

Søren Wichmann

## 6 Statistical observations on implicational (verb) hierarchies

### 1 Introduction

Implicational hierarchies have been one of the key ingredients in linguistic typology for around half a century, i.e., ever since the discovery of Berlin & Kay (1969) that the presence of a certain color term in a language may imply the presence of others, Silverstein's (1976) observations on animacy scales, and the formulation of the Noun Phrase Accessibility Hierarchy by Keenan & Comrie (1977). The following passage from Corbett (2010: 191) is worth quoting in full because it clearly states why such hierarchies are important, and also because the last sentence reflects an assumption which is worth dwelling upon as the point of departure for the present paper:

Hierarchies are one of the most powerful theoretical tools available to the typologist. They allow us to make specific and restrictive claims about possible human languages. This means that it is easy to establish what would count as counterexamples, and as a result there are relatively few hierarchies which have stood the test of time.

The assumption hinted at, although easy to overlook since it is not stated explicitly, is a strong one: a hierarchy must be discarded if exceptions to its predictions are found. This assumption or practice cannot, however, be considered uncontroversial. In fact, it is at variance with both current tendencies in typology and normal practice in the social sciences. In typology we are increasingly accustomed to approach matters quantitatively, looking for statistically supported patterns rather than exceptionless universals. For instance, implications of the sort first suggested by Greenberg (1963) were from the outset formulated as statistical implications, and such implications (*if A then B*) are, in fact, but a special case of more extended hierarchies (*[if A then B] and [if B then C] ...*) (Cysouw 2003). It seems contradictory to admit for exceptions in the case of pairwise implications but then to not accept them in the case of chained pairwise implications, i.e., implicational hierarchies. Moreover, in the social sciences at large the necessity for statistical approaches is taken for granted. Although the notion of an implicational hierarchy was introduced into social psychology a couple of decades before it became current in linguistics (Guttman 1944), linguists to this day seem to have been unaware of this, and are therefore also unaware of how social psychologists as a matter of course have been dealing quantitatively with *exceptions* to implicational hierarchies.

The purpose of this paper is twofold. First I would like to take some steps towards the introduction of a new methodological approach to implicational scales

with special attention to how to test the significance of such scales. Next, I apply the methodology to subsets of the data of Hartmann et al. (2013) (henceforth ValPaL) on alternations across languages in order to propose implicational scales among the verb meanings included in the database and to provide estimates of their statistical significance. Finally the implicational scales are compared.

Some words about how this paper relates to previous work are in order. The paper is a sort of sequel to Wichmann (forthc.), which introduced Guttman scaling as a way to measure the strength of an implicational hierarchy and used Neighbor-Net graphs as a tool for visual inspection of the degree of unidimensionality of hierarchies. Five types of alternation were investigated: antipassive, passive, reciprocal, reflexive, and causative, and it was found that the hierarchies of verb meanings (to some extent) governing the applicability of each of the first four of these alternations across languages were inter-correlated, whereas the causative responds to its own, separate hierarchy. Setting the stage for the entire analysis was the transitivity hierarchy of Tsunoda (1985). The present paper can be read independently of the previous one and has changes in emphasis and much new material. Here Guttman scaling will also be applied, but since Wichmann (forthc.) I have created a computer program for carrying out the analyses faster and more rigorously and have also developed a significance test. As for the data, I am now drawing upon a more recent, improved version of the database, and I will also be looking at different selections of the data in addition to the ones previously analyzed.<sup>1</sup> Finally, less emphasis will be placed on Tsunoda (1985). Thus, this paper in several ways supersedes Wichmann (forthc.) and can be read independently of it, but the reader may nevertheless find potentially useful additional information in Wichmann (forthc.) and is encouraged to also consult that paper for more background on the topics treated here.

## 2 Guttman scaling: the basics

Guttman scaling is a method for measuring the one-dimensionality of a dataset—in other words the degree to which it conforms to an implicational hierarchy. The method is named after its inventor, who proposed it in Guttman (1944). In presenting it, I will stick to the original model because of its conceptual simplicity, ignoring more recent derivatives which have been developed for the same purpose as the Guttman scale but are more complicated.

Following Guttman's method, the values (here: 1 or 0, corresponding to presence vs. absence) of an attribute (here: a certain alternation in a certain language)

---

<sup>1</sup> The new selections were suggested by Andrej Malchukov, who also formulated the criteria for the selections found in Section 5 and Appendix 2.

of a given individual (here: a certain verb) are first ordered in a ‘scalogram’, where individuals and attributes are in rows and columns, and where these rows and columns are ordered such that the row with the most frequently occurring instance of a value is at one extreme of the scalogram, the next row, placed adjacent to the first, is the one with the next-most frequent instance of the value, and so on. Subsequently, the number of deviations from the pattern of a perfect scale can be counted. Let me first illustrate in (1) what a scalogram corresponding to a perfect implicational scale would look like.

(1) Verb A	11111
Verb B	11110
Verb C	11100
Verb D	11000
Verb E	10000
Verb F	00000

Example (2) reproduces (1), but with the introduction of two changes, in the rows for Verb A and Verb E.

(2) Verb A	11101
Verb B	11110
Verb C	11100
Verb D	11000
Verb E	10010
Verb F	00000

The changes introduced in (2) illustrate two cases of deviation from perfect scalarity. The 0 value in the first row (for Verb A) is an error, since a 0 is included in a sequence of 1’s; and so is the second 1 value in the row for Verb E since it is included in a sequence of 0’s. The Guttman Coefficient (GC) is a measure of scalarity. It is calculated in a simple manner, by dividing the total number of errors by the total number of data points, expressing this as a percentage and finally subtracting it from 100 %. In (2), there are 2 errors among the 30 data points, so  $GC = 100 - (100 \cdot 2/30) = 93.33\%$ .

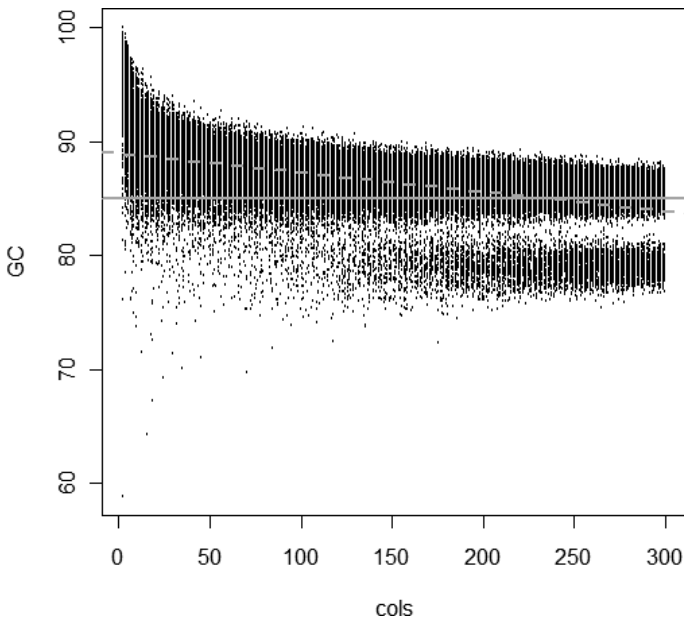
Appendix 1 reproduces the R script created for calculating a GC and outputting a scale to a file. It serves to make the algorithm more explicit and the present study more easily replicated.

As far as I am aware, no statistical evidence has been brought to bear on the question of just how much deviation can be deemed acceptable, which is a weakness of the Guttman Coefficient; but Guttman found, based on practical experience, that “85 percent perfect scales or better have been used as efficient approximations to perfect scales” (Guttman 1944: 140). In Wichmann (forthc.) the 85 % criterion was tentatively adopted as a rule of thumb for assessing the acceptability of a GC

although it was stated that “in its application to linguistic data the Guttman scale needs to be tested more before we can place much confidence in such estimates.”

Figure 1 shows that for the data at hand it is in fact not warranted to accept an 85% cut-off as an indication that a meaningful implicational hierarchy has been found. To produce Figure 1, 100,000 subsets of the ValPaL database were selected with equal probabilities of picking from 3 to 300 random columns. No other criteria were applied. Thus, each subset is completely random. Some alternations included in a given sample may be related in function, others may not. In spite of the random selection, the vast majority (81.7%) of GC's are 85% or greater. Another interesting result of this exercise is that an inverse correlation between the number of randomly picked columns and the GC's emerges, indicating that the GC tends to increase when the matrix size decreases ( $\rho = -0.581$ ,  $p < 2.2e-16$ ). There is also an inverse correlation between the number of empty cells and GC ( $\rho = -0.626$ ,  $p < 2.2e-16$ ). The fact that most GC values are in the high ~77–90% range has to do with the small proportion of 1's to 0's in the dataset. The percentage of 1's out of the total number of filled cells rounds up to 21%. I return to this factor in the next section.

All these findings motivate the construction of a significance test which is based directly on a given dataset with a certain size, a certain number of empty



**Fig. 1:** GC as a function of the number of columns for 100,000 randomly selected subsets of the entire ValPaL database with a linear regression line (dashed) and a line demarcating the conventional 85% cut-off.

cells, and a certain proportion of 1's to 0's. In the next section such a test is developed, and in section 4 it is applied to the ValPaL data.

### 3 A significance test for Guttman scales

In order to probe further into the properties of the GC and construct an appropriate significance test I conducted a simulation experiment where 10,000 matrices were constructed, varying the size and the distribution of 1's and 0's as follows. For each matrix the number of columns was set to a number randomly chosen between 3 and 100, and similarly for the number of rows. Thus, the size of a matrix can vary from 9 to 10,000 cells. Once each matrix was constructed it was filled with 1's and 0's with the probability of picking a 1 for each cell being randomly chosen from the following probability set: 0.1, 0.2, 0.3, ..., 0.9. For each matrix the GC was computed. Additionally, a *p*-value for each of these GC's were computed in the following way. For a given matrix 10,000 randomizations were made without changing the sum of 1's in each row and column. The number of GC's in the 10,000 randomizations greater or equal to the GC of the original matrix were counted and the *p*-value was calculated by dividing this number by 10,000. The *p*-value, then, represents the probability for getting a GC equal to or better than the original one by chance. The less often this happens, the more confidence one can invest in the GC. Conventionally  $p < 0.05$  is taken to be a reasonable significance cut-off. Matrix randomization is a standard statistical procedure, and has also been recommended for linguistic typological data (Janssen et al. 2006). The simulated matrices were constructed using the base package of R (R Development Core Team 2008) and for the randomization keeping row and column sums constant the function `permatfull` of the `vegan` package (Oksanen et al. 2011) was used.

The simulation experiment serves to investigate the relationships among the matrix size (*N* number of cells), GC, the distribution of 1's and 0's, and *p*. Table 1 reports the relevant correlations among these entities. As regards the distribution of 1's and 0's what matters is not so much the proportion of one to the other, but rather the deviation from a fifty-fifty distribution. This deviation is obtained by taking the absolute value (converting minuses to pluses) of  $p(1) - 0.5$ , where  $p(1)$  is the preset probability for filling a cell with a 1. This is designated  $p(1)^{DEV}$ . It takes values between 0 and 0.4.

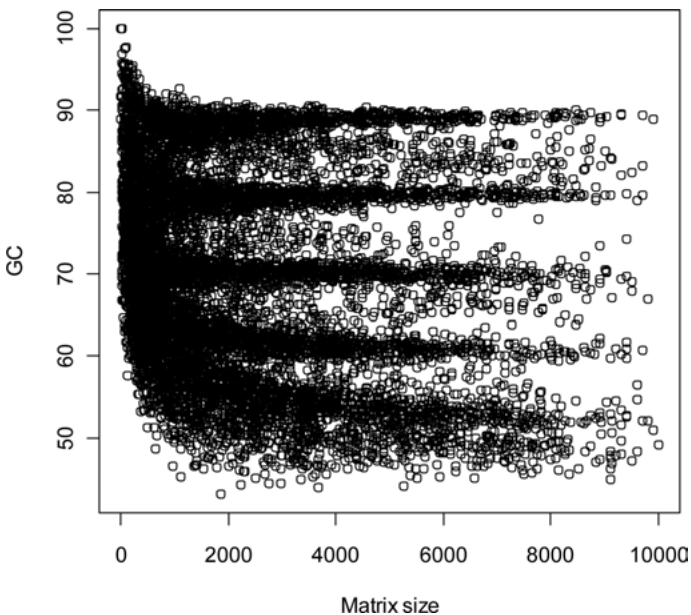
The numbers in Table 1 tell us that there is a small negative correlation between the matrix size and GC. Although it is smaller than in the real data (cf. Figure 1 above) it cannot be neglected. There is a very strong positive correlation between  $p(1)^{DEV}$  and GC, indicating that the more skewedness there is in the proportion between 1's and 0's – whether one or the other dominates – the easier it is to get high GC values. The other correlations serve as a test of the *p*-values. These values are almost totally uncorrelated with either GC, the size of the matrix or  $p(1)^{DEV}$ .

**Tab. 1:** Correlations among entities in the simulation experiment. Values of Spearman's  $\rho$  are in the lower left triangle (dark shade) and  $p$ -values in the upper right triangle (light shade).

$\rho \setminus p$	N	GC	$p(1)^{\text{DEV}}$	$p$
N		< 2.2e-16		0.0288
GC	-0.2152		< 2.2e-16	0.3471
$p(1)^{\text{DEV}}$		0.8154		8.737e-14
$p$	-0.0219	-0.0094	0.0745	

This is as it should be: we do not want the  $p$ -values, which help us to estimate the validity of a GC, to be predictable from the value of GC or its correlates. As for the distribution of  $p$ -values, it is important to report that 4.2% of these are smaller than or equal to 0.05. In other words, close to 5% of the simulations gave significant results at the 5% level, which again is as it should be for a meaningful significance test.

In addition to the development of a significance test the simulation also serves to provide some insights into the distribution of GC values seen in the empirical dataset in Figure 1. In Figure 2 GC is plotted as a function of matrix size for the simulation. There are five horizontal bands of dots. The bottom-most band corresponds to  $p(1)^{\text{DEV}} = 0$ , the one above it to  $p(1)^{\text{DEV}} = 0.1$ , and so on, with the top-most band corresponding to  $p(1)^{\text{DEV}} = 0.4$ . Thus, a certain distribution of 1's and



**Fig. 2:** GC as a function of the number of cells for 10,000 simulated matrices.

0's leads to a certain characteristic distribution of GC values. This tells us that the overall concentration around high values in Figure 1 must clearly be related to the skewedness in the proportion of 1's to 0's. The percentage of 1's out of the total number of filled cells, which is ~21%, corresponds to  $p(1)^{DEV} = 0.5 - 0.21 = 0.29$ . In the simulated data  $p(1)^{DEV} = 0.3$  gives GC values concentrated around 80%. This is similar to what we find in the empirical data. The only issue which is presently not clear is why, in Figure 1, there are two horizontal bands rather than just one, but it would take us too far astray to investigate this matter further.

## 4 Testing for significant implicational hierarchies for subsets of the ValPaL data

Two selections of subsets of the ValPaL data were made for the purpose of studying implicational verb hierarchies: a *restricted* selection, where alternations were required to be quite similar; where at least 10 languages should be attested for a given alternation; and where only one language is represented per alternation. This selection, which will be called the R-selection, is essentially<sup>2</sup> the one used in Wichmann (forthc.). It features five alternations: causative<sup>R</sup>, passive, antipassive, reciprocal<sup>R</sup>, and reflexive<sup>R</sup>. None of the subsets contains more than 15 columns. Another selection was made for the present paper which is more *inclusive* in the sense that the categories are sometimes broader and also less strict since it was not required that only one alternation should be represented per language. This will be called the I-selection. It includes causative<sup>I</sup>, object-demoting/deleting, reciprocal<sup>I</sup>, reflexive<sup>I</sup>, and subject-demoting/deleting. The number of columns for each category in the I-selection ranges from 16 to 117. Appendix 2 shows which language-specific alternations were assigned to the different categories in the R- and I-selections and also gives short descriptions of the criteria for including different alternations in the larger categories. Guttman coefficients and *p*-values were computed for each category in the two selections, cf. Table 1. The *p*-values are based on a set of 10,000 matrices, where 99,999 are randomizations of the original matrix and 1 is the original matrix itself (including the original matrix among the randomizations is a way of avoiding the possibility of a 0 *p*-value and is recommended by North et al. 2002). Lamentably, I have not been able to produce *p*-values that reflect the data with complete accuracy because the data contain empty cells (ranging from 0.6% to

---

<sup>2</sup> During the period between the writing of Wichmann (forthc.) and the present paper the database underwent some minor changes. Of relevance for the R-selection are especially the facts that it was decided to remove the Ket and Mandarin object omission alternations as well as the German reciprocal and reflexive from the database. Moreover Wichmann (forthc.) draws upon 87 verb meanings, whereas the present paper draws upon 80.

12.6% for the different subsets of the R- and I-selections), and empty cells cannot be dealt with by the randomization algorithm implemented in the function `permat-full` of the `vegan` package, which is the function used for producing the randomized matrices. Creating an efficient algorithm for randomizing a matrix with the row and column sums kept constant, and additionally keeping the number of empty cells constant, is not a trivial problem. Because of time constraints I have not been able to solve it yet. Instead, two  $p$ -values were produced, one for the matrix where all missing values were replaced by 1's and one where all missing values were replaced by 0's. While it is unclear what effect these imputed values have on the  $p$ -value, it is reasonable to assume that a GC which receives a significant  $p$ -value in both situations (imputation of 1's and imputation of 0's) is far more trustworthy than one which does not receive a significant  $p$ -value in any of the two situations. It is unclear how to interpret cases where a  $p$ -value is significant in just one of the two situations. Here such cases are given the benefit of the doubt and are included in further analyses.

The procedure just described provides the motivation for the inclusion of the different numbers displayed in Table 2. It shows, for each subset, the number of columns (which for the R-selection is equivalent to the number of languages), the percent empty cells (which adversely affects the precision of  $p$ -values), the GC for the original matrix, and the GC's and  $p$ -values for respectively the matrix with 1's and 0's imputed.

**Tab. 2:** GC's and significance tests for selections of the data. Legend: Cols. = the number of columns; %NA = percent empty cells; GC = Guttman coefficient (in percent) for the data including empty cells; GC(1) = Guttman coefficient (in percent) for the data with 1's imputed into empty cells; GC(0) = Guttman coefficient (in percent) for the data with 0's imputed into empty cells;  $p(1)$  =  $p$ -value for GC(1);  $p(0)$  =  $p$ -value for GC(0). Significant  $p$ -values are highlighted.

Alternation	Cols.	%NA	GC	GC(1)	GC(0)	$p(1)$	$p(0)$
causative <sup>R</sup>	15	4.5	88.74	87.08	87.17	<b>0.0005</b>	<b>0.0001</b>
passive	11	4.3	89.35	90.00	85.80	0.3009	0.7429
antipassive	10	5.6	89.14	87.25	89.12	0.5209	0.5953
reciprocal <sup>R</sup>	10	4.1	92.44	89.38	91.62	0.7521	0.2066
reflexive <sup>R</sup>	11	5.7	86.99	85.80	87.50	0.4794	0.6039
four previous combined	57	4.8	83.10	81.84	82.28	<b>0.0001</b>	<b>0.0001</b>
causative <sup>I</sup>	38	7.9	87.20	88.12	85.13	<b>0.0001</b>	0.9643
subject-demoting/ deleting	117	8.7	86.64	86.63	86.62	<b>0.0004</b>	0.2890
object-demoting/ deleting	58	8.0	89.43	86.96	89.94	<b>0.0001</b>	<b>0.0001</b>
reciprocal <sup>I</sup>	16	7.8	93.14	89.84	92.42	<b>0.0053</b>	0.4491
reflexive <sup>I</sup>	21	12.6	85.85	85.00	89.05	0.4376	<b>0.0002</b>



For the R-selection the results in Table 2 serve as a replication and further test of the findings of Wichmann (forthc.). As far as the Guttman coefficients are concerned, the present automated results show values that are consistently a little (2–4 %) higher than the hand-produced GC's of Wichmann (forthc.), but the earlier and the present values are almost perfectly correlated ( $r = 0.982$ ).<sup>3</sup> It is difficult to track the sources of the differences. As mentioned in footnote 2, the database has undergone some changes, and another source of differences may be inconsistencies in the hand-produced GC's. All in all, the high correlation can be taken as a confirmation that the computational implementation of Guttman scaling works as intended. The  $p$ -values for the five implicational scales in Wichmann (forthc.) are an important new result. Interestingly, only the hierarchy for the causative turns out to be significant. This does not necessarily mean that the verb hierarchies for the four other alternations – passive, antipassive, reflexive, reciprocal – included in Wichmann (forthc.) are wrong, it only means that they are very hypothetical and that data from more languages would be needed to support them. An indication that the results are at least meaningful is the fact that a positive rank correlation was found among all of the hierarchies, except that of the causative; and when the data for the passive, antipassive, reflexive, and reciprocal are combined the GC becomes significant.

In spite of the shortcomings of the  $p$ -values they greatly help us to direct the focus when selecting data for further analysis. The alternations (or alternation categories) for which  $p(1) \leq 0.05$  or  $p(0) \leq 0.05$  are: subject-demoting/deleting (henceforth 'subject demdel'), object-demoting/deleting (henceforth 'object demdel'), reciprocal<sup>l</sup> (henceforth simply 'reciprocal'), reflexive<sup>l</sup> (henceforth simply 'reflexive'), causative<sup>R</sup>, and causative<sup>l</sup>. Between the two last I choose causative<sup>R</sup> (henceforth simply 'causative'), since for this both  $p(1)$  and  $p(0)$  are significant, whereas for causative<sup>l</sup> only  $p(1)$  is significant.

## 5 NeighborNets and Implicational verb hierarchies for different alternation categories

This section presents NeighborNets and Guttman scales for the different alternation categories selected. NeighborNet (Huson and Bryant 2006) is useful for clustering and also for showing how tree-like the data is. If verbs are ordered in a perfect implicational scale, the NeighborNet will simply be a string connecting the verbs. A non-tree-like behavior is shown in a NeighborNet by boxes. These boxes along

---

<sup>3</sup> The GC's reported in Wichmann (forthc.) are as follows. Causative: 85.4 %, passive: 86.6 %, antipassive: 86.0 %; reciprocal: 88.8 %; reflexive: 84.7 %.

with the structure as a whole of the network, indicating whether or not it has an approximately unidimensional behavior, provide visual clues as to how far from an implicational hierarchy a dataset strays. Moreover, the SplitsTree software has a function to calculate values of  $\delta$ , which is a measure of the amount of reticulate behavior corresponding to the visual impression given by the boxes.  $\delta$  takes values between 0 and 1 (see Holland et al. 2002 for the first description of this measure and Wichmann et al. 2011 for discussion and application to linguistic phylogenetic data). A  $\delta$ -score around 0.5 or greater indicates strong non-tree-likeness. There is no cut-off point for what can be considered a tree and what not, but we know, for instance, that values in the vicinity of 0.3 are typical for lexically-based linguistic phylogenies (Wichmann et al. 2011). The  $\delta$ -values for the NeighborNets shown below range between 0.27 and 0.39, which is the sort of range expected when the behavior is relatively tree-like. It should be stressed that in the present context NeighborNets only serve as a convenient visualization of the data, which is why they are simply displayed without further ado. The hierarchical relations among verbs will be based on Guttman scaling alone.

In my implementation of Guttman scaling (Appendix 1) the required input is a matrix with data for verbs (rows) and columns (language-specific alternations). The matrix can contain 1's, 0's and empty cells, which should be encoded as "NA" in the input format. The same goes for the SplitsTree input file (except that here "NA" is encoded as a hyphen). These requirements necessitate some simplifications of the data. The ValPaL data has the options "Never", "Marginally", and "Regularly" for indicating whether a given alternation applies to a given verb. "Never" was converted to "0", whereas "Marginally" (which is quite infrequent) and "Regularly" were merged to "1". ValPaL, moreover, has two different ways of encoding missing information about whether a given verb takes a given alternation: empty cell and "No data". These are merged into "NA". After these conversions (which were also applied in the analyses in sections 2 and 4 above) the matrix is first inverted so that the alternations become rows and the verbs columns. Then the columns are rearranged such that the verbs with the highest number of 1's are to the left, the rows are rearranged such that the languages with the highest number of 1's are towards the top. The Guttman scale can now be read off from the order of the verbs from left to right. Normally there will be ties such that some verbs have an equal number of 1's. My script outputs the verbs in their rank order (with the order of verbs whose ranks are tied being arbitrary) along with numbers indicating their rank. For instance, for the object demdel hierarchy the verbs that top the list are EAT, WASH, and GIVE, which then all get a "1". Next come STEAL, TEACH, SHAVE, and COOK, which all get a "4" since there were three verbs competing for rank 1 and these next four verbs now compete for rank 4. And so on. When reporting the hierarchies a comma is used to separate verbs which compete for the same rank and a greater-than sign to separate verbs or groups of verbs having different, adjacent ranks.

The following subsections on each alternation category are structured as follows. First a NeighborNet is displayed and then the hierarchy resulting from Guttman scaling is given. The hierarchy is followed by small paragraphs introduced by the heading ‘some explanatory factors’ where I try to highlight possible explanations for the particular organization of a hierarchy found. These explanations will be quite mundane, sometimes bordering on truisms.

## 5.1 Subject demdel

In the way it is defined for the purposes of this project (cf. Malchukov, this volume, for further detail), Subject demdel refers to a number of alternations, both marked and unmarked, where the subject is demoted or deleted. Thus, it covers not only passives and anticausatives but also ambitransitives, for which it is assumed that the intransitive variant is derived and the transitive one is basic. The NeighborNet is displayed in Figure 3.

Hierarchy:

CUT > BREAK, TEAR, POUR > FILL > PEEL > COVER, BUILD > COOK, TAKE > HIDE, LOAD > SHOW > TIE > WASH, KILL, SHAVE, SEND > THROW > GRIND, BEAT, TEACH > CARRY, PUT > DRESS, FRIGHTEN, WIPE > STEAL, GIVE > HIT, HUG > EAT > BRING > LOOK AT, PUSH, TELL > DIG, ASK FOR > SEE, NAME, THINK > SMELL > HELP, SAY, TOUCH, SING > BLINK > SEARCH FOR, BURN > KNOW > HEAR, SHOUT AT, CLIMB, LIVE > LIKE > MEET, FEAR, ROLL, TALK > FOLLOW, SIT > SIT DOWN > LEAVE, PLAY > RUN, COUGH, SINK, JUMP, FEEL COLD > BE DRY, LAUGH, BE HUNGRY > FEEL PAIN > DIE, BOIL > GO > BE SAD > SCREAM > RAIN, BE A HUNTER.

Explanatory factors:

- for subject demoting or deletion to be possible a verb normally needs to be transitive;
- semantic transitives (the Effective Action verbs of Tsunoda 1985) tend to occur towards the top of hierarchy, followed by two argument verbs, which do not conform to the transitivity prototype, and intransitive (monovalent) verbs cluster at the bottom of the hierarchy (it also seems that on balance agentive intransitives – ‘unergatives’ – are higher on the hierarchy than patientive intransitives – ‘unaccusatives’ – although the pattern is not fully consistent).

## 5.2 Object demdel

Object demdel covers a family of alternations where an object is demoted or deleted. These include cases of object omission (cf. *He reads too much*) and object incorporation, but also marked alternations including different varieties of antipassive





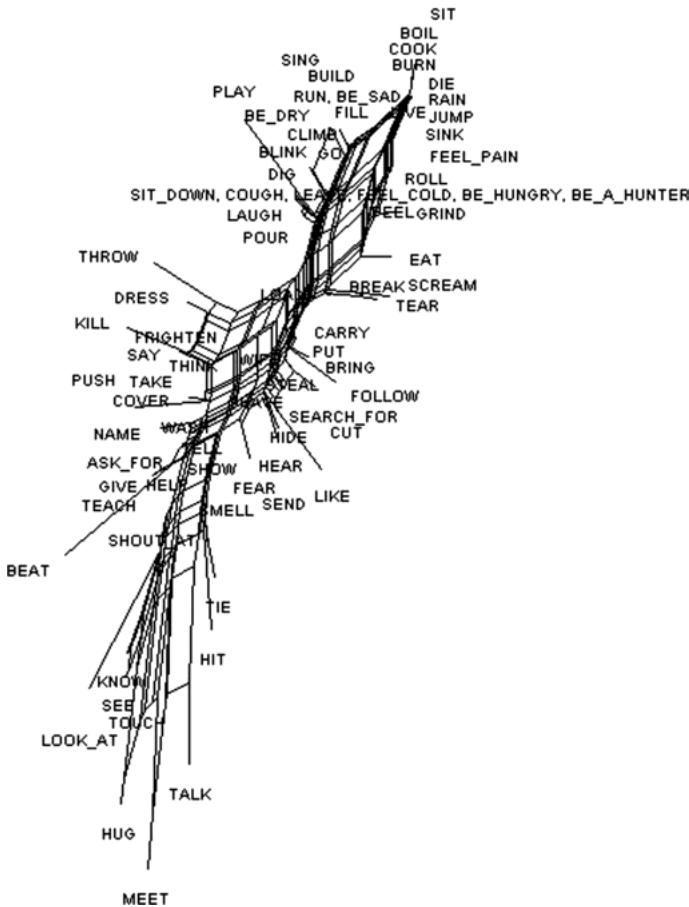


Fig. 5: NeighborNet of data for the reciprocal alternation category  $y$ .  $\delta = 0.2790$ .

e.g., *I met him* vs. *We met*). Reciprocals constructed solely by means of pronouns, such as *to see each other* are not counted as true alternations in the ValPaL database. The NeighborNet is displayed in Figure 5.

### Hierarchy

MEET, HUG, SEE, TOUCH, TALK, KNOW, LOOK AT, BEAT, HIT, TIE, TEACH, GIVE, SHOUT AT, HELP, SHOW, TELL > LIKE, PUSH, ASK FOR, WASH, FRIGHTEN, SMELL, COVER, TAKE > KILL, NAME, HIDE, SHAVE, SEARCH FOR, THINK, FEAR > THROW, SAY, BRING, PLAY > FOLLOW, HEAR, STEAL, WIPE, CUT, PUT, TEAR, DRESS, SCREAM > CARRY, SEND, LOAD, BREAK, EAT, LAUGH, DIG, GRIND, POUR, BUILD > SING, BLINK, BE DRY, ROLL, FILL, PEEL, BE SAD, GO, RUN > CLIMB, BOIL, COOK, BE A HUNTER, RAIN, SINK, BE HUNGRY > BURN, DIE > FEEL COLD, LIVE > FEEL PAIN, LEAVE, JUMP > SIT DOWN > SIT > COUGH.

Some explanatory factors:

- for a reciprocal to apply a verb normally needs to be transitive;
- reciprocals are most likely to be employed when the action described by the verb is typically carried out in a reciprocal manner, which is why we find ‘symmetric predicates’ like MEET and HUG at the top of the hierarchy, followed by transitive verbs taking an animate object, followed by a heterogeneous group of verbs which are less likely to be either transitive or to take an inanimate object, and monovalent (intransitive) verbs are expectedly at the bottom of the hierarchy.

## 5.4 Reflexive

The reflexive alternation includes constructions involving reflexive voice (cf. Russian: *myt’* ‘wash’ vs. *myt’-sja* [wash-REFL] ‘wash oneself’), and also their unmarked counterparts involving reflexive labiality (cf. English *he shaved the customer* vs. *he shaved*). The NeighborNet is displayed in Figure 6.

### Hierarchy

WASH, COVER, SHAVE, SHOW, CUT, SEE, HIDE, DRESS, GIVE, TOUCH > LOOK AT, HEAR, PUT, BEAT, HUG, SMELL, TIE, THROW, HIT, KILL, LIKE, FEAR, WIPE > KNOW, PUSH, ASK FOR, TEAR, NAME, HELP > SEARCH FOR, THINK, TEACH, TAKE, SAY, CARRY, TELL, BREAK, SEND > FRIGHTEN, TALK, LOAD > BUILD, STEAL > BRING, PEEL, COOK, FOLLOW, EAT > FILL, MEET, GRIND, SING, BURN, DIG, BE SAD, POUR, ROLL > SHOUT AT, BE DRY, SCREAM, LAUGH, RUN, PLAY, FEEL PAIN, LEAVE, GO > JUMP, SIT, BLINK, BOIL, BE A HUNTER > LIVE, RAIN, SINK, BE HUNGRY, DIE, FEEL COLD, CLIMB > SIT DOWN > COUGH.

Some explanatory factors:

- for a reflexive to apply a verb normally needs to be transitive;
- reflexives are most likely to be employed when the action described by the verb is amenable to be carried out such that the agent is at the same time the undergoer, which is why verbs denoting grooming actions such as WASH and SHAVE head the hierarchy;
- among transitive verbs, the ones that can take an animate object are more likely to undergo the reciprocal alternation.

## 5.5 Causative

Causative is an A-adding operation, which can operate on intransitive verbs and in some languages also on transitive verbs. With intransitive verbs the underlying S becomes a P, and with transitives an A can be demoted to direct or indirect object

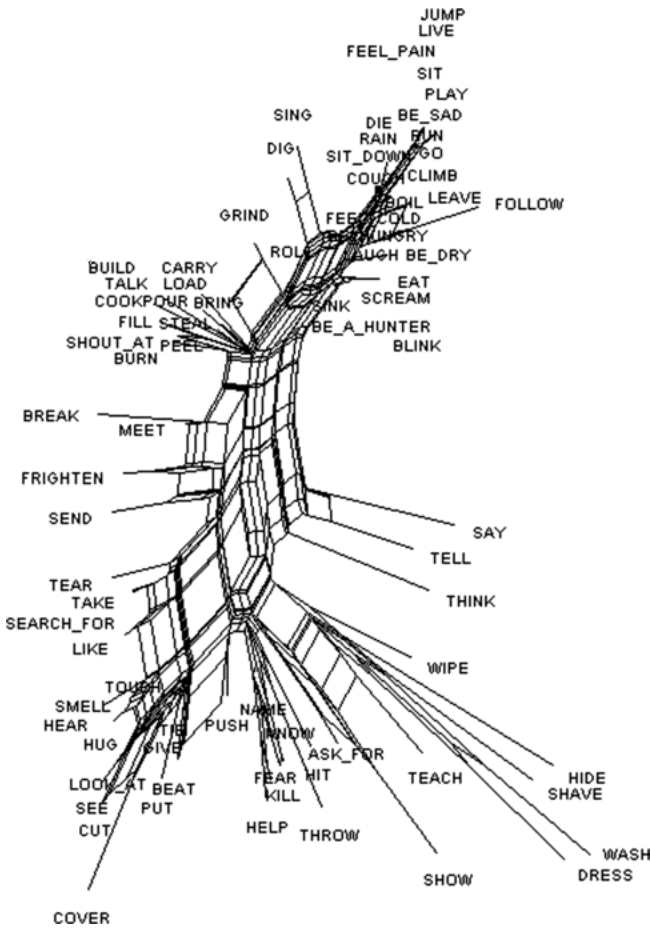


Fig. 6: NeighborNet of data for the reflexive alternation category  $y$ .  $\delta = 0.3252$ .

(cf. Malchukov, this volume, for more detail). The NeighborNet is displayed in Figure 7.

### Hierarchy

JUMP > LAUGH > FEAR > BOIL, SINK, SIT DOWN, RUN, CLIMB, SING, ROLL, PLAY > LIVE, SHOUT AT, SCREAM, SIT, EAT, FEEL COLD, BE DRY, LEAVE, STEAL, PUSH, BE SAD, DIE, FEEL PAIN, GO, COUGH, PUT, HIDE, MEET, KNOW, SMELL, BE HUNGRY, LOAD, POUR, CARRY, TEAR > HEAR, COVER, CUT, TOUCH, DIG, WIPE, GRIND, BURN, BLINK, TALK, FOLLOW, LIKE, HUG, PEEL, TAKE > TEACH, BRING, FILL, TIE, HIT, BUILD, SAY, SHAVE, RAIN, THROW, BEAT, KILL, COOK, SEARCH FOR > BREAK, ASK FOR, HELP, WASH, NAME, FRIGHTEN > SEE, LOOK AT, SEND, TELL > THINK, DRESS, GIVE, SHOW > BE A HUNTER.



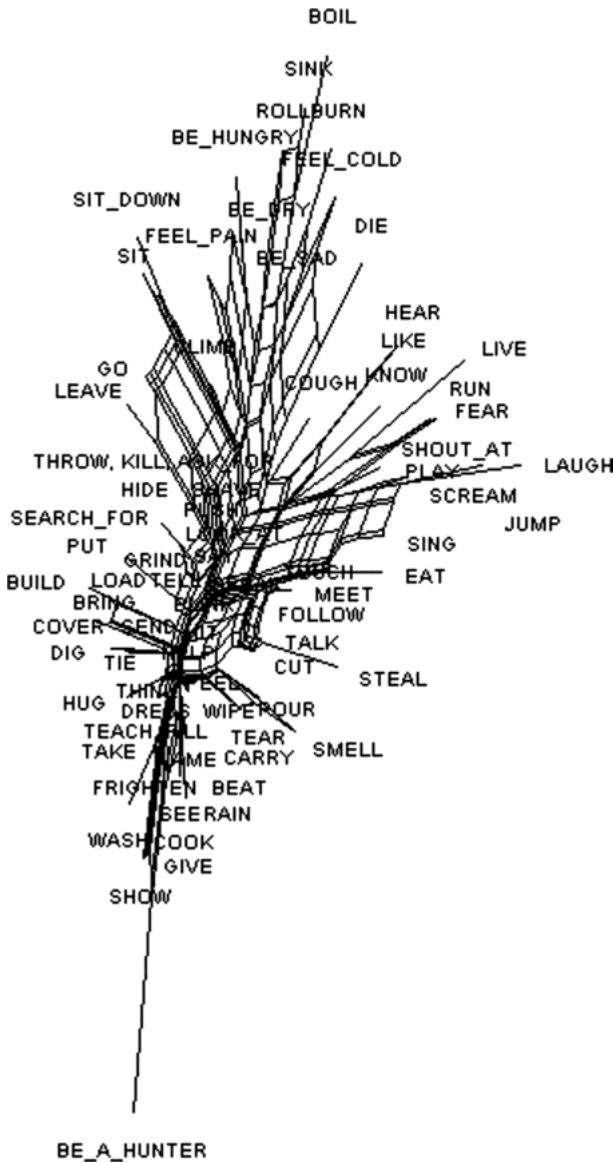


Fig. 7: NeighborNet of data for the causative alternation category.  $\delta = 0.3382$ .

Some explanatory factors:

- intransitives are more likely to undergo causativization since some languages prefer periphrastic constructions for causativizing a transitive;
- verbs describing actions amenable to external control, such as actions that are often involuntary (LAUGH, FEAR, SINK, FEEL COLD, etc.), are more likely to undergo causativization than verbs describing actions which are typically volun-

tary (SEND, ASK FOR, BEAT, SAY, TEACH, etc.) or by their nature not easily controlled by an external agent (e.g., RAIN, BE A HUNTER).

## 5.6 Comparison of hierarchies

The ranks of the two hierarchies were correlated through a Spearman Rank Correlation, yielding the results shown in Table 3. The correlations ( $\rho$ ) are shown in the lower left (dark shaded) triangle and the  $p$ -values in the upper right (light shaded) triangle.

Tab. 3: Spearman Rank Correlations for the hierarchies for the five alternation categories.

$\rho$ / $p$	S demdel	O demdel	Recipr	Reflex	Caus
S demdel		1.067e-13	0.0002	1.542e-10	1.727e-05
O demdel	0.7138		1.695e-07	1.400e-12	1.851e-05
Recipr	0.4067	0.5453		5.752e-17	5.680e-06
Reflex	0.6407	0.6903	0.7713		3.339e-07
Caus	-0.4605	-0.4590	-0.4830	-0.5342	

All the correlations in Table 3 are significant even if none is particularly strong. While all the hierarchies except that of the causative are *positively* correlated, the causative is *negatively* correlated with all the others. This means that the hierarchy for the causative weakly tends to be ordered opposite to the others. This result is quite expected, mainly because transitivity is a factor that has opposite effects on the causative hierarchy and the others while, at the same time, other unrelated factors are at work in the different cases. For instance, habituality is one of the factors at work in the case of object demdel whereas external controllability influences the causative hierarchy.

We might consider whether there is some universal transitivity scale at work behind the subject demdel, object demdel, reciprocal, and reflexive hierarchies as claimed in Tsunoda (1985) for the passive, antipassive, reciprocal, and reflexive. To the result that the four categories from the I-selection in Table are positively correlated we can add the result of Wichmann (forthc.) that the corresponding categories from the more narrowly defined R-selection – antipassive, passive, reflexive, and reciprocal – are also intercorrelated; and, as has been shown here (cf. Table 2 above), the 83.10 % GC for the combination of these four R-selection categories is significant even if low. Nevertheless, not all factors that explain the behavior of verbs with regard to a particular alternation are relevant for the others even if all hierarchies are correlated, and the respective influences of these various factors would no longer be distinguishable if the hierarchies were merged. For instance,

among the highest verbs on the hierarchy for the reciprocal are HUG and MEET. That is not surprising, since to hug and to meet are actions that are often, if not mostly, carried out in a reciprocal manner. In the object demdel hierarchy HUG is in the upper end of the scale but not particularly high-ranking and MEET is in the lower end of the scale. If we merged the two hierarchies it would be hard to interpret whatever placement HUG and MEET somewhere towards the top of the hierarchy would get. Similarly, when we here find that the causative hierarchy is inversely correlated with the four others there might be a temptation to merge a reversed version of it with the others, claiming the existence of an underlying transitivity hierarchy governing all five alternations cross-linguistically. But, again, it would just be hard to interpret the result and no real insights would be gained.

Thus, in all cases we clearly are dealing with separate hierarchies which are determined by the function of each individual alternation. Yet, they do correlate to some extent (positively or negatively), and this is due to transitivity. Ultimately, the correlation is expected to be stronger with syntactic transitivity (that is, the transitivity values of individual verbs for particular languages), but we also expect a correlation with semantic transitivity, as already anticipated by Tsunoda (1985).

## 6 Conclusions

In the quote given in the introduction to this paper, Corbett (2010: 191) states that implicational hierarchies can potentially “allow us to make specific and restrictive claims about possible human languages.” If there are highly restrictive hierarchies among verbs with respect to the sorts of alternation they can undergo across languages the ValPaL cross-linguistic data on alternations across would have shown this. Instead, the data suggest that when we do find evidence for hierarchies they only constitute tendencies. Thus, for instance, we should not interpret the placement in the causative hierarchy of such items as COUGH and JUMP to mean that if a language can apply a causative formation to its verb for ‘to cough’ then it can also apply it to ‘to jump’. Rather, we may allow ourselves to entertain a certain expectation that this would be the case. An implicational hierarchy can serve as a *descriptive summary* of a certain typological distribution, and such a summary can be highly useful for identifying possible factors affecting grammatical phenomena cross-linguistically. Often such factors will be rooted in the nature of human action, interaction, and cognition.

If implicational hierarchies are often, or perhaps even usually, imperfect, how can we distinguish between a hierarchy that says something interesting and one that does not? The methodological parts of this paper have been concerned with the assessment of the predictive powers of such hierarchies and also the reliability of these predictive powers. The Guttman Coefficient (GC) can be equated with the predictive power of a hierarchy. A 100 % GC would correspond to a linguistic uni-

versal. An example of an alternation with a relatively high GC would be the passive, which showed a GC of 89.35%. But is this GC reliable? Our significance test indicates that it is not. The GC was based on data from just eleven languages, so this may be the reason why the significance test fails. It cannot be excluded that more data would give a better result. In any case, the significance test helps to avoid making claims about a pattern which is only apparent since it could as easily have been obtained by chance.

The results hinting at effects of the size of the matrix can be drawn upon to suggest some guidelines for typological sampling. Table 1 in Section 4 shows that the smallest dataset for which significance is reached consists of 15 columns (i.e.,  $15 \cdot 80 = 1200$  data points, not taking into account empty cells). This suggests that a cross-linguistic database for studying implicational hierarchies should have at least around 1000 data points per category to be useful. The utility obviously increases with the size of the database, but since the cost of gathering data is always a concern one might wonder whether there is a size at which the returns begin to diminish. Figure 1 in Section 2 shows that there is a consistent dependency on the size of the database of the GC for random selected subsets, but that this dependency is stronger in the range from 3 to around 50 columns. Size-dependency becomes less of an issue with 50 or more columns, i.e.,  $50 \cdot 80 = 4000$  data points. Based on this study, then, I can recommend a matrix size of about 4000 data points per category as one that maximizes reliability of results while minimizing the practical efforts, and I can also recommend avoiding anything less than 1000 data points per category.

I believe that this paper constitutes an important step forwards as regards the development of methods for dealing with implicational hierarchies, and the study has also produced some potentially interesting empirical observations. Nevertheless, there are still fertile areas left for the future harvesting of insights. They include the following:

- the significance test of GC's needs to be refined such that it can deal more adequately with missing data;
- alternations that only fail one of the two parts of the significance test (reciprocal, reflexive, subject-demoting/deleting) or for which the ValPaL dataset includes too few representatives deserve further study through additional data;
- the synchronic results for implicational scales imply diachronic predictions about how, i.e., to which verbs, a given alternation will extend its domain, something that invites further investigation;
- the particular procedure for Guttman scaling applied here possibly needs further consolidation through tests and comparisons with practices in the social sciences.

## Appendix 1: An R script for constructing an implicational scale and calculating GC

The script takes a matrix with 0, 1, and NA's (empty cells) as input. It is called by writing "Gm(x)", where x is the matrix to be analyzed, and writes the GC to the console and the scale to a file called GSout.txt. If a scale is to be output the matrix should have row names indicating individuals (in this case verbs) and column names indicating attributes (in this case alternations). But it will output a GC in the absence of row and column names.

```
Gm <- function(x) {
  x <- t(x) # rearrange the matrix
  rl <- length(x[1,]); cl <- length(x[,1])
  horsum <- function(y) { sum(x[y,],na.rm=T) / (rl - length(which (is.na(x[y,1:rl])))) }
  versum <- function(z) { sum(x[,z],na.rm=T) / (cl - length(which (is.na(x[1:cl,z])))) }
  hsum <- c(); vsum <- c()
  for (i in 1:cl) {hsum[i] <- horsum(i)}
  for (i in 1:rl) {vsum[i] <- versum(i)}
  hs <- rev(order(hsum)); vs <- rev(order(vsum))
  left <- x[,vs]
  x <- left[hs,]
  for (i in 1:length(x[,1])) {
    if (is.na(x[i,1])==T) {
      x[i,1] <- 0
    }
  }
  indiv <- colnames(x) # produce the Guttman scale
  sums <- c()
  for (k in 1:length(x[,1])) {
    sums[k] <- sum(x[,k],na.rm=T)
  }
  ranking <- rev(rank(sums, ties.method="min"))
  GS <- cbind(indiv,ranking)
  l <- L <- length(x[1,]); H <- length(x[,1]) # calculate the GC
  vp <- c(); err <- c()
  for (i in 1:H) {
    I <- c(); O <- c()
    for (j in 1:l) {
      if ( is.na(x[i,j])==F ) {
        I[j] <- length(x[i,1:j][which(x[i,1:j])==0])
        O[j] <- length(x[i,j:L][which(x[i,j:L]==1)])
        if ( x[i,j]==1 ) {O[j] <- O[j]-1}
      }
    }
    IO <- I + O
    l <- vp[i] <- max(which(IO==min(IO,na.rm=T)))
    err[i] <- IO[vp[i]]
  }
}
```

```
GC <- 100-round(100*sum(err)/(length(x)-length(which(is.na(x))),2)
write.table(GC,file="GSout.txt",quote=F,sep="\t",row.names=F) # output results
return(GC)
}
```

## Appendix 2: Assignment of language-specific alternations to different categories

### *Causative*<sup>R</sup>

Description: An A-adding alternation usually, but not always involving a morphological causative. In this restrictive selection only one alternation per language – the more common or more proto-typical one – has been chosen.

Inventory: Arabic Stem II Causative, Balinese Causative with Theme causee, Bora Causative derivation, Chintang Causative, Eastern Armenian Causative, Hoçak Coercive/default causative (+*hii*), Italian Causative, Japanese Hokkaido Causative, Japanese Mitsukaido causative, Ket Causative, Mandinka Causative Derivation 1, Mapudungun Causative 1, Yaqui Causative, Yucatec Maya Causative, Zenzontepec Chatino Causative of active verb.

### *Passive*

Description: An A (or S) argument of a corresponding active construction is left unexpressed or is demoted (and optional), but is semantically implied. The alternation is most commonly verb-marked (by a voice marker), but not necessarily so (cf. Mandinka and Mandarin Chinese).

Inventory: Arabic Stem VII Passive, Balinese Passive -*a* alternation, German Passive with *werden*, Icelandic Nominative Passive, Japanese Hokkaido DO passive, Japanese Mitsukaido Direct passive, Mandarin BEI Alternation, Mandinka Active/Passive Alternation, Russian Participial Passive, Yaqui Passive, Yucatec Maya Passive.

### *Antipassive*

Description: A P argument of a corresponding active transitive clause is left unexpressed or is demoted (and optional). The construction may be marked (antipassive voice) or unmarked.

Inventory: Ainu Antipassive, Arabic Object Omission Alternation, Bezhta Antipassive 1 Alternation, Eastern Armenian Object Omission, Even Object deletion, German Object Omission Alternation, Italian Object Omission, Mandinka Antipassive Middle, Russian Object deletion, Zenzontepec Chatino Object omission alternation.

*Reciprocal*<sup>R</sup>

Description: Includes constructions with a subject which is cross-coreferential with an object argument. In the case of transitive verbs this involves detransitivization. Pronominal reciprocals (involving a reciprocal pronoun) do not count as an argument alternation, since they do not affect verbal valency. The alternation can be either marked or unmarked (as in English *he met the friend* vs. *they met*).

Inventory: Arabic Stem VI Reciprocal, Bora Reciprocal derivation, Chintang Reciprocal, Eastern Armenian Reciprocal, Even Direct Reciprocal, Hoocak Reciprocal (+*kiki*), Icelandic Reciprocal Alternation, Italian Direct Reciprocal Reflexive, Ket Reflexive/Reciprocal Alternation, Russian Reflexive Reciprocal.

*Reflexive*<sup>R</sup>

Description: The constructions with a subject which is coreferential with an object argument. In the case of transitive verbs this involves detransitivization. Pronominal reflexives (involving a reflexive pronoun) do not count as an argument alternation, since they do not affect verbal valency.

Inventory: Ainu Reflexive, Arabic Stem V Reflexive, Bora Reflexive derivation, Chintang Reflexive, Even Reflexive deleting alternation, Hoocak Reflexive (+*kii*), Italian Direct Reflexive, Ket Reflexive/Reciprocal Alternation, Mapudungun Reflexive, Russian Semantic reflexive, Yucatec Maya Reflexive.

*Causative*

Description: An A-adding alternation (usually involving a morphological causative).

Inventory: Arabic Stem II Causative, Arabic Stem IV Causative, Arabic Stem X Causative, Bezhta Causative with *-k'*, Bezhta Causative with *golal*, Bezhta Causative with *yowal*, Bora Causative derivation, Chintang Causative, Eastern Armenian Causative, Evenki Adversative Passive, Hoocak Coercive/default causative (*hii*), Hoocak Permissive causative (*gigi*), Hoocak Possessive reflexive causative (*karagi*), Hoocak Reflexive causative (*kij*), Italian Causative, Japanese (standard) Argument-Increasing Alternation, Japanese Hokkaido Causative, Japanese Hokkaido Lexical argument increasing alternation, Japanese Mitsukaido causative, Japanese Mitsukaido lexical argument increasing alternation, Ket Causative, Mandinka Causative Derivation 1, Mandinka Causative Derivation 2, Mapudungun Causative 1, Mapudungun Causative 2, N||ng Coded causative, N||ng Coded causative + addition of oblique, Ojibwe Causative Simple, Sliammon Active-intransitive-causative, Sliammon Causative, Sliammon Causative with Active-intransitive, Sliammon Control transitive, Sliammon Noncontrol transitive, Xârâcûù Causative alternation, Yaqui Causative, Yucatec Maya Causative, Zenzontepec Chatino Causative of active verb, Zenzontepec Chatino u-Causative alternation.

*Object-demoting/deleting*

Description: A P argument is deleted or demoted. The alternation may be verb-marked (by an antipassive marker) or unmarked, involving lexically restricted P-omission or P-incorporation.

Inventory: Ainu Antipassive, Ainu Noun incorporation, Arabic Goal Alternation – Prepositional (*'ilâ+* ‘to’)/Accusative Object, Arabic Object Omission Alternation, Arabic Prepositional (*bi+*)/Accusative Object Alternation, Arabic Prepositional (*min* ‘from’)/Accusative Object Alternation, Arabic Theme/Source Alternation – Prepositional (*min+* ‘from’) Accusative Object, Bezhta Antipassive 1 Alternation, Bezhta Antipassive 2 Alternation, Bezhta Object Incorporation Alternation, Chintang S/A detransitivisation, Eastern Armenian Internal Object, Eastern Armenian Object Omission, Emai Object omission, English Conative, English Understood Omitted Object, Even Object deletion, German Object Omission Alternation, Hoocak Detransitive/slot filler (*wa-*), Italian Conative, Icelandic Obj.dat vs. Obj.acc, Italian Object Omission, Jakarta Indonesian Prenasalized Ditransitive, Jakarta Indonesian Prenasalized Intransitive, Jakarta Indonesian Prenasalized Transitive, Jaminjung Transitivity alternation of activity predicates, Jaminjung Transitivity alternation of motion predicates, Jaminjung Transitivity alternation of stative predicates, Ket Incorporation, Mandinka Object/Oblique Permutation, Mandinka Oblique/Object Alternation, N||ng Deagentive, N||ng Demotion of direct object to oblique, N||ng Promotion of direct object to oblique, N||ng Promotion of oblique to indirect object, N||ng Uncoded causative, Ojibwe Antipassive *-ige*, Ojibwe Antipassive *-iwe*, Ojibwe Body Part Incorporation, Russian Accusative-Genitive alternation, Russian Cognate Object alternation, Russian Negative Accusative-Genitive alternation, Russian Object deletion, Sliammon Active-intransitive, Xârâcùù Conative alternation, Xârâcùù Object Omission alternation, Yaqui Equipollent Applicative, Yaqui Intransitivizer incorporation: *ji'i*, Yaqui Locative, Yaqui Object incorporation, Yaqui Undetermined object, Yoruba Preposition-dropping, Yoruba Theme-dropping alternation, Yucatec Maya Incorporative, Yucatec Maya Introversive, Zenzontepec Chatino Object omission alternation, Zenzontepec Chatino Object/instrument incorporation.

*Reflexive<sup>I</sup>*

Description: See the description of reflexive<sup>R</sup> above. The reflexive<sup>I</sup> selection differs only by allowing for more than one alternation per language.

Inventory: Ainu Reflexive, Arabic Stem V Reflexive, Arabic Stem VI Reflexive, Arabic Stem VIII Reflexive, Bora Reflexive derivation, Chintang Reflexive, English Understood Reflexive Object, Even Reflexive deleting alternation, Hoocak Reflexive *kii*, Icelandic Reflexive (mediopassive form), Italian Direct Reciprocal Reflexive, Italian Direct Reflexive, Italian Indirect/Dative Reflexive, Jaminjung Reflexive-reciprocal alternation of IVs, Mapudungun Reflexive, Ojibwe Reflexive, Russian Re-



flexive Middle (impersonal), Russian Semantic reflexive, Sliammon Causative with Reflexive, Xârâcùù Reflexive/Reciprocal omission alternation, Yucatec Maya Reflexive.

### *Subject-demoting/deleting*

Description: The A or S-argument is deleted or demoted (it may or may not be implied). The alternation covers different kinds of passive, anticausative, resultative and ‘middle’ constructions. The construction may be either marked or unmarked: the S=P labile verbs also belong to this type.

Inventory: Ainu Anticausative, Arabic Stem V Anticausative, Arabic Stem VI Passive, Arabic Stem VII Anticausative, Arabic Stem VII Passive, Arabic Stem VIII Anticausative, Balinese Passive *-a* alternation, Balinese Passive *ka-* alternation, Balinese Resultative *ma-* alternation, Bezhta Accidental/Potential Transitive Alternation, Bezhta Ambitransitive Alternation, Bezhta Causative with *-l/-ll*, Bezhta Potential Transitivity Alternation, Chintang Passive participle, Chintang S/P ambitransitivity, Eastern Armenian Mediopassive, English Causative-Inchoative, English Inchoative-Causative, English Middle, Even Adversative Passive, Even Mediopassive, Even Resultative, Evenki Decausative, Evenki Personal passive, German Anticausative Alternation, German Instrument Subject Alternation, German Passive with *werden*, Hoocək Facilitative alternation, Hoocək Resultative alternation, Icelandic Accusative Subject Ambitransitive vs. Transitive, Icelandic Ambitransitive, Icelandic Ambitransitive alt. with *-ja*, Icelandic Ambitransitive alternation with umlaut and *-na*, Icelandic Caused-motion from intransitive, Icelandic Dat-Acc Ambiditransitive, Icelandic Dative passive, Icelandic Genitive Passive, Icelandic Genitive Subject Ambitransitive vs. Transitive, Icelandic Impersonal mediopassive, Icelandic Impersonal passive, Icelandic Mediopassive (middle), Icelandic Middle, Icelandic Nominative Passive, Icelandic Transitive from intransitive, Italian Anticausative (coded), Italian Anticausative (uncoded), Italian Impersonal of Reflexives, Italian Impersonal Passive, Italian Impersonal Reflexive, Italian Middle Reflexive, Italian Oblique Subject, Italian Passive, Italian Reflexive Passive, Jakarta Indonesian Accidental Passive, Jakarta Indonesian Adversative Passive, Jakarta Indonesian Impersonal Passive, Jakarta Indonesian Oblique Subject Passive Agentive by-phrase, Jakarta Indonesian Oblique subject Passive Applicative Ditransitive (A P), Jakarta Indonesian Oblique subject Passive Applicative Ditransitive Agentive by-phrase, Jakarta Indonesian Oblique Subject Passive Ditransitive, Jakarta Indonesian Passive Transitive, Jakarta Indonesian Patient Subject Passive Ditransitive Applicative Agentive by-phrase., Jakarta Indonesian Thematic Adversative Passive, Jaminjung Ambitransitive alternation, Jaminjung Causative alternation of internally caused change of state predicates and non-locative states, Jaminjung Causative-inchoative alternation of “cooking” predicates, Jaminjung Causative-inchoative alternation of change of location predicates, Jaminjung Causative-inchoative alterna-

tion of externally caused change of state predicates, Jaminjung Causative-inchoative-stative alternation of predicates of location, Japanese (standard) Argument-Decreasing Alternation, Japanese (standard) Direct Passivization Promoting Ni-marked Argument, Japanese (standard) Direct Passivization Promoting O-marked argument, Japanese (standard) Potential Formation, Japanese Hokkaido Anticausative with /rasar/, Japanese Hokkaido DO passive, Japanese Hokkaido Lexical argument decreasing alternation, Japanese Mitsukaido lexical argument decreasing alternation, Japanese Mitsukaido potential, Ket Anti-Causative Alternation, Ket Resultative, Mandarin BA Alternation, Mandarin Neutral Ditransitive Alternation, Mandinka Causative/Noncausative Alternation, Mandinka Active/Passive Alternation, Mandinka Antipassive Middle, Mandinka Intransitive/Middle Alternation, Mandinka Subject/Oblique Alternation, N||ng Addition of direct object, N||ng Deletion of direct object, Ojibwe Lexical passive, position and configuration, Russian Participial Passive, Russian Reflexive Anticausative, Russian Reflexive Middle, Russian Reflexive Passive, Sliammon Indirective applicative with Passive, Sliammon Lexical suffix with Middle, Sliammon Middle, Sliammon Middle with Indirective applicative, Sliammon Middle with Relational applicative, Sliammon Passive, Sliammon Relational applicative with Passive, Xârâcùù Ambitransitive Alternation 1, Xârâcùù Applicative alternation, Xârâcùù Impersonal alternation, Xârâcùù Middle alternation, Yaqui Equipollent Causal, Yaqui Equipollent Noncausal, Yaqui Middle, Yaqui Passive, Yaqui Resultative, Yoruba Inchoative-causative, Yucatec Maya Deagentive, Yucatec Maya Passive, Zenzontepec Chatino Active/inactive alternation, Zenzontepec Chatino Ambitransitive alternation, Zenzontepec Chatino Equipollent causative/intransitive alternation, Zenzontepec Chatino Intransitive alternation

## Acknowledgements

This paper has profited from many suggestions by Andrej Malchukov, including suggestions regarding classes of alternations to be analyzed. It has also benefited from comments by Bernard Comrie on previous versions. Roger Mundry and Damian Blasi provided helpful discussion of statistical issues (and are not to be blamed for any shortcomings). Johanna Nichols provided very useful comments on an oral presentation. Last but not least, I am indebted to Eric W. Holman for originally having drawn my attention to Guttman scaling as well as for comments on the paper.

## References

- Berlin, Brent & Paul Kay. 1969. *Basic Color Terms: Their Universality and Evolution*. Berkeley: University of California Press.

- Corbett, Greville G. 2010. Implicational hierarchies. In Jae Jung Song (ed.), *The Oxford Handbook of Linguistics Typology*, 190–205. Oxford: Oxford University Press.
- Cysouw, Michael. 2003. Against implicational universals. *Linguistic Typology* 7. 89–101.
- 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*, 40–70. Cambridge, Mass.: The MIT Press.
- Guttman, Louis. 1944. A basis for scaling qualitative data. *American Sociological Review* 9. 139–150.
- Hartmann, Iren, Martin Haspelmath & Bradley Taylor (eds.). 2013. *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://valpal.info>, Accessed on 2014-08-15).
- Holland, Barbara R., Katharina T. Huber, Andreas Dress & Vincent Moulton. 2002.  $\delta$  plots: A tool for analyzing phylogenetic distance data. *Molecular Biology and Evolution* 19. 2051–2059.
- Huson, Daniel & David Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23. 254–267.
- Janssen, Dirk P., Balthasar Bickel & Fernando Zúñiga. 2006. Randomization tests in language typology. *Linguistic Typology* 10. 419–455.
- Keenan, Edward L. & Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8. 63–99.
- North, B. V., D. Curtis & P. C. Sham. 2002. A note on the calculation of empirical *P* values from Monte Carlo procedures. *American Journal of Human Genetics* 71. 439–441.
- Oksanen, Jari, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens & Helene Wagner. 2011. *vegan: Community ecology package*. R package version 2.0–2. <http://CRAN.R-project.org/package=vegan>
- R Development Core Team. 2008. *R: A Language and Environment for Statistical Computing. Version 2.13.1. R Foundation for Statistical Computing*. Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Silverstein, Michael. 1976. Hierarchy of features and ergativity. In Robert M. W. Dixon (ed.), *Grammatical Categories in Australian Languages*, 112–171. Canberra: Australian Institute of Aboriginal Studies.
- Tsunoda, Tasaku. 1985. Remarks on transitivity. *Journal of Linguistics* 21. 385–396.
- Wichmann, Søren. Forthc. Quantitative tests of implicational verb hierarchies. In Taro Kageyama & Wesley M. Jacobsen (eds.), *Transitivity and Valency Alternations: Studies on Japanese and Beyond*. Berlin: De Gruyter Mouton.
- Wichmann, Søren, Eric W. Holman, Taraka Rama & Robert S. Walker. 2011. Correlates of reticulation in linguistic phylogenies. *Language Dynamics and Change* 1. 205–240.

