Understanding Typological Distributions

Balthasar Bickel

University of Leipzig

	20th century "Universalist T."	21st century "Distributional T."
nature	method	discipline
goal	formulating UG	explaining distributions
interest in	implicational universals	distributional probabilities
part of	CogSci	anthropology, including psychology
core question	What's a Possible Human Language?	What's Where Why?

- Distributional Typology: what's where why?
- Studying typological distributions:
 - *what:* typology as a method of comparing structures
 - why: typology as a historical discipline
 - where: typology as statistical modeling
- Core factors and their methodological problems
 - Inheritance: how can we detect inheritance patterns?
 - Contact: what are linguistic areas?
 - Social structure: how does social structure affect grammar?
 - Cognition: how does processing affect grammar?

- If you want to compare two structures, you helps to know what they are...
- The Descriptive A Priori: analytical metalanguages ('theoretical frameworks')
- Two ingredients:
 - 1. Absolute Universals = primitive concepts that are required for the analysis of every imagineable language
 - 2. Variables = sets (inventories and scales) of primitive concepts that are required for the analysis of at least one language (most popular in typological metalanguages, less so in generative metalanguages)

Other terms: *tertium comparationis,* comparative notions, metrics, parameters

Too many options to take: how to justify metalanguages?

- One popular answer: claim the metalanguage to be psychologically or biologically real
- Problems:
 - if the metalanguage also includes variables ('has vs. does not have clicks'), all values of the variable must be universally present in the brain, i.e. even in languages that don't have evidence for them.
 - there is no biological evidence
 - language acquisition does not rely on any of our linguistic metalanguages but rely on general learning mechanisms (Tomasello 2006)

A pragmatic approach to metalanguages

- The working linguist's everyday answer: justify the metalanguage by:
 - logical consistency and simplicity
 - availability of 'proofs' and argumentation
 - replicability
 - universal descriptive success: all structures can be analyzed consistently (cf. "Subject" vs. "A")
 - typological success: analysis allows comparison
- Compare this to the use of the metrical system as justified by its consistency, symplicity, replicability, and descriptive and 'typological' success!

A pragmatic approach to metalanguages

- Conceived therefore as analytical instruments, the metalanguage
 - captures brain processes only very indirectly (ERP signals vs. distributional signals of structures)
 - makes no claim on the distribution of its denotata in the world (they could be all over, like "verbs", or rare like "clicks")
 - therefore, it provides the right means to study statistical universals, i.e. trends in real-life distributions!
- A worked example: phonological words

A universal, accepted by virtually all theories

- "All languages have words."
 - More precisely: "All languages have exactly one phonological domain between feet and phrases. This domain is the word."
 - Or: "In all languages, any (general) phonological pattern refers to one of {segment, mora, syllable, foot, word, phrase, utterance}."

A challenge, posed by many languages

- Limbu (Sino-Tibetan, Nepal; joint work with René Schiering, U. Leipzig and Kristine Hildebrandt, U. Manchester):
 - pf-[stem-sf-cl], domain of Liquid Alternation and other rules kε-[Li'-Le=Lo] > kε[li'rero] 'of your bow' 2sPOSS-bow-GEN=PTCL
 - [pf-stem-sf-cl], domain of Coronal Assimilation and other rules
 [mε-n-mεt-paŋ] > [mεmmεppaŋ] 'We did not tell him' nsA-NEG-tell-1>3.PST
- What's the word in Limbu? The red one or the blue one? Does the word in Limbu include or exclude the prefix?

A challenge, posed by many languages

Apparently, languages differ as to which and how many units they have between the foot and the phrase, i.e. there is variation:



1. REDUCE the variation and maintain an absolute universal:

- There are no possible exceptionless (in the same sense as a Neo-Grammarian sound law is exceptionless)
- The universal is necessarily true; it follows from our metalanguage.
- 2. MEASURE the variation and examine distributional skewings
 - Need variables for comparing words within and across languages

- 'Limiting the Scope', for example:
 - limit the scope to rules that are general across the lexicon. (We did this, but it didn't help.)
 - 'if a language has X, then Word 2, etc.' or 'Word 2 is enforced by higher-ranking constraint X'. (No known X.)
- 'Reanalyzing the Data', for example:
 - analyze prefixes as clitics and posit a special 'clitic domain' C (Vogel 2007). Doesn't help because that necessitates two C domains, one with, one without proclitics.
 - analyze prefixes on a different stratum than suffixes.
 Doesn't help because clitics are in both domains.
 - allow recursive words: $[\omega [\omega]]$. But that predicts shared properties, which is not what we find.

- One way of doing this:
 - code each domain in each language for the nature of its defining phonological pattern p (e.g. 'tone bearing unit', 'segmental assimilation', 'alternation' etc.) and the morphemes types that are included (e.g. prefix, suffix, enclitic etc.)
 - start with those patterns that are delimited by morphological information (thereby excluding mora, syllable and feet) and include at most one stem (thereby excluding compounds and phrases)

Strategy 2: MEASURE

00	00				word_	type	(xserve2.	.rz.ur	128 1	hese	e two coda	consonant	s alte	r in voic	ing f	eatu	res in this	domain
Brow	wse	Rec	cord ID 377	,					129 /	/r/ is inserted in this domain, not sure of syllable membership								
) 9	16	674	674 Limbu							Nasal coda may not occur in this domain								
Lavou	t:		The ee	nonnant Islin nat narmitt	ad/daaa a		un in Abia al-		131 /	/e, o/ are banned here/do not occur in this domain								
essent	tials 🕨	132	2 The co	nsonant /r/ is not permitt	ea/aoes no	ot occ	ur in this de	omain	132 1	The consonant /r/ is not permitted/does not occur in this domain								
									133 Secondary Stress is realized in the specified place in this domain								omain	
	4	Lexica stratum specification				d	omain_de	ef (xse	rve2.rz.uni-lei	pzig.de)						omam		
41	(a)				Browse											this dom	1.0	
					<u>)</u> ? 1 🖸	ID	173 Do	omain si	tem ± prefi	x ± suffi	x ± postposed pa	rticle		^		III II	i this doma	ain
-			Source	of lean phonology (LID	Layout:									-				
Record	d:		Source	s or loan phonology (LID	Domain 🕨	Pos1	0 n/a		T	(p1 12	stem	Restr1	3 u	nrestricted		ble	boundary	in this
5		Dor	main refere	enced by ppattern:								Result				ble	boundary	in this
Found	:	173	stem ±	prefix ± suffix ± postpos	<1. →	Rel1	MAY also in	clude										
12				,		Pos2	1 prae	9	т	yp2 1	formative	Restr2	1 re	estricted		his	s domain	
Total:		2	left edg	e	Record:	Rel2	MAY also in	clude		1						of	the cuffix	changes
855					122	Neiz		chado		_								
Unsor	ted	M	Morpheme		Def	inition				p3	RestrValue	Concernation	ana hui	ant all DOG/h	Restr	Def		a find but with a
		1	formative	Marker of an inflectional cate	SEA versati	nnot oc le verb	cur as an ind	ependen	t part of sp	B 2 3	unrestricted	Can occur with s	anything	(e.g. Turkish	mi)	ments	(so, NOT a phrasa	al amix , but rather
		0	n/a	In domain_def, use if no oth	er morph don	nain pa	arts relevant in	n domain	definition	4	n/a	not applicable						
		7	unknown	type is unknown at this time						5	unknown Part circum/simulf	degree of restric	tion unkn	own at this tim	ne od with	eimulfi	vation & circumfive	tion
	Dom.	13	stem part	An element of a stem (as de	fined under I	D 12) t	hat is not deli	mited by	general	4 7	restricted: Head	Something like a	ffix only f	to a head elem	nent (in	a narro	wed sense of Res	trID 1
	162	Do	duplicent	There is no overt mannestat	Ion of the exe	Rel4	INCLUDES	NO OTH	IER	8	restricted: Phrase	Can occur with v	hatever	POS element	of that p	ohrase	it is adjacent to like	e Manange NP
	164	St	euuplicant em + circun	ofiv2 + cuffix + postposo												TE	1	
	165	Ste	em + Prefix	es + Suffix		Pos5			T	/p5		Restr5				UE:	1	
	168	Ste	em + Prefix	es + Circumfix + Suffix +		Date						Restro					1	
	169	ste	em ± suffix :	± postponed particle		Reis											1	
	170	Ste	em ± Prefix	± Suffix ± Semirestricte													1	
100	171	Ste	em ± Prefix	± Circumfix ± Suffix ± C		Pos6			T)	/p6		Restre					1	
	172	ste	em ± prefix	± entire circumfix		Ral6											1	
	173	ste	em ± prefix	± suffix ± postposed par		Reio											1	
	174	ste	em ± prefix	± suffix ± preposed parti		Pos7			Т	yp7		Restr7					1	
	175	Ste	em ± Prefix	± Circumfixes ± Suffix ±	100	Brow	150 C									•	1	
	176	ste	em ± prefix	± infix ± postposed parti		BIOW		-	-	-			0			.22	0	
	177	ste	em ± prefix	± infix ± suffix			0		0		0	1	0	1		1	1	
	178	pre	eposed par	ticle ± stem ± suffix			1		0		1	0	0	0		1	1	
	179	ste	em ± prefix	± suffix ± simulfixes/circi	umfix parts	±	U	_	U	_	U	1	U	1	_	1	1	

Bickel, Hildebrandt, & Schiering, in press.

- For much of the 20th century, theoretical linguistics, including typology, has been dominated by the REDUCE VARIATION strategy. But:
 - Most other sciences chose the MEASURE VARIATION strategy (e.g. biology, psychology etc) — why should linguistics be different?
 - The REDUCE VARIATION strategy does not fit well into research integrating other anthropological disciplines (esp genetics and psychology) in modeling and explaining the distribution of structures over the world.
- One reason why the MEASURE VARIATION strategy has been neglected is, I suspect, that on top of detailed work on specific languages, it requires probabilistic methods uncommon in linguistics.

- The pragmatic approach to metalanguages:
 - seek variables instead of absolute universals
 - think of variables as instruments of analysis and comparison, with no necessary ressemblance to patterns in the brain (no 'mental grammar')
 - don't reduce but measure variation
- Two common problems in the pragmatic approach

*Solution 1: treat each structure as a datapoint

- enter all p-words into the analysis, e.g. their degree of 'compactness' (how much of each grammatical word is inside the p-domain?)
- control for the fact that they are related in the same way as we control for historical relatedness between languages (cf. later)
- This approach is standard when taking measurements from multiple speakers per language (e.g. experiments), but very sofar very rare when raking measurements from multiple structures per language

*Solution 2: aggregate data

degrees of ergativity, of OV-order, etc.

Within-language variation

language	alignment	Referential conditions	Structural conditions
Bartangi	0	5	PTC-based
Bartangi	S=A	0	PTC-based
Bartangi	0	2	PTC-based
Bartangi	A	82, 105	PTC-based
Bartangi	A	94, 108	other
Bartangi	A=O	2	other
Bartangi	S	0	other
Beja	S=A	0	
Веја	0	0	
Belhare	A	108, 94, 10, 2, 92, 82	
Belhare	S=0	0	
Belhare	A	105	
Basque	A	0	
Basque	S=0	0	
Dimili	A	82, 2, 105	PTC-based
Dimili	A	108, 94	PTC-based
Dimili	S=0	0	PTC-based
Dimili	0	10, 82, 121, 105, 92	other
Dimili	0	108, 122, 94	other
Dimili	S=A	0	other

ERG/ACC ratios

Bartangi	0
Basque	1
Веја	0
Belhare	1
Dimili	0.5

Within-language variation

*Solution 3: choose exemplars

- basic word order (Dryer 2005): main clause, declarative, pragmatically neutral, lexical arguments, most frequent
- case exemplar (Bickel & Nichols 2005):

Table 1 Select Brahui declension forms (xal 'stone')									
	SINGULAR	PLURAL							
NOMINATIVE	xal	xal-k							
ACCUSATIVE	xal-ē	$xal-t-\overline{e}$							
DATIVE	xal-ki	xal-tē-ki							
ABLATIVE	xal-ān	xal-tē-ān							

- (i) If there is any difference in the morphological type across case formatives, pick the grammatical cases. Within grammatical cases, pick accusative or ergative or agentive (or whatever is chiefly used on A or P arguments). If there is none of these, pick nominative or absolutive (if these are at all marked overtly). If neither the A nor the P argument of transitive clauses is identified as such by overt marking, or if case-marking is restricted to pronouns, assume the language has no "case".
- (ii) If there is any difference in the morphological type across tense-aspect-mood formatives, pick tense. Within tenses, pick past (or whatever is chiefly used for simple past time reference); if there is none, pick future; if there is none, pick present. If there is no tense, pick the closest aspect equivalent of past tense as a proxy. If there is no aspect, pick that mood, status, or evidentiality formative that is mostly used for past tense narration. If there is no grammatical marker for any of these notions, assume the language has no "tense-aspect-mood".
- (iii) For both case and tense-aspect-mood: if the marking is zero, pick the overtly marked opposite value of the category (e.g. the plural of nominatives, if the singular is zero-marked; or the future tense, if the nonfuture is zero-marked).
- (iv) For both case and tense-aspect-mood: if categories differ in their degree of grammaticalization, pick the most nearly grammaticalized one. Pick synthetic tense formatives over periphrastic ones.

Within-language variation: summary

- 1. Each structure as a datapoint:
 - advantage: no data loss; needs good control of relatedness
- 2. Aggregate data
 - advantage: some data loss; simpler statistics

- 3. Taking exemplars
 - advantage: very fast data collection; simple statistics

Another frequent problem: 'many similarities, few identities'

- An example: cosubordination
- Universal definition: clause linkage with conjunct illocutionary scope (operator dependency) (e.g., Foley and Van Valin 1984; Van Valin 2005; Bickel 1998)

Amele (Trans-New Guinea: Madang; PNG; Roberts 1988)

ho busale-?e-b dana age gbo-ig-a fo? pig run.out-DS-3s man 3p hit-3p-T.PST Q 'Did the pig run out and did the men kill it?'

'many similarities, few identities'

 Frequent problem: constructions that do not match the definition but are similar in all other respects (e.g. they are used for narrative chaining, are different from symmetrical coordination, have a fixed 'chaining' position)

Kâte (Trans-New Guinea: Finisterre-Huon; PNG; Suter 1992:25ff)

a. Kpende? go opâ-nane fia-râ opa? foŋ-nane Kp. 2s water-1sPOSS get-SEQ nettle bundle-1sPOSS
hâpo-râ dzâhe? fâ-tseme? carry-SEQ finally follow-2sHORT.FUT
'Kpende! Get water for me, bring me a bundle of nettle and follow me!'

b. *su? kpeue-me natsa-ndzepieŋ.* banana ripen-SEQ3sDS 1P.tell-2pHORT.FUT

'Let me know when the bananas are ripe!'

- Solution 1: posit a structure ("cosubordination 2") without a scope constraint
 - Problem: the structures are so similar to each that one reading of cosubordination 1 is the sole reading of cosubordination 2.
- Solution 2 ("functional exemplar"): Assume that in Kâte etc. one reading 'really' reflects cosubordination while the other reflects something else (probably subordination, with disjunct scope) (Bickel 1998).
 - Problem: No independent evidence for this within Kâte etc.

'many similarities, few identities'

- Solution 3: Revise the definition, e.g. define cosubordination by constraint-free scope (Bickel 1991)
 - Problem: We can base the definition on any variable we want (e.g. finiteness, tense scope, extraction possibilities etc.) but we may always run into the same problems!
- Solution 5: Ignore the non-matching languages!
- Solution 6: Give up comparative research!

'many similarities, few identities'

- The underlying problem of this example is that things are mostly similar but never identical, and that we want to measure the similarity by means (universal definitions) that require identity.
- Linguists are not the only scientists with this problem, but, curiously, for most other scientists, this is not a problem...
- ... because similarity is a straightforward issue:
 - similar = identical re X and non-identical re Y
 - where X, Y, ...: variables of comparison / measurement tools
- Let's do the same!

Multivariate typology

000											clause_li	inkag	e (xserve2.rz.uni-l	eipzig	g.de)						
R LID	::Language	Ш.,	. ::Scope	J	. ::Marking	T	. ::Scope	T	::Marking	Foc	. ::Marking	Cat	::categorical_sym	Cla	. ::clause_position	Emb	::embedding	Jun.	:juncture	Interpr	:interproposition
1 14	Alambiak	4	disjunct	1	ok	1	constraint-free	1	ok			1	symmetrical	4	flexible-relational	2	adjoined	2	clause	2	conditional
2 20	Amele	2	blocked	2	banned	1	constraint-free	1	ok			1	symmetrical	1	fixed-final	2	adjoined	2	clause	4	inconsequential
3 20	Amele	3	conjunct	2	banned	3	conjunct	3	harmonic	2	banned	2	asymmetrical	1	fixed-final	2	adjoined	2	clause	1	narrative
4 20	Amele	3	conjunct			1	constraint-free	1	ok	2	banned	1	symmetrical	1	fixed-final	2	adjoined	2	clause	3	disjunction
5 20	Amele	4	disjunct	1	ok	1	constraint-free	1	ok	1	ok	3	constraint-free	4	flexible-relational	2	adjoined	2	clause	2	conditional
52 20	Amele	4	disjunct	1	ok			1	ok			1	symmetrical	1	fixed-final	2	adjoined	2	clause	6	purposive
7 35	Belhare	1	constraint-free			1	constraint-free	1	ok	2	banned	1	symmetrical	2	flexible-adjacent	2	adjoined	2	clause	1	narrative
8 35	Belhare	1	constraint-free	1	ok	1	constraint-free	1	ok	1	ok	2	asymmetrical	2	flexible-adjacent	2	adjoined	2	clause	1	narrative
9 35	Belhare	-		1	ok	1	constraint-free	2	banned	2	banned	2	asymmetrical	2	flexible-adjacent	2	adjoined	2	clause	4	inconsequential
10 35	Belhare	4	disjunct			1	constraint-free	1	ok	2	banned	2	asymmetrical	4	flexible-relational	2	adjoined	2	clause	5	concessive
11 35	Belhare	4	disjunct	1	ok	1	constraint-free	1	ok	1	ok	3	constraint-free	4	flexible-relational	2	adjoined	2	clause	9	topic
12 40	Burushaski	1	constraint-free			1	constraint-free	2	banned	2	banned	2	asymmetrical	2	flexible-adjacent	2	adjoined	2	clause	1	narrative
13 629	Chantyal	-				1	constraint-free	2	banned			2	asymmetrical	2	flexible-adjacent	2	adjoined	2	clause	1	narrative
51 2862	Chintang	1	constraint-free	1	ok	1	constraint-free	1	ok	1	ok	2	asymmetrical	2	flexible-adjacent	2	adjoined	3	core	1	narrative
54 71	Dvirbal	-		-		1	constraint-free	1	ok	-		1	symmetrical	5	fixed-initial	2	adjoined	-		-	
55 71	Dvirbal	-		-		3	conjunct	2	banned			1	symmetrical	5	fixed-initial	2	adjoined			6	purposive
57 71	Dvirbal	-		-	-	-		2	banned				-,	-		2	adjoined	4	NP	-	
14 74	English	1	constraint-free	1	ok	1	constraint-free	1	ok	1	ok	1	symmetrical	2	flexible-adjacent	2	adjoined	2	clause		
15 74	English	1	constraint-free		ok	1	constraint-free		ok	1	ok.		symmetrical	2	flexible-adjacent	2	adjoined	2	clause	6	numosiva
16 1404	Eore	1	constraint-free		ok	9	conjunct	3	harmonic		<u> </u>	2	asummatrical	2	flexible.adjacent	2	adjoined	2	clause	1	parposive
17 87	German		disjunct	2	banned	1	constraint-free	1	ok	1	ok.	9	constraint free	4	flexible-selational	2	adjoined	2	clause	7	causal
18 87	German	-	disjunct	1	ok		constraint-free		ok	2	banned	1	summetrical	4	flexible-relational	2	adjoined	1	causo	7	causal
10 1556	Godiá	-	conjunct	2	banned	2	consummente	2	bacced	0	banned	2	agrimmetrical	-	flexible-relational	e 2	adjoined	2	clause	'	Causai
20 1556	Godié	3	disjunct	6	banned	1	conjunct	6	oarmed	6	Garmed	6	constraint free	6	flexible-adjacent flexible-selational	e 0	adjoined	6	clause	0	topic
20 1556	Goole (Anniant)	-	disjunct	2	banned		constraint-free	-	OK beened	-		3	constraint-free	4	flexible-relational	2	adjoined	2	clause	9	topic
0 2001	Greek (Ancient)	-	aisjunia	2	banned	-	constraint-free	2	banned	0	hanned	2	asymmetrical	4	flexible ediacont	2	adjoined	2	clause	9	topic
21 103	hua	3	conjunct	2	banned	3	conjunct	2	banned	2	banned		symmetrical	2	flexible-adjacent	2	adjoined	2	clause	0	
22 103	Hua	3	conjunct	2	banned	3	conjunct	2	banned	2	banned	1	symmetrical	2	fiexible-adjacent	2	adjoined	2	clause	•	manather
23 103	Hua	1	constraint-free	0	hound	3	conjunct	3	narmonic	0	housed	2	asymmetrical	1	rixeo-rinal flouible seletionel	2	adjoined	6	clause	1	narrauve
24 103	Hua	4	disjunct	2	banned	1	constraint-free	1	ok	2	banned	3	constraint-free	4	flexible-relational	2	adjoined	2	clause	9	topic
25 103	Hua	4	disjunct	2	banned	1	constraint-free	1	ok	2	banned	2	asymmetrical	4	flexible-relational	2	adjoined	2	clause	4	inconsequential
26 103	Hua	4	disjunct	2	banned	1	constraint-free	1	OK	-		2	asymmetrical	4	flexible-relational	2	adjoined	2	clause	2	conditional
27 126	Kate	1	constraint-free	2	banned	3	conjunct	2	banned	2	banned	2	asymmetrical	1	fixed-final	2	adjoined	2	clause	1	narrative
28 126	Kate	3	conjunct	1	OK	1	constraint-free	2	banned			2	asymmetrical	1	fixed-final	2	adjoined	2	clause	1	narrative
29 1806	Korate	1	constraint-free			1	constraint-free	1	ok	_		2	asymmetrical	2	flexible-adjacent	2	adjoined	2	clause	1	narrative
30 483	Lenakel	_				1	constraint-free	1	ok			2	asymmetrical	2	flexible-adjacent	2	adjoined	2	clause	1	narrative
31 206	Nepali	1	constraint-free	2	banned	3	conjunct	1	ok	1	ok	2	asymmetrical	2	flexible-adjacent	2	adjoined	2	clause	1	narrative
32 206	Nepali	4	disjunct	2	banned	1	constraint-free	1	ok	1	ok	3	constraint-free	4	flexible-relational	2	adjoined	2	clause	9	topic
33 468	Newar (Dolakha)	1	constraint-free					2	banned	1	ok	2	asymmetrical	2	flexible-adjacent	2	adjoined	2	clause	1	narrative
34 2866	Pali	1	constraint-free							1	ok	2	asymmetrical	2	flexible-adjacent	2	adjoined	2	clause	1	narrative
56 340	Russian	1	constraint-free	1	ok	1	constraint-free	1	ok	1	ok	1	symmetrical	2	flexible-adjacent	2	adjoined				
35 361	Swahili	3	conjunct			3	conjunct	2	banned	2	banned	2	asymmetrical	2	flexible-adjacent	2	adjoined	2	clause	1	narrative
36 361	Swahili	3	conjunct	2	banned	3	conjunct	2	banned	2	banned	2	asymmetrical	2	flexible-adjacent	2	adjoined	2	clause		
37 361	Swahili	3	conjunct	1	ok	1	constraint-free	1	ok	2	banned	1	symmetrical	2	flexible-adjacent	2	adjoined	2	clause		
38 361	Swahili	4	disjunct							1	ok	2	asymmetrical	4	flexible-relational	2	adjoined	2	clause	9	topic
39 367	Tauya	3	conjunct	2	banned	3	conjunct	2	banned	2	banned	1	symmetrical	2	flexible-adjacent	2	adjoined	2	clause	8	
40 367	Tauya	1	constraint-free	1	ok	3	conjunct	2	banned			2	asymmetrical	1	fixed-final	2	adjoined	2	clause	1	narrative
41 367	Tauya	2	blocked									2	asymmetrical	4	flexible-relational	2	adjoined	2	clause	4	inconsequential
42 367	Tauya	2	blocked	1	ok							3	constraint-free	4	flexible-relational	2	adjoined	2	clause	9	topic
43 2864	Toura	3	conjunct	2	banned	3	conjunct	2	banned	2	banned	2	asymmetrical	2	flexible-adjacent	2	adjoined	2	clause	1	narrative
44 502	Turkish	3	conjunct	2	banned	3	conjunct	2	banned	2	banned	2	asymmetrical	1	fixed-final	2	adjoined	2	clause	1	narrative
45 502	Turkish	3	conjunct	1	ok					2	banned	2	asymmetrical	1	fixed-final	2	adjoined	2	clause	1	narrative
46 393	Usan	1	constraint-free									2	asymmetrical	2	flexible-adjacent	2	adjoined	2	clause	1	narrative
100	Browse)	-	

- How many structures? As many as are distinct in var1... varN
- How many variables? As many as are of interest to the research question, e.g. all variables needed to capture crosslinguistic differences in the syntax of clause linkage; or in morphological coding.
- Which variables? Developed as needed for distinguishing structures during data collection (Autotypologizing Method: Bickel & Nichols 2002), or pre-determined by the research question.
 - structural variables: formal properties in which structures are alike or differ.
 - denotation variables: denotations (stimuli, contexts, functions) in which structures are alike or differ

- Are there any structures across languages that are more similar to each other than to others? (a standard dissimilarity problem)
- Here, one question concerns the degree to which sets of variable values (feature choices) cluster on sets of languagespecific structures
- One suitable tool for this, with a long tradition in sociology:
 - Multiple Correspondence Analysis (MCA)

- similarity between variables ('factors') = covariation of values
- similarity between individual structures = same values in many variables
- measure both and superpose them in the same coordinate system
- assess fit of model by % of covered χ^2/N deviations ('inertia' or ' λ -value')

Multivariate typology





Multivariate typology

- Another example where we need a multivariate typology:
- Classical morphological typology
 - "isolating": Chinese
 - "agglutinating": Turkic
 - "inflectional": Latin
 - "root-inflectional": Arabic

Another example of multivariate typology

- Variables following Bickel & Nichols 2007
- Flexivity: ± lexical Allomorphy (Inflectional classes)
- Fusion types (simplifyng)
 - concatenative: phonological dependency
 - isolating: relative number of phonological domains that separate morphemes
 - nonlinear: simultaneous realization (Ton, Ablaut)
- Degree of synthesis: N(categories/g-word)
- Exponence: realization of categories
 - cumulative: several in one
 - separative: one-to-one
 - distributive ("extended"): one across several, e.g. Kind-er-n: -PL-DAT.PL

Another example of multivariate typology

	Flexivity	Fusion	Synthesis	Exponence
agglutinating	non-flexive	concatenative	high	separative
isolating	non-flexive	isolating	low	separative
root-inflectional	flexive	nonlinear	mid	distributive
inflectional	flexive	concatenative	mid	distributive
Otomí	flexive	weakly isolating	high	cumulative
Lai Chin	non-flexive	strongly isolating	high	separative
Kinyarwanda	non-flexive	nonlinear	high	separative
etc.				

Consequence: there are not only 4 types, but at least

2 • 3 • 3 • *N* (degres of synthesis)

- One final advantage: given a multivariate typology, traditional terms ("agglutination", "subordination", etc) can always be defined as specific constellations (with options) and so
 - past knowlegde is not lost
 - no 'paradigm change' in terminologies is needed
 - grammars are compatible across theories and
 - 'sustainable', i.e. readable even in the future!