Against implicational universals: a statistical critique of typological methods

(Note: VERY basic statistics only!)

Michael Cysouw, ZAS Berlin
(cysouw@zas.gwz-berlin.de)

The problem

(1) *An implicational universal A ⇒ B (e.g. Croft 1990: 47-49)*

		A: +	A: –
B	+	**X_1**	**X_2**
	–	**Ø**	**X_3**

Can we interpret non-occurrence as a universal fact, or only as an empirical finding? If the zero is an empricial fact, the value of this zero should be statistically tested.

(2a) *Apparently an implication A ⇒ B*

		A: +	A: –	total
B	+	**10**	**31**	41
	–	**2**	**12**	14
	total	12	43	55

(2b) *The expected values*

		A: +	A: –	total
B	+	$\frac{41}{55} \cdot \frac{12}{55} \cdot 55 = 8.9$	$\frac{41}{55} \cdot \frac{43}{55} \cdot 55 = 32.1$	41
	–	$\frac{14}{55} \cdot \frac{12}{55} \cdot 55 = 3.1$	$\frac{14}{55} \cdot \frac{43}{55} \cdot 55 = 10.9$	14
	total	12	43	55

(2c) *The difference between the actual and the expected values*

		A: +	A: –	total
B	+	**+ 1.1**	**- 1.1**	41
	–	**- 1.1**	**+1.1**	14
	total	12	43	55

What do typologists say?

Smallest number	Kind of universal	Hypothetical distributions of a 100-language sample							
Zero	Exceptionless universal	33	34	26	48	20	60	14	72
		0	33	**0**	26	**0**	20	**0**	14
Five	Strong tendency	36	23	31	33	27	41	22	51
		5	36	**5**	31	**5**	27	**5**	22
Ten	Statistical tendency	38	14	33	24	30	30	25	40
		10	38	**10**	33	**10**	30	**10**	25
Fifteen	Maybe something			35	15	31	23	28	29
				15	35	**15**	31	**15**	28
Nineteen	Nothing					31	19	27	27
						19	31	**19**	27

What do statisticians say?

Hypothetical distributions of a 100-language sample

33	34	26	48	20	60	14	72
0	33	**0**	26	**0**	20	**0**	14
36	23	31	33	27	41	22	51
5	36	**5**	31	**5**	27	**5**	22
38	14	33	24	30	30	25	40
10	38	**10**	33	**10**	30	**10**	25
		35	15	31	23	28	29
		15	35	**15**	31	**15**	28
				31	19	27	27
				19	31	**19**	27

Kind of interaction	Very strongly significant	Strongly significant	Significant	No interaction
Fisher's Exact two-tailed	$p < 0.000001$	$p < 0.001$	$p < 0.05$	$p > 0.2$

There are almost never statistics used to validate frequencies in typological investigations, except in work by Issac Kozinky (cited in Testelets 2001: 314-317) and Joan Bybee (e.g. Bybee et al. 1990; 1991, cf. Perkins 2001). Johanna Nichols (1986; 1992) uses lots of statistics, though here are some nice examples of wrong usage of numbers here.

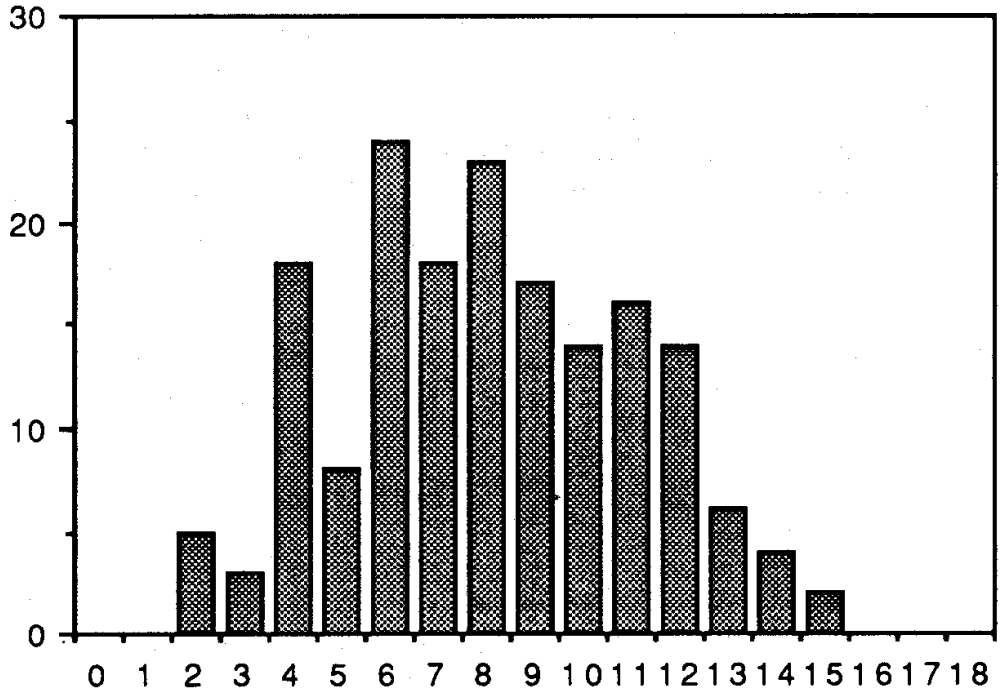
Nichols' Head/Dependent marking typology

Nichols (1986; 1992) measured Head and Dependent marking as a priori independent variables in a sample of 172 languages. In each language, each of the following constructions was scored on there being head and/or dependent marking present. Maximally, a language could score 9 H(ead) points and 9 D(ependent) points. English scored 0 H points and 4 D points:

Noun phrase possession (maximal two H and two D points):		
Pronominal:	*my book*	(English: one D point, as *my* is marked)
Nominal	*John's book*	(English: one D point, as *John* is marked)
Noun phrase modification (maximal one H and one D point):		
	the red book	(English zero points, no marking)
Sentence arguments (maximal six H and six D points)::		
Pronominal:	*I gave it to you.*	(English two D points, as *I* and *it/you* are case marked)
Nominal:	*John gave the book to Mary*	.(English zero points, as there is no case marking on nouns)

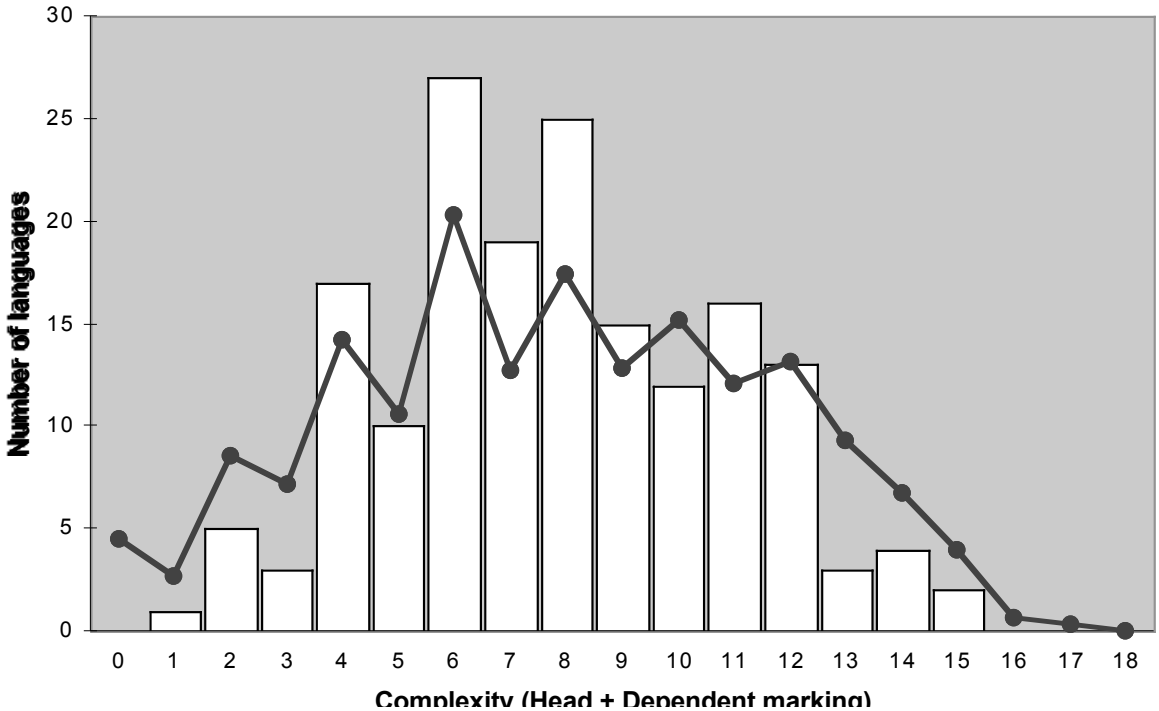
- Nichols does not include the English third person singular present tense *-s* as an example of head marking.
- Nichols also scored Adpositional Phrases on their Head/Dependent marking, but the did not use these counts in her analyses.
- Nichols also scored F points (for floating markers), but as there were just a few, she also let them out of most her analyses. I also ignored them, which leads to slight differences between my graphs and Nichols' graphs.
- Nichols uses the D/H measure to argue for areal dissimilarities. This argument is not disqualified by the following criticism (cf. Cysouw 2002).

Sum of Head and Dependent marking: 'complexity':

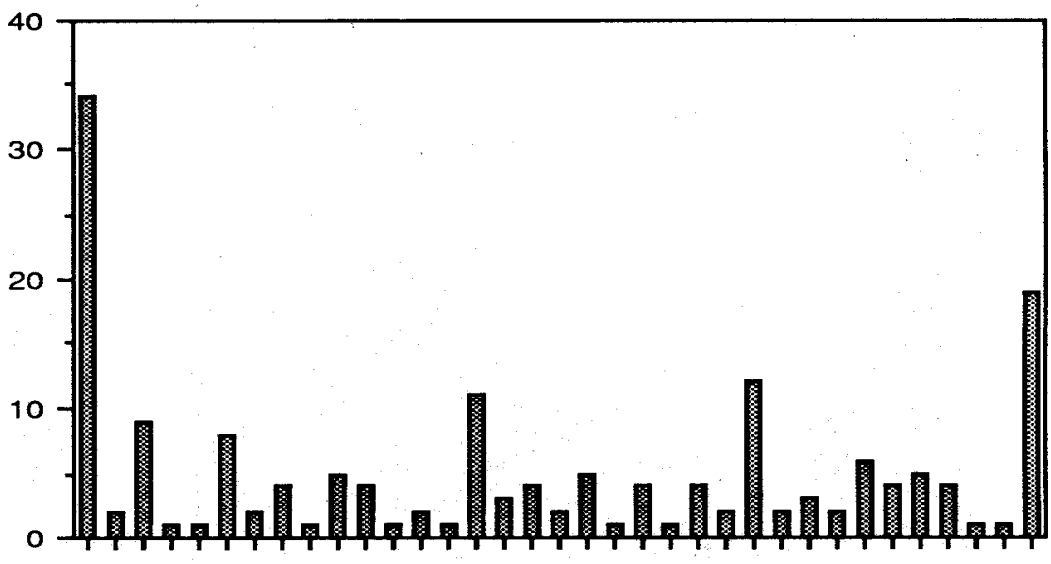


'… the complexity (Dependent points plus Head points …) has a roughly normal distribution. Neither zero complexity nor the theoretical maximum complexity of [18] points (9 Head points plus 9 Dependent points …) occurs. the highest attested complexity is 15, found in only two languages. Figure 4 shows the complexity values attested in my sample. … The normal distribution and preference for moderate complexity shown in the overall sample are echoed in most … areas, with high complexity predominating in only two.' (Nichols 1992: 88-89)

However: actual values (bars) and expected (line) are highly alike. There is a slight tendency for the extremes to be less common than expected, and for the moderate complexity to be more common than expected.



Ratio of Dependent and Head points: indicating the relative strength of head or dependent marking in a language.

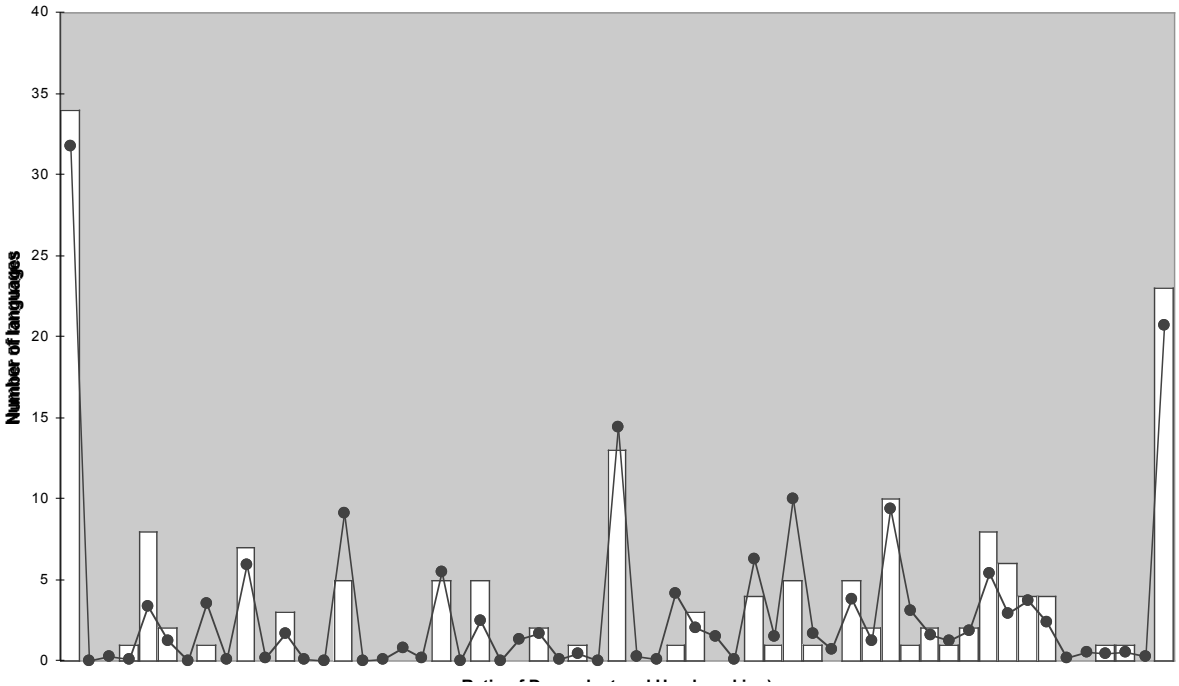


'… computing the ration of dependent to head marking … gives us 35 different ratios among the 174 sample languages. Their distribution is shown in figure 1. It is bimodal, with the greatest peaks at the extremes of exclusive head marking (ration of zero since D = 0) and exclusive dependent marking (since H = 0, an actual ratio cannot be computed as it has a zero denominator). The other ratios, whose without zeroes, run from 0.14 (two languages) to 8.00 (one language). The highest frequencies are:

0.00	34 languages (radically head marking)	
0.17	9 languages	
0.50	8 languages	[should be '0.33', MC]
1.00	11 languages	
2.00	12 languages	
H = 0	19 languages (radically dependent marking)	

… The other three frequency peaks suggest that preferred patterns cluster at perceptually simple ratios: two to one, one to one, and one to two. Overall, then , we have a preferecne for neatness of some sort: polar types, two-to-one ratios and even splits.' (Nichols 1992: 72-73)

However: actual values (bars) and expected values (line) match almost precisely (note 57 different theoretically possible ratios - no continuum):



Implicational hierarchies

The problems of evaluation of frequencies becomes even more prominent in the case of implicational hierarchies. An implicational hierarchy is a combination of implications (3a). This is not equivalent to nested implications, which are logically equivalent to a conjuncting of features in the implicatum (3b). Such nested implications are sometimes used, cf. Hawkins 1983, Pericliev 2002. However, they are statistically very suspect, cf. Dryer 1997. Also note that logically they are very weak universals. Only one combination of feature-values is claimed to be non-existent. In the present case of four features, only one of 2^4=16 logical possibilities is excluded. Logically, an implicational hierarchy looks like (3c). For easier reading, one might write it like (3d). However, The most common depiction of a hierarchy is by using a table as in (3e) (the unattested types are not shown here).

(3a) A → B
B → C
C → D

(3b) A → (B → (C → D)) ≡ (A ∧ B ∧ C) → D

(3c) (A → B) ∧ (B → C) ∧ (C → D)

(3d) A > B > C > D

(3e)

	A	B	C	D
type 1:	+	+	+	+
type 2:	–	+	+	+
type 3:	–	–	+	+
type 4:	–	–	–	+
type 5:	–	–	–	–

In my dissertation, I investigated the internal structure of paradigms of person (Cysouw 2001). I found a hierarchy of four characteristics of person paradigms. The positive values of the four parameters in this hierarchy represent the following characteristics:

A) minimal inclusive vs. augmented inclusive
(i.e. inclusive dual ≠ inclusive plural in languages without other dual marking)

B) inclusive vs. exclusive person marking
(i.e. inclusive *we* ≠ exclusive *we*)

C) no syncretism in the non-singular person marking
(at least three different person in the non-singular: *we* ≠ *you* (plural) ≠ *they*)

D) no syncretism in the singular person marking
(three different persons in the singular*: I* ≠ *you* (singular) ≠ *he/she/it*)

One might expect that B+ is necessarily implied by A+, but there is one counterexample to this implication in my sample (see case 9 in the following table) and some special structures in which the minimal or augmented inclusive is identical to the exclusive (see case 6 in the following table). These cases indicate that this implication is not necessary – though highly significant.

(4a) *Apparently an implicational hierarchy A > B > C > D*

	A	B	C	D	
1	+	+	+	+	**26**
2	–	+	+	+	**78**
3	–	–	+	+	**99**
4	–	–	–	+	**20**
5	–	–	–	–	**21**
6	+	–	+	+	**3**
7	–	+	–	+	**12**
8	–	–	+	–	**4**
9	+	–	–	+	1
10	–	+	+	–	0
11	+	+	–	+	0
12	+	–	+	–	0
13	–	+	–	–	0
14	+	+	+	–	1
15	+	+	–	–	0
16	+	–	–	–	0
Total +	31	117	211	239	

(4b) *The apparent hierarchy, ordered to the relative deviation from the statistical expectation*

	A	B	C	D	deviation / standard dev.	
1	+	+	+	+	**+ 5.2**	more common than expected
5	–	–	–	–	**+ 11.5**	
2	–	+	+	+	+ 0.5	no significant deviation from expectation
3	–	–	+	+	+ 0.7	
4	–	–	–	+	- 0.9	
14	+	+	+	–	- 0.1	
15	+	+	–	–	- 0.6	
16	+	–	–	–	- 0.6	
12	+	–	+	–	- 1.2	less common than expected
7	–	+	–	+	- 1.4	
9	+	–	–	+	- 1.5	
13	–	+	–	–	- 1.5	
11	+	+	–	+	- 1.6	
8	–	–	+	–	**- 2.0**	
6	+	–	+	+	**- 2.8**	
10	–	+	+	–	**- 2.9**	

From implications to markedness

The notion of an implicational hierarchy is not completely wrong. Rescue can be found in the notion of markedness. However, the remaining result will turn out to be much less strong as is implied by the implicational hierarchy. Note that the four parameters in isolation have different frequency-distributions (5a). This is necessarily the case in data that would traditionally be interpreted as a hierarchy by typologists. These frequencies in isolation show a hierarchy (5b) that can be interpreted as a markedness hierarchy (5c) if high frequency is taken as a sign of low markedness.

(5a) *Independent frequency of the four parameters from (4a)*

A+	31	A-	234
B+	117	B-	148
C+	211	C-	54
D+	239	D-	26

(5b) *Hierarchy of frequencies*

A+ < B+ < C+ < D+

(5c) *High frequency interpreted as low markedness*

A > B > C > D

This interpretation is only allowed iff the parameters are correlated. Concluding:

(6) ***There is a markedness hierarchy iff***

- There is a significant interaction between the parameters, AND
- The differences in frequency between the independent frequencies are large

The large differences in the second condition are important. For example, Hawkins' word order data (Hawkins 1983) show a significant interaction (7a), but no markedness hierarchy because the differences between the frequencies are too small (7b). It is always possible to order parameters in order of frequency, but this order does not necessarily mean something (7c).

(7a) *Significant interaction*

(VO ~ Pr ~ NG ~ NA) versus (OV ~ Po ~ GN ~ AN)

(7b) *Independent frequencies*

VO	162	OV	174
Pr	148	Po	188
NG	145	GN	191
NA	187	AN	149

(7c) *Hierarchy of frequencies, but no markedness hierarchy!*

NG < Pr < VO < NA

Reinterpreting implicational universals

Implicational universals can be seen as small hierarchies, and thus reinterpreted as small markedness hierarchies.

(8a) *An implicational universal A ⊃ B*

		A: +	A: –
B	+	**X_1**	**X_2**
	–	**Ø**	**X_3**

(8b) *An implicational universal as a hierarchy with two parameters*

	A	B	
type 1:	+	+	attested with frequency X_1
type 2:	–	+	attested with frequency X_2
type 3:	–	–	attested with frequency X_3
type 4:	+	–	unattested

(8c) *An implicational universal A ⊃ B is a markedness hierarchy A > B iff:*

- There is a significant interaction between the parameters A an B, AND
- A+ is much smaller than B+.

The typological view versus the statistical view

The traditional logic of the implicational universal stressed the frequency difference and therefore wrongfully interpreted (2a), repeated here as (9a), as an implicational universal. The statistical interpretation stresses the significant interaction, and thereby declares a distribution like (9b) as interesting, although there is no frequency difference and thus no implicational universal.

(9a) *No significant interaction, but a large frequency difference A+ << B+*

		A: +	A: –	total
B	+	**10**	**31**	41
	–	**2**	**12**	14
	total	12	43	55

(9b) *A significant interaction, but no frequency difference between A+ and B+*

		A: +	A: –	total
B	+	**17**	**10**	27
	–	**9**	**19**	28
	total	26	29	55

A problem for the interpretation of data

Data from WALS, correlating an inclusive/exclusive opposition in the independent pronouns with an inclusive/exclusive opposition in the verbal inflection. There appear to be four major types (dark grey in the figure) and four minor types (light grey in the figure). For a theory of linguistic structure, it has to be explained why exactly these types are more common than the others are; at least, so it might seem.

(10a) *Typological distribution with apparently 4 major types (dark grey) and 4 minor types (light grey).*

Verbal inflection	Independent pronouns: no *we*	*we* identical to *I*	unified *we*	only inclusive *we*	inclusive+ exclusive *we*	
no person marking	1	5	36	1	27	70
we identical to *I*	1	1	9	0	1	12
unified *we*	0	2	75	0	2	79
only inclusive *we*	0	0	0	4	5	9
inclusive and exclusive *we*	0	2	0	0	28	30
	2	10	120	5	63	200

Actually, the type unified pronouns/no inflectional persn marking is the odd one out. The actually attested 36 cases are less than expected by chance. Except for this type, all the major and minor types in Figure 3 are exactly those cases that are more frequent than the chance expectation.

(10b) *Major deviations from expectation. The positive deviations are shaded dark grey (highly significant) and light grey (slightly significant)*

Verbal inflection	Independent pronouns: no *we*	*we* identical to *I*	unified *we*	only inclusive *we*	inclusive+ exclusive *we*
no person marking		+ 1.5	- 6.0		+ 5.0
we identical to *I*			+ 1.8		- 2.8
unified *we*		- 1.9	+ 27.6	- 2.0	- 22.9
only inclusive *we*			- 5.4	+ 3.8	+2.2
inclusive and			- 18.0		+ 18.5

References:

Bybee, J. L., Pagliuca, W. and Perkins, R. D. 1990. "On the asymmetries in the affixation of grammatical material". In *Studies in Typology and Diachrony*, W. Croft (ed.), 1-42. [Typological Studies in Language 20]. Amsterdam: John Benjamins.

—— 1991. "Back to the future". In *Approaches to grammaticalization*, E. C. Traugott and B. Heine (eds.), 17-58. [Typological Studies in Language 19]. Amsterdam: John Benjamins.

Croft, W. 1990. *Typology and Universals*. Cambridge: Cambridge University Press.

Cysouw, M. 2001. *The Paradigmatic Structure of Person Marking*. Ph.D. thesis, University of Nijmegen

—— 2002. "Interpreting typological clusters". *Linguistic Typology* 6 (1): 69-93.

—— forthcoming. "Against implicational universals". *Linguistic Typology* 7.

Dryer, M. S. 1997. "Why Statistical universals are better than absolute universals". *CLS* 33 (2): 123-145.

Hawkins, J. A. 1983. *Word Order Universals*. New York: Academic Press.

Nichols, J. 1986. "Head-Marking and Dependent-Marking Grammar". *Language* 62 (1): 56-119.

—— 1992. *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.

Pericliev, V. 2002. "Economy in formulating typological generalizations". *Linguistic Typology* 6 (1): 49-68.

Perkins, R. D. 2001. "Sampling procedures and statistical methods". In *Language Typology and Language Universals: An International Handbook*, M. Haspelmath, E. König, W. Oesterreicher and W. Raible (eds.), 419-434. [Handbücher zur Sprach- und Kommunikationswissenschaft 20]. Berlin: De Gruyter.

Testelets, Y. G. 2001. "Russian works on linguistic typology in the 1960-1990s". In *Language Typology and Language Universals: An International Handbook*, M. Haspelmath, E. König, W. Oesterreicher and W. Raible (eds.), 306-323. [Handbücher zur Sprach- und Kommunikationswissenschaft 20]. Berlin: De Gruyter.