Pro and Con

Against implicational universals

MICHAEL CYSOUW

Abstract

Ever since Greenberg's (1963) seminal paper, the implicational universal has been a standard tool for expressing typological generalisations. However, there is a major statistical problem with this tool, and also with implicational hierarchies, that has not attracted the attention that it deserves and that often leads to erroneous interpretations of data suggestive of implicational relationships. The problem is that a frequency in a sample that appears to be remarkably high or low does not necessarily mean anything. The saliency of a frequency in a typological sample depends on the deviation from the statistical expectation, not on the absolute number of occurrences. I will argue that implications should be interpreted as bidirectional statistical correlations with skewed initial parameters. Assuming directionality of interaction, as suggested by the use of arrows or other symbols expressing an asymmetry, is not warranted. However, some implications can be seen as markedness clines by reinterpreting the relative frequency of occurrence of the individual parameters.

Keywords: *implication, hierarchy, methodology, statistics*

1. The implicational universal

An implicational universal states a dependency between two logically independent parameters. In the case of two parameters A and B, both having two possible values, the salient feature for an implicational universal is that one of the four logically possible combinations does not occur. A distribution as shown in Table 1 is the ideal example of an implicational universal A → B (Croft 1990: 47–49).

Linguistic Typology 7 (2003), 89–101 1430–0532/2003/007-089

Table 1. *An implicational universal A → B*

		A	
		+	−
B	+	×	×
	−	Ø	×

However, in reality universals are based on empirically observed frequencies, and such data have to be statistically evaluated.[1] Statistically, it turns out that there is no justification for an asymmetric dependency between the parameters in an implicational universal as is implied by the direction of the arrow.

Distributions as found in a typological sample should not be taken as exact reflections of linguistic possibilities but should be interpreted as statistical deviations from chance (Comrie 1989: 19–20; Cysouw 2002: 74–79; *contra* Pericliev 2002). On this assumption, the absence (or near absence) of one particular combination of values is not sufficient to warrant a notable observation – let alone an asymmetric dependency as expressed by an implication.[2] I will illustrate this problem with some distributional patterns found for various aspects of person marking systems as discussed in my thesis (Cysouw 2001: 168–169).[3] These patterns show that something might look like an implica-

1. All statistical analyses given in this paper have a long scientific history. In fact, I do not present anything new, I only apply some basic statistical insights to the field of typology. Statistical awareness should be much more widespread in typology, which is, after all, an explicitly empirical and very often even quantitative field of investigation.
2. I believe non-occurrence in typological samples to be accidental. Combinations that are not attested, like (A+, B−) in Table 1, will be attested if more languages are investigated and/or more languages would exist. The experience of a few decades of empirical typological research shows us that non-trivial gaps in a series of theoretically possible types will eventually be filled (Dryer 1997: 124–125).
3. In my thesis, devoted to the structure of person marking paradigms, one of the various characteristics that I investigated was whether or not non-singular person marking shows any syncretism. For example, such a syncretism could be identical marking for 'we' and 'they'. I was hoping to find other features that would correlate with the presence or absence of such syncretisms. Playing around with the data, I divided the sample in two sets, one set with syncretism (55 cases) and one set without (210 cases). At one point I found a difference. In the set of 55 cases without syncretism, I found a distribution that suggested an implicational universal (the data are given in Table 2 in this paper). This implicational universal was not found in the set of 210 cases with syncretism (the data are given in Table 3). The apparent universal said that that if there is an inclusive/exclusive opposition then the paradigm is inflectionally marked. I was stunned by this universal because it did not fit in with many other results that I had already found. If anything, I had expected the opposite: an implicational universal between the presence of an inclusive/exclusive opposition and non-inflectional person marking. When I looked again at the numbers in my sample, I recognized that the apparent universal was a chimera – a recognition that lead to the present paper.

tional universal, but is actually meaningless. Conversely, something that looks like a meaningless distribution can sometimes be a significant effect that has to be explained in a theory of linguistic structure.

The first distribution, as given in Table 2, comes close to the ideal case of typological textbooks: it yields a tetrachoric table with one (almost) empty cell. This distribution might suggest an implication A $\rightarrow$ B, but this is an unwarranted inference. The cell (A+, B−) is indeed almost empty, but the value for this cell as expected by pure chance is also very low. Both A+ and B− are relatively rare. The chance that the combination (A+, B−) will occur is thus also very low. Calculating the chances, only 14 cases (= 25.5 % of the total 55) are of type B− and only 12 cases (= 21.8% of the total 55) are of type A+. This results in a chance expectation of 3.1 cases (= $0.255 \times 0.218 \times 55$) for the combination (A+, B−). The two cases that are attested are thus not indicative of an almost empty cell. Two cases are roughly just as much as would be expected from the distribution of the two parameters, already intrinsically skewed. The intersection of the two parameters does not show any supplementary dependency.

The same calculations are performed for the other cells in Table 2. In Table 3, the resulting frequencies are shown as expected by chance alone. The deviation of these expectations from the actual values is added in parentheses. The deviation is 1.1 in all cases, being positive in one diagonal, but negative in the other. For a 2×2 table, it is always the case that the deviations are the same in all cells, except for the plus/minus sign. This can easily be acknowledged by realising that the sum of the deviations has to be zero for every row and column in the table. The deviation from expectation is thus always symmetrically distributed in a 2×2 table. This means that any interaction between the two parameters in a 2×2 table is also always symmetrical. If there is only a slight deviation from chance, this might be purely accidental.

There are various statistical measures that can be used to show whether or not there is a real dependency between A and B. In the case of a 2×2 table, Fisher's Exact test can best be used.[4] This test gives a value between 0 and 1 that expresses how likely it is for the distribution to be the result of pure chance. The higher the value, the more likely it is to be the result of chance. In the distribution as shown in Figure 2, a two sided Fisher's Exact gives a

4. Fisher's Exact test is calculated by the following formula, in which a_1, a_2, a_3, and a_4 are the four cells in a 2×2 table: $\frac{(a_1+a_2)!\cdot(a_3+a_4)!\cdot(a_1+a_3)!\cdot(a_2+a_4)!}{(a_1+a_2+a_3+a_4)!\cdot a_1!\cdot a_2!\cdot a_3!\cdot a_4!}$. It is also possible to use a χ^2-test, but this test is normally not used for distributions that have cells with a frequency of less than five. However, in the case of tables that are larger than 2×2, Fisher's Exact cannot be used, so one has to resort to a χ^2-test. It is currently not clear which test can best be used when there are empty cells in larger tables. D. Jannssen (p.c.) suggest that the χ^2-test might be more robust than often assumed, also in the case of (almost) empty cells.

Table 2. *An apparent implication A → B, but actually there is no interaction*

		A		
		+	−	total
B	+	**10**	**31**	41
	−	**2**	**12**	14
	total	12	43	55

Table 3. *The statistically expected values for the data from Table 2*

		A		
		+	−	total
B	+	**8.9** (+1.1)	**32.1** (−1.1)	41
	−	**3.1** (−1.1)	**10.9** (+1.1)	14
	total	12	43	55

Table 4. *An apparently meaningless distribution, but actually there is a significant interaction*

		A		
		+	−	total
B	+	**61**	**42**	103
	−	**44**	**63**	107
	total	105	105	210

chance of 0.71, which means that the odds are seven to ten that this distribution is due to pure chance. This chance is so high that this distribution is useless for a theory of linguistic structure.

The opposite situation is attested in a distribution as shown in Table 4. No meaningful interaction between A and B appears to exist as all four cells in the table are filled with a high number of cases. However, Fisher's Exact test gives a chance of 0.013 for this distribution, meaning that the odds are about one to one hundred for such a distribution to occur. It is thus highly unlikely for this distribution to be the result of chance. In this case, the intersection of the two parameters shows a significant interaction that has to be explained. On closer inspection, it can be observed that the combinations on the top-left to bottom-right diagonal are clearly more frequent than the combinations on the other diagonal, although the totals of the rows and columns suggest that by chance all cells should have been equally common. Calculating like above, the deviation from chance is +9.5 for the cells in the top-left to bottom-right diagonal and −9.5 for the cells in the other diagonal. This difference between the two diagonals is the phenomenon to be explained.[5]

To summarise, it is not enough for one cell to be empty for there to be a distribution that is of linguistic interest (cf. Table 2). Nor is it necessary for any cell to be empty for there still to be an interaction that has to be explained (cf. Table 4). Whether there is an interaction between parameters should be checked by testing its significance as, for example, identified by Fisher's Exact (cf. Dryer 1997: 141–143). If there is a significant interaction in the data, it is always bidirectional A ↔ B and never unidirectional like an implication A → B (cf. the symmetric deviation from chance in Table 3). Finding exactly one empty cell, as for ideal implicational universals, is not the result of a interaction between A and B but of a skewing of both parameters A and B in isolation.

Thus, the implicational universal should not be used as a tool for the analysis of typological data for two reasons. First, this notion implies a unidirectional dependency between parameters that cannot be extracted from the data. Second, the emphasis on implicational universals wrongly values a (significant interaction in a) distribution with one empty cell as more important than a (significant interaction in a) distribution without any empty cells.

2. The implicational hierarchy

The implicational hierarchy is an extension of the implicational universal, and therefore the same statistical problems are encountered. I will argue that hierarchies as used in typology are significant interactions between the endpoints of the hierarchy. The apparent intermediate cline is a result of the asymmetric distribution of the individual variables.

An implicational hierarchy consists of "a 'chain' of implicational universals, so that the implicatum of the first universal is the implicans of the second, the implicatum of the second universal is the implicans of the third, and so on" (Croft 1990: 96–98). A set of chained universals is shown in (1a). Such a chain is not equivalent (in the mathematical sense) to the chain as shown in (1b). The logically accurate way to formulate an implicational hierarchy is shown in (1c). As this notation is rather cumbersome and uninformative, a hierarchy will normally be summarised by using another symbol instead of the implicational arrow, as shown in (1d). Finally, another equivalent way to depict a hierarchy is the matrix in (1e).

5. A different approach to show that a distribution without empty cells can be linguistically meaningful is the relation between OV/VO and the position of the article relative to the noun, as discussed by Dryer (1989: 273; 1992: 103–104).

(1) a. $A \rightarrow B$
$B \rightarrow C$
$C \rightarrow D$

b. $A \rightarrow (B \rightarrow (C \rightarrow D))$

c. $(A \rightarrow B) \wedge (B \rightarrow C) \wedge (C \rightarrow D)$

d. $A > B > C > D$

e.

	A	B	C	D
type 1:	+	+	+	+
type 2:	−	+	+	+
type 3:	−	−	+	+
type 4:	−	−	−	+
type 5:	−	−	−	−

The matrix in (1e) shows that five different types of languages are attested (out of $2^4 = 16$ logically possible types). The parameter settings intuitively reveal the hierarchical structure. This final presentation of a hierarchy is visually appealing and most directly insightful, and will therefore be used for the discussion below.

The statistical problems with the concept of the implicational hierarchy will again be illustrated with distributional patterns as described in my thesis (Cysouw 2001: 178).[6]

The frequencies of four parameters observed in a sample of 265 languages are summarised in Table 5. This distribution shows a classic case of an implicational hierarchy. There are five combinations of parameter settings that are clearly more common than the other logical possibilities (numbered 1 to 5 in Table 5), and these five combinations show an implicational hierarchy as explained in (1e) above. I also analysed the hierarchy as a combination of three implicational universals, as per (1a), and all three universals individually showed a significant interaction (Cysouw 2001: 176–178, $p \leq 0.001$ for each interaction). I interpreted the occurrence of three minor types (numbered 6 to 8 in Table 5) as belonging to a secondary hierarchy, strengthening the existence of a directional dependency between the four parameters. However, this

6. The positive values of the four parameters in this hierarchy represent the following characteristics:
(A) no syncretism in singular person marking
(B) no syncretism in non-singular person marking
(C) inclusive vs. exclusive person marking
(D) minimal inclusive vs. augmented inclusive
One might expect that C+ is necessarily implied by D+, but there is one counterexample to this implication in my sample (see case 9 in Table 5). This case indicates that this implication is not necessary, though highly significant.

Table 5. *An apparent implicational hierarchy A > B > C > D*

	A	B	C	D	observed frequencies
1	+	+	+	+	**26**
2	−	+	+	+	**78**
3	−	−	+	+	**99**
4	−	−	−	+	**20**
5	−	−	−	−	**21**
6	+	−	+	+	**3**
7	−	+	−	+	**12**
8	−	−	+	−	**4**
9	+	−	−	+	1
10	−	+	+	−	0
11	+	+	−	+	0
12	+	−	+	−	0
13	−	+	−	−	0
14	+	+	+	−	1
15	+	+	−	−	0
16	+	−	−	−	0
total +	31 (11.7 %)	117 (44.2 %)	211 (79.6 %)	239 (90.2 %)	
total −	234 (88.3 %)	148 (55.8 %)	54 (20.4 %)	26 (9.8 %)	
total	265	265	265	265	265

interpretation is statistically incorrect. As argued above, it is not the absolute frequency that should be interpreted, but the deviation from the frequency as expected by pure chance.

In Table 6 various measures have been computed to evaluate the frequencies from Table 5. The second column of Table 6 presents the expectation of each parameter setting on the basis of pure chance. These values have been computed on the basis of the frequencies of occurrence of the four parameters A, B, C, and D in isolation. For example, considered in isolation, parameter A is positive in 11.7 % and negative in 88.3 % of all 265 cases, as shown at the bottom of Table 5. Using the individual distributions of all four parameters, the expected value of each parameter setting can be calculated. For example, the expected frequency of the sixth combination in the table (A+, B−, C+, D+) is 12.4. This value is calculated by multiplying the chance of occurrence of the individual parameter settings with the total of 265 cases

Table 6. *Statistical expectations and deviations form expectation for the hierarchy in Table 5*

	statistically expected frequencies	deviation from expectation	standard-deviation from expectation	deviation/ standard-deviation
1	9.8	+**16.2**	3.1	+**5.2**
2	74.2	+**3.8**	7.3	+0.5
3	93.8	+**5.2**	7.8	+0.7
4	24.0	−4.0	4.7	−0.9
5	2.6	+**18.4**	1.6	+**11.5**
6	12.4	−9.4	3.4	−**2.8**
7	19.0	−6.0	4.2	−1.4
8	10.2	−6.2	3.1	−**2.0**
9	3.2	−2.2	1.5	−1.5
10	8.1	−8.1	2.8	−**2.9**
11	2.5	−2.5	1.6	−1.6
12	1.4	−1.4	1.2	−1.2
13	2.1	−2.1	1.4	−1.5
14	1.1	−0.1	1.0	−0.1
15	0.3	−0.3	0.5	−0.6
16	0.3	−0.3	0.5	−0.6

$(12.4 = 0.117 \times 0.558 \times 0.796 \times 0.902 \times 265)$. The actually attested three cases are thus clearly less than expected (contrary to the interpretation in my thesis). The deviation from the statistical expectation is shown in the third column of Table 6. These values appear to corroborate the existence of the hierarchy to some extent. The only frequencies attested that are above expectation by pure chance are part of the hierarchy (types 1, 2, 3, and 5, given in boldface in Table 6). However, one type of the hierarchy (type 4) is less common than expected by pure chance alone. Also, types 6, 7, and 8, which I analysed as a secondary hierarchy, are all clearly less common than expected by chance.

But matters will get even worse. It has to be made explicit what "clearly more common" or "clearly less common" mean. These intuitions can be evaluated by considering the standard deviation. The standard deviation is a statistical measure to evaluate the range of variation that is still within the limits of pure chance. For the current case, the standard deviation is defined as $\sqrt{Np(1-p)}$, where N is the total amount of cases (265) and p the expected chance of occurrence of a type. The exemplary type 6 has an expected chance of occurrence p

Table 7. *The apparent hierarchy from Table 5, ordered by the relative deviation from the statistical expectation*

	A	B	C	D	deviation/standard deviation	
1	+	+	+	+	+5.2	more common than expected
5	−	−	−	−	+11.5	
2	−	+	+	+	+0.5	no significant deviation from expectation
3	−	−	+	+	+0.7	
4	−	−	−	+	−0.9	
14	+	+	+	−	−0.1	
15	+	+	−	−	−0.6	
16	+	−	−	−	−0.6	
9	+	−	−	+	−1.5	less common than expected
10	−	+	+	−	−2.9	
6	+	−	+	+	−2.8	
7	−	+	−	+	−1.4	
8	−	−	+	−	−2.0	
11	+	+	−	+	−1.6	
12	+	−	+	−	−1.2	
13	−	+	−	−	−1.5	

of 0.0468 (= $12.4/265$) and thus a standard deviation of 3.4. As a general rule of thumb, the actual deviation from the expectation has to be greater than two times the standard deviation for it to be statistically significant.[7] This means for type 6 that every frequency within a range of $6.8 (= 2 \times 3.4)$ from the expected frequency 12.4 is no significant deviation, hence every actual frequency between 5.6 and 19.2 is still within the limits of chance. Type 6 was actually attested in three cases, which is thus significantly less than expected by chance. In the last column of Table 6, the ratio of the deviation by the standard deviation is shown. Only those ratios that are larger than 2 or smaller than -2 are significant deviations from expectation. Note that only the first and the last type of the hierarchy (types 1 and 5) show a significant positive deviation. The intermediate steps on the hierarchy do not show a deviation that is significantly different from the chance expectation.

7. Roughly speaking, this amounts to saying that the chances for the deviation are less that 5 %.

A better way to interpret the apparent hierarchy is shown in Table 7. The sixteen parameter settings from Table 5 are here ordered according to the relative strength of the deviation from the expected frequency. The first two types (with only pluses or only minuses) are the only ones that are clearly more common than expected. This indicates that the four parameters A, B, C, and D show an interaction, although there is no direction of interaction. The next six types remain within the limits of just one standard deviation from the expected frequency. This means that these types are attested just as often as one would expect on the basis of pure chance. The first three of these six types are precisely the intermediate cases of the implicational hierarchy, and the other three the intermediate cases of the opposite hierarchy. The fact that all these cases do not deviate from chance strengthens the observation that there is no directionality in the interaction of the four parameters. The remaining parameter settings all show a more or less significant negative deviation from chance.

Statistical tests, like that described in this section, should always be performed to interpret frequencies.[8] In some cases, hierarchies will pass the statistical test. However, experimenting with other hypothetical frequency distributions strongly suggests that the above conclusions very often hold for data that would traditionally be interpreted as an implicational hierarchy. In almost all hypothetical distributions that I have tested, an apparent hierarchy could only be interpreted statistically as a significant interaction between the endpoints of the hierarchy. The stages in between these endpoints occurred roughly as often as would be expected by pure chance.

3. From hierarchy to markedness

Among all possible statistically significant interactions, there is something special about those distributions that are traditionally interpreted as implicational universals and implicational hierarchies.

Looking back at the example given in Table 5, a frequency cline can be found at the bottom of this table. As repeated in (2a) below, the frequency of A+ in the complete sample is less than the frequency of B+, which is in turn less frequent than C+, which is in turn less frequent than D+. The existence of this hierarchy of frequency is not accidental. Every distribution that would

8. In discussing this paper with a few statisticians, many other possible statistical tests were suggested to me. D. Janssen suggested using partial c2-tests and trend tests and R. van Hout suggested using Mokken scales with Green and Loevinger coefficients. First comparisons indicated that there are no big differences between the results of these tests for the interpretation of the frequencies as discussed in this paper. However, future comparison of the many possible statistical tests has to decide which of them are best used in typology.

traditionally have been interpreted as an implicational hierarchy shows such a frequency hierarchy.[9]

(2)	a.	A+	<	B+	<	C+	<	D+
		31 cases		117 cases		211 cases		239 cases
	b.	range: 21–41		range: 101–133		range: 199–224		range: 230–248
		(st.dev. = 5.2)		(st.dev. = 8.1)		(st.dev. = 6.6)		(st.dev. = 4.8)

It is important to note the difference between the hierarchy in (2a) and the alleged implicational hierarchy from Table 5. The hierarchy in (2a) simply says that D+ is more frequent than C+, which in turn is more frequent than B+, which in turn is more frequent than A+. It does not say anything about the relative interaction between these parameters. However, if the parameters show a significant interaction (and this is a necessary condition), then the relative frequency of their occurrence can be interpreted as relative markedness. Low frequency is generally seen as an indication of high markedness, hence the hierarchy in (2a) reflects a markedness cline: A is more marked than B, which in turn is more marked than C, which in turn is more marked than D (cf. Croft 1990: 84–89). Of course, all differences in frequency on this markedness cline have to be tested for their significance. This test is shown in (2b). For each frequency, the standard deviation is calculated, and then a range of 2 times the standard deviation around each frequency is computed. These ranges should not overlap (which indeed they do not in this case) for them to be significantly different stages on the markedness cline. This is an extremely strong test, meaning that only strong markedness clines will pass this statistical test. However, my impression is that frequency distributions that would be interpreted by typologists as a hierarchy will indeed show such an extreme markedness cline.

To summarise, implicational hierarchies are a special kind of significant interactions, namely such interactions in which the frequency of the parameters in isolation shows a markedness cline. The same holds for implicational universals, as they are implicational hierarchies with only two parameters.

4. Conclusion

A skewed distribution is an observation that should be explained; it is no explanation in itself (Nichols 1992: 42). In this sense, an implicational universal in itself does not mean anything – it is an observation that has to be interpreted. In this paper, I have summarised various problems and pitfalls with such interpretations.

---

9. This is easily shown by considering a hierarchy as a strong prominence of the parameter settings as shown in (1e). If these five types are common, then B+ is more common than A+ because of the common occurrence of type 2. C+ is more common than B+ because of the ubiquity of type 3, and so on.

First, absolute frequencies have to be compared with values as expected by pure chance. This implies that an unattested type is not necessarily unusual, let alone impossible. Conversely, a commonly attested type does not necessarily ask for an explanation: only if there is a significant deviation from expected values has this deviation to be explained in a theory of linguistic structure. Second, as there are only (significant) interactions, this implies that no justification exists to deduce from data any unidirectional dependency between parameters. A significant interaction is always bidirectional. Third, the focus of typologists on finding implicational universals (with exactly one zero in any of the parameter settings) is misguided because there are other equally interesting (because also significantly skewed) distributions that cannot be expressed as implications.

For these reasons, I propose not to use the implicational universal anymore, but only tests of statistical significance. However, the traditional implicational universal and implicational hierarchy belong to a special class among all possible significant interactions. This special class is characterised by a frequency cline in the occurrence of the individual parameters. Because the parameters show a significant interaction, this frequency cline can be interpreted as a markedness cline. So, the proper way to analyse a (significant) implicational universal as in (3a) is shown by the two conjoined statements in (3b).

(3) a. If A then B
b. A and B show a significant interaction,
and A is more marked than B

Received: 28 January 2002
Revised: 9 October 2003

Zentrum für Allgemeine Sprachwissenschaft

Correspondence address: Zentrum für Allgemeine Sprachwissenschaft, Jägerstraße 10/11, D-10117 Berlin, Germany; e-mail: cysouw@zas.gwz-berlin.de

Acknowledgements: I thank Laura Downing, Volker Gast, Roland van Hout, Sabine Zerbian, and an anonymous referee for *LT* for their comments and suggestions and Frans Plank for encouragement to rewrite and finalise this paper. Dirk Janssen provided indispensable help with the statistical analyses. Of course, I alone remain responsible for its present content.

References

Comrie, Bernard (1989). *Language Universals and Linguistic Typology.* Oxford: Blackwell.

Croft, William (1990). *Typology and Universals.* (Cambridge Textbooks in Linguistics.) Cambridge: Cambridge University Press.

Cysouw, Michael (2001). The paradigmatic structure of person marking. Doctoral dissertation, Katholieke Universiteit Nijmegen.

— (2002). Interpreting typological clusters. *Linguistic Typology* 6: 69–93.

Dryer, Matthew S. (1989). Large linguistic areas and language sampling. *Studies in Language* 13: 257–292.
— (1992). The Greenbergian word order correlations. *Language* 61: 80–138.
— (1997). Why statistical universals are better than absolute universals. *Chicago Linguistic Society* 33-2: 123–145.
Greenberg, Joseph H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of Language*, 73–113. Cambridge, Mass.: MIT Press.
Nichols, Johanna (1992). *Linguistic Diversity in Space and Time.* Chicago: University of Chicago Press.
Pericliev, Vladimir (2002). Economy in formulating typological generalizations. *Linguistic Typology* 6: 49–68.

A case for implicational universals

by ELENA MASLOVA

In his paper "Against implicational universals" (2003), Michael Cysouw argues that the class of typological phenomena referred to as implicational universals cannot be established by the analysis of statistical data. On the one hand, the canonical defining feature of an implicational universal A → B – that is, exactly one empty (or nearly empty) cell [+A,−B] in the tetrachoric table that cross-classifies a representative sample **S** according to parameters A and B – is not a reliable criterion for ANY type of interaction between parameters, since it does not guarantee that the parameters involved are not independent in the first place.[1] On the other hand, if the hypothesis of independence is rejected by a reliable statistical test – that is, the number of languages in the cell [+A,−B] is not only close to zero, but also LESS THAN EXPECTED under the hypothesis of independence –, then the number of languages in the "diagonal" cell

1. In this paper, I use the canonical concept of independent events: events A and B are INDEPENDENT if the probability p(A,B) of A and B occurring together is equal to the product of probabilities p(A) and p(B). This is equivalent to saying that the conditional probability p(A|B) of A under the condition that B occurs is equal to the conditional probability p(A|−B), hence, to the unconditional probability p(A). In the present context, A and B are values of binary linguistic parameters, so such parameters are independent if and only if A and B are independent. The notion of INDEPENDENT LANGUAGES, as commonly used in the typological literature, is only remotely related to this concept; it is not invoked in the present paper. A dependency between linguistic parameters, in the sense defined above, can be established only with regard to a given population **W** of languages (which can be the total set of modern languages, languages of one geographical area, a population containing a single language from each genetic group of a given time depth, etc.). It is clear that the plausibility of possible INTERPRETATIONS of such dependencies depends on the properties of **W**; this problem is not discussed in the present paper (see, however, Footnote 2).