A case for implicational universals

A response to Michael Cysouw

Elena Maslova

In his paper "Against implicational universals", Michael Cysouw argues that the class of typological phenomena referred to as implicational universals cannot be established by analysis of statistical data. On the one hand, the canonical defining feature of implicational universal $A \rightarrow B$, 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 [–A,+B] is also less than expected; in this simple sense, any statistical dependency is bidirectional. It follows, according to Cysouw, that, statistically, "there is no justification for an asymmetric dependency between the parameters in an implicational universal". The goal of this response is to show that this is not the case, that is, a dependency between linguistic parameters can be asymmetric in an explicitly definable and statistically testable sense that can be viewed as a formalization of the original concept of a Greenbergian implicational universal.

The tetrachoric table that cross-classifies languages from a sample **S** according to two categorical parameters of linguistic variation (A and B) is intended to test linguistically interesting hypotheses about the (joint) distribution P(A,B) represented by this sample. One such hypothesis is the hypothesis of independence, which states that the probability of any combination of values of A and B is just the product of the probabilities of these values taken in isolation (see Note 1), i.e., there is no interaction between A and B. Statistical tests for independence, like Fisher Exact or the more widely known χ^2, can be used to reject this hypothesis.[2] It seems worth noting that if such a test is applied to a given sample **S** and fails to reject the hypothesis of independence, this does not entail that the parameters are independent: it can also be the case that the sample is too small to reveal the dependency.[4] In any event, if the hypothesis of independence cannot be rejected on the basis of sample **S**, then this sample can hardly reveal anything about HOW the parameters interact (even if such an interaction actually takes place). If, however, this hypothesis is rejected, then we can try to ask further questions about the detected dependency, that is, to test more specific hypotheses about the interaction between parameters A and B.

The concept of implicational universal can be constructed as a hypothesis about the type of dependency between A and B, which, since Greenberg's seminal paper (1963), is widely considered as particularly linguistically interesting. In the canonical case, a tetrachoric table reveals an implicational universal if one of its cells is empty and the numbers in the other three cells are large enough to reject the hypothesis of independence. This situation is opposed to the canonical case of "logical equivalence", where two diagonal cells are empty, so that all languages are distributed between the other two cells. This difference reveals a significant distinction between two types of interaction between parameters of linguistic variation:

1. In the case of implication, one value of A constrains the variation along the other parameter, whereas the other does not. In other words, the hypothesis of the ABSOLUTE implicational universal $A \rightarrow B$ states that the CONDITIONAL probability of B for A languages $P(+B|+A)$ is equal to 1, whereas the distribution of B among non-A languages is even, that is, the conditional probability of B for non-A languages $P(+B|–A)$ is close to 0.5.
2. In the case of (absolute) logical equivalence, both values of one parameter constrain the variation along the other parameter, so that $A \leftrightarrow B$ means that $P(+B|+A) = 1$ and $P(+B|–A) = 0$.

The challenge is to extend these notions to the case of statistical (distributional) universals, that is, to “relax” the requirement that one or two conditional probabilities must be equal to 1 in such a way as to save the linguistically significant distinction between implicational and bidirectional dependencies. In my view, a natural and useful extension of the notion of implicational universal can be obtained if the requirements that $P(+B|+A)$ is equal to 1 and $P(+B|–A)$ is close to 0.5 are replaced by the single requirement that the distribution of B among A languages is MORE STRONGLY SKEWED than its distribution among non-A languages, where the “strength” of skewing can be measured simply as deviation of $P(+B|+A)$ from 0.5. This extension retains the basic semantics of the original notion: +A constraint the value of B stronger than –A; in other words, if we know that a language is +A, it gives us more information about the value of B than the knowledge that a language is –A.[4]

Let us say that one value of A is MARKED with respect to B if it imposes stronger constraints on B than the other value; this concept of (relative) markedness can be expressed mathematically as in the following definition:

Definition 1. One value of a binary parameter A is marked with respect to another binary parameter B if one of the following conditions is met:

(i) [+A] is marked if $|P(+B|+A) - 0.5| > |P(+B|-A) - 0.5|$
(ii) [–A] is marked if $|P(+B|+A) - 0.5| < |P(+B|-A) - 0.5|$

If $|P(+B|+A) - 0.5| = |P(+B|-A) - 0.5|$, then neither value of A is marked.

Then, a dependency between A and B can be said to be SYMMETRIC with respect to A if neither value of A is marked with respect to B. This is the “null hypothesis” that has to be rejected to establish ANY implicational universal with (a value of) A in the antecedent. The null hypothesis is true in two cases:

1. $P(+B|+A) = P(+B|-A)$, or
2. $P(+B|+A) = 1 - P(+B|-A)$

Figure 1. Change of parameters: from B to B = A.

	+A	–A			+A	–A
+B	n_{11}	$\boldsymbol{n_{01}}$	➔	B = A	n_{11}	$\boldsymbol{n_{00}}$
–B	n_{01}	$\boldsymbol{n_{00}}$		B ≠ A	n_{01}	$\boldsymbol{n_{01}}$

Figure 2. A symmetrical dependency

Table 2a. Original distribution / Table 2b. Distribution for A and A = B

	+A	–A				+A	–A	Total
+B	12	60	**72**	➔	B = A	12	30	**42**
–B	40	30	**70**		B ≠ A	40	60	**100**
Total	**52**	**90**	**142**		Total	**52**	**90**	**142**
Fisher Exact: p < 0.000001					$p = 0.253$			

Figure 3. An implicational dependency

Table 3a. Original distribution / Table 3b. Distribution for A and A = B

	+A	–A				+A	–A	Total
+B	40	50	**90**	➔	B = A	40	60	**100**
–B	6	60	**66**		B ≠ A	6	50	**56**
Total	**46**	**110**	**156**		Total	**46**	**100**	**156**
Fisher Exact: p < 0.000001					$p < 0.00001$			

The first case is equivalent to the hypothesis of independence (see Note 1), i.e., this case can be ruled out by one of the standard statistical tests for independence (e.g., by Fisher Exact). The second equation defines the case of genuine symmetry: assuming that the notion of “positive” value (+) can be given the same linguistic sense for both parameters (for example, “head-final” for OV and GenN), this equation means that the probability that both parameters HAVE THE SAME VALUE does not depend on the value of A. This hypothesis can be falsified by means of the same test for independence applied to a slightly modified tetrachoric table; namely, the parameter B must be replaced with a new parameter, B = A, as shown in Figure 1. The hypothesis of symmetry is rejected if the hypothesis of independence is rejected for the new pair of parameters, A and B = A. An example of symmetrical dependency is given in Figure 2: Fisher Exact rejects the hypothesis of independence for the original tetrachoric table (Table 2a) but not for the new parameters, A and B = A (Table 2b). Figure 3 illustrates an asymmetric distribution: the hypothesis of independence is rejected for both pairs of parameters, more specifically, the positive value of A is shown to be marked with respect to B.

Now a specific hypothesis of implicational universal can be formulated as follows:

Definition 2. A joint distribution of linguistic parameters A and B counts as an IMPLICATIONAL UNIVERSAL A → B if the following conditions are simultaneously met:

(i) [+A] is marked with respect to B.
(ii) P(+B|+A) > 0.5
(iii) [+B] is not marked with respect to A.

Figure 4. A weak implicational dependency

Table 4a. Original distribution

	+A	–A	
+B	90	35	**125**
–B	10	28	**38**
Total	**100**	**63**	**163**

Fisher Exact: $p < 0.00001$

➔

Table 4b. Distribution for A and A = B

	+A	–A	Total
B = A	90	28	**118**
B ≠ A	10	35	**45**
Total	**100**	**53**	**163**

$p < 0.0000001$

Table 4c. Distribution for B and B = A

	+B	–B	
A = B	90	28	**118**
A ≠ B	35	10	**45**
Total	**125**	**38**	**163**

Fisher Exact: $p = 1$

The first condition ensures the P(B|+A) is more strongly skewed than P(B|–A); the second simply establishes the direction of skewing, that is, it distinguishes A → (+)B from A → –B. And finally, condition (iii) rules out the reverse implication B → A; note that this condition is met in the example distribution of Figure 3, since the distribution of A among B languages is less skewed than the distribution of A among non-B languages. If only conditions (i)-(ii) are met, the distribution of A and B is the statistical counterpart of EQUIVALENCE (A ↔ B).

It seems to me that Definitions 1 and 2 describe a natural extension of the concept of (absolute) implicational universal to the case of distributional universals. On the other hand, conditions (i)-(ii), hence, the hypothesis of implicational universal as a whole, can be easily tested against statistical data: if the data are sufficient to reject the “null” hypotheses of independence and harmony, this very fact guarantees that the values in our tetrachoric table are significantly different, so that the marked value of A (condition (i)) and the direction of skewing (condition (ii)) can be established in a statistically justified fashion (see example distributions of Figures 2-4). More specifically, the implicational universal A → B can be established if the number of languages in the cell [+A,–B] is much smaller than in the cell [+A,+B] and this difference (divided by the total number of A languages), is significantly greater than the difference between the two other cells (divided by the total number of non-A languages). In effect, this is just a more precise and explicit description of the criterion of “exactly one nearly empty cell”.

There is an important distinction between the concept of absolute implicational universal and its extension to the case of statistical universals, which deserves to be briefly mentioned here. An absolute implicational universal A → B is always equivalent to –B → –A; if P(+B|+A) = 1, then P(–A|–B) = 1. For a statistical implicational universal, this is not the case (which is why the requirement of unidirectionality had to be stated separately in Definition 2). More specifically, there are three statistically distinguishable types of dependencies that count as implicational universals A → B according to Definition 2:

1. STRONG unidirectional implication: A → B and –B → –A

2. WEAK unidirectional implication: A $\rightarrow$ B, but neither value of B is marked with respect to A in the sense of Definition 1.

The distribution represented in Figure 3 shows a strong implicational dependency: the conditional distribution of A for the negative value of B is obviously more skewed than for the positive value, i.e., –B is marked with respect to A. Figure 4 gives an example of a weak implication: the Fisher Exact *p*-value for the distribution of A and B reveals the implication A $\rightarrow$ B in the sense of Definition 2 (see Tables 4a and 4b). However, neither value of B is marked with respect to A: as shown by the Fisher Exact value for Table 4c, events B and A = B are independent. This means that the conditional probability P(+A|+B) is equal to P(–A|–B), i.e., both conditional distributions are equally skewed in favor of A = B. Thus, this dependency is asymmetrical with respect to A, but symmetrical with respect to B.

On the other hand, a (statistical) equivalence A $\leftrightarrow$ B is not identical to bidirectional symmetrical dependency: the former implies that both [+A] and [+B] are marked with respect to each other, and the latter, that neither value of either parameter is marked with respect to the other parameter. Thus, the framework proposed here allows for a more elaborate classification of dependencies between linguistic parameters than the original concept of implicational universal, and all these types of dependencies can be distinguished by analysis of statistical data. A further discussion of this classification is beyond the scope of this paper.

To conclude, the inadequacy of the criterion of “exactly one nearly empty cell” (without further qualifications), correctly detected by Cysouw, does not entail that the concept of implicational universal is statistically unwarranted, only that this concept must be explicated in a more mathematically precise way; once such an explication is in place, it is not a problem to develop an appropriate statistical test. It must be stressed that the definition of statistical implicational universal proposed here is not the only possible one; for example, one may want to confine the notion of implicational universal to “strong” implicational universals. My main goal in this paper has been to demonstrate that such a definition is possible in principle, i.e., the concept of implicational universal CAN be extended to the case of statistical universals and supplied with adequate statistical criteria.

Notes

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, Note 2)**.**

2. Contrary to Dryer's claim (1989, 2003), the existence of genetic and areal relations between languages does not make standard statistical tests for independence inapplicable or useless in typological studies. The conditions on applicability of such tests are formulated in terms of SAMPLING procedure, not in terms of properties of the population **W** from which the sample is drawn (roughly speaking, the fact that one language from **W** is present in the sample must not change the probability of selection of any other language). Of course, such a test does not guarantee that its output is determined by universal constraints on language variation, it just helps to examine the properties of **W** (in particular, of the population of world's languages). Further explanatory hypotheses can, in their turn, be testable by statistical methods (for example, by comparing different subsets **W** of existing languages, as in Dryer's (1989) method).

   Fisher Exact is likely to be the best choice for samples that appear to reveal an implicational universal, since χ^2 is not reliable if some cells are nearly empty. Cysouw writes, somewhat misleadingly, that the value given by this test "expresses how likely it is for the distribution to be a result of pure chance"; this interpretation is based on the unwarranted implicit assumption that the distributions of A and B taken in isolation also reflect "pure chance", which is particularly controversial if A and/or B are strongly skewed.

3. This can be the case in the example of "apparent implicational universal" discussed by Cysouw (his Figure 2), that is, Cysouw's data do not demonstrate that there is no dependency since the sample is very small. Notably, if the sample for analysis of the second distribution (Cysouw's Figure 3), where a correlation is shown to exist, had been as small is that of Figure 2, Fisher Exact test would have been unable to reject the hypothesis of independence.

4. This formulation can be given a precise mathematical sense if the entropy E(B) of B (the amount of unknown information about the value of B) is calculated according to the usual formula from the information theory:

   $$E(B) = P(+B)\boldsymbol{log_2}(1/P(+B)) + P(–B)log_2(1/P(–B))$$

   The value of E(B) ranges from 1 (if P(+B) = 0.5) to 0 (if P(+B) = 1 or P(+B) = 0). Thus, if P(B|+A) deviates from 0.5 more than P(B|–A), then E(B|+A) is less than E(B|–A).

References

Cysouw, Michael (2003). Against implicational universals. *Linguistic Typology* 7-1.

Dryer, Matthew S. (1989). Large linguistic areas and language sampling. Studies in Language 13: 257-292.

Dryer, Matthew S. (2003). Significant and non-significant implicational universals. *Linguistic Typology* 7-1.

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*, pp. 73-113. Cambridge, MA: MIT Press.