Leipzig Spring School on Linguistic Diversity
The Analysis of Typological Data
Michael Cysouw
cysouw@eva.mpg.de

Empirically Specifying Type-similarities

Data on word order, map 81 from the *World Atlas of Language Structures* (Dryer 2005). The total number of genera for which more than one language is available is 179. For each value, the number of genera having a language with this value is given, and also the number of genera having both types in different languages. Similarities between types is established by:

$$Similarity = 1 - \frac{(Genera1 - Both + 1)(Genera2 - Both + 1)}{(Genera1 + 1)(Genera2 + 1)}$$

Map	Value 1	Value 2	Genera 1	Genera 2	Both	Total	Similarity
81	SOV	SVO	105	67	17	179	0.37
81	SOV	VSO	105	29	7	179	0.28
81	SOV	VOS	105	11	3	179	0.27
81	SOV	OVS	105	5	3	179	0.51
81	SOV	OSV	105	3	3	179	0.76
81	SOV	variable	105	73	36	179	0.66
81	SVO	VSO	67	29	14	179	0.58
81	SVO	VOS	67	11	9	179	0.78
81	SVO	OVS	67	5	2	179	0.35
81	SVO	OSV	67	3	2	179	0.51
81	SVO	variable	67	73	29	179	0.65
81	VSO	VOS	29	11	6	179	0.60
81	VSO	OVS	29	5	3	179	0.55
81	VSO	OSV	29	3	1	179	0.28
81	VSO	variable	29	73	17	179	0.67
81	VOS	OVS	11	5	0	179	0.00
81	VOS	OSV	11	3	2	179	0.58
81	VOS	variable	11	73	8	179	0.70
81	OVS	OSV	5	3	0	179	0.00
81	OVS	variable	5	73	4	179	0.68
81	OSV	variable	3	73	2	179	0.51

Hierarchy of person marking

I have proposed a hierarchy of four characteristics of person paradigms (Cysouw 2001). 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*)
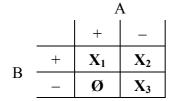
One might expect that B+ is necessarily implied by A+, but there is one counterexample to this implication in my sample (see case 9 in the following table) and some special structures in which the minimal or augmented inclusive is identical to the exclusive (see case 6 in the following table). These cases indicate that this implication is not necessary – though highly significant.

(1)   *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	

Statistical view of implicational universals

(2)   *An implicational universal A $\rightarrow$ B*

A

	+	–
+	X_1	X_2
–	Ø	X_3

B

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.
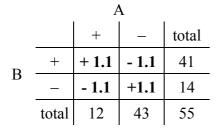
(3a)   *Apparently an implication A $\rightarrow$ B*

A

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

B

(3b)   *The expected values*

A

	+	–	total
+	$\frac{41}{55} \cdot \frac{12}{55} \cdot 55 = 8.9$	$\frac{41}{55} \cdot \frac{43}{55} \cdot 55 = 32.1$	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

B

(3c)   *The difference between the actual and the expected values*

A

	+	–	total
+	**+ 1.1**	**- 1.1**	41
–	**- 1.1**	**+1.1**	14
total	12	43	55

B

This difference is not statistically significant (e.g. Fisher's Exact $p = 0.71$)

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.

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

	Independent pronouns					
Verbal inflection	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.

(5b) *Major deviations from expectation. The positive deviations are shaded dark grey (highly significant) and light grey (slightly significant) as measured by Pearson Residuals*

	Independent pronouns				
Verbal inflection	no *we*	*we* identical to *I*	unified *we*	only inclusive *we*	inclusive+ exclusive *we*
no person marking	+ 0.3	+ 1.5	- 6.0	- 0.8	+ 5.0
we identical to *I*	+ 0.9	+ 0.4	+ 1.8	- 0.3	- 2.8
unified *we*	- 0.8	- 1.9	+ 27.6	- 2.0	- 22.9
only inclusive *we*	- 0.1	- 0.5	- 5.4	+ 3.8	+2.2
inclusive and exclusive *we*	- 0.3	- 0.5	- 18.0	- 0.8	+ 18.5

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).

Haspelmath's Indefinite Pronouns Typology

To approach the linguistic diversity, Haspelmath distinguishes nine typologically primitive functions of indefinite pronoun encoding, as shown here in (1) to (9). An opposition between two of these nine functions is crucial for at least some indefinite pronouns in some languages. There are even more possible functions of indefinite pronouns, but they are left aside for unexplained – yet probably practical – reasons.

(1)   specific, known to the speaker ('**Somebody** called while you were away: guess who!')
(2)   specific, unknown to the speaker ('I heard **something**, but I couldn't tell what it was.')
(3)   non-specific, irrealis ('Please try **somewhere** else.')
(4)   polar question ('Did **anybody** tell you anything about it?')
(5)   conditional protasis ('If you see **anything**, tell me immediately.')
(6)   indirect negations ('I don't think that **anybody** knows the answer.')
(7)   direct negation ('**Nobody** knows the answer.')
(8)   standard of comparison ('In Freiburg, the weather is nicer than **anywhere** in Germany')
(9)   free choice ('**Anybody** can solve this simple problem.')

In 40 languages, Haspelmath identifies 133 differnt indefinite pronouns. The similarity between two primitives (as described above) can be measured by counting how often two of these primitives are coded by the same indefinite pronoun.

	1	2	3	4	5	6	7	8	9
1	133	34	27	25	25	14	5	1	0
2	34	133	34	31	31	16	6	3	2
3	27	34	133	44	44	26	12	10	6
4	25	31	44	133	62	48	26	28	13
5	25	31	44	62	133	46	22	31	17
6	14	16	26	48	46	133	39	39	20
7	5	6	12	26	22	39	133	23	12
8	1	3	10	28	31	39	23	133	32
9	0	2	6	13	17	20	12	32	133

Typology of Person Marking

Person Categories

Primitive	Translation	Referential meanign
1	'I'	first person singular
2	'you'	second person singular
3	's/he, it'	third person singular
12	'we'	inclusive dual
123	'we'	inclusive plural
13	'we'	exclusive
23	'you'	second person plural
33	'they'	third person plural

Combinations of person categories

With eight categories, there are theoretically 2^8-1-8-1= 246 different combinations possible (minus one taking none, minus eight for taking only one, and minus one for taking all; these are all not considered combinations of person markers here). There are in total 35 different combinations of the basic eight person attested in 325 person paradigms as described in Cysouw (2003: Ch. 3-4). The most frequent combinations are readily interpretable referentially.

Person categories	Approx. meaning	Frequency	Person categories	Frequency
3/33	'third'	125	123/13	3
12/123/13	'first plural'	100	1/2	3
12/123	'inclusive'	97	1/2/3	3
2/23	'second'	84	12/13	2
1/12/123/13	'first'	35	13/23	2
1/13	'exclusive'	29	3/23	2
12/123/13/23	'non-third plural'	18	12/123/23	2
23/33	'non-first plural'	17	1/12/123/13/23	2
12/123/13/33	'non-second plural'	11	123/13/23	1
1/3	'non-second singular'	10	13/33	1
2/3	'non-first singular'	7	1/12	1
2/3/23/33	'non first'	6	1/23	1
3/13/33		5	12/123/33	1
2/12/123/13		5	1/12/123	1
12/123/13/23/33		5	3/12/123/33	1
2/13/23		4	1/2/12/123/13/23	1
2/12/123/23		4	2/12/123/13/23/33	1
			1/2/12/123/13/23/33	1

Number of pairwise combination of person categories

	1	2	3	12	123	13	23	33
1	325	8	13	41	40	68	5	1
2	8	325	16	12	12	4	101	8
3	13	16	325	1	1	5	8	137
12	41	12	1	325	286	181	34	20
123	40	12	1	286	325	184	35	20
13	68	4	5	181	184	325	35	24
23	5	101	8	34	35	35	325	30
33	1	8	137	20	20	24	30	325