* very stable genealogically, as witnessed by the general absence of random fluctuations of values among related languages. Nevertheless, it is also amenable to diffusion. While the degree of diffusability is perhaps impossible to measure, the overall effect of internal language change appears to be greater than that of diffusion, since otherwise all languages would quickly end up having similar typological profiles. This is why it makes sense to study the different rates of change for typological features, leaving aside the issue of diffusability.

Moreover, although it is disquieting not to have a way to control for convergence, there is evidence to suggest that we need not worry too much about this factor when making historical inferences on the basis of typological features. Leafing through WALS quickly shows that no two distributional maps are quite the same. Furthermore, it is well known that even the best established ‘linguistic areas’ are defined on the basis of just a small handful of features, and even these areas have fuzzy boundaries since the isoglosses that define them rarely align neatly. If genealogies were to be established on the basis of a small handful of typological features, areal effects could skew the results heavily, but the more features that are used, the more such effects should be expected to cancel each other out due to the differential geographical distributions.

## **2. The quantitative measure of diachronic stability: background**

A starting assumption for our approach is that the feature value that is most favored in a given genus is the one that should be reconstructed for the proto-language of the genus, and that languages exhibiting other values must have undergone changes. There are obviously counterexamples to this, but we think that at a shallow time depth this assumption holds true in the great majority of cases. Moreover, in most cases we simply do not know what the true history was, so we need this kind of a working assumption. On the basis of such considerations we can make the crucial deduction that *the better represented the best represented feature value in a given genus is, the more stable that feature may be assumed to be within*

*the genus*. In other words, the more widespread the feature value that we consider to be inherited from the proto-language is, the more stable the feature. Following this logic we may study the distribution of values of a given feature for each genus and then calculate an average of how well represented the best-represented value is throughout all genera in the WALS sample. It is irrelevant to what degree ‘the best represented value’ varies across genera—when each genus has a high degree of consistency of one particular value then the feature as a whole should be considered highly stable.

A number of problems arise in implementing this kind of evaluation strategy. One issue is how to compare features that have different numbers of values. The features of WALS may have from 2 to 9 values. Another issue is that the number of languages sampled per genera varies. For some features particular genera are only represented by one language. At the other extreme, there is the case of the map on the order of object and verb (DRYER 2005*b* = WALS 83) for which as many as 95 languages of the Bantoid genus have been coded. Thus, we have to be able to deal with a number of languages per genus ranging from 1 to 95. In short, the question is how are we to compare the stability of features, considering the three variables that are involved: the number of occurrences of the best represented feature value, the number of possible feature values, and the number of languages within a genus for which the feature is attested in the WALS sample?

For an illustration of this question, suppose that we are testing which of the two following features are most stable: the order of subject, object, and verb (DRYER 2005*c* = WALS 81) or the associative plural (DANIEL & MORAVCSIK 2005 = WALS 36). Even if we just limit the test to Germanic, we still have the problem that the numbers of languages sampled for the two features are different, and the number of possible values is also different. For the order of subject, object, and verb the best-represented value (‘SVO’) occurs in 5 of the 8 Germanic languages in the sample and it is one out of 7 possible feature values. How may we now compare this to the situation for the associate plural? Here the best-represented feature value (‘unique periphrastic associative plural’) is one out of 4 possible values and occurs 5 times in a sample of 7 Germanic languages. Put schematically, we have the situation shown in Table 1.

Table 1. The distribution of values for two features in the Germanic genus

	<i>Associative Plural</i>	<i>Order of Subject, Object and Verb</i>
No. of occurrences of best represented value	5	5
No. of languages in sample	7	8
No. of possible feature values	4	7

To explain our approach, consider a translation of the problem into a situation that might be more familiar: the random drawing of cards. In the first case we have to decide what the probability is of drawing the same card 5 times when there are 7 cards to draw from and each may have 4 different values. This can only be decided if we know what the probability of each of the 4 different values is. For the present

paper, we simply assume that these probabilities are equal. With a bit of mathematical intuition, it is easy to see that the probability of drawing the same card 5 times when there are 7 cards to draw from and each may have 4 different values is higher than the probability of drawing the same card 5 times when there are 8 to draw from and the number of (equally) possible values is 7. Obviously, all else being equal, it is a bit less likely to draw the same card 5 times out of 7 than 5 times out of 8. But when there are several more equally possible values involved, as in the second situation, it is going to become *very* improbable to draw 5 that are the same out of 8.

In the example from Germanic, then, the probability that the number of occurrences of the best-represented feature value could be due to sheer chance is lower for the order of subject, object, and verb than for the associate plural. So the Germanic evidence suggests that the order of subject, object, and verb is the more stable feature of the two.<sup>2</sup> But we need to quantify such probabilities—henceforth referred to as *p*-values—such that we may calculate them for all features and all genera. In section 3 we will describe our procedure for the quantification of the *p*-values. As alluded to above, when computing *p*-values we assume that the occurrence of each value of any feature is equally probable. This is potentially a problematic assumption. An example that may illustrate this issue is the feature on the presence of uncommon consonants (MADDIESON 2005*b* = WALS 19). This feature has a highly skewed distribution. One value, ‘none’, accounts for 79% of the languages, while the other six values together account for only 21%. Such a distribution may produce a much lower *p*-value than for features whose values are more equally distributed. We discuss this issue further in section 4.

### 3. A procedure for measuring diachronic stability

The variables involved in calculating *p*-values are the different possible values of a feature (we label these *a*, *b*, *c*, ...), the number of possible values, *k*, and the number of languages in the set, *n*. The number of times that the best-represented feature-value occurs is labeled *r*. As an illustration of our procedure, we provide an example in Table 2. In this example we have a feature with 2 possible values (*a* or *b*). Thus *k* = 2. There is a set of 4 languages (*n* = 4). Table 2 provides all the logically possible distributions and the corresponding value of the best-represented feature values, *r*. Since there are 16 logically possible distributions, the different values of *r* have the probabilities attached to them as shown in Table 3.

---

<sup>2</sup> The example used here is perhaps not the best, as pointed out to us by B. COMRIE, p.c. In actual fact, proto-Germanic had SOV word order. So in this case using the best represented feature value (SVO) as the criterion for reconstruction gives the wrong result. However, in a large statistical investigation where only a small minority of the features considered have been reconstructed by proper methods such errors are unavoidable and, as already mentioned, hardly damaging overall. We could have chosen a different example, but this would not change the methodology.

Table 2. An example of the distributional possibilities when  $k = 2$  and  $n = 4$

<i>Distribution</i>	<i>r</i>	<i>Distribution</i>	<i>r</i>
aaaa	4	bbab	3
bbbb	4	bbba	3
aaab	3	aabb	2
aaba	3	abab	2
abaa	3	abba	2
baaa	3	baab	2
abbb	3	baba	2
babb	3	bbaa	2

Table 3. Probabilities of the different possible values of  $r$  for the data in Table 2

<i>r</i>	<i>Probability</i>
4	2/16
3	8/16
2	6/16

There are two ways that a  $p$ -value could be calculated. One is to generate a table like table 2 for the different values of  $k$  and  $n$  and then go through the table for given values of  $r$  and calculate the  $p$ -value. We might call this the ‘brute force’ approach. Another way would be to derive a general mathematical formula for calculating these values. We have chosen the former, less elegant approach. Thus, one of us (KAMHOLZ) wrote a computer program in Perl, which simply calculated the  $p$ -values for each feature, and genus in WALS.<sup>3</sup> By averaging the  $p$ -values found for each feature, we generate a ranked list of features where the (averaged)  $p$ -value is inversely proportional to the rank-order of the corresponding feature in terms of its usefulness for genealogical analyses. The results are shown in the Appendix. In the Appendix, the column ‘No. of genera’ indicates how many genera were available for the measurement of  $p$ -values (genera with just one language obviously cannot be used). This should be taken into account when assessing the validity of the  $p$ -values. It is hard to say exactly how much data it takes for the figures to be reliable, but one should certainly not place too much confidence in a  $p$ -value based on just some 30 genera or less.

Our approach to simply take the average of the  $p$ -values was licensed by a two-way ANOVA (analysis of variance). This statistical test was used to estimate whether there are significant overall differences in variance between  $p$ -values within a given feature as opposed to between features (and it also tests for significant overall differences across the groups of languages compared). In order to avoid the controversial issue of how to deal with empty cells when doing this test, we reduced the data set to the maximum size possible that did not contain empty cells. In this set, there were  $p$ -values for only 66 features and just 7 genera. The genera considered were Bantoid, Bodic, Germanic, Gur, Indic, Oceanic, and Se-

<sup>3</sup> The program as well as data and instructions may be downloaded from [http://email.eva.mpg.de/~wichmann/pcalculator/pcalculator\\_index.htm](http://email.eva.mpg.de/~wichmann/pcalculator/pcalculator_index.htm)

mitic, and the features considered were the following: 1-9, 11, 12, 13, 18, 19, 26, 30-33, 37, 38, 41-44, 48, 51, 55, 57, 69, 70, 71, 73, 77, 78, 81, 82, 83, 85-88, 90-97, 100-105, 107, 111, 112, 115, 116, 118, 119, 120, and 124. The result was highly statistically significant, as witnessed by the low probability for the differences to obtain ( $p < 0.00000001$ ) and a value of  $F$ , which is higher than  $F_{crit}$ . There were also significant differences in averages of  $p$ -values across genera, but this only tells us that some genera have undergone more changes than others. Table 4 summarizes the results of the ANOVA. The test is not ideal because it assumes normally distributed data and the independence of genera, neither of which hold. Regarding the latter, Bantoid and Gur both belong to the Niger-Congo family and Germanic and Indic are both Indo-European. Moreover, several pairs among the genera are within a distance of 5000 km, where we expect the effects of diffusion to be found (HOLMAN *et al.* 2006). An improved sample, however, would require us to reduce the amount of data so much that a statistical test would not be meaningful anymore.

Table 4. Results of a two-way ANOVA of the stability indices across 7 genera

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Features	5.7023	65	0.0877	2.6493	3.86E-09	1.3408
Genera	2.4983	6	0.4164	12.5743	5.64E-13	2.1218
Error	12.9145	390	0.0331			
Total	21.1152	461				

By doing so-called *post-hoc* tests it would be possible to say something about the significance in differences between average  $p$ -values for individual features. We suspect that most adjacent features in the ranking of the features (as shown in the Appendix) are not significantly different from each other, and that it requires a certain number of steps within the ranking for features for them to be significantly different. The point of the rank ordering is not, however, to suggest fine-grained differences in the first place. The utility of the list is that it gives a general idea of what sorts of features are at the top as opposed to the bottom of the scale. The ANOVA test suggests that the differences among the extremes are significant.

In the following three sections we discuss various potential problems for our approach.

#### 4. Skewed distributions

As mentioned above, it is potentially a problem for our approach that features whose distribution of values happens to be skewed may result in very low average  $p$ -values. The application of Herfindahl-Hirschman index (HHI) was suggested as a measurement of skewedness that could be used to test for an inverse correlation with our  $p$ -values (M. CYSOUW, p.c.). The HHI has its origin in economics to measure the concentration of a market (HERFINDAHL 1950, HIRSCHMAN 1945). It is calculated by squaring the market share of each firm competing within a market

(in percentages) and then summing the resulting numbers. The maximum concentration is 10,000, i.e., the situation where one firm has 100% market share.

The WALS database contains information about how many languages are in the sample for a given feature and how many languages in that sample have each of the feature values. Percentages can be derived from these numbers, which can be treated in the same way as one would treat market shares when calculating the HHI. The HHIs for WALS features all lie in the range of 1,605 to 8,747. If features were markets, values were firms, and frequencies among the world's languages were market shares, then the most 'unmonopolized' feature would be the feature on action nominal constructions (KOPTJEVSKAJA-TAMM 2005 = WALS 62) for which the HHI is 1,605. All other features have higher HHIs (see the Appendix for the HHI of each WALS feature). We used the HHI to check whether there is an inverse correlation with our average  $p$ -values. There indeed turns out to be a significant negative correlation (Pearson's  $r = -0.30$ ,  $p < 0.001$ ), though with an explained variance of less than 10% ( $r^2 = 0.09$ ) this is only of minor influence on the actual  $p$ -values.

[For the next theme: insert a few sentences explaining that although the number of values has been included in the computation of the  $p$ -values, it is still interesting to check whether there is any effect with the \*average\*  $p$ -values. Maybe you should move this discussion to the end of the previous paragraph, because it is a control of a dependent variable. This is different from the discussion on the HHI, which is a test of an independent variable.] SW: I have now added a couple of sentences, marked in yellow below

## 5. Variation in the number of possible features

Another factor, which one might suspect could influence our results, is the sheer number of possible values that a feature may have. The number of possible values feeds into the calculation of  $p$ -values for individual features and genera and is thus a dependent variable, but our method would be problematical if the number of possible values determined the *average*  $p$ -values. We checked for a correlation between the number of values for each feature and the corresponding average  $p$ -value and found a weakly negative correlation ( $r = -0.195$ ;  $p < 0.05$ ), meaning that the average  $p$ -values have a slight tendency to grow when the number of values diminishes. The correlation, however, disappeared when binary values were eliminated ( $r = -0.092$ ; n.s.). This means that the ten binary features of WALS are responsible for this correlation because they tend to have high average  $p$ -values. It is an interesting question whether this is an artifact of our method or whether these ten binary features actually do tend to be less stable. We believe that the latter is a real possibility.

One should keep in mind that a major motivation behind the selection and formulation of features leading to the WALS maps was to produce striking maps. For binary features in the WALS there is perhaps a tendency that the two values are more evenly distributed than one might expect from a randomly chosen binary fea-

ture. An example of a feature that is formulated as many-valued, but could equally well have been formulated as binary, is the feature on reciprocal constructions (MASLOVA & NEDJALOV 2005 = WALS 106). This feature has four values, where one is the absence of reciprocal constructions and the others are different kinds of such constructions. If it were formulated as a binary feature—the values being presence vs. absence of reciprocal constructions—then there would only be 16 languages having the value ‘absence’ and 159 having the value ‘presence’. In contrast, the feature on passive constructions is binary (SIEWIERSKA 2005 = WALS 107) and the distribution within the sample is more equal: 162 languages having and 211 languages not having a passive. If the feature on reciprocal constructions had been encoded as a single binary feature like the feature on passive constructions, then it would have come out as highly stable, although the map would have been somewhat uninteresting. Based on such examples, we believe that it is the urge to produce striking maps that ultimately explains the special behavior of binary features.

[Add some more explanation on the argumentation here, something like this: “If a binary feature would be strongly skewed, it would not produce a striking map, and consequently the author of the feature would attempt a further differentiation of the features. Because of such differentiation, the stability of the feature is reduced as small changes already count as a typological change. It is thus not the case that binary features in general are less stable. However, in typological practice, strongly skewed binary features will normally be split up into several subtypes by the authors. In a sense, the attested correlation is thus an artifact of the design of the maps being included in the WALS] SW: I don’t think this adds anything to what was already said so I prefer to not have more text here.

In any case, if there really were a strong effect of the sheer number of values on our stability indices, the correlation should also hold when binary values are excluded. So we do not see any problem for our method regarding the slightly significant correlation of the number of feature-values with the average  $p$ -values.

## 6. Variation in the number of languages attested

A final factor potentially influencing our results is the number of languages for which a given feature is attested (see the sixth column in the Appendix for these numbers). Like the number of possible features this is a dependent variable, feeding into the calculation of  $p$ -values, and it is potentially problematical if it affects the averaged values significantly. A test of a correlation between averaged  $p$ -values and the number of languages for which a given feature is attested (ranging from 112 to 1370) gives  $r = -0.533$ , indicating a moderate negative correlation. On a first consideration this correlation seems very problematical since it explains as much as 28.4% of the variance. It turns out, however, that the correlation is mainly due to the relatively few features which are very well attested. If the 24 features for which more than 600 languages are attested are excluded, the coefficient of correlation drops down to a mere  $r = -0.190$ . The 24 features in question nearly all relate to either word order or the ordering of affixes. The fact that the correlation largely disappears when such features are excluded suggests that it could be the stability

inherent in these features rather than the sheer high degree to which they are attested that mainly contributes to the low  $p$ -values. In other words, we propose that the moderate negative correlation between  $p$ -values and the number of languages for which features are attested is a secondary effect of a certain type of feature happening to be both particularly well attested and relatively stable.

## 7. Conclusion: on the utility of WALS features for genealogical analyses

The Appendix provides the rank ordering by average  $p$ -values of the features in WALS. We propose that this ranking is indicative of the genealogical stability of these features. Naturally, this stability ranking has to be taken with a grain of salt. A feature such as the one on the word for *tea* in the world's languages (DAHL 2005 = WALS 138) ranks in the middle of the list, but cannot be used for any genealogical analyses whatsoever since it is a feature which maps the distribution of a loanword. Other features are potentially problematic because one of their values is defined negatively, i.e. as the non-applicability of any of the other possible values. An example of this is the high-ranking feature on the absence of common consonants (MADDIESON 2005c = WALS 18). Out of a sample of 566 languages, 502 (88.7%) exhibit the same value corresponding to the presence of all three kinds of common consonants considered (bilabials, fricatives, and nasals). There is no reason to assume that the presence of common consonants is indicative of a shared history. On the other hand, there may be good reasons to believe that other values in the set, such as the absence of fricatives, which is relatively widespread among Australian languages and certain languages of New Guinea, are, indeed, valid genealogical markers. In the end, it may turn out that such features are not very damaging to the genealogical analysis. In an analysis carried out by the use of current computational phylogenetic techniques, the highly frequent value 'presence of common consonants' would likely be harmless, since it would only produce some noise, not reinforcing nodes motivated by other features. The different values for absences of common consonants, however, may reinforce nodes motivated by other data, and thus the overall effect would be a contribution in the right direction. It would require more empirical data to decide for given cases whether features that allow for an 'other' value are damaging or productive.

In WICHMANN & SAUNDERS (2006), average  $p$ -values were used for selecting a small set of features used for making genealogical inferences concerning a sample of Native American languages. When a selection of the features with low average  $p$ -values was used, better results were obtained than when using a selection of the features with high  $p$ -values. That is, when the former selection was used as data-input to various phylogenetic algorithms, known genealogical relations among languages were brought out; when the latter selection was used, the genealogical trees were erroneous. This type of test cannot fully confirm that the average  $p$ -values represent an adequate measure of stability, but it does bring support in their favor. A fully adequate test of the stability metric proposed here, as well as any other metric that one might come up with, would require an artificial data set where the actual rates of change of different features were known. Such a data set could be created by means of computer simulations in which a data set similar to that of WALS

was generated, but where the designer of the simulation determined the different rates of change for each feature. Subsequently our proposed stability metric would be applied to this data set (perhaps iteratively to different data sets generated by one and the same model in order to get statistical support). If the metric gave stability values correlating with the rate of change pre-specified for each feature, then its validity would be confirmed within the assumptions of the simulation.

## References

- DAHL, ÖSTEN. 2005: Tea, in: HASPELMATH *et al.*, 554-557.
- DANIEL, MICHAEL & EDITH MORAVCSIK. 2005: The associative plural, in: HASPELMATH *et al.*, 150-153.
- DRYER, MATTHEW S. 1989: Large linguistic areas and language sampling, in: *Studies in Language* 13: 257-292.
- . 2005a: Genealogical language list, in: HASPELMATH *et al.*, 584-644.
- . 2005b: Order of object and verb, in: HASPELMATH *et al.*, 338-341.
- . 2005c: Order of subject, object, and verb, in: HASPELMATH *et al.*, 330-333.
- DUNN, MICHAEL, ANGELA TERRILL, GER REESINK, ROBERT A. FOLEY, AND STEPHEN C. LEVINSON. 2005: Structural phylogenetics and the reconstruction of ancient language history, in: *Science* 309: 2072-2075.
- HERFINDAHL, ORRIS C. 1950: *Concentration in the steel industry*. Ph.D. dissertation, Columbia University.
- HIRSCHMAN, ALBERT O. 1945: *National Power and Structure of Foreign Trade*. Berkeley: University of California Press.
- HASPELMATH, MARTIN, MATTHEW S. DRYER, DAVID GIL, & BERNARD COMRIE. (eds.) 2005: *The World Atlas of Language Structures*. Oxford: Oxford University Press.
- HOLMAN, ERIC W., CHRISTIAN SCHULZE, DIETRICH STAUFFER, & SØREN WICHMANN. 2006: On the relation between structural diversity and geographical distance among languages: observations and computer simulations. Manuscript under review.
- KOPTJEVSKAJA-TAMM, MARIA. 2005: Action nominal constructions, in: HASPELMATH *et al.*, 254-257.
- MADDIESON, IAN. 2005a. Front rounded vowels, in: HASPELMATH *et al.*, 50-53.
- . 2005b. Presence of uncommon consonants, in: HASPELMATH *et al.*, 82-85.
- . 2005c. Absence of common consonants, in: HASPELMATH *et al.*, 78-81.
- MASLOVA, ELENA & VLADIMIR P. NEDJALKOV. 2005: Reciprocal constructions, in: HASPELMATH *et al.*, 430-433.
- SIEWIERSKA, ANNA. 2005: Passive constructions, in: HASPELMATH *et al.*, 434-437.
- WICHMANN, SØREN & ARPIAR SAUNDERS. 2006: How to use typological databases in historical linguistic research. Revised submission under review for *Diachronica*.

## Correspondence address

Søren Wichmann  
 Max Planck Institute for Evolutionary Anthropology  
 Deutscher Platz 6  
 D-04103 Leipzig, Germany  
 wichmann@eva.mpg.de

## Appendix: Rank-ordering by $p$ -values of the features in WALS

No.	Description	Mean $p$ -value	No. of values	No. of genera	No. of langs.	HHI
85	Order of Adposition and Noun Phrase	12.6	5	155	1074	4266
18	Absence of Common Consonants	13.4	6	107	566	7945
90	Order of Relative Clause and Noun	16.8	7	107	705	5505
88	Order of Demonstrative and Noun	17.5	6	158	1085	4070

No.	Description	Mean <i>p</i> -value	No. of values	No. of genera	No. of langs.	HHI
11	Front Rounded Vowels	18.5	4	106	561	8747
51	Position of Case Affixes	18.7	9	146	934	3560
89	Order of Numeral and Noun	19.7	4	148	1001	4521
95	Order Object/Verb and Order Adposition/NP	20.2	5	156	1033	3539
33	Coding of Nominal Plurality	20.5	9	142	957	3199
87	Order of Adjective and Noun	20.9	4	169	1213	4865
86	Order of Genitive and Noun	21.6	3	169	1105	4493
81	Order of Subject, Object and Verb	22.2	7	179	1228	3142
96	Order Object/Verb and Order Relative /Noun	22.4	5	115	756	3307
30	Number of Genders	22.5	5	40	256	3773
54	Distributive Numerals	22.6	7	41	250	2161
118	Predicative Adjectives	22.6	3	77	386	3412
7	Glottalized Consonants	22.9	8	107	566	5518
19	Presence of Uncommon Consonants	23.3	7	107	566	6400
39	Inclusive/Exclusive Distinction in Pronouns	23.4	5	41	200	4625
82	Order of Subject and Verb	23.4	3	195	1344	6459
116	Polar Questions	23.6	7	131	842	4426
94	Order of Adverbial Subordinator and Clause	23.7	5	103	611	4123
57	Position of Pronominal Possessive Affixes	23.9	4	126	795	3280
46	Indefinite Pronouns	23.9	5	50	326	4317
83	Order of Object and Verb	24.0	3	197	1370	4402
6	Uvular Consonants	24.0	4	107	566	6936
69	Position of Tense-Aspect Affixes	24.1	5	166	1062	4625
122	Relativization on Subjects	24.5	4	24	166	5941
93	Position of Interrogative in Content Questions	24.7	3	124	803	5463
84	Order of Object, Oblique, and Verb	25.1	6	81	449	3113
115	Negative Indefinite Pronouns and Negation	25.3	4	39	206	6913
97	Order Object/Verb and Order Adjective/Noun	25.4	5	169	1170	2394
111	Nonperiphrastic Causative Constructions	25.7	4	49	310	6837
135	Red and Yellow	25.8	5	22	119	6839
134	Green and Blue	26.0	7	22	119	4001
44	Gender Distinctions in Personal Pronouns	26.0	6	61	378	4922
80	Verbal Number and Suppletion	26.0	5	23	193	6894
101	Expression of Pronominal Subjects	26.8	6	102	674	4029
28	Case Syncretism	26.8	4	15	197	4378
40	Inclusive/Exclusive in Verbal Inflection	27.1	5	15	200	3067
31	Sex-based and Non-sex-based Gender Systems	27.1	3	40	256	4370
55	Numeral Classifiers	27.6	3	58	400	4846
17	Rhythm Types	27.7	5	43	323	3389
26	Prefixing vs. Suffixing in Inflection	28.0	6	137	894	2528
61	Adjectives without Nouns	28.1	7	18	124	3867
8	Lateral Consonants	28.2	5	107	566	5060
112	Negative Morphemes	28.5	6	154	1011	3458
130	Finger and Hand	29.0	2	90	593	7867
27	Reduplication	29.2	3	61	367	5988
123	Relativization on Obliques	29.6	5	16	112	3101
98	Alignment of Case Marking of Full NP	30.0	6	16	190	3549
9	The Velar Nasal	30.1	3	75	468	3832
121	Comparative Constructions	30.4	4	33	167	3160
20	Fusion of Selected Inflectional Formatives	30.4	7	10	162	5910
70	The Morphological Imperative	30.6	5	87	547	3671
79	Suppletion According to Tense and Aspect	30.6	4	23	193	4591
29	Syncretism in Verbal Person/Number Marking	30.7	3	15	197	3421

No.	Description	Mean <i>p</i> -value	No. of values	No. of genera	No. of langs.	HHI
14	Fixed Stress Locations	30.8	7	64	500	2856
92	Position of Polar Question Particles	30.9	6	122	777	3092
45	Politeness Distinctions in Pronouns	31.1	4	25	207	4941
32	Systems of Gender Assignment	31.2	3	40	256	4136
15	Weight-Sensitive Stress	31.5	8	64	500	3558
66	The Past Tense	31.5	4	39	222	3658
72	Imperative-Hortative Systems	31.7	4	61	375	4191
48	Person Marking on Adpositions	31.8	4	62	378	3854
5	Voicing and Gaps in Plosive Systems	31.9	5	107	566	3954
78	Coding of Evidentiality	32.0	6	65	418	3140
138	Tea	32.0	3	39	230	3789
99	Alignment of Case Marking of Pronouns	32.3	7	14	172	3515
136	M-T Pronouns	32.3	3	23	230	7701
21	Exponence of Selected Inflectional Formatives	32.7	5	11	160	4104
50	Asymmetrical Case-Marking	32.8	6	30	261	2422
104	Order of Person Markers on the Verb	32.8	5	62	379	3355
42	Pronominal and Adnominal Demonstratives	32.9	3	36	201	5510
137	N-M Pronouns	33.1	3	23	230	7256
41	Distance Contrasts in Demonstratives	33.5	5	40	234	4383
91	Order of Degree Word and Adjective	33.9	3	84	437	4000
13	Tone	33.9	3	98	526	4300
128	Utterance Complement Clauses	33.9	3	24	143	6573
73	The Optative	33.9	2	45	319	7443
124	'Want' Complement Subjects	34.0	5	52	283	3522
53	Ordinal Numerals	34.4	8	51	321	2107
132	Number of Basic Colour Categories	34.5	7	22	119	3029
129	Hand and Arm	34.8	2	91	617	5340
117	Predicative Possession	35.3	5	46	240	2177
119	Nominal and Locational Predication	35.3	2	77	386	5775
74	Situational Possibility	35.4	3	34	234	5315
100	Alignment of Verbal Person Marking	35.5	6	62	380	3736
109	Applicative Constructions	36.1	8	26	183	3811
71	The Prohibitive	36.2	4	78	495	2858
102	Verbal Person Marking	36.4	5	61	378	3493
77	Semantic Distinctions of Evidentiality	36.5	3	65	418	3741
16	Weight Factors in Weight-Sensitive Stress	36.6	7	64	500	3151
125	Purpose Clauses	36.6	3	26	170	4411
25	Locus of Marking: Whole-language Typology	36.9	5	17	235	3458
62	Action Nominal Constructions	37.1	8	27	168	1604
75	Epistemic Possibility	37.4	3	32	240	3396
2	Vowel Quality Inventories	37.5	3	107	563	3932
108	Antipassive Constructions	37.7	3	31	194	5989
64	Nominal and Verbal Conjunction	37.8	3	48	301	4610
37	Definite Articles	38.0	5	93	566	2685
10	Vowel Nasalization	38.1	2	25	243	6130
103	Third Person Zero of Verbal Person Marking	38.2	6	62	380	3168
36	The Associative Plural	38.3	4	41	237	2990
106	Reciprocal Constructions	38.4	4	28	175	4000
60	Genitives, Adjectives and Relative Clauses	38.9	6	21	138	3844
127	Reason Clauses	39.2	3	30	169	3933
12	Syllable Structure	39.4	3	88	485	4301
105	Ditransitive Constructions: The Verb 'Give'	39.8	4	66	378	3405
52	Comitatives and Instrumentals	39.8	3	64	566	5038

No.	Description	Mean <i>p</i> -value	No. of values	No. of genera	No. of langs.	HHI
38	Indefinite Articles	40.4	5	88	473	2629
47	Intensifiers and Reflexive Pronouns	40.5	2	24	168	5071
120	Zero Copula for Predicate Nominals	40.6	2	77	386	5043
56	Conjunctions and Universal Quantifiers	40.8	3	22	116	3372
4	Voicing in Plosives and Fricatives	40.9	4	107	566	2963
131	Numeral Bases	41.3	6	30	196	4412
35	Plurality in Independent Personal Pronouns	41.4	8	32	260	2538
49	Number of Cases	41.5	9	30	261	2038
110	Periphrastic Causative Constructions	41.5	3	16	118	4362
107	Passive Constructions	41.6	2	60	373	5086
34	Occurrence of Nominal Plurality	41.8	6	53	290	2801
63	Noun Phrase Conjunction	41.8	2	46	234	5072
43	Third Person Pronouns and Demonstratives	42.3	6	42	225	2783
3	Consonant-Vowel Ratio	42.6	5	107	563	2617
68	The Perfect	42.7	4	39	222	4035
126	'When' Clauses	43.3	3	30	174	3692
114	Subtypes of Asymmetric Standard Negation	44.3	7	38	297	2535
58	Obligatory Possessive Inflection	44.7	2	27	244	7097
65	Perfective/Imperfective Aspect	46.0	2	39	222	5041
67	The Future Tense	46.5	2	39	222	5000
59	Possessive Classification	47.3	4	26	243	4213
113	Symmetric and Asymmetric Standard Negation	47.5	3	38	297	3708
1	Consonant Inventories	47.7	5	106	562	2281
23	Locus of Marking in the Clause	47.7	5	17	235	2539
76	Situational and Epistemic Modal Marking	48.2	3	30	207	3892
24	Locus of Marking in Possessive Noun Phrases	48.7	5	17	235	3094
133	Number of Basic Colour Categories	50.4	7	22	119	1871
22	Inflectional Synthesis of the Verb	56.7	7	10	145	2328