Søren Wichmann

14 Quantitative tests of implicational verb hierarchies

1 Introduction

This chapter will begin by discussing the implicational verb hierarchy of Tsunoda (1985) as a convenient starting point for looking at what happens when a relatively large dataset and a principled, quantitative approach to their analysis are brought to bear on a linguistic typological hypothesis. After introducing new methods for assessing the validity of an implicational hierarchy, I go on to inquire into the presence of implicational hierarchies governing the distribution of 5 different alternation types across 87 verb meanings and 22 languages (Ainu, Balinese, Bezhta, Bora, Chintang, Eastern Armenian, Even, German, Hokkaido Japanese, Hoocąk, Icelandic, Italian, Ket, Mandarin Chinese [henceforth 'Mandarin'], Mandinka, Mapudungun, Mitsukaido Japanese, Modern Standard Arabic [henceforth 'Arabic'], Russian, Yaqui, Yucatec Maya, and Zenzontepec Chatino).[1]

The data used are from the database of the Leipzig Valency Classes Project (Hartmann et al. 2013) in the state it was in as of July 17, 2012, although the names used to designate different alternations have been updated. Contributors were asked to supply information about the presence or absence of different alternations for a fixed set of 87 verb meanings specified through a set of 'meaning labels', e.g. EAT, a 'role frame', e.g. 'A eats P', and 'typical contexts', e.g. 'the boy ate the fruit'. The alternations vary from highly language-specific ones, such as the '*be*-alternation' in German, to alternations that are more comparable across several languages, such as the 'passive' in Yaqui and Yucatec Maya, to

1 An earlier analysis included 7 additional alternations (ambitransitive, anticausative, applicative, impersonal passive, locative, mediopassive, and resultative), but since comparable instances of all of these are represented by few languages (7 or less) and since, moreover, some of the results for these alternations were problematic in certain respects, possibly due to the scarcity of data, I chose to only include alternations for which 11 or more languages were attested. I would like to draw the reader's attention to Wichmann (2015), which is a sequel to the present paper, even though it was published earlier.

Søren Wichmann, Leiden University Centre for Linguistics and Kazan Federal University

which can be added alternations that are given language-specific designations by contributors but can still be considered instances of the general category of passive, such as the Mandarin ‘BEI alternation’, the Balinese ‘passive *-a* alternation’, the German ‘passive with *werden*’, the Russian ‘participial passive’, etc. (see Appendix 1 for a mapping between alternations as named in this study and designations in individual languages; readers are strongly encouraged to also consult the online database of Hartmann et al. 2013 for more information about individual alternations, should questions about these arise).

Christian Lehmann has expressed skepticism both about the viability of larger studies of implicational hierarchies among verbs and also about outcomes showing neat results:

> An empirically-based survey, no matter whether of predicate meanings or of situations functioning in linguistic structure, presupposes comprehensive research into the whole verbal and adjectival vocabulary. This has occasionally been tried for one language. It seems plainly impossible to do such research in depth on a cross-linguistic scale (...). Moreover, it should be clear from the outset that this kind of research cannot be expected to yield clear-cut cross-linguistic generalizations, to come up with regularities structuring the grammars of all languages. (Lehmann 1991: 187)

One of the aims of the Leipzig Valency Classes Project is precisely to overcome the practical problem of labor-intensity by distributing work over many contributors. It has not aimed to achieve the kind of in-depth coverage of individual languages found in works such as Ballmer & Brennenstuhl (1986) on German or Levin (1993) on English, for instance, but the coverage within and across languages is good enough to enable a better assessment of the degree to which we can expect to find regularities and make generalizations.

The availability of relatively abundant and systematic data invites the application of quantitative methods. The methods will be introduced by way of studying a familiar example, the hierarchy of verb meanings of Tsunoda (1985), in the next section.

2 Tsunoda’s hierarchy

In his well-known 1985 paper, following earlier work from 1981, Tasaku Tsunoda introduced the transitivity hierarchy of verb types displayed in (1).

(1) Direct Effect > Perception > Pursuit > Knowledge > Feeling > Relationship > Ability

Two-place predicates farther to the left in the hierarchy are more prone to take transitive case marking (ergative-absolutive or nominative-accusative, depending on the language type), and as one moves towards the right other types of case marking increasingly appear, with ergative-absolutive or nominative-accusative being completely absent for the Ability category. Moreover, Tsunoda (1985: 391) predicts that the following four types of construction, often regarded as morphosyntactic correlates of transitivity, will apply increasingly less often as one moves towards the right in the hierarchy: passive, antipassive, reflexive, reciprocal. This latter prediction is of special interest here since it can be tested through the data gathered under the auspices of the Leipzig Valency Classes Project. I wish to make clear from the outset that the evidence for Tsunoda's hierarchy which comes from case marking patterns is not addressed here, only the evidence that comes from alternations.

Not all verb meanings listed by Tsunoda are covered in the database, but the following representatives of the five leftmost types are available (unfortunately none are available for Relationship and Ability). Direct Effect: KILL, BREAK, HIT, EAT; Perception: SEE, HEAR, LOOK AT; Pursuit: SEARCH FOR; Knowledge: KNOW; Feeling: LIKE, WANT, FEAR. The dataset available for the behavior of alternation types across 12 verbs from 22 languages is provided in Appendix 2.

An example of the prediction of the transitivity hierarchy would be that if the reflexive applies to the Pursuit verb SEARCH FOR, then it should also apply to the various Perception and Direct Effect verbs. Moreover, the hierarchy predicts that a verb pertaining to a given type should have a behavior more similar to that of the other members of its type than to verbs pertaining to other types. I will test both predictions, starting with the latter.

NeighborNet (Huson & Bryant 2006) is useful for clustering[2] and also for showing how tree-like the data is. If verbs are ordered in an implicational scale the structure should be highly tree-like, ideally with the verbs ordered on a string, as in Figure 1.

2 Clustering algorithms can either be character- or distance-based. Since we have characters at our disposal it may be argued that a character-based method is the best choice. However, since the distance-based NeighborNet algorithm is also useful for inspecting the tree-likeness of a dataset, I prefer to use it. From the characters the program computes Hamming distances and takes those as input to the algorithm. Since this potentially means a loss of information, I have also checked the results of a character-based method, as mentioned in note 4.

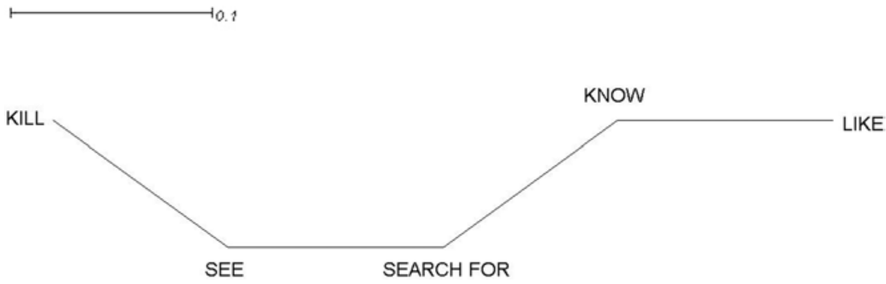

Figure 1: What a NeighborNet graph for five verbs showing perfect conformity with the Tsunoda hierarchy would look like.

Non-tree-like behavior is shown in NeighborNet by boxes. These boxes, along with deviation from the unidimensional case, where verbs are ordered in a single, particular direction, can 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 δ, which is a measure of the amount of reticulate behavior corresponding to the visual impression given by the boxes. δ 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 data). The results for the dataset pertaining to Tsunoda's hierarchy are shown in Figure 2.

Looking for clusters that confirm Tsunoda's types is disappointing. The Direct Effect type comprises KILL, BREAK, HIT, and EAT, but we find WANT interspersed among them in the lower part of the graph. SEE and LOOK AT are supposed to belong with HEAR in a Perception category, but the three verbs do not cluster, although they do appear in the same general region of the network.[3] In any case, they do not form a clean cluster, as is evident from the boxes connecting them, which indicate conflicts between different choices of how they could be clustered. LIKE and FEAR are supposed to group with WANT in a Verbs of Feeling category, but WANT strays far from the two others. Thus, there is not strong support to be found for the types proposed by Tsunoda, suggesting that typologists should be careful about making a priori assumptions about what constitutes a semantic class or 'type'.[4]

3 With respect to the visual interpretation of the network it is important to realize that two taxa can be relatively closely grouped even if they appear on opposite sides of the network. That is, the distance between taxa is not determined by the length of the path on the periphery of the network, but by the shortest path between them. For instance, LOOK AT and SEE are not a lot closer to one another than either is to SEARCH FOR although LOOK AT and SEE are neighbors along the periphery, whereas SEARCH FOR is on the other side of the network.
4 Using a character-based clustering algorithm such as Wagner Parsimony as implemented in the pars.exe program of Felsenstein (2009) does not improve these results.

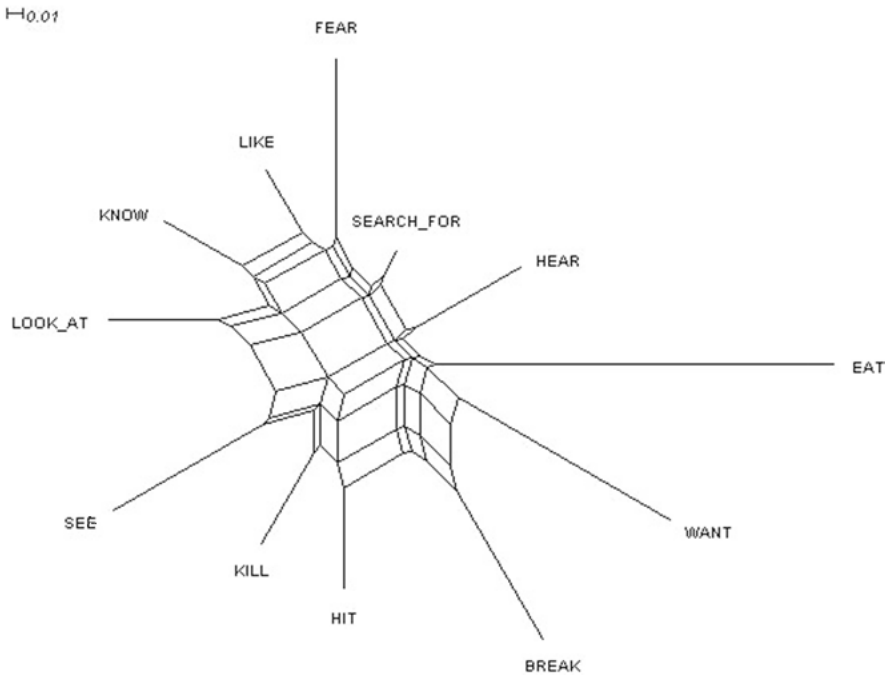

Figure 2: NeighborNet of 12 verbs argued by Tsunoda to form an implicational hierarchy

Having checked whether Tsunoda's types can be supported, we now go on to the perhaps more pressing issue of whether there is some support for an implicational hierarchy. The graph in Figure 2 can provide some visual leads in this regard, further enhanced by the δ-score. A δ-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). From the non-tree-like configuration and the δ-score, which is 0.45, we should not expect a perfect implicational hierarchy. On the other hand, there seems to be a tendency for the ordering of the verbs predicted by the Tsunoda hierarchy to recur in the structure of the graph. Thus, KILL, BREAK, HIT, and EAT do occur at one end of the graph, LOOK AT and SEARCH FOR and KNOW are in the middle region and LIKE, WANT, and FEAR are towards the other end. It is clearly worthwhile investigating how strong the evidence is for a hierarchy and what the best supported hierarchy would look like, so we need a method for this.

There is a method for measuring the one-dimensionality of a dataset – in other words, the degree to which it confirms to an implicational hierarchy –

which has been around for more than half a century although it has largely been ignored by language typologists,[5] namely the Guttman scale. The method is named after its inventor, who proposed it in Guttman (1944). In presenting it, I will follow 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: 'yes' or 'no') of an attribute (here: a certain alternation in a certain language) of a given individual (here: a certain verb) are first ordered in a 'scalogram', where individual 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 following the first is the one with the next-most frequent instances of the value, and so forth. Consequently, the number of deviations from the pattern of a perfect scale can be counted. Let me first illustrate in (3) what a scalogram corresponding to the perfect implicational scale in Figure 1 would look like.

```
(3) LIKE        ynnnn
    KNOW        yynnn
    SEARCH_FOR  yyynn
    SEE         yyyyn
    KILL        yyyyy
```

Example (4) reproduces (3), but with the introduction of two changes, in the rows for LIKE and SEE.

```
(4) LIKE        ynnyn
    KNOW        yynnn
    SEARCH_FOR  yyynn
    SEE         ynyyn
    KILL        yyyyy
```

The changes I introduced in (4) illustrate two cases of deviation from perfect scalarity. The 'yes' value in the fourth column of LIKE is an error since it is found within a sequence of 'no's; the 'no' value in the second column of SEE is also an

5 Sasse (2001) is the only work which I am aware of that applies Guttman scaling to any larger extent, but it is limited to showing how a matrix structure can reveal an implicational scale and does not make use of the method's ability to numerically measure the deviation from scalarity. Croft & Poole (2008:7) mention Guttman scaling, but only in passing.

error since it is included among 'yes' values, where it does not belong. Errors involving a 'yes' value in the wrong place and a 'no' value in the wrong place count as having equal weight. In general, both for Guttman's method and for the construction of NeighborNets the availability of a construction carries the same weight as non-availability.

The Guttman Coefficient (GC) is a measure of scalarity. It is calculated in a simple manner, by subtracting the total number of errors from the total number of datapoints, T, then dividing by T, and finally expressing the result as a percentage. In (4), there are 2 errors among the 25 datapoints, so GC = (25 – 2)/ 25 = 92%. 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). Thus, if we find that the Guttman Coefficient is 85% or greater we can probably regard this as good evidence for an implicational hierarchy. Nevertheless, in its application to linguistic data the Guttman scale needs to be tested more before we can place much confidence in such estimates. The following is therefore a somewhat tentative exploration of its application.

Having illustrated the Guttman scale by means of a toy example, we now go on to use it to test the Tsunoda hierarchy. To this end, the data in Appendix 2 are reordered such that the 'yes's are concentrated in the lower left corner and the 'no's in the upper right corner.[6] The arrangement producing the highest Guttman Coefficient implies the scale shown in (5). The subscript numbers indicate where the items fall in Tsunoda's hierarchy, a subscript 1 corresponding to the group of highest transitivity.

(5) Results of the test for Tsunoda's hierarchy
SEE_2 > KILL_1 > HIT_1 > LOOK AT_2 > KNOW_4 > EAT_1 > HEAR_2 > BREAK_1 > SEARCH FOR_3 > LIKE_5 > FEAR_5 > WANT_5

In general, the way that a hierarchy such as the one in (5) is read is that a higher position on the scale indicates regular participation of a particular verb in alternations across languages, while a low position indicates that the verb participates in few or no alternations across languages.

6 I am not aware of non-commercial software implementations of Guttman's method, but it can be carried out relatively simply by using sorting and counting functions in spreadsheet software. Here and elsewhere missing values are not counted as errors, but they are also not counted in the denominator of the Guttman Coefficient formula.

Unfortunately for Tsunoda's hypothesis, there are discrepancies between his hierarchy and the one in (5), with several displacements. Direct Effect and Perception type verbs are mixed among each other, and Knowledge is also displaced, whereas Pursuit and Feeling verbs behave as they should. The Guttman Coefficient is 85.6%, just enough to justify calling (5) an implicational hierarchy. This is not the place to inquire into possible explanations for the failure of the data to conform neatly to Tsunoda's hierarchy, and much less to propose or analyze alternatives.[7] The point of the exercise is to illustrate how quantitative methods such as graphic networks, δ scores, and Guttman scaling can be used to inquire into the existence of implicational hierarchies among verbs. This is meant to set the stage for the next section, where 5 different types of morphosyntactic alternation are investigated for the purpose of uncovering underlying hierarchies among the 87 verbs sampled in the Leipzig Valency Classes Project.

3 Implicational hierarchies among verbs across languages for different alternations

This section presents basic empirical findings on each of 5 different alternations. For each, the specific alternation is briefly introduced, and then a NeighborNet, a Guttman Coefficient, a δ score, and a hierarchy emerging from scalogram analysis (Guttman scaling) are provided. The hierarchies are produced by observing for each verb the length of the part of the row containing 'yes's. The procedure roughly amounts to counting the number of languages in which a given verb enters a given alternation, but since the scalogram analyses correct for missing values by treating them as being in conformity with the overall configuration, a simple count of languages per verb does not always yield the same results as the scalogram analyses. Very often several verbs have the same range of application in different languages for a given construction, in which case they are merged into clusters. The 'greater than' symbol (>) separates clusters, and within clusters verbs are listed in alphabetical order separated by commas.

NeighborNets, Guttman Coefficients, δ scores, and verb hierarchies are not further introduced in the subsections below, but at the end of each subsection a few brief comments on the findings are provided.

7 Malchukov (2005) proposes a semantic map, i.e. a whole network of implicational relations, as an alternative to Tsunoda's hierarchy.

3.1 Antipassive

In the constructions gathered under the label of antipassive, objects are omitted or demoted. Complete omission is required in Ainu, Arabic, German, Italian, Ket, Mandarin, Zenzontepec Chatino, Even, and Russian, whereas Bezhta, Eastern Armenian, and Mandinka allow for the expression of the P (patient) marked as an oblique. In some languages the alternation incurs changes in the meaning of the predicate. Thus, in the Bezhta antipassive the predicate acquires a durative sense, in Eastern Armenian the focus shifts to the state or activity of the agent, in German the predicate acquires a more generic sense, etc.

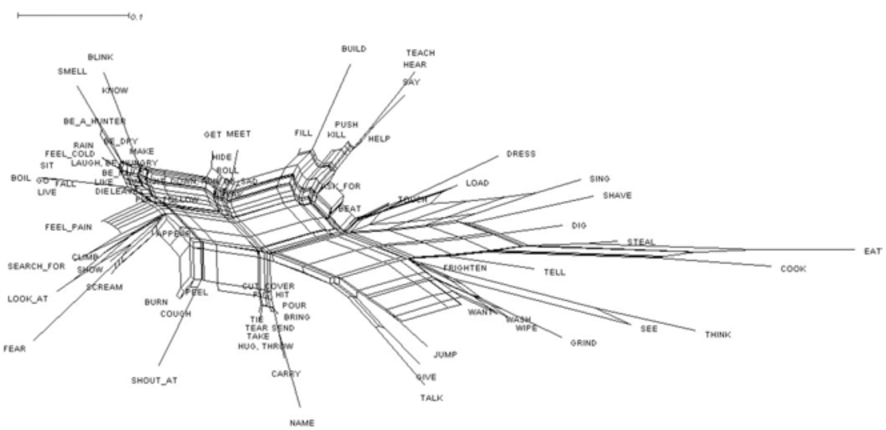

Figure 3: NeighborNet for the antipassive alternation

Guttman Coefficient: 86.0%

δ: 0.354

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

Comments: The NeighborNet here suggests that one-dimensionality is not a very good approximate description for this alternation. On the other hand, it would probably take many dimensions to get a better fit. Increasing the parameter space drastically is not necessarily preferable, so one-dimensionality may still be a preferred description, and it finds mild support from the Guttman Coefficient, which is just above the conventional 85% cut-off point.

3.2 Causative

Most of the languages (Bora, Chintang, Eastern Armenian, Hokkaido Japanese, Ket, Mitsukaido Japanese, Yaqui, Yucatec Maya) add a suffixed element to the verb to derive a causative, whereas Arabic changes the stem shape from CaCaCa / CaCiCa / CaCuCa (stem I) to CaCCaCa (stem II) and Italian uses a periphrastic construction with either *fare* 'make' (factitive/coercive) or *lasciare* 'let' (permissive). The alternation introduces a causer and the S or A of the corresponding active construction becomes a P and is generally marked as

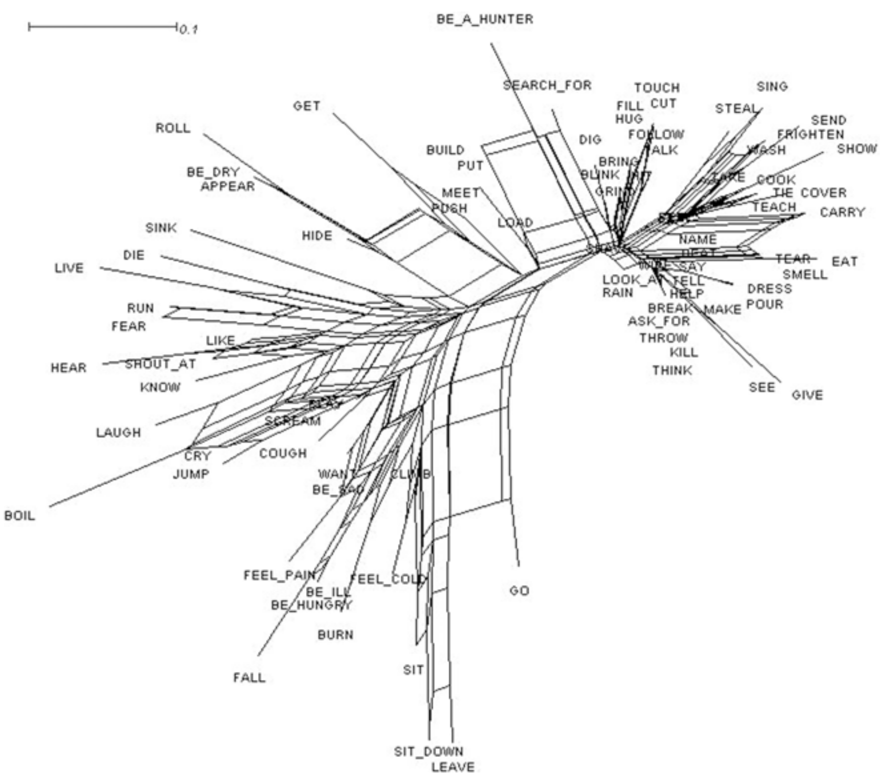

Figure 4: NeighborNet for the causative alternation

such in the languages in which grammatical relations are expressed morphologically through indexing (Chintang, Ket) or case (Bora, Hokkaido and Mitsukaido Japanese[8]). In some cases the causative introduces additional semantic effects. Thus, in Arabic there is sometimes an added intensive meaning, and in Ket typically an added inceptive meaning.

Guttman Coefficient: 85.4%
δ: 0.379

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

Comments: The NeighborNet suggests a relatively sharp demarcation between clusters attracted by opposite poles at the left end of the network where verbs are typically semantically intransitive and the right end where they are typically semantically transitive.

3.3 Passive

Common to the alternations considered here is the promotion of the object in a transitive clause to subject in an intransitive clause. In several languages it is possible to optionally express the agent. In Balinese and German the agent is introduced in a prepositional phrase; in Yucatec by a preposition-like relational noun; in Hokkaido Japanese, Mitsukaido Japanese, and Russian it is marked by a non-core case. Mandinka does not allow for the expression of the agent. The Mandarin construction is accompanied by a sense of adversity, and appears to be the only one introducing a semantic affect. Some languages (Arabic, Icelandic, Russian) have more than one passive construction. In these cases the one that seems to be more frequent and/or general is selected here.

8 In Mitsukaido Japanese case-marking of the causee is dative when the corresponding non-causative sentence is transitive and accusative when the corresponding non-causative sentence is intransitive. This also seems to be the general pattern for Hokkaido Japanese.

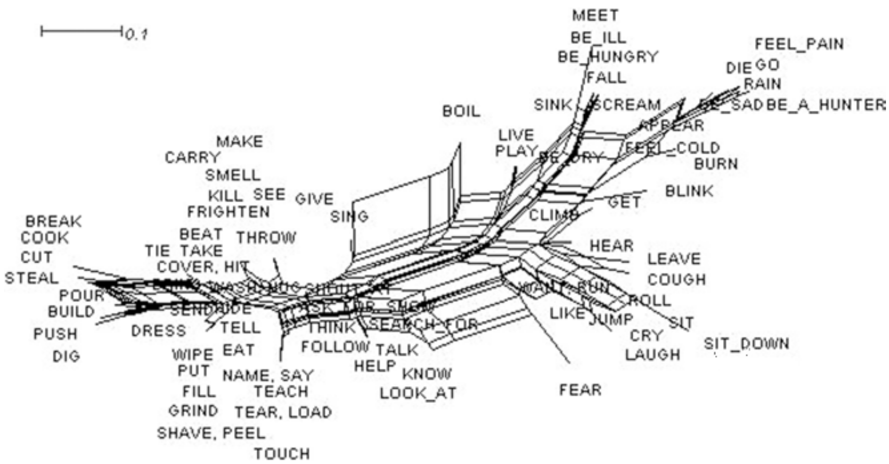

Figure 5: NeighborNet for the passive alternation

Guttman Coefficient: 86.6%
δ: 0.309

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

Comments: Here is another case where unidimensionality is suggested both by the NeighborNet and the Guttman Coefficient, albeit weakly.

3.4 Reciprocal

In all the languages the reciprocal is typically used in a situation where two or more participants are at the same time agents and patients of the action.

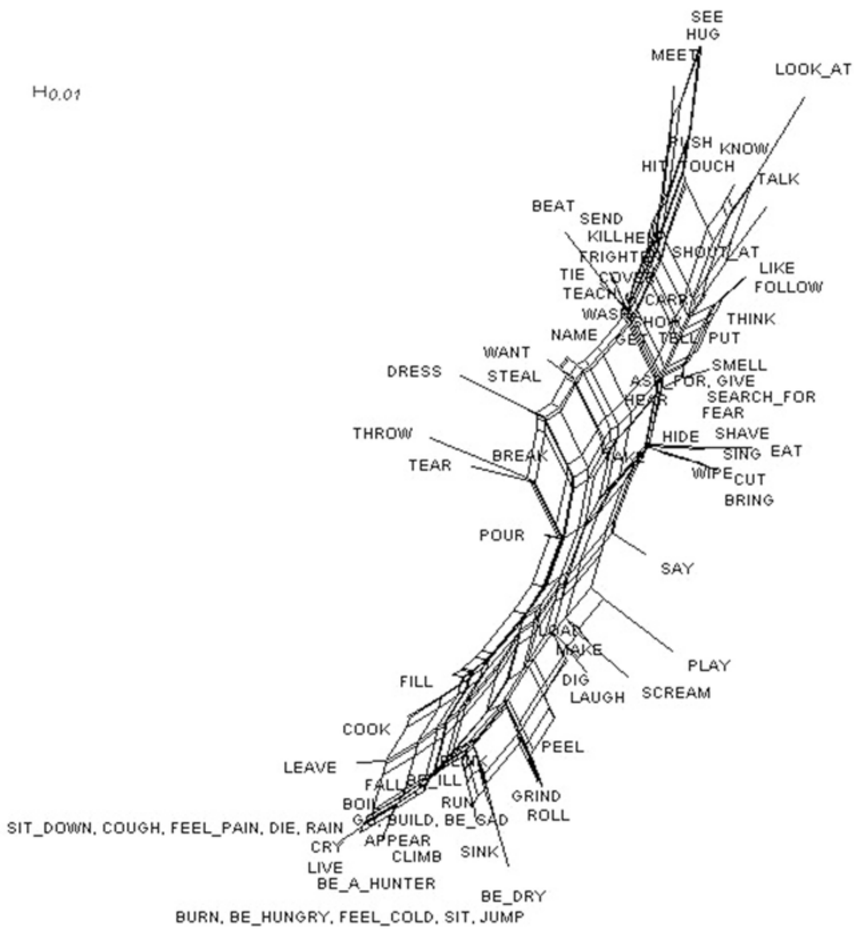

Figure 6: NeighborNet for the reciprocal alternation

Guttman Coefficient: 88.8%
δ: 0.285

Hierarchy: HUG, SEE > HIT, KNOW, LOOK AT > LIKE, MEET, TOUCH > HELP > BEAT, FRIGHTEN, WASH > ASK FOR, CARRY, COVER, DRESS, FEAR, FOLLOW, GIVE, KILL, NAME, PUSH, SEARCH FOR, SHAVE, SHOUT AT, SHOW, SMELL, TALK, TELL, TIE > BREAK, BRING, CUT, EAT, HEAR, HIDE, POUR, PUT, SAY, SEND, SING, STEAL, TAKE, TEACH, TEAR, THINK, THROW, WIPE > GET, PLAY, SCREAM > LAUGH, LOAD, WANT > BE DRY, BLINK, DIG, FILL, GRIND, MAKE, PEEL, ROLL > BE A

HUNTER, BE SAD, BUILD, CLIMB, GO, LEAVE, RUN, SINK > APPEAR, BE HUNGRY, BE ILL, BOIL, BURN, COOK, COUGH, CRY, DIE, FALL, FEEL COLD, FEEL PAIN, JUMP, LIVE, RAIN, SIT, SIT DOWN

Comments: Both the NeighborNet and the Guttman Coefficient suggest unidimensionality.

3.5 Reflexive

Across languages the reflexive is used to express a situation where the agent does something to him- or herself. For some languages (Ainu, Arabic, Hoocąk, Even) it is sometimes described as a valency-decreasing operation, whereas for some other languages, for which a reflexive pronoun is available (German, Yucatec Maya, Italian), it is explicitly or implicitly regarded as a subtype of a transitive construction. For some languages (German, Ket, Mapudungun) the construction overlaps with the reciprocal when there are multiple participants.

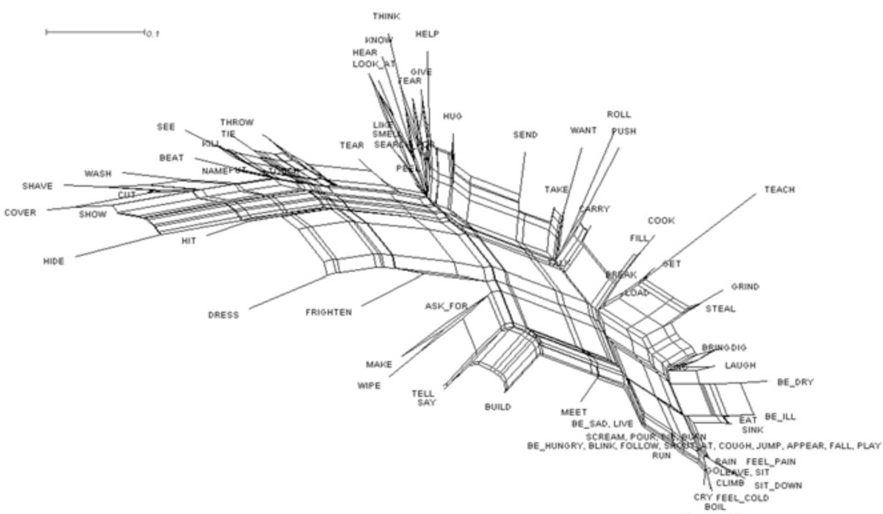

Figure 7: NeighborNet for the reflexive alternation

Guttman Coefficient: 84.7%
δ: 0.324

Hierarchy: COVER, SHAVE > HIDE, SEE, SHOW > CUT, WASH > BEAT, DRESS, FEAR, GIVE, HELP, HIT, KILL, KNOW, LOOK AT, NAME, PUT, THINK, THROW, TIE, TOUCH > HEAR, HUG, SMELL > ROLL, TEAR > TEACH, WIPE > FRIGHTEN > ASK FOR, BUILD, LIKE, PUSH, SAY, SEARCH FOR, SEND, TELL > BREAK, CARRY, FILL, LOAD, MAKE, MEET, PEEL, TAKE, TALK > BE DRY, COOK, GRIND, LAUGH, STEAL,

WANT > APPEAR, BE A HUNTER, BE HUNGRY, BE ILL, BE SAD, BLINK, BOIL, BRING, BURN, CLIMB, COUGH, CRY, DIE, DIG, EAT, FALL, FEEL COLD, FEEL PAIN, FOLLOW, GET, GO, JUMP, LEAVE, LIVE, PLAY, POUR, RAIN, RUN, SCREAM, SHOUT AT, SING, SINK, SIT, SIT DOWN

Comments: The Guttman Coefficient is just short of the threshold for unidimensionality, but the overall shape of the NeighborNet indicates that this is nevertheless a good approximate description.

3.6 Concluding comments on individual alternations

Overall the evidence presented shows that the assumption of single implicational hierarchies (unidimensionality) underlying the distribution of syntactic alternations across verbs and languages is supported in the majority, if not in all cases. We have observed relatively tree-like NeighborNets with relatively low δ-scores and acceptable Guttman coefficients. Interestingly, when we correlate the δ-scores and Guttman coefficients we find a high negative correlation of Pearson's $r = -0.707$ ($p = 0.182$, which is too high for significance, but p is not expected to reach significance with just five datapoints).[9] This justifies the use of NeighborNets as part of the toolkit for estimating unidimensionality.

The antipassive behaves somewhat multidimensionally but is just on the right side of Guttman's threshold for unidimensionality. Each alternation is worthy of more detailed study, leaving plenty of opportunities for future research. Here, however, we are interested in the larger picture. This has shown that effects of implicational hierarchies among verbs are ubiquitous. For all alternations there is some justification for assuming single underlying hierarchies, even if the justification is better in some cases than in others. This raises the issue of whether the hierarchies are the same or different across alternations. This is the topic of the next section.

4 Correlations among verb hierarchies

The previous section presented hierarchies among the verbs for each alternation. These were produced from scalogram analyses, which often yield a group-

9 Pearson's r can take values between 0 and 1, with higher values corresponding to better correlations. It can moreover be tested for significance, with the significance expressed as p-values, which get lower as significance increases, with the conventional thresholds of significance being <.05 or, more conservatively, <.01. The same goes for the Spearman Rank Correlation Coefficient, used in the next section.

ing of verbs at different steps of a hierarchy. To take the example of the reflexive (the last of the cases presented in the previous section), a hierarchy was found where COVER & SHAVE comprise a single group at the top of the hierarchy, and where the next group is HIDE, SEE & SHOW and the one following that CUT & WASH, and so on. For the purpose of investigating whether the hierarchies for different alternations are correlated, we can assign numbers to groups of verbs in each hierarchy according to their rank. In the case of the reflexive hierarchy we would assign a '1' to COVER and SHAVE, a '2' to HIDE, SEE, and SHOW, and so on. Given rank numbers for each verb in each alternation hierarchy, all the various hierarchies can be tested for intercorrelations by means of the Spearman Rank Correlation Coefficient. The significance test assumes independence of variables. This is a questionable assumption in this case because our sample includes related languages, even very closely related ones (different variants of Japanese). However, if a verb behaves identically in two languages, as will be the tendency if the languages are closely related, this does not affect the hierarchy, but simply adds redundant evidence. What matters is difference in behavior. So it does seem that we can rely on *p*-values in the present context. The results of testing for correlations among all pairs of verb hierarchies for the 5 alternations are given in Table 1.

The results in Table 1 are quite unambiguous. The hierarchy for the causative is not correlated with any of the other hierarchies. If this hierarchy is set apart, a very clear picture emerges according to which everything correlates with everything else! In other words, the distributions of the antipassive, passive, reciprocal, and reflexive across verbs and languages all appear to adhere to the same (loosely) implicational hierarchy. This finding supports the part of the claim of Tsunoda (1985) stating that these four alternations operate on the same hierarchy, even if we found above that this hierarchy is not quite the same as the one proposed by Tsunoda.

Table 1: Spearman's Rank Correlation Coefficient (lower left triangle) and p-values (upper right triangle) among all pairs of verb hierarchies for different alternations.

	apas	caus	pass	reci	refl
apas		0.124	<0.001	<0.001	<0.001
caus	−0.166		0.329	0.185	0.256
pass	0.608	−0.156		<0.001	<0.001
reci	0.590	−0.143	0.608		<0.001
refl	0.588	−0.123	0.621	0.745	

Correlations for which $p < 0.001$ are marked in grey.

This hierarchy can be extracted by adding up the numbers referring to positions in the individual hierarchies for each verb. When verbs are ranked by these sums a combined hierarchy emerges which is displayed in (6).

(6) A combined loosely implicational hierarchy partly governing the distribution of 4 alternations over 87 verbs in 22 languages

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

Verbs within one and the same group are not necessarily significantly more similar in their behavior than neighboring verbs, so the 'greater than' symbols do not represent sharp borders, but the existence of many groups nevertheless indicates that a high degree of granularity is necessary for capturing similarities and differences among verbs – in order words, one should avoid defining large groups such as 'effective action' or 'perception' a priori. Such groupings should clearly emerge from the data or else be discarded. If nothing else, the hierarchy in (6) can serve as a warning against relying on intuitions.

Calling (6) a 'transitivity hierarchy' would immediately invite questions like why, say, SEE is more 'transitive' than, say, KILL. It has become commonplace since Hopper & Thompson (1980) to view affectedness of the undergoer as part of the definition of transitivity. Nevertheless, the hierarchy is certainly relevant to the definition of transitivity since it underlies morphosyntactic transitivity-changing alternations. With the caveat in mind that the hierarchy might have deserved to be called 'the transitivity hierarchy' if the notion of transitivity was not already so loaded, I choose to not call it anything other than 'a combined loosely implicational hierarchy partly governing the distribution of 4 alternations over 87 verbs in 22 languages.'

5 Conclusions

This is a somewhat preliminary study of how the distribution of morphosyntactic alternations across languages is at least to some degree governed by implicational hierarchies among verbs. What I hope to have achieved is to demonstrate the following points:

- quantitative methods are available for detecting the existence of implicational hierarchies;
- deviation from the unidimensionality in implicational hierarchies can be measured;
- implicational hierarchies, although they are never completely perfect, are nevertheless ubiquitous with regard to the distribution of morphosyntactic alternations over verbs and languages;
- the causative responds to its own individual implicational hierarchy among verbs which is uncorrelated with other such hierarchies;
- in contrast, the antipassive, passive, reciprocal, and reflexive across verbs and languages all appear to adhere to the same or at least similar (loosely) implicational hierarchies.

Appendix 1: Mapping between general and language-specific alternation names

Name used here	Language	Name used in database
Antipassive	Ainu	Antipassive
	Bezhta	Antipassive 1
	Arabic, Eastern Armenian, German, Italian, Ket, Mandarin, Zenzontepec Chatino	Object omission
	Even, Russian	Object deletion
	Mandinka	Antipassive middle
Causative	Arabic	Stem II causative
	Balinese, Bora, Chintang, Eastern Armenian, Hokkaido Japanese, Italian, Ket, Mitsukaido Japanese, Yaqui, Yucatec Maya	Causative
	Hoocąk	Coercive/default causative (*hii*)
	Mandinka	Causative derivation 1
	Mapudungun	Causative 1
	Zenzontepec Chatino	Causative of active verb

Passive	Arabic	Stem VII passive
	Balinese	Passive *-a* alternation
	German	Passive with *werden*
	Hokkaido Japanese	DO passive
	Mandarin	BEI alternation
	Mitsukaido Japanese	Direct passive
	Yaqui, Yucatec Maya	Passive
	Icelandic	Nominative passive
	Mandinka	Active / passive alternation
	Russian	Participial passive
Reciprocal	Arabic	Stem VI reciprocal
	Bora	Reciprocal derivation
	Even	Direct reciprocal
	Ket	Reflexive/reciprocal alternation
	Chintang, Eastern Armenian, German, Icelandic	Reciprocal
	Hoocąk	Reciprocal (+*kiki*)
	Italian, Russian	Reciprocal reflexive
Reflexive	Ainu, Chintang, German, Ket, Mapudungun, Yucatec Maya	Reflexive
	Arabic	Stem V reflexive
	Bora	Reflexive derivation
	Even	Reflexive deleting alternation
	Hoocąk	Reflexive (+*kii*)
	Italian	Direct reflexive
	Russian	Semantic reflexive

Appendix 2: Data of relevance for the Tsunoda hierarchy

The numbers cross-reference the different columns. The capital letters in the second and third lines abbreviate languages, and should be read top down. The abbreviations are as follows. AR: Arabic; BA: Balinese; GE: German; HO: Hokkaido Japanese; IC: Icelandic; MA: Mandarin; MD: Mandinka; MI: Mitsukaido Japanese; RU: Russian; YA: Yaqui; YM: Yucatec Maya; AI: Ainu; BE: Bezhta; EA: Eastern Armenian; EV: Even; IT: Italian; KE: Ket; ZC: Zenzontepec Chatino; BO: Bora; CH: Chintang; HC: Hoocąk; MP: Mapudungun; YM: Yucatec Maya. Within the matrix, a y encodes presence of the availability of an alternation for a given verb (either marginally or regularly), an n the absence, and – indicates lack of data. Passives are in columns 1–11, antipassives in 12–23, reflexives in 24–35, and reciprocals in 36–47.

	1 2 3 4
	12345678901234567890123456789012345678901234567
	ABGHIMMMRYYAABEEGIKMMRZAABCEGHIKMRYABCEEGHIIKRY
	RAEOCADIUAMIREAVETEADUCIROHVECTEPUMROHAVEOCTEUM
EAT	nyyyyyyyyyyyyyyyyyyynyynnn-nnnnnnnynn-nyyynnnny
LOOK AT	nyyyn-yynyynnnnnnyn-nnynnyynyyynynyyyyynyyyynny
SEE	nyyyn-ynyyyynnnnyyyyyny--yynyyyyynyyyynyyyyyyyy
FEAR	nynyn-nynyn-ynnnnnnynnyynynnnyynynynynnyyynynny
LIKE	nyyynyn-nyn-nnnnnnn-nnn-ny-nyynnynnyyynyyynynny
KNOW	nyyynyyynyn-nnnynnnynnnyny-nynynynyyyynyyyyynny
SEARCH FOR	nyyynyy-nyyynnnnnyn-nnn-ny-nnynnynyny-nyyynynny
BREAK	yyyyyyyyyyy-nnnnnny-nnnnny-nnynynnnnynnynynyynn
KILL	nyyyyyyyyyy-nnnnnyy-nn-yny-nynyyynyny-nyynnyyny
HIT	nyyyyyyyyyy-nynnnny-nnn-ny-nyynyyyyyyynyyyyyyny
HEAR	n-ynnyyny-yynnn-nyyynny-nyy-yynnynyny-n-yynynny
WANT	n-y---n---n-n-n-yy--n-n-n---nyn-y-nn--n-yy-y--n

Acknowledgments

Thanks go to Eric W. Holman for important input on this work and to Andrej Malchukov for many stimulating discussions and comments on a previous draft. My research was supported by an ERC Advanced Grant (MesAndLin(g)k, proj. no. 295918) and by a subsidy of the Russian Government to support the Program of Competitive Development of Kazan Federal University.

References

Ballmer, Thomas A. & Waltraud Brennenstuhl. 1986. *Deutsche Verben. Eine Sprachanalytische Untersuchung des deutschen Verbwortschatzes*. Tübingen: G. Narr.

Croft, William & Keith T. Poole. 2008. Inferring universals from grammatical variation: multidimensional scaling for typological analysis. *Theoretical linguistics* 34. 1–37.

Felsenstein, Joseph. 2009. PHYLIP. Phylogeny inference package. Version 3.69. http://en.biosoft.net/tree/Phylip.html.

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).

Holland, Barbara R., Katharina T. Huber, Andreas Dress & Vincent Moulton. 2002. δ 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.

Lehmann, Christian. 1991. Predicate classes and Participation. In Hansjakob Seiler & Waldfried Premper (eds.), *Partizipation. Das sprachliche Erfassen von Sachverhalten*, 183–239. Tübingen: G. Narr.

Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.

Malchukov, Andrej L. 2005. Case pattern splits, verb types and construction competition. In Mengistu Amberber & Helen de Hoop (eds.), *Competition and variation in natural languages*, 73–117. Amsterdam: Elsevier.

Sasse, Heinz-Jürgen. 2001. Scales between nouniness and verbiness. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and universals. An international handbook*, 495–509. (Handbücher zur Sprach- und Kommunikationswissenschaft 20.1). Berlin & New York: Walter de Gruyter.

Tsunoda, Tasaku. 1981. Split case-marking patterns in verb-types and tense/aspect/mood. *Linguistics* 19. 389–438.

Tsunoda, Tasaku. 1985. Remarks on transitivity. *Journal of linguistics* 21. 385–396.

Wichmann, Søren. 2015. Statistical observations on implicational (verb) hierarchies. In Andrej Malchukov & Bernard Comrie (eds.), *Valency classes in the world's languages*, 155–181. Berlin & Boston: 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.