Significant and Non-significant Implicational Universals[1]

Matthew S. Dryer

University at Buffalo

1. Introduction

Cysouw (this volume) argues against the use of implicational universals in linguistic typology. While some of the points he makes are valid, his argumentation is in other places somewhat confused. I attempt in this paper to give some clarity to some of the issues, drawing on examples based on my typological database (Dryer 1989, 1991, 1992, 1997, 1998), which currently contains data for over 1200 languages.

The discussion below will support the following observations of Cysouw's (though the wording in some cases here is not consistent with other claims he makes):

- It does not follow from the fact that one of the four types in a typology defined by two two-valued parameters is rare or unattested that the pattern reflects a significant implicational universal.

- Conversely, even though all four types are well-attested, there may still be a significant correlation between two typological parameters.

---

[1] The research for this paper was supported by Social Sciences and Humanities Research Council of Canada Research Grants 410-810949, 410-830354, and 410-850540, and by National Science Foundation Research Grant BNS-9011190.

- It is necessary to use statistical tests to determine whether a universal generalization is statistically significant.

I have made similar observations in various publications (Dryer 1989, 1991, 1992, 1998), though perhaps not as clearly as Cysouw.

However, I will argue that the following claims of Cysouw's are *not* correct:

- Cysouw implies that what is important is that the pattern be statistically significant *rather than* that there be an implicational universal. I will argue that it is confused to imply that we need to choose between statistically significant generalizations and implicational universals. Rather, we must determine whether the pattern described in an hypothesized implicational universal is statistically significant or not.

- Cysouw claims that there are not unidirectional dependencies of the sort implied by unidirectional implicational universals. I will argue that this claim is simply confused.

- Cysouw uses the Fisher Exact Test as a statistical test to determine whether typological generalizations are statistically significant. But for reasons discussed in Perkins (1989, 1992) and Dryer (1989), tests like the Fisher Exact Test cannot be used to test most typological generalizations because these tests require that the events being counted be independent, a condition not met by most typological samples.

In the discussion below, I will discuss four different sorts of relationships involving implications or dependencies that can hold between two typological parameters:

(1) unidirectional implicational universals (i.e. with the form "If P, then Q") which are true in what I will call a weak sense, which obtains if languages with property P more often have property Q regardless of whether languages with property not-P also tend to have property Q;

(2) unidirectional implicational universals which are true in what I will call a strong sense, which obtains if languages with property P more often have property Q and if the tendency for languages with property P to have property Q is significantly stronger than the tendency for languages with property not-P to have property Q;

(3) bidirectional implicational universals (i.e. those of the form "If P, then Q; and if Q, then P") (these, if true, will always be true in a strong sense);

(4) statistically significant dependencies that do not involve significant implicational universals.

2. Testing hypotheses of implicational universals

On its most literal interpretation, an implicational universal of the form "If a language has property P, then it also has property Q" means that every language with property P will also have property Q. As Cysouw argues, and as I argue in Dryer (1998), and as in practice is the primary assumption in linguistic typology, implicational universals are of interest not only if they are true of all languages but also if there exist some exceptions. At the very least, as is often assumed in linguistic typology, a relatively small number of exceptions does not alter the fact that a generalization for which there are exceptions may express a significant generalization about language.

But how many exceptions are permissible? As Cysouw argues, what is crucial is that the generalization be statistically significant, and a generalization with very few exceptions may not involve a statistically significant generalization while one with many exceptions may. In order to make the discussion more concrete, I will discuss a number of implicational universals, based on data from my typological database. The first is the implicational universal in (5) (see Hawkins 1983: 74-75 for some similar universals); (1a) is the simple form of this implicational universal, while (5b) spells out more explicitly what generalizations like this are intended to mean.

(5) a. If Prep, then NRel

b. If a language has adpositions and has a dominant order of adposition and noun phrase, and if the language has externally-headed relative clauses and one order of noun and relative clause is dominant, then if the dominant order of adposition and noun phrase is prepositional, then it will employ NRel order as the dominant order for noun and relative clause.

Evidence in support of (5) is given in Table 1.

PUT TABLE 1 NEAR HERE

The form of the data given in Table 1 is discussed at greater length in Dryer (1989, 1992).[2] Briefly, the numbers represent numbers of genera containing languages of the

---

[2] I assume here six continental areas, as in Dryer (1992) and differing from the five areas in Dryer (1989). These six areas are Africa (including the Semitic languages of the Middle East), Eurasia (excluding Sino-Tibetan, Tai-Kadai, and Mon-Khmer languages of southeast Asia), Southeast Asia & Oceania (including the languages in the three families

given type in the geographical area listed, where a genus is a genetic group roughly comparable to the subfamilies of Indo-European.[3] For each pair of numbers in a given area, the larger number, denoting the more frequent type, is enclosed in a box. For example, the first two figures under Africa indicate that my database contains languages from 25 genera in Africa that are prepositional and NRel and languages from 1 genus in Africa that are prepositional and RelN. The "25" is enclosed in a box, indicating that NRel is represented by more genera in Africa containing prepositional languages than RelN is.

For reasons explained in depth in Dryer (1989), I consider a generalization to be valid if it is reflected independently in all six geographical areas. The logic behind this is that there is only one chance in 64 (i.e. 2^6) that all six areas will exhibit a given preference.[4] Since Table 1 shows that NRel outnumbers RelN among prepositional languages in all six areas, we can conclude that a prepositional language is more likely to be NRel than RelN, and that the implicational universal in (5) expresses a valid generalization.

---

just mentioned plus languages in the Andamanese and Austronesian families), Australia-New Guinea (excluding the Austronesian languages of the New Guinea area, which are counted in the previous area), North America (defined as including all languages in Greenland, Canada, the United States, Mexico and the Mayan and Uto-Aztecan languages of Guatemala and El Salvador but no other languages of Central America), and South America (including languages of Central America except those included in North America).

[3] At the time of this publication, a list of genetic groups I assume to be genera is available on the web at http://linguistics.buffalo.edu/people/faculty/dryer/dryer/genera

[4] More conservatively, there is one chance in 32 of all six areas being the same.

Despite what I have said in the last paragraph, there is a sense in which the evidence provided so far fails to provide adequate evidence for the implicational universal in (5), for reasons discussed by Cysouw in his discussion of the data in his Figure 2. The data in Table 1 does not say anything about what happens in languages which are not prepositional, which for present purposes I will treat as languages which are postpositional (ignoring languages lacking adpositions or languages lacking a dominant order of adposition and noun phrase). Suppose it were the case that postpositional languages exhibited the same overwhelming preference for NRel order that we find in prepositional languages. Then, while it might be strictly speaking true to say that if a language is prepositional then it is NRel, it would be very misleading. In other words, the implicature of an implicational statement of the form "If P, then Q" is not just that languages with property P are more likely to have property Q than property not-Q, but also that languages with property P are more likely to have property Q than languages with property not-P are.

Some of Cysouw's criticisms of implicational universals can be understood better if we are careful to distinguish these two interpretations of implicational statements. Let me introduce some terminology to distinguish these two interpretations. I will say that an implicational universal of the form "If P then Q" is true *in a weak sense* if languages with property P more often have property Q regardless of whether languages with property not-P also tend to have property Q. And I will say that an implicational universal of the form "If P then Q" is true *in a strong sense* if languages with property P more often have property Q *and* the tendency for languages with property P to have property Q is significantly stronger than the tendency for languages with property not-P to have property Q.

While the data in Table 1 is sufficient to demonstrate that (5) is true in a weak sense, we need to look at data for languages with postpositions to determine whether it is also true in a strong sense. The relevant data is given in Table 2.

PUT TABLE 2 NEAR HERE

Informally, the data in Table 2 shows that postpositional languages are quite unlike prepositional languages. While NRel order is somewhat more common by 59 genera to 41, this preference is found in only 4 of the 6 continental areas, and is thus well within the realm of accident.

But how do we explicitly test the hypothesis that the preference for NRel order among prepositional languages is stronger than the same preference in postpositional languages? Using a method employed in Dryer (1992), we can compare the proportions of genera in each of the six areas, as in Table 3.[5] The figure .96 under Africa for prepositional languages is computed by taking the number of genera containing prepositional languages which are NRel, namely 25, and dividing that by the sum of the same number and the number of genera containing prepositional languages which are RelN, namely the sum of 25 and 1, or 26, yielding .96. The proportion that is highest for each area is enclosed in a box.

---

[5] As discussed in Dryer (1992), it is possible for a single genus to be represented in more than one cell in tables like Table 1, if more than one type is represented in a single genus. To take an extreme example of this, the Semitic branch of Afro-Asiatic is represented in all four types in Table 1 above: Amharic is AdjN&DemN, Chaha is AdjN&NDem, Iraqi Arabic is NAdj&DemN, and Hebrew is NAdj&NDem. Strictly speaking, the proportions of genera cited in this paper represent proportions of *subgenera*, where a subgenus is defined as a subset of languages in a genus which are identical with respect to the typological parameters under consideration. However, for ease of exposition, I will continue to refer to proportions of genera.

PUT TABLE 3 NEAR HERE

Table 3 shows that the proportions of genera with NRel languages are higher among prepositional languages in all six areas. Since the probability of this occurring due to chance is only 1 in 64, we can conclude that prepositional languages exhibit a significantly stronger preference for NRel order than postpositional languages do. Hence, we can say that (5) is true, not only in a weak sense, but also in a strong sense.

The principle of accepting a generalization as statistically significant only if it is independently supported in all six areas is a very conservative test in the sense that it is more likely that a real linguistic preference will fail to satisfy this test than it is that a nonpreference will pass the test. There are often situations in which one area does not quite satisfy the test, where it seems likely the linguistic preference is real. Consider the data in Table 4 on the two orders of genitive and noun in verb-initial languages (defined here as languages in which both arguments in transitive clauses follow the verb in the dominant order, if there is one).

PUT TABLE 4 NEAR HERE

The pattern in Table 4 shows that NGen order is more common among verb-initial languages in five of the six continental areas, but in one area, Australia-New Guinea, GenN order is more common, by two genera to one. However, the preference for NGen order among verb-initial languages is quite strong outside Australia-New Guinea: there are only two other areas with any verb-initial GenN languages and in these two areas NGen order is much more common (in North America by 16 to 3, and in South America by 4 to 1). As I rule of thumb, I adopt the practice of tentatively accepting a pattern as reflecting a real linguistic preference if a type is more common in 5 out of the 6 areas, if the preference for

that type is quite strong in those other 5 areas, and if the greater number of genera in the one exceptional area is by a relatively small margin.

An example of a pattern that falls short of this (discussed by Dryer 1992) is shown in Table 5, on the order of numeral and noun in VO languages.

PUT TABLE 5 NEAR HERE

Table 5 shows that numeral-noun order is more common among VO languages in five of the six areas. However, in the sixth area (Africa), the opposite order, noun-numeral order, is not only more common, but much more common (31 genera to 6). Furthermore, while there are three areas in which all the VO languages in my database are NumN, there are two other areas in which NumN is only slightly more common that NNum. I therefore do not take the evidence as providing a basis for saying that there is a preference for NumN order among VO languages.

Let us return to the implicational universal in (5). There is one further point to be made about this implicational universal. Namely, it is a *unidirectional* implicational universal rather than a bidirectional one; in other words, it has the form "If P, then Q" rather than the form "If P, then Q; and if Q, then P". If (5) were bidirectional, then it would not only be the case that prepositional languages tend to be NRel, but it would also be the case that NRel languages tend to be prepositional. The data in Table 6 shows that this is not the case: there are three areas in which NRel languages are more often prepositional, two areas in which they are more often postpositional, and one area in which the number of genera are the same.

PUT TABLE 6 NEAR HERE

It is worth comparing the unidirectional implicational universal in (5) with a pattern that reflects a bidirectional universal. The data in Table 7 shows the relationship between the order of object and verb and the order of noun phrase and adposition.

PUT TABLE 7 NEAR HERE

The data in Table 7 shows that in all six areas, VO languages are more often prepositional than postpositional; it also shows that OV languages are more often postpositional than prepositional. To illustrate the bidirectional nature of the pattern in Table 7, we need to show that prepositional languages are more often VO. To do this, we can extract the first and third lines of Table 7, as in Table 8.

PUT TABLE 8 NEAR HERE

Table 8 shows that in all six areas, prepositional languages are more often VO, and hence that we have evidence for the bidirectional implicational universal in (6).

(6) If VO, then Prep; and if Prep, then VO.

Note that bidirectional implicational universals are necessarily true in a strong sense, so the difference between weak and strong interpretations does not arise: given that (6) is true in a weak sense, it follows that VO languages are prepositional significantly more often than OV languages since the second part of (6) implies that OV languages tend not to be prepositional.

3. Discussion of Cysouw's claims

With the examples above as a basis for discussion, let us turn to some of Cysouw's claims about implicational universals. First, he makes the following claim (p. 13??):

> For all these reasons, I propose not to use the implicational universal anymore, but only tests for statistical significance.

I have illustrated above how we can determine whether an implicational universal is statistically significant. Cysouw's suggestion that we use tests for statistical significance rather than implicational universals is simply confused, because implicational universals and tests for statistical significance are not the same sort of thing. An implicational universal is an hypothesis about the relationship between two typological parameters. Tests for statistical significance are tools that we use to test whether we should accept a particular hypothesis or not. The tools we use to test hypotheses and the form that such hypotheses take are different kinds of things and thus not things we need to choose between.

One of Cysouw's objections to implicational universals is that even when one cell in a two-by-two table is zero or close to zero, there may not be a statistically significant dependency between the two typological parameters. He is quite right about this. However, this issue is covered by the distinction between implicational universals that are true in a weak sense and implicational universals that are true in a strong sense. Some of Cysouw's objections to implicational universals are legitimate objections to the use of implicational universals interpreted in a weak sense, and in so far as implicational universals are sometimes interpreted in a weak sense, his criticisms are well-taken.

However, if one insists on always interpreting implicational universals in a strong sense, then at least some of Cysouw's objections to implicational universals lose their force.

The issues can be illustrated by the following example. Of 421 languages in my database that contain data on the relevant features, there is only a single exception to the generalization in (7).[6]

(7) If a language has plural prefixes on nouns, then the dominant order of subject and object will be subject before object.

The number of genera containing languages of each of the four types defined by these two parameters is given in Table 9.

PUT TABLE 9 NEAR HERE

The data in Table 9 does not, however, provide any evidence for any dependency between the order of subject and object and the position of plural affixes. The rarity of OS languages with plural prefixes simply reflects the rarity of each of these two types. Cysouw proposes applying the Fisher Exact Test to data like that in Table 9, but for reasons discussed below in section 4, this test is not appropriate here. If we apply the test that involves comparing proportions of genera, we find no difference between languages with plural suffixes and languages with plural prefixes as far as the frequency of SO as opposed to OS order is concerned: both exhibit a strong preference for SO order. As shown in Table 10, the proportion of SO languages is higher among languages with plural

---

[6] Excluded from consideration are languages that lack a dominant order of subject and object and languages that lack a dominant type among plural prefixes and plural suffixes on nouns, including languages that lack plural affixes on nouns altogether.

suffixes in some areas, while in other areas, it is higher among languages with plural prefixes. The cells with "--" in Table 10 represent areas in which there are no languages in my database with plural prefixes, so a proportion cannot be computed.

PUT TABLE 10 NEAR HERE

The mean of the proportions over the six areas is .95 for both languages with plural suffixes and languages with plural prefixes, meaning that both types of languages exhibit a 95% preference for subject before object.

What this example shows is that although the implicational universal in (7) is true in a weak sense, it is not true in a strong sense. But the issue has nothing to do with choosing between implicational universals and tests for statistical significance. The weak and strong interpretations of implicational universals still both involve implicational universals. And both can be tested by means of tests for statistical significance, as illustrated above. The implicational universal (7) is statistically significant in a weak sense while it is not statistically significant in a strong sense. Cysouw's arguments imply that implicational universals are only interesting in a strong sense, and I am inclined to agree with him about this, but it is worth bearing the weak interpretations in mind, if only to avoid confusion between the two interpretations.

Cysouw's objection to implicational universals because they can assign significance to cells with low values even when there is no dependency between the two parameters can now be more clearly understood. It is really an objection to implicational universals that are true only in a weak sense. It is not an objection to implicational universals which are true in a strong sense, since to be true in a strong sense requires that there be a dependency between the two parameters.

Cysouw also objects to implicational universals on the grounds that even when all four types defined by the interaction of two typological parameters are well-attested, there may still be a statistically significant dependency between them. Again, he is quite correct about this. An example illustrating this possibility is discussed by Dryer (1992) and involves the relationship between the order of object and verb and the order of article and noun. Relevant data is given in Table 11.

PUT TABLE 11 NEAR HERE

The first two lines of Table 11 show that while ArtN order is more common than NArt order among VO languages in five areas, one area (Africa) does not conform to this, where NArt is more common by 17 genera to 7. Hence the evidence falls short of statistical significance for a generalization "If VO, then ArtN". The last two lines of Table 11 show the opposite trend in OV languages, but again, it is not true in all six areas, being more common in only four areas.

However, if we compare proportions of genera over areas, as in Table 12, we find that the proportions of genera that are ArtN are higher among VO languages than among OV languages in all six areas.

PUT TABLE 12 NEAR HERE

We can therefore conclude that there is a statistically significant dependency between the order of object and verb and the order of article and noun, even though there is not a statistically significant implicational universal relating the two parameters. This illustrates Cysouw's point that there can be a statistically significant dependency between two typological parameters without there being an implicational universal involved.

But is this a problem for implicational universals? The examples above illustrate four types of relationships between typological parameters: ones that involve unidirectional implicational universals that are true only in a weak sense, ones that involve unidirectional implicational universals that are true in a strong sense, ones that involve bidirectional implicational universals, and ones where there is a dependency without there being a significant implicational universal. Cysouw is right in saying that if one restricts attention to implicational universals one will fail to observe statistically significant dependencies that do not involve significant implicational universals. However, it may make sense to distinguish different sorts of dependencies, and it is not clear why one should be satisfied with identifying a dependency without identifying which sort of dependency is involved.

A further argument of Cysouw's against implicational universals is that they imply "a unidirectional dependency between the parameters that can not be extracted from the data". Although he does not make completely clear what he means by this, there are a number of more specific claims he makes that seem to be covered by this general statement. First, he appears to be claiming that the Fisher Exact Test can test for whether there is a dependency between two parameters but cannot test for whether there is a unidirectional dependency between them. But that at most reflects a property of that particular statistical test. Quite apart from the fact that Cysouw's use of the Fisher Exact Test is not valid in the uses he makes of it (as discussed in section 4 below), it does not follow from his argument that there might not be some other statistical test from which one can draw an inference of a unidirectional implication, and the argument given above for the generalization in (5) based on the data in Tables 1 to 3 illustrates a statistical test for which such an inference can be drawn.

Cysouw claims that significant correlations are always bidirectional, and that where there is the appearance of a unidirectional implicational universal, this is only because one of the values is marked. Now it may well be the case that underlying every unidirectional

implicational universal are two distinct typological forces, one expressible by a bidirectional implicational universal, and one that involves a preference for one value of a parameter over another. For example, it is plausible that underlying the implicational universal in (5) is a principle whereby the order of adposition and noun phrase correlates with the order of noun and relative clause and a second principle whereby NRel order is preferred over RelN order (independently of the adposition type). This sort of idea is proposed, in fact, by Greenberg (1963), in terms of dominance principles and harmony principles. A harmony principle is one which represents a dependency between two typological parameters, where one value for one parameter occurs with one value of the second parameter significantly more often than the second value of the first parameter does. A dominance principle is one whereby for a single typological parameter, there is a linguistic preference for one value over the other value (or other values, if there are more than two). If there is a harmony principle relating adposition type to the order of noun and relative clause and if there is a dominance principle favouring NRel order, then the fact that prepositional languages are usually NRel would reflect the fact that a language which is prepositional and NRel would conform to both principles, while a language which is prepositional and RelN would conform to neither. A postpositional language, in contrast, cannot conform to both principles: if it is RelN, then it conforms to the harmony principle but not the dominance principle, while if it is NRel, it conforms to the dominance principle, but not the harmony principle. Dryer (1992) offers suggestions as to deeper explanations for the harmony principle here, while Hawkins (1990, 1994) proposes deeper explanations for both the harmony principle and the dominance of NRel.

If the implicational universal in (5) reflects two such underlying principles, does that mean that Cysouw is correct is saying that the implicational universal therefore "does not meaning anything" (p. 12)? Surely this is something of an overstatement. Even if, as I have suggested, the implicational universal in (5) reflects two such deeper principles, it hardly follows that it does not mean anything. What it means is that prepositional

languages are more likely to be NRel than RelN (and, if interpreted in the strong sense, to an extent that is greater than postpositional languages). The idea that it reflects two deeper principles is an hypothesis, and one that is much more difficult to test than the generalization expressed by (5). The crosslinguistic evidence supporting the generalization in (5) is quite strong and involves hard empirical evidence that one can apply statistical tests to, to determine whether the pattern is likely to be due to chance. This evidence and the results of the statistical test give us a fairly high degree of confidence that the empirical generalization expressed in (5) is valid. But there is not really any hard empirical evidence that this generalization reflects two deeper principles of the sort suggested. The existence of plausible deeper explanations for typological generalizations like (5) does not render such generalizations any less meaningful or useful. While I share the interest that others have in explaining crosslinguistic generalizations, there is a sense in which such generalizations are more valuable than the hypothesized explanations, since we can often have a much greater degree of confidence in the validity of the generalizations themselves than we can have in the explanations that have been hypothesized for them.

The point can be further illustrated by considering the implicational universal in (8) (see Hawkins 1983: 81 for an equivalent universal).

(8) If AdjN, then DemN

Evidence for (8) is presented in Table 13.

PUT TABLE 13 NEAR HERE

The first two lines of Table 13 show that AdjN languages exhibit a clear tendency to be DemN. AdjN&DemN order outnumbers AdjN&NDem order in five areas, and is

overwhelmingly more common in those six areas, and the two types are equally common in the sixth area (Africa). This shows that (8) is true in a weak sense. To show that (8) is true in a strong sense, we need to compare the proportions of genera over areas, as in Table 14.

PUT TABLE 14 NEAR HERE

Since the proportions of genera that are DemN is higher for AdjN languages in all six areas, we can conclude that (8) is a statistically significant implicational universal in a strong sense.

Now can we analyse (8) in terms of underlying harmony and dominance principles? There is a plausible harmony principle linking the order of adjective and noun with the order of demonstrative and noun that would prefer the two types AdjN&DemN and NAdj&NDem, since each of the two types places the two sorts of modifiers on the same side of the noun. There are also two plausible candidates for dominance principles, one favouring NAdj order over AdjN order, the other favouring DemN order over NDem order. For each of the four types defined by the order of these two modifiers of nouns, (9) shows which principle that type conforms to.

(9) AdjN&DemN: harmony, dominance of DemN
NAdj&DemN: dominance of NAdj, dominance of DemN
NAdj&NDem: harmony, dominance of NAdj
AdjN&NDem: none

We can see that each of the three types that are widely attested conform to two of the three principles, while the rare AdjN&NDem type conforms to none of these principles. These principles would thus explain the implicational universal in (8).

There are two points to be made here. First, how confident can we be of the validity of each of these three principles? We can test them each individually to see whether there is statistically significant evidence for them. Evidence for the harmony principle is given above in Table 14, since that shows that AdjN languages are DemN significantly more than NAdj languages are. To test for the dominance principles, we simply need to examine the distribution of the two orders of adjective and noun independently of the order of demonstrative and noun, and the two orders of demonstrative and noun independently of the order of adjective and noun. Table 15 gives evidence concerning the relative frequency of the two orders of adjective and noun.

PUT TABLE 15 NEAR HERE

Table 15 shows that although NAdj order is overall considerably more frequent than AdjN order, it is more frequent in only four areas and in one of the two areas AdjN order is considerably more frequent (Eurasia, by 30 genera to 7). Similarly, Table 16 shows that while DemN order is more frequent than NDem overall, this higher frequency is reflected in only three areas.

PUT TABLE 16 NEAR HERE

This does not mean that there is no preference for NAdj or DemN order, but it does mean that we have insufficient evidence that there are such preferences. The overall larger numbers could simply reflect historical factors (areal and possibly genetic) that have caused the more frequent types to be more frequent overall rather than a linguistic preference. Hence of the three principles hypothesized to underlie (8), only one achieves statistical significance. Note further that one can in principle explain the implicational universal in (8) by appealing to only two of these three principles, since for any two of

these principles, the three common types will conform to at least one of those two principles while the fourth will conform to neither of them. What this shows is that even if we believe that (8) is reducible to harmony and dominance principles, we cannot be sure which of the dominance principles is relevant.

Nor is it clear that it really is true that all implicational universals are reducible to harmony and dominance principles. How do we know that there isn't some property of AdjN languages that causes them to be DemN, while there is simply no relevant property of NAdj languages leading to some order of demonstrative and noun? How do we know that the greater number of DemN languages is not due to a principle favouring this order in AdjN languages? In other words, is there an overall preference for DemN order that contributes to the low frequency of AdjN&NDem languages? Or is there a specific preference for DemN order in AdjN languages that leads to DemN languages being more common? While on a priori grounds, I suspect that Cysouw (p. 4??) is right that implicational universals do reflect underlying harmony and dominance principles, I know of no convincing argument that this must be the case.

The general point is that the possibility of reducing implicational universals to harmony principles and dominance principles does not detract from the value of implicational universals. Implicational universals represent descriptive generalizations that are neutral with respect to competing hypotheses as to their explanation. And the distinction between unidirectional and bidirectional implicational universals captures a descriptive difference between two sorts of patterns, even if this distinction is ultimately due to the presence versus absence of dominance principles.

4. The inapplicability of the Fisher Exact Test

Cysouw proposes testing for dependencies between typological parameters by using the Fisher Exact Test. However, the Fisher Exact Test requires that the elements being counted be independent of each other, and except for very small samples, the languages in a typological sample that are being counted are not independent of each other.[7] The issues here are discussed at greater length by Perkins (1989, 1992) and Dryer (1989), so the discussion of this here will be brief.

There are three questions that might arise here. The first is: what does it mean for the languages or genera in a sample to be *independent?* The second is: how do we determine whether the languages in a sample are sufficiently independent of each other to make the use of the Fisher Exact Test appropriate? And third: why does it matter whether the languages are independent of each other?

Rather than attempt to give a general explanation of what it means for elements to be independent, I will discuss what it means in the context of language samples. What it means is that if languages from the same language family occur in the sample, they must be sufficiently distant that they are no more likely to be the same with respect to the typological parameters being examined that any random pair of languages, and similarly for languages from the same geographical area. It also depends on the particular typological parameters being examined. For example, if the parameters being examined involve features that are rarely borrowed, then it might be acceptable to have multiple languages from the same general area but different families without independence being compromised. And even if a feature can be borrowed, it might be acceptable to have multiple languages from the same continental area as long as they are from different

---

[7] The same condition applies to the Chi-Square Test.

regions within that area. However, as I argue in Dryer (1989), for some typological parameters, there is evidence that languages within large continental areas may not be sufficiently independent.

So how can we determine whether a set of languages is sufficiently independent? In order to make the discussion more concrete, I will discuss evidence for a possible relationship between the order of adjective and noun and the order of negative word and verb, given in Table 17.

PUT TABLE 17 NEAR HERE

The first two lines in Table 17 provide evidence for the implicational universal in (10) under a weak interpretation: NegV order is more common among AdjN languages in five of the six areas, and this preference is very strong in most of these areas, and in the one area in which VNeg is more common (Africa), the difference is small.

(10) If AdjN, then NegV

However, (10) is not valid in a strong sense, since it is not clear that languages which are AdjN are more likely to be NegV than languages which are NAdj: the last two lines of Table 17 show that NAdj languages show a similar preference for NegV order, and in fact that preference is found in all six areas. When we compare proportions of genera over the six areas, as in Table 18, we find that although the preference for NegV order is stronger among AdjN languages in four areas, it is weaker in the other two.

PUT TABLE 18 NEAR HERE

Hence, we have no reason to believe that (10) is valid in a strong sense or that there is any dependency between the order of adjective and noun and the order of negative and verb.

Suppose we attempt to apply the Fisher Exact Test to the total number of genera of each of the four types in Table 17, as given in Table 19.

PUT TABLE 19 NEAR HERE

The Fisher Exact Test yields a result of .0025, implying that there are only two and a half chances in a thousand of obtaining results like those in Table 19 if there is no dependency between the order of adjective and noun and the order of negative word and verb.[8] However, as noted above, that probability assumes that the 200 genera given in Table 19 are independent of each other.[9] Saying that they are independent means that two languages in different genera but in the same family are no more likely to have the same orders with respect to these two parameters that any random pair of languages, and that two languages spoken in the same geographical area are no more likely to have the same orders than any random pair of languages. However, it is easy to show that they are not independent. In fact, we can use the Fisher Exact Test to show that languages within the same continental area are more likely to be the same than random pairs of languages. We

---

[8] The figures cited here for the Fisher Exact Test are the probabilities of obtaining numbers that deviate from what one expects due to chance *or which deviate to a greater extent*.

[9] Actually, following footnote 4, there are 200 subgenera represented in Table 19. However, since languages within different subgenera are necessarily different with respect to the typological parameters in question, the question of whether subgenera within the same genus are independent does not arise, although they decrease the likelihood of the figures deviating from the norm due to chance.

can do this if we focus on individual cells in Table 17 and reduce Table 17 to a two-by-two matrix, as in Table 20, which focuses on the cell for AdjN&NegV languages in Africa, where the value is 4.

PUT TABLE 20 NEAR HERE

This cell is lower than what might be expected given the other three cells: outside of Africa this type represents almost 40% of the genera (60 out of 153), while in Africa, this type represents less than 10% of the genera (4 out of 47). The Fisher Exact Test can be used to determine the likelihood of this happening due to chance if languages within an area are no more likely to be the same type than languages in different areas. The result of the Fisher Exact Test for Table 20 is .000025. This means that there are only 25 chances in a million of Africa being this different from the rest of the world if pairs of languages within Africa are no more likely to be of the same type than pairs of languages in the entire world. This provides us with good reason to conclude that the genera providing the basis for Table 20 are not independent and hence that the condition for the Fisher Exact Test is not met.

We can do a similar operation to test the likelihood of having as many as 24 genera in North America containing AdjN&NegV languages, as in Table 21.

PUT TABLE 21 NEAR HERE

The result for the Fisher Exact Test for Table 21 is .000001, meaning that there is only one chance in a million of such results obtaining if languages within North America are no more likely to be of the same type as languages in different continental areas. In other words, we can use the Fisher Exact Test to show that there is a dependency between area and type, thus showing that we cannot use the Fisher Exact Test to determine whether

there is a dependency between the two typological parameters defining the types in Table 17.

Why does it matter that the genera in Table 17 are not independent? Why do they have to be independent in order to use the Fisher Exact Test? It is not possible to explain this fully here. The reasons are obvious if one understands the probability theory underlying the test. However, I can attempt to give an intuitive idea. The relative frequency of AdjN&NegV languages varies considerably from area to area in Table 17. It is the least frequent of the four types in Africa, but the most frequent of the four types in North America. What this means is that if a few areas coincidentally have considerably more of one type, simply due to historical factors in those areas, then the relative frequency of that type will be much higher than it might have been if historical events in these areas had not coincidentally led to this type being more common. In other words, the probability of getting figures like those in Table 17 due to chance is actually much higher than the probabilities obtained by applying the Fisher Exact Test. Hence the Fisher Exact Test does not give an accurate probability of the likelihood of obtaining the results one obtains due to accident.

A further way of making the same point in an intuitive way is to see the results of leaving out languages in one part of the world. The idea behind this is that if one finds an apparent preference for one type over another, this apparent preference should not disappear if one excludes languages from one part of the world.[10] If it does disappear, then the apparent preference can be explained in terms of historical accident in that one area. Suppose we remove the data for languages from North America from Table 17. The result is given in Table 22.

---

[10] This idea was originally suggested to me by Tom Givón.

PUT TABLE 22 NEAR HERE

If we apply the Fisher Exact Test to the figures in Table 22, the result is .064, which means that there are 64 chances in a 1000 of obtaining results like those in Table 22 (or which deviate from the norm to a greater extent) due to chance, which does not fall within the conventional figure of .05 for statistical significance. In other words, if we include North America, we get a figure of .0025, which is highly statistically significant, while if we exclude North America we obtain a figure of .064, which is 24 times greater and which falls short of statistical significance.

Now it is not surprising that the probability goes up when we exclude languages from the sample, since we are making the sample smaller. But we do not expect the amount of the increase to go from a figure that is highly significant statistically to a figure that falls short of statistical significance. The size of the increase reflects the fact that languages within an area can be similar to each other for historical reasons and reflects the reasons why the Fisher Exact Test cannot be used for data like this. In other words, the figure of .0025 which we got by applying the Fisher Exact Test is not an accurate estimate of the probability of achieving the results obtained due to accident, and the figure suggests that the probability is much less than it actually is.

The general conclusion is that while Cysouw is quite correct in claiming that it is important to test whether typological generalizations are statistically significant, it is important that one apply tests that are appropriate to the data being examined. Applying statistical tests to data which fail to meet the conditions for the test is dangerous because one can create the illusion that the statistical tests provide good reason to conclude that a typological generalization is valid, when in fact that pattern in the data may simply reflect the effects of areal phenomena in a few areas of the world. It is much better not to apply

tests for statistical significance than to apply them to sets of data which do not meet the conditions for the test.

It is precisely because one cannot apply standard statistical tests like the Fisher Exact Test and the Chi-Square Test to typological data that I proposed the method of counting genera within continental-sized areas in Dryer (1989). In some respects, however, it is a crude test, and is subject to legitimate questions about how confident we can be about the decisions as to which genetic groups we should count as genera. There are probably ways to test typological hypotheses that are more powerful than the method I have illustrated here. I personally prefer to use tests where the probability theory underlying the test is clear to me and can be explained to other linguists. But I hope that others will work on this problem to come up with more powerful methods.

References

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

Dryer, Matthew S. (1991). SVO languages and the OV:VO typology. *Journal of Linguistics* 27: 443-482.

Dryer, Matthew S. (1992). The Greenbergian word order correlations. *Language* 68: 81-138.

Dryer, Matthew S. (1997). On the six-way word order typology. *Studies in Language* 21: 69 - 103.

Dryer, Matthew S. (1998). Why statistical universals are better than absolute universals. In *Chicago Linguistic Society 33: The Panels*, 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.

Hawkins, John A. (1983). *Word Order Universals*. New York: Academic Press.

Hawkins, John A. (1990). A parsing theory of word order universals. *Linguistic Inquiry* 21: 223-261.

Hawkins, John A. (1994). *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.

Perkins, Revere D. (1989). Statistical techniques for determining language sample size. *Studies in Language* 13: 293-315.

Perkins, Revere D. (1992). *Deixis, Grammar, and Culture*. Amsterdam: John Benjamins.

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
Prep&NRel	25	9	14	9	12	6	75
Prep&RelN	1	0	1	0	0	0	2

Table 1

Order of noun and relative clause in languages with prepositions

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
Postp&NRel	19	5	4	9	14	8	59
Postp&RelN	4	17	8	7	2	3	41

Table 2

Order of noun and relative clause in postpositional languages

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Mean
Prep	.96	1.00	.93	1.00	1.00	1.00	.98
Postp	.83	.23	.33	.56	.88	.73	.59

Table 3

Proportions of genera which are NRel

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
V-1&NGen	7	1	8	1	16	4	37
V-1&GenN	0	0	0	2	3	1	6

Table 4

Order of noun and genitive in verb-initial languages

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
VO&NumN	6	9	14	7	22	8	66
VO&NNum	31	0	12	6	0	0	49

Table 5

The order of numeral and noun in VO languages

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
NRel&Prep	25	9	14	9	12	6	75
NRel&Postp	19	5	4	9	14	8	59

Table 6

Order of adposition and noun phrase in NRel languages

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
VO&Prep	24	8	20	13	22	7	94
VO&Postp	6	2	0	1	3	5	17
OV&Prep	3	2	0	2	0	0	7
OV&Postp	20	26	12	45	26	28	157

Table 7

Order of verb and object and order of adposition and noun phrase

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
Prep&VO	24	8	20	13	22	7	94
Prep&OV	3	2	0	2	0	0	7

Table 8

Order of verb and object in prepositional languages

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
PluralPref&SO	12	0	4	1	8	0	25
PluralPref&OS	0	0	1	0	0	0	1
PluralSuff&SO	33	32	10	22	35	18	150
PluralSuff&OS	2	0	0	1	3	3	9

Table 9

Position of plural affixes and order of subject and object

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Mean
PluralPref	1.00	--	.80	1.00	1.00	--	.95
PluralSuff	.94	1.00	1.00	.96	.92	.86	.95

Table 10

Proportion of genera containing SO languages

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
VO&ArtN	7	6	12	5	20	5	55
VO&NArt	17	0	9	2	4	0	32
OV&ArtN	2	7	2	9	4	4	28
OV&NArt	10	3	6	17	7	1	44

Table 11

Order of object and verb and order of article and noun

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Mean
VO	.29	1.00	.57	.71	.83	1.00	.73
OV	.17	.70	.25	.35	.36	.80	.44

Table 12

Proportions of genera that are ArtN among VO and OV languages

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
AdjN&DemN	7	29	7	15	24	9	91
AdjN&NDem	7	0	1	0	3	1	12
NAdj&DemN	11	4	15	18	17	18	83
NAdj&NDem	47	2	19	31	9	6	114

Table 13

Order of adjective and noun and order of demonstrative and noun

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Mean
AdjN	.50	1.00	.88	1.00	.89	.90	.86
NAdj	.19	.67	.44	.37	.65	.75	.51

Table 14

Proportions of genera that are DemN

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
AdjN	14	30	8	17	31	12	112
NAdj	52	7	28	52	28	26	193

Table 15

Order of adjective and noun

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
DemN	17	32	18	35	51	30	183
NDem	50	2	20	35	15	8	130

Table 16

Order of demonstrative and noun

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total
AdjN&NegV	4	17	6	9	24	4	64
AdjN&VNeg	6	4	0	0	0	2	12
NAdj&NegV	21	4	19	20	9	8	81
NAdj&VNeg	16	0	7	13	2	5	43

Table 17

Order of adjective and noun and order of negative word and verb

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Mean
AdjN	.40	.81	1.00	1.00	1.00	.67	.81
NAdj	.57	1.00	.73	.61	.82	.62	.72

Table 18

Proportions of genera which are NegV

	NegV	VNeg
AdjN	64	12
NAdj	81	43

Table 19

Total number of genera

	Africa	Other Areas
AdjN&NegV	4	60
Other 3 types	43	93
Total	47	153

Table 20

Interaction of area and type

	NAmer	Other Areas
AdjN&NegV	24	40
Other 3 types	11	125

Table 21

Interaction of area and type

	NegV	VNeg
AdjN	40	12
NAdj	72	41

Table 22

Total number of genera, excluding North America