

LINFER: inferring implications from the WALS database

Dik Bakker
University of Amsterdam

To appear in: Stuff (special issue edited by Michael Cysouw)

Abstract

In linguistic typology, implications typically emerge from work on a specific subdomain as a result of in depth study of a few related linguistic parameters. The Konstanz Universals Archive has been built up mainly from such contributions over the last decade. Another way is inferring implications from a typological database with a large number of variables. In this contribution I discuss the LINFER computer program which was developed precisely for this goal. I applied the program to the database containing the around 140 variables on which the WALS Atlas was based, arguably one of the larger typological databases available. I show that from such a database, around one in every 2000 value pairs lead to an acceptable and potentially interesting implication. A simulation shows that this is 4 times as many as a purely random dataset of the same size would produce.

1. Introduction

In Bakker (1994) a package of computer programs is introduced called Computer Programs for Language Typology (CPLT). Apart from applications for the representation, storage and retrieval of typological data, the package contains a program for the construction and evaluation of language samples and one for the automatic inference of linguistic implications, LINFER. This contribution deals with the application of the latter to the database underlying the WALS Atlas (Haspelmath et al (eds) 2005). In section 2 I will briefly discuss the formal aspects of the database containing the WALS data. Section 3 explains the technique used by the program for finding the implications lingering in the database and for assessing the relevance and potential linguistic interest of those found. Section 4 discusses some problems with the WALS database. In section 5 we will see what kind of implications were actually found by the program among the WALS variables. We will have a closer look at a few of them. However, I will mainly concentrate on the more formal aspects of the inference process. An in depth linguistic interpretation of the resulting implications falls outside the scope of this short contribution. Sections 6 and 7 show that a balanced sub sample from the large WALS database fares at least as good as the complete database in terms of the derivation of reliable and linguistically interesting implications. Finally, in section 8 I will use a manipulated version of the same WALS database in order to get an estimate of how many implications we might expect from such data on the basis of pure chance, when no underlying relations between variables may be assumed. Section 9 will present some conclusions.

The aim of this contribution lies not so much in presenting some new, so far unknown universals of language. Rather, it seeks to discuss some empirical observations that may further serve the discussion, raised in Linguistic Typology 7-1, about the pros and cons of implicational universals (Cysouw (2003); Dryer (2003); Maslova (2003); and Plank (2003)).

2. The data

The programs of the CPLT system expect their input data in the so-called comma delimited format (.CSV). The underlying shape of these files is the well-known rectangular data matrix, with the linguistic variables on the columns and the languages on the rows. Via an interface module the WALS data were reshaped into this format. A fragment is given in table 1. Column 1 in this matrix contains the language name, column 2 the 3-letter abbreviation for the language as defined by the WALS project, and based on the Ethnologue code. From column 3 onwards one finds the actual WALS variables, labelled with a 6-character code. The first three letters of this code are derived from the name of the author of the corresponding WALS chapter; the last three relate to the variable. E.g. in column three we find the values for Siewierska's map on the Passive. In the cells, values 1 and higher represent the respective symbols on the map. E.g. BAKADP = 4 means that the language, in this case Abkhaz has person marking on adpositions for both pronouns and nouns. An empty cell codes the fact that for the language concerned there is no value for the corresponding variable.

LANGUAGE	CODE	SIEPAS	DAHTEA	CYSIND	BAKADP	AUWEPI
Abaza	ABZ		1			
Abenaki (Western)	ABW					
Abidji	ABD					
Abipon	ABI	2		3	2	
Abkhaz	ABK	2		5	4	1
Abun	ABU	2			2	
Acehnese	ACE	2	2		2	1
Acholi	ACL					
Achuar	ACU					
Achumawi	ACM	2			1	
Acoma	ACO	2		1	2	2
...						

Table 1. Fraction of WALS data matrix

The complete WALS dataset contains 139 variables for a total of 2558 languages.¹ For each of the variables, the number of different values ranges between 2 and 9, with a mean of 4.7 values per variable. However, this 139 by 2558 matrix is rather empty in the sense that just over 16% of the cells have a value. This means that per variable, there is a score for 417 languages on average, with a minimum of 116 and a maximum of 1343 languages. And per language in the database, there is a value for an average of 23 variables. The sparseness of the matrix has consequences for the inference process. I will come back to this later.

3. The program and the method

The LINFER program operates according to the generate-and-test principle. It systematically produces all logically possible combinations of two ordered variable-value pairs. An example is the pair in (1) below, to be interpreted as 'IF a language has person marking on adpositions for both pronouns and nouns THEN it has no passive'.

$$(1) \quad \text{BAKADP} = 4 > \text{SIEPAS} = 2$$

During the inference process it is first assessed for how many languages the pair is found in the database. If there are any languages at all, then the number of counterexamples is

established, i.e. those with BAKADP = 4 but SIEPAS ≠ 2. If one is only interested in true universals, then any counterexample means rejection of the implication. If one also accepts statistical universals, the user will have to supply a lower limit for the percentage of counterexamples (the applicability rate, or APL).

Furthermore, two tests are built in which implement the notion ‘(much) better than chance’ already found in most of the Greenberg (1963) universals. In the case of binary variables, with values of the yes-no type, Fisher’s Exact test (FE) is applied, which considers the pure chances on the co-occurrence of the respective values. Traditionally, rejection is established on the basis of a value anywhere between 0.5% and 5%. For variables with more than 2 values, the Chi Square test (χ^2) is employed, which gives an approximation of the chances. The user of the LINFER program may opt for any of the pre-established probabilities 0.5, 1.0, 2.5 and 5.0%. In case the value matrix contains too many low frequencies for χ^2 to be reliable the program automatically applies the Fisher exact test also for variables with more than two values. In such cases, all values other than the two under consideration are subsumed under ‘not X’ for both variables.

Several other criteria are built in to further restrict the acceptance of implications. One is relevance (REL), which is the relative proportion of the values of the premise for the corresponding variable. With this criterion one may filter out implications that are valid for very small subsets only, e.g. one or two languages, and for which the usual statistical tests do not apply. Another criterion is the coverage (COV), which is the relative number of languages that have the value of the conclusion but not of the premise (in terms of example (1): SIEPAS = 2 but BAKADP ≠ 4). Requiring a relatively high coverage, e.g. > 0.75 restricts the effect of highly skewed value distributions, which combine with virtually all values of other variables. Finally, a very strict requirement one may opt for is that for any implication $p \rightarrow q$ to be acceptable, its negation (NEG), $\neg p \rightarrow \neg q$ should also hold.²

These are the most important instruments for the control of what will be considered an acceptable implication. After the selection process, the remaining implications are written to a file for further inspection, both linguistic and statistical. In the output, they are provided with explanatory labels for both variables and values. They are sorted according to descending strength in terms of the constraining factors. Optionally, one may display the languages that fulfil the requirements, the counterexamples and genetic and areal information about the languages involved. A (slightly simplified) example of such output is found in (2) below.

(2) 9. DRYCOQ=2 => NICOBL=2 n=89

Position of Interrogative Phrases in Content Questions: Not obligatorily initial

=>

Obligatory Possessive Inflection: No obligatorily possessed nouns

[REL=0.64, APL=0.93, COV=0.54, NEG=0.24, $\chi^2 < 0.5\%$] STAT

LGS: ABK ALW AME AMH AEG ARP ARM AMP ...

EXC: ASM BUR CAX HAT LIM PMS WAP

This implication was accepted as valid for 89 languages with 7 counterexamples – making it a statistical universal rather than an absolute one - under the (default) criteria REL = 0.2, APL = 0.9, COV = 0.5 and NEG = 0.0. Note that this implication is accepted as such since the values in (2) are all higher than these default values. Furthermore, its χ^2 value is better than 0.5%.

4. Preparing the WALS database

Before we have a look at some of the implications that appeared in the WALS database, two points will have to be made here. The first point, already announced above is the missing values. For the treatment of these, several methods have been developed in the domain of statistics. However, they presuppose a more or less coherent database. The WALS database, however, is not very coherent. The data stem from around 40 different authors, each having designed their own sample of between 116 and 1343 languages, not necessarily the same set even for their own variables. In that sense the empty cells cannot be seen as real missing values at all. For practical purposes, I decided to simply ignore empty cells, and take for the relevance set of a potential implication only the languages which have a (non-missing) value for both variables involved.

A second point, related to the first is the nature of the sample constituted by the WALS database. Obviously, it is not a sample in the sense that it has been drawn from one population under the application of a fixed principle. Rather, we can see it as a collection of samples, roughly one per author, each with its own internal logic. Nevertheless, given the availability of reliable descriptive linguistic data, we may expect a bias in the direction of the better known languages. I first assessed the sample on the basis of the Diversity Value (DV) sampling technique discussed in Rijkhoff & Bakker (1998), using the 13th edition of the Ethnologue for a genetic classification (Grimes (ed.) 1996).³ Of the 2558 languages in the WALS, 381 have codes that were not found in the Ethnologue. Ignoring these for the purpose of this contribution, I found the following for the remaining 2177 languages. At the level of the major phyla, 262 languages were dislocated. This means that the sample could be made more felicitous in terms of the Diversity Value technique by replacing around 12% of the languages in phyla that are overrepresented by languages from phyla which are currently underrepresented. The most important deviations are found in table 2.

Overrepresented	n of lgs	should be	Underrepresented	n of lgs	should be
Niger Congo	320	266	Tr. New Guinea	137	183
Sino Tibetan	105	80	Austro-Asiatic	41	73
Oto-Manguean	49	25	Nilo-Saharan	77	95
Arawakan	40	19	Macro-Ge	15	32
Indo-European	122	103	Tupi	19	31

Table 2. Overrepresentation an underrepresentation in WALS

Interestingly, for Austronesian (the largest phylum considered) the number of languages is exactly right at 289, although the allocation to the many subgroups is not always ideal. Especially the smaller groups, often consisting of only 1 or 2 languages, are underrepresented while the Central Philippines group, with 22 representants out of its total of 46 languages, should have 2 languages only.

The program that implements the DV technique has an option to determine the largest sub sample that does fulfil the criteria. Applying this option to the 2177 subset resulted in a 359-language sample, which is genetically as close to the ideal DV sample as possible. Unlike the 2177 one this sample may be considered representative for the world's languages, at least genetically.⁴ This sub sample was optimised in the sense that, whenever there was a choice between some of the original languages to be kept in the sample, the ones were selected with the highest number of variable values. The resulting sample, referred to as W359, will be used as a control sample for the exercise of the next section, which will take the complete database W2558 as a point of departure.

5. Some results from the WALS database

As discussed above in section 3, the LINFER program has a number of built in criteria that create a threshold for accepting potential implications. A first run was done with the applicability rate APL fixed at 1.0, i.e. only absolute implications were admitted. χ^2 and FE were established at 5.0%. Out of the 413.886 implications that are logically possible in W2558 1385 passed this test (I will use the notion of reduction factor for the proportion of accepted implications out of the overall potential; the value for the reduction factor is here $1385/413886 = 0.0033$). Looking more closely at the remaining implications, 132 turn out to be equivalences (both APL and NEG have value 1.0). In many of these cases, both variables stem from the same author. Rather than pointing out some new discoveries in the domain of language universals, they suggest a high degree of consistency in the coding. (3) below is an example of such a (trivial) equivalence.

- (3) 1. CORSEX=1 <=> CORNUM=1 n=144
[REL=0.56, APL=1.00, COV=1.00, NEG=1.00, chi2 < 0.5%] EQUIV
Sex-based and Non-sex-based Gender System: No gender system
<=>
Number of Genders: None

The strongest implication with variables from different authors is found at place 15 in the list. It is shown here in (4). Note that its frequency is extremely low – 2 occurrences – which renders the χ^2 test unreliable and forces a Fisher test.

- (4) 15. KAYBCC=4 <=> VESTAM=2 n=2
[REL=0.13, APL=1.00, COV=1.00, NEG=1.00, chi2 < 0.5%***, fe=0.00] EQUIV
Basic Colour Categories: Between 7 and 8 categories
<=>
Suppletion According to Tense and Aspect: According to aspect

Over 90% of these 1385 implications stem either from one author (and are then often tautological, as the one in (3)) or have extremely low frequencies (as the one in (4)). An implication that includes variables from two authors and a reasonable number of languages is found in (5).

- (5) 392. HASNPL=4 => SIEGEN=6 n=21
[REL=0.07, APL=1.00, COV=0.25, NEG=0.43, chi2 < 0.5%] ABS
Occurrence of Nominal Plurality: Plural in all nouns, always optional
=>
Gender Distinctions in Indep Pro: No gender distinctions

Although this implication looks quite robust on formal grounds, it is not directly obvious what deeper linguistic factor binds these two values.

One way to repair the overkill in very low frequencies is to put the coverage high, e.g. at 0.5. This would mean that at least 50% of the languages with the conclusion should be in the implication. This results in a dramatic reduction in implications: we go down from 1385 to 47, of which 14 are equivalences. The latter are all tautologies. Furthermore, there are only a total of 15 implications that involve variables from different authors. Of these, 11 have the dominant value ‘the language has more than 10 basic colour categories’ as one of its factors. This does not seem to have a transparent relationship with any of the other factors. Of the remaining 4 implications, the one in (6) below suggests an absolute relationship between the

obligatoriness of plural marking on the noun and the availability of a separate word for ‘finger’.

- (6) 42. HASNPL=6 => BROFIN=2 n=49
 [REL=0.49, APL=1.00, COV=0.53, NEG=0.14, chi2 < 1.0%] ABS
 Occurrence of Nominal Plurality: Plural in all nouns, always obligatory
 =>
 Finger and Hand: Differentiation: one word denotes 'hand'

Whether or not there are any interesting implications hidden in the database, it seems to be clear that we need another way of looking at the inference results in order to find them. Since many of the original Greenbergian universals were not exceptionless, fixing the applicability criterion at a lower value may open the door to some interesting implications, be it that they will have counterexamples. This is indeed what happens when we change APL to 0.9, instructing the program to accept all implications with an applicability rate of 0.9 or higher, i.e. with a maximum of 10% counterexamples. The output increases to a total of 202 implications, the original 47 plus 155 statistical ones. Indeed, some of these are old friends. (7) is related to Greenberg’s universals #3 and #24.

- (7) 57. DRYRPO=4 => DRYREL=1 n=291
 [REL=0.49, APL=0.98, COV=0.69, NEG=0.52, chi2 < 0.5%] STAT
 Relationship between the Order of Object: Verb-object and prepositional (VO&Prep)
 =>
 Order of Relative Clause and Noun:Relative clause follows noun (NRel)
 EXC: CNT HAK MND SQU TUK

We may also welcome some new friends. The implication in (8) is a case in point. Given the historical relationship between pronouns and agreement markers, this is definitely an implication with a linguistic explanation. The counterexamples may have retained the clusivity distinctions while they have gone lost in the case of the verbal agreement markers.

- (8) 99. CYSVRB=3 => CYSIND=3 n=75
 [REL=0.4, APL=0.95, COV=0.63, NEG=0.63, chi2 < 0.5%] STAT
 Inclusive/Exclusive Distinction in Verba: No inclusive/exclusive opposition
 =>
 Inclusive/Exclusive Distinction in Indep Pro: No inclusive/exclusive opposition
 EXC: ABK CLE MAP MRD

6. Inferences from the subsample

At this point, it may be interesting to see how the genetically balanced sub sample W359 fares. I ran the program with the same constraints as the previous run, which are in fact the default minimum values for the four factors that the program applies. This gave the result in table 3. The middle column presents the figures for the database without selection on the languages. The righthand column presents the figures for the 359 language sample.

	FULL DATABASE	SAMPLE W359
Number of implications	202	190
Equivalence	14	9
Absolute	33	38
Statistical	155	143

Table 3. Comparison between overall sample and balanced sub sample

Note that only slightly fewer implications (94%) are found for the genetically balanced W359 sample than for the original dataset of 2558 languages under these parameters. The distribution over the three types of implications – equivalence, absolute and statistical - is also roughly the same. Although the number of languages has been dramatically reduced to around 14% of the total, the mean number to which the implications apply is much higher at 65.9, or 18.4% of the 359 languages (was only 6.4 %). The relative number of counterexamples is about the same for both sets: 8.7 on 162.6 (5.3%) and 3.7 on 65.9 (5.6%). This is undoubtedly a reflection of the fact that from the respective families and groups, the languages with the highest number of variables were selected.

Looking at the mean values for the four constraining factors, we get the results given in table 4 below.

	DEFAULT (MINIMUM)	FULL DATABASE (MEAN)	SAMPLE W359 (MEAN)
Relevance	0.2	0.50	0.50
Applicability	0.9	0.95	0.95
Coverage	0.5	0.70	0.71
Negation	0.0	0.48	0.80

Table 4. Comparison between the constraining factors

The mean values for relevance, applicability and coverage are all virtually the same in both cases. However, the value for NEG is almost doubly as high for W359 as for the total set. This means that in the balanced sub sample it is much more often the case that an implication has the ‘ideal’ characteristic of dividing the languages in [p & q] and [-p & -q] ones. Although the eventual number of implications resulting from an analysis is not directly affected in this way, one could say that the implications which come out of a balanced database tend to be much more ‘telling’ in the sense that they also say something about the behaviour of the variable value in the conclusion.

Comparing the two sets of implications, we find that the overlap in terms of implications is 136. In other words, 67% of the implications found in the complete database are also found in the W359 sample. Obviously, only a more in depth comparison of the two sets may tell us whether the sub sample gives us a more reliable impression of the interaction between the WALS variables than the set at large. That exercise falls outside the scope of this short contribution. What seems to be the case, though, is that the genetically balanced sample has at least the same potential in terms of providing us with implications as the much larger unbalanced data set. Overall, the ones we find tend to be stronger than those from the larger set in terms of their negation and proportion of the languages involved.

7. Grouping the variables

Whatever our conclusion would be, the majority of the implications found in both sets W2558 and W359 are still between variables of the same author. They are often lying within the same linguistic sub domain and often are of a tautological character. A further option that has been built into the program with which this phenomenon may be eliminated is to gather the respective variables in groups. Optionally then, no interactions between the variables within a group will be taken into consideration. I selected 15 authors who contributed several variables each, typically around 6. In all, this left us with 89 out of 139 variables (64%). The

applicability value was lowered to 0.8, allowing for more counterexamples. For this fourth run the number of resulting implications over the total database was considerably lower than for the third run: 67 instead of 202, be it out of a potential of only 143.956 implications rather than the original 413.886. The reduction factor, at 0.0005 is therefore roughly the same in both cases under these constraints. Furthermore, there were no absolute universals, only statistical ones. With the absolute, but mainly tautological implications out of the way, it seems that what comes to the fore now is slightly more interesting from a linguistic perspective. An example is shown in (9).

- (9) 28. BICHDC=1 => COMALN=1 n=35
 [REL=0.29, APL=0.9, COV=0.5, NEG=0.64, chi2 < 0.5%] STAT
 Locus of Marking in the Clause: P is head-marked
 =>
 Alignment of Case Marking of Full Noun Phrase: Neutral
 EXC: COO DNI GUA SNM

Thus, a potential statistical universal could be that languages that mark the Patient on the verb almost always have neutral alignment in terms of case marking on their NPs.

Activation of the grouping option for the W359 sample led to the following results. I found 75 implications, i.e. slightly more than the 67 of the full database. Since grouping itself only eliminates potential implications, this rise is caused exclusively by the lower value of the applicability parameter. Of the 75 implications, 43 were also found in the previous run, without the grouping, and 32 are new. As for the other scores, the same picture arises as for the previous comparison. The mean values for the criteria are virtually the same again, with the exception of negation, which at 0.73 is much higher for W359 than the 0.48 of the complete database. And again, the average number of languages to which the implications apply is relatively high for W359 in comparison to the sample size. Here it is even 30.3% of the figures for the total set as opposed to the 18.4% of the non-grouped versions.

Many of the implications found for this run are statistically convincing but still seem to lack obvious linguistic relevance, and may simply be caused by areal or genetic effects. (10) is a potential example of such an implication.

- (10) 95. KAYBCC=7 => SIEPAS=1 n=7
 [FREL=0.28, APL=1.0, COV=0.58, NEG=0.72, chi2<0.5%] ABS
 Number of Basic Colour Categories: More than 10 categories
 =>
 Passive Constructions: There is a passive construction
 Indo-European(5) Korean-Japanese (2)

Interestingly, however, most of the 75 implications found under these constraints include languages from a considerable range of language families, typically between 25 and 70 of the 204 distinguished in the Ethnologue version I used.

However, in a database that would contain thousands rather than hundreds of variables, and that would have few missing values, there may turn out to be paths that establish links between variables that now provide us with puzzling pseudo-implications. The program has a build in option to reveal such paths, and to construct a network of concatenated implications. In order for this option to reveal deeper relationships we would need to repair the gaps in the database in terms of missing values.⁵ Another option helps explaining counterexamples by searching the database for deviant behaviour of the relevant languages in terms of the other variables and implications in the database. For these, and further options of the LINFER program I refer to Bakker (1994). In the next, and last section I will try and find

an answer to the question of how many implications we may expect to find in a large database such as WALS purely on the basis of chance.

8. How many inferences may we expect?

A main objective of statistics is to help distinguish between accidental and ‘real’ relationships among phenomena. Given the relatively high number of potential implications in the WALS database – almost half a million – one wonders what is the real chance on an implication fulfilling the requirements in this mass of data even if there would be no underlying relationship at all. Therefore, I did an experiment that seeks to give an answer to precisely that question. The LINFER program has a simulation option under which the real variable values in a database are distributed randomly over the languages. Since this is done in a completely haphazard way, the result will be that existing relationships between values of pairs of variables will largely disappear. New ones, of course may come into existence, but these will be due to chance only, and become ‘visible’ to the extent that they surpass the respective thresholds created by the four selection criteria, and the χ^2 and FE tests. Missing values are not replaced under this option. As far as the real values are concerned, I opted for leaving the original value distributions intact rather than assigning the values in a completely random fashion. This gives the real variable-value combinations the same chance of appearing. Thus, for variable SIEPAS the value distribution will remain as in figure 1 below, i.e. 43.4% YES vs. 56.6% NO, be it that the two values are assigned to the 373 languages of the relevant sub sample in the database completely randomly. Note that this will not affect the theoretical number of implications, only the potentially existing underlying relationships between any two variable-value pairs.

<u>Values for variable SIEPAS</u>		
PASSIVE CONSTRUCTION	162	(43.4%)
NO PASSIVE CONSTRUCTION	211	(56.6%)
TOTAL	373	
MISSING	2185	(85.42% OF 2558)

Figure 1. Frequencies for variable ‘Passive Constructions’

I ran 100 simulations, each with a new random distribution for the values of the 139 variables. The acceptability criteria for implications were kept at the default values presented in the previous section. Table 5 gives the results of these simulations. In the middle column are the figures for the run on the unaffected database. The column on the right-hand side gives the mean results over the hundred simulations.

	Normal inference Process	Simulations
Number of runs	1	100
Number of implications mean	202	52.3
minimum		19
maximum		91
st. dev.		15.2

Table 5. Results for the simulation

Thus, it is certainly the case that some implications are found in the database after the randomisation of the original data. However, their number is considerably lower, at around a quarter of the original number on average, and sometimes as low as 19 (i.e. 10% of 202) and never higher than 91 (45%). For most of these runs, there was hardly any overlap whatsoever in terms of the implications that were found. As to be expected, these ‘fantasy’ implications were typically between the high frequency values of the variables with the highest number of languages. Often, the few implications found for the simulated versions almost all had the same variable-value combination for their conclusion.

We may conclude that, at least for the WALS database, only a minority of the implications that were found are a side effect of large numbers, and due to chance only. The majority deserve to be analyzed, and are of potential linguistic interest, especially since there seems to be a genetic (or areal) explanation for only very few of them.

9. Conclusions

In the previous sections, I subjected the variables of the WALS database to the blind inference machine contained in the LINFER program. Under the default constraints applied for this exercise, which allow some room for counterexamples but are rather strict in terms of several other acceptability criteria, I found that around 1 in 2000 logically possible variable-value combinations are actually attested, and are strong enough to pass as implications. However, the WALS database is not very representative for what is the usual linguistic database. It contains an extremely high percentage of missing values. Unlike regular databases, which are often the product of one author, and stem from a relatively small and well-defined linguistic sub domain, the WALS data originate from a large number of authors and from very diverse sub domains. Each of these contains a small number of variables with a rather restricted number of different values. Within the sub domains, there is a large amount of overlap between the variables, which leads to the presence of many tautologies among the implications that are found. If we filter these out by blocking implications between variables stemming from one author, the reduction is roughly to 1 in 5000 implications. These, however, seem to be linguistically more interesting, at least potentially. Neither the reduction figures nor the implications found change dramatically when we select a sub sample of 359 out of the 2559 languages, which form a genetically representative sample of languages with a much lower number of missing values. The only factor that changes considerably is the negation value. If we would attribute considerable weight to the negation factor then the implications that are found in the W359 sub sample would be more representative for language as a whole than those found for the complete database. Finally, applying the inference procedure to a number of randomised versions of the database suggested that a relatively small proportion of the implications found for the complete WALS and the W359 sub sample stem from pure chance. At least some 75% should then be of potential linguistic interest and worth exploring. What is an interesting implication, however, is ultimately in the eye of the linguistic beholder and determined by the theory in which it may be successfully embedded. LINFER can only help finding them in a sea of non-implications, semi-implications and tautologies.

References

- Bakker, D. (1994). *Formal and computational aspects of Functional Grammar and Language Typology*. Amsterdam: IFOTT.
- Cysouw, M. (2003). 'Against implicational universals'. *Linguistic Typology* 7-1, 89-100.
- Dryer, M.S. (2003). 'Significant and non-significant implicational universals'. *Linguistic Typology* 7-1, 108-1127.
- Greenberg, J.H. (1963). 'Some universals of grammar with particular reference to the order of meaningful elements'. In: Joseph H. Greenberg (ed.) *Universals of Language*. London: MIT Press, pp. 73-113.
- Grimes, B.F. (ed.) (1996). *Ethnologue: Languages of the World*. Dallas: SIL.
- Haspelmath, M., M. Dryer, D. Gil & B. Comrie (eds) (2005). *The World Atlas Of Language Structures*. Oxford: Oxford University Press.
- Haspelmath, M. (2003). 'The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison'. In: M. Tomasello (ed.) *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure*. Mahwah: Erlbaum, 211-242.
- Maslova, E. (2003). 'A case for implicational universals'. *Linguistic Typology* 7-1, 101-107.
- Plank, F. (2003). 'There is more than one way to make sense of one-way Implications – and sense they need to be made of'. *Linguistic Typology* 7-1, 89-100.
- Rijkhoff J. & D. Bakker (1998). 'Language sampling'. *Linguistic Typology* 2-3, 263-314.
- Universals Archive. <http://ling.uni-konstanz.de/pages/proj/sprachbau.htm>

¹ I eliminated three variables that had data for a very restricted number of languages only.

² In other words: the implication is bidirectional, $p \leftrightarrow q$. The NEG criterion is also expressed as a value between 0.0 and 1.0. The maximum value is reached when all languages with a premiss $\neg p$ have a conclusion $\neg q$.

³ It is important to note that the DV technique is defined independently from the type of classification used, be it a genetic classification or an areal one. Given a predetermined sample size and any tree-like classification, it will seek to find the maximal overall distance between the languages in the sample in terms of the classification.

⁴ The considerable reduction from 2177 to 359 is caused by the fact that some of the families are severely underrepresented. Since in this procedure only the languages already in the sample may be kept, and no new languages added, the result is that the sample size is scaled down to the level of the family with the smallest number of languages in the original sample.

⁵ Such paths between the appropriate type of variables may lead to the discovery of semantic maps, as discussed in Haspelmath (2003).