To appear in: Isabelle Bril [ed.] *Clause-hierarchy and clause-linking: the syntax and pragmatics interface*, Amsterdam: Benjamins

Capturing particulars and universals in clause linkage: a multivariate analysis*

Balthasar Bickel

University of Leipzig

Cross-linguistic variation in adjoined clause linkage is higher than what is allowed by universal concepts like 'coordination' or 'subordination' which entail sets of strictly correlated properties. This chapters uses statistical techniques to uncover probabilistic correlations and clusters in a pilot database. For this, a set of variables is developed that ranges in coverage from the scope of illocutionary force operators to extraction constraints and that allows both detailed qualitative analysis of language-specific clause linkage structures and large-scale quantified measurement of the similarities of such structures within and between languages. The study tentatively suggests that there is a prototype of subordination which is closer to 'and'-like than to 'chaining' constructions, and that there is a continuum between more vs. less tightly constrained types of converb and chaining constructions, but no general prototype of 'cosubordination'.

1 Introduction

The analysis of individual, language-specific structures normally starts with a set of terms that are defined, or assumed to be defined, in a cross-linguistic way, and are taken as such from field manuals, handbooks, formal theories, or reference grammars of other languages. A case in point is the term 'clausal cosubordination', which was introduced by Olson (1981) and Foley & Van Valin (1984) and is defined by conjunct illocutionary scope: a clause that is cosubordinate to a main clause

*Versions of this paper were presented at the International Symposium on the Grammar and Pragmatics of Complex Sentences (LENCA 3) in Tomsk, June 29, 2006, at the International Conference on Role and Reference Grammar in Leipzig, September 30, 2006, at the Syntax of the World's Languages conference in Berlin, September 26, 2008, and as a guest lecture at the University of Zürich, December 8, 2008. I thank all audiences for stimulating questions. Many thanks also go to Robert Van Valin for discussing issues of focus and extraction with me and to Zarina Molochieva and Alena Witzlack-Makarevich for discussing the Chechen and Russian data with me. I am also indebted to Volker Gast and the two non-anonymous reviewers Jeff Good and Michael Cysouw for very helpful comments on an earlier draft. This chapter is dedicated to the memory of Mickey Noonan (1947-2009).

obligatorily falls under the scope of illocutionary operators in the main clause. An example of this is the 'medial form' or 'converb' construction, as it is found for example in the Papuan language Amele or the African language Swahili:[1]

(1) Amele (Trans-New Guinea: Madang; Papua New Guinea; Roberts 1988)
ho busale-ʔe-b dana age gbo-ig-a fo? ('CHAIN')
pig run.out-DS-3s man 3p hit-3p-T.PST Q
'Did the pig run out and did the men kill it?'

(2) Swahili (Niger-Congo: Bantu; Bickel 1991)
je, u-li-baki nyumba-ni u-ka-tayarisha ch-akula ch-etu? ('CHAIN')
Q 2s-PST-stay home-LOC 2s-SEQ-prepare VII-food VII-our
'Did you stay home and prepare our food?'

In both cases, the interrogative marker in the main clause (final *fo* in Amele, initial *je* in Swahili) has scope over both clauses so that the only possible reading is one in which the speaker inquires about the truth value of both propositions.

The definition sets cosubordination apart from coordination, where the scope of such markers does not necessarily extend over both clauses, and also from subordination, where it is impossible to have conjunct scope (cf. Foley & Van Valin 1984, Tikkanen 1995, Van Valin 2005, among others). When one takes the term 'cosubordination' further to the field, however, one quickly runs into structures that look very similar to the data in Amele or Swahili, but do not entirely fit the definition. Such structures are found for example in South Asian languages, such as Belhare:[2]

(3) Belhare (Sino-Tibetan: Kiranti; Nepal)
a. *khar-e ki jutta ŋŋ-in-ghutt-he-ga i?* ('CHAIN')
[3sS]go-PST SEQ shoes[NOM] 3sA-buy-bring.for-PST-2sP Q
'Did she go [there] and buy you shoes?'
or 'Did she buy you shoes when she went [there]?' (presupposing either 'she went' or 'she bought')
b. *ne-e yuŋ-sa mundhupt-he i?* ('CVB')
here-LOC sit-CVB [3sS]chat-PST Q
'Did he sit here and chat (with you)?'

[1] Where constructions figure in the pilot study described in Section 4, I include in brackets the (relatively arbitrary) identification label used in the Appendix. Glossing follows the Leipzig Glossing Rules (`http://www.eva.mpg.de/lingua/resources/glossing-rules.php`), with the addition of ADD 'additive (focus)', ASS 'assertive', B 'B gender (in Chechen)', CESS 'cessative', CONC 'concessive', COND 'conditional', DECL 'declarative', DEP 'dependent', DS 'different subject', F.PST 'far past', HORT 'hortative', ILL 'illocutionary', J 'J gender (in Chechen)', PRED 'predicate marker', PURP 'purposive', PTCL 'particle', SEQ 'sequential', SS 'same subject', TEMP 'temporary (aspect)', T.PST 'today's past', V 'V gender (in Chechen)', W.PST 'witnessed past', and Y.PST 'yesterday's past'. Roman numerals indicate noun classes.

[2] Data without a source specification are from my own fieldnotes.

or 'Did he chat with you when sitting here?' (presupposing either 'he chatted' or 'he sat')

In these structures, the scope of the interrogative marker in the main clause (*i*), is indeterminate: depending on the context of utterance, the sentences may be interpreted as having conjunct or disjunct scope. This indeterminacy can be found both with finite (3a) and nonfinite (3b) forms (cf. Section 3.3 on finiteness). The same pattern can also be observed in the Indo-European (Indo-Aryan) language Nepali:

(4) Nepali
yahɑ̃̄ ā-era khānā khā-yo? ('CHAIN')
here come-CVB food[NOM] eat-3sM.PST
'Did he come here and eat?'
or 'Did he eat after coming here?' (presupposing either 'he came here' or 'he ate')

Data similar to these can be found in many other languages of South Asia (e.g. in Kathmandu Newar: Hale & Shrestha 2006, Dolakha Newar: Genetti 2005, Burúshaski: Tikkanen 1995, Sanskrit: Tikkanen 1987, or Pali: Bickel 1991), and also in Papuan languages — even in languages of the same family as Amele, e.g. in Tauya:

(5) Tauya (Trans-New Guinea: Madang, Papua New Guinea; MacDonald 1990:226)
tepau-fe-pa yate fitau-a=nae? ('CHAIN')
break-PRF-SS go throw-2=POLAR.Q
'Did you break it and go away?'
or 'Did you go away after breaking it?' (presupposing either 'you went away' or 'you broke it')

Other examples from Papuan languages include Hua (Haiman 1980:400), Usan (Reesink 1987:297f), Kâte (Suter 1992:25ff), and Korafe (Farr 1999).

The question that arises is how to analyze structures like (3) - (5). There is a number of possibilities. First, one could posit a second analytical term ("cosubordination 2"), defined without a scope constraint. But this would miss the fact that the structures are so similar to each other that one reading of "cosubordination 2" (namely the one with conjunct scope) is the sole reading of "cosubordination 1". Second, one could try and argue that in Belhare and similar languages, one reading 'really' reflects cosubordination while the other reflects something else — presumably subordination, with disjunct scope (Bickel 1998). While this may be a viable solution in some cases, at least in Belhare and Nepali, I am not aware of any independent evidence for assuming structural ambiguity: the possible readings can only be resolved pragmatically, and it seems unjustified to posit differentiated syntactic representations for this (at least under a parsimonious approach to syntax that does not try to resolve in the syntax what can just as well be left to pragmatics). Third, one could revise the definition of cosubordination, for example by defining the term without any syntactic constraint on illocutionary scope (as is done by Bickel 1991 or Croft 2001). However, this may not solve the problem once and forever because ultimately, we can base the definition on any property we want (e.g. non-assertion, finiteness, tense scope, extraction possibilities, etc.) and always run into the same problem when analyzing other languages: if we define 'cosubordination' without

a scope constraint, the term would no longer capture the distinctive properties of 'cosubordination' in Amele and Swahili, and we would miss again the overall similarity between these structures and those in the other languages. Similarly, if we define a notion like 'subordination' via 'non-assertion' (Cristofaro 2003), we will run in structures that look very similar to 'subordinate' structures, yet are asserted (e.g. with imperatives in an *although*-clause such *This is true, although don't expect examples!*, cf. Green 1976, Lakoff 1984, Takahashi 2008, among many others). Any property that is picked as definitional will favor one type and make it the model for others. The fundamental problem is that there is no non-arbitrary choice (cf. Croft 2001): should Amele provide the model, or Tauya, or English? Any answer seems wrong. Finally, one could follow Lazard's (2006) or Haspelmath's (2007) suggestions and keep issues of language-specific analysis free of comparative notions: we could set up entities like 'Amele *ʔe*-construction' or 'Nepali *era*-construction', describe their properties and leave it to typological research to compare these entities on the basis of some comparative concept like 'cosubordination^{comp}', defined without regard to the language-specific details and independently of their analysis. While this may seem to solve the problem of how to classify language-specific constructions, it does not address (and is not intended to address) the comparative problem of just what the 'right' definition of the comparative concept might be. Yet comparative notions are often of critical help in language-specific analyses and when positing constructional entities, and it is one of the great steps of progress that typological and theoretical knowledge increasingly informs such descriptive work: for example, without the publication of Foley & Van Valin (1984), the issue of scope properties would have had little chance of being addressed in descriptive grammars. In fact, as many early descriptions in the American structuralist tradition testify, any attempt at describing languages purely 'in their own terms' risks missing important analytical questions. Moreover, unless the analysis is coupled with an explicit metalanguage of description, the range of properties that are taken to be relevant for a given construction remains arbitrary (or even opportunistic, as Croft (2001) puts it). To most empirically-minded linguists, however, such a descriptive metalanguage is the more appealing the more it is informed by knowledge of typological variation and not just by meta-principles such as elegance in theory design. But then, we are back to the problem of defining the 'right' comparative concepts for analyzing language-specific structures.

Underlying all these solutions and their problems is a general attempt to try and reduce the observed diversity — here between languages like Amele and Swahili on the one hand, and languages like Belhare, Nepali or Tauya on the other hand — to one or two universal structures or comparative concepts. In this paper, I propose an alternative, based on standard methods used in other disciplines when confronted with diversity: this alternative consists in measuring instead of reducing diversity. I describe the general ideas behind this in Section 2. In Section 3 I review some of the key structural properties that lead to the diversity in clause linkage noted above and in general. Section 4 presents ways in which the diversity can be measured and discusses cross-linguistic and possibly universal patterns emerging from this, based on a pilot database of 69 constructions from 24 languages.

2 Multivariate analysis

When confronted with diversity, most other disciplines try to measure it. The key to making this possible is that structural similarity needs to be understood as what it is: structures $\mathcal{S}_1$ and $\mathcal{S}_2$ are similar iff they are identical in some variables (also known as 'properties', 'parameters', or

'features') *A*...*M*, but different in other variables *N*...*Z*. Therefore, we need to decompose terms like 'cosubordination' (or 'subject', 'word', 'sentence', 'antipassive' etc.) into sets of variables that capture all dimensions *A*...*Z* in which any given pair of structures may be identical or different — whether between languages or within languages. I call such decompositions 'multivariate analysis', extending the use of the term from its statistical meaning of simultaneously analyzing entire sets of variables to the development of these sets itself.[3]

The set of variables must be large enough so as to capture all known variation, and in principle could extend to the minutest phonetic differences. Obviously, practical choices in research interests and time budgeting dictate upper limits, as always. The choice of variables is determined by similar questions of research planning, but if one subscribes to standard principles of economy in theory design, variables need to be logically independent of each other and, if one subscribes to empiricist principles of theory design, variables should also be developed inductively (e.g. using the AUTOTYP method of Bickel & Nichols 2002) rather than exclusively on the basis of *a priori* assumptions about the nature of grammar. Ideally, the set of variables is large enough to capture the full range of known diversity, so the logically possible combinations of all levels in this set allow a precise description of each known structure. Full development of such a system of variables is clearly a long-term goal and must proceed in tandem with progress in the detailed analysis of many different languages.

Two kinds of variables are relevant in multivariate typology: structural and denotational variables. Structural variables are defined in syntactic or semantic terms and their levels capture specific syntactic or semantic properties, e.g. properties like 'conjunct scope of illocutionary operators' or 'conditional'. Denotational variables are defined in terms of extra-linguistic stimuli or contexts, to which language-specific structures may respond in the same or in different ways: e.g. a narrative context may elicit one kind of structure in one language and another structure in another language. In the following, I will limit myself to structural variables. Within these, I will furthermore mostly concentrate on aspects of syntax and issues of semantic scope. I have nothing to say in this paper on the semantic relations between propositions or usage patterns, although there is no doubt that the relevant variables are important for understanding the distribution of clause linkage structures in the languages of the world.

3 Some variables in clause linkage, with particular attention to adjoined structures

In this section, I review some of the better-known ways in which clause linkage structures differ from each other within and across languages. To keep the scope of the discussion manageable, I focus on adjoined clauses and disregard clauses that are subcategorized by main clauses (i.e. that are embedded in the sense of complement clauses). The results of the discussion are summarized in Section 4, where the variables are applied to a pilot database.

[3] For an earlier proposal moving in a similar direction, but assuming pre-defined 'ideal' types, see Lehmann (1988). For more general discussion of the multivariate approach, see Bickel (2007).

3.1 Illocutionary scope

As noted in the introductory discussion, a key variable is the scope of illocutionary operators[4] like question or imperative markers that occur in the main clause. In principle, each of these operators could have their own scope properties, but for present purposes I simplify matters by collapsing them.

A first possibility is for structures not to impose any syntactic constraint. This was illustrated by the introductory examples in (3) through (5). If a structure constrains the scope, there appear to be at least three options. First, in some structures, the scope is always conjunct. This is what was illustrated by the Amele and Swahili data in (1) and (2). Second, some clause linkage structures obligatorily impose disjunct scope so that only one of the linked clauses can be in the scope of the illocutionary marker. This behavior is traditionally associated with the notion of 'subordination' (e.g. Tikkanen 1995) and is instantiated here by an example from Belhare (and its English translation):

(6) Belhare
ne-e *yuŋ-a=naa* *mundhupt-he* *i?* ('SUB')
DEM-LOC [3sS]sit-SBJV.PST=TOP [3sS]chat-PST Q
'When he was here, did he say something?' (or was he silent?)
or 'Did he say something when he was here?' (or later only?)
but not 'Was he here, and did he say something?'

In structures involving the topic marker *=naa* in Belhare (or *when* in English), only one of the linked clauses can be questioned; the other clause must be interpreted as presupposed.

A further option is for main clause illocutionary scope to be limited to its clause — I call this 'local' scope in the following. Structures with local scope often correspond to what is traditionally called 'coordination', but translations do not always involve coordination in English and linked clauses may show asymmetrical dependencies that one would normally take to indicate subordination in traditional terms. Obviously, the variable of illocutionary scope is logically independent of variables having to do with symmetry or dependency. The following structure in Amele involves two independent clauses, linked by the conjunction *gba* 'but':

(7) Amele (Roberts 1987)
ho *busale-i-a* *gba* *dana* *age* *gbo-i-ga* *fo?* ('BUT')
pig run.away-3s-T.PST but man 3p hit-3p-T.PST Q
'The pig ran away but did the men kill it?'

In Tauya, local scope is associated with 'topic' clauses, more akin to adverbial subordination in English:

[4] Throughout this chapter, I use the term 'operator' for any grammatical category that takes scope over some other linguistic object. Thus, illocutionary force markers are operators, while for example person markers are not.

(8) Tauya (MacDonald 1990)
nen mei momune-i-nani=ra pofei-ti nen=tu-e=nae? ('TOPIC')
3p here sit-3p-ASS=TOP talk-CONJ 3p=give-2=POLAR.Q
'They sat here and/but did you talk to them?'
or 'Since they sat here, did you talk to them?'

In some languages, local scope optionally extends to dependent clauses, a pattern which will be called 'extensible scope' (as opposed to 'local', 'conjunct', and 'disjunct' scope). For the structure in the following example from Usan, Reesink (1987:239) notes that "it is not automatically clear whether the interrogative marked on the final verb also applies to the preceding predicate:"

(9) Usan (Trans-New Guinea: Madang; Papua New Guinea, Reesink 1987)
munon iyau wârâm-a um-orei ʔiyo? ('CHAIN.SWR')
man dog him.hit-3sDS die-3sREM.PT Q
'Did the man hit the dog and it died?'

A similar pattern has recently been noted by Molochieva (2008) for the Nakh-Daghestanian language Chechen and by Forker (2009) for Hinuq, another language of the same family. In Chechen, clause linkage with the general-purpose converb in *-na* allows main scope illocutionary markers to scope over the main clause or over both the main clause and the dependent clause, but not over the dependent clause alone:

(10) Chechen (Nakh-Daghestanian; Good 2003)
Maliika tyka-na='a j-ax-na c'a-j-e'a-r=ii? ('CHAIN')
M.(J).NOM store-DAT=SS J-go-CVB home-J-come-W.PST=Q
'Did Malika come home, having gone to the store?' (presupposing 'having gone to the store')
or 'Malika went to the store, but did she come home?'
or 'Did Malika go to the store and come home?'
but not: 'Did Malika go to the store, having come home?'

Structures with extensible scope are almost as flexible as structures with unconstrained scope, except that extensible scope requires the main clause to always fall into the scope of main clause illocutionary operators.

3.2 Scope of negation, tense and other main clause operators

While the scope of illocutionary markers is particularly prominent in differentiating clause linkage types, the scope of other main clause operators can also provide fundamental distinctions, and each of these can in principle be subject to their own scope constraints.

One important such operator is negation. Turkish converb clauses in *-Ip*, for example, have been noted to have conjunct negation scope, while converb clauses in *-Ip de* or *-ince* have local scope, i.e. are limited to the main clause:

(11) Turkish (Johanson 1995)

a. *ev-e gel-ip el-ler-in-i yıka-ma-dı.* ('CHAIN')
house-DAT come-SEQ hand-p-3sPOSS-ACC wash-NEG-PST[3s]
'He did not come home and [did not] wash his hands.'

b. *ev-e gel-ip de el-ler-in-i yıka-ma-dı.*
house-DAT come-SEQ AND hand-p-3sPOSS-ACC wash-NEG-PT[3s]
'When he came home, he did not wash his hands.'

c. *ev-e gel-ince el-ler-in-i yıka-ma-dı.*
house-DAT come-WHEN hand-p-3sPOSS-ACC wash-NEG-PST[3s]
'When he came home, he did not wash his hands.'

As the data in (11b-c) show, *-Ip de* and *-ince* block the scope of the negation marker (*-ma*) in the main clause. This contrasts with *-Ip* alone in (11a), where the scope of the negation marker is conjunct.

Converbs in Puma, a Kiranti language, impose disjunct scope:

(12) Puma (Sino-Tibetan: Kiranti; Schackow et al. in press)
gaph mu-so kama pʌ-mu-e-min
talk do-SIM work NEG-do-1p-NEG
'Chatting, we do not work.'
or 'We work without talking.'
but not: 'We neither work nor talk.'

Burúshaski converb constructions do not constrain the scope of negation, and sentences like the following can be understood with either conjunct or disjunct scope:

(13) Burúshaski (Tikkanen 1995)
khíiruman sis majít-ar n-úu-nin nimáaz ay-é-č=á-am. ('CHAIN')
some people mosque-DAT CVB-3hum.pSBJ[-go]-CVB prayer NEG-do-DUR=AUX-3hum.pSBJ
'Having gone to the mosque some people do not pray' (but read)
or 'Some people do not pray after getting to the mosque.' (but after getting up)
or 'Some people do not go the mosque and do not pray.'

While I am not aware of a structure with negation having extensible scope of the kind discussed earlier for illocutionary force, negation operators allow for yet another possibility, not attested for illocutionary force: some clause linkage structures require the scope to extend exclusively to the dependent clause. This is well-known from studies of complement clauses, where the phenomenon has been dubbed 'NEG-transport' (e.g. Horn 1989) and is exemplified by expression like *I don't think that p* which are regularly interpreted as 'I think that *p* is not the case'. The following examples illustrate this for adjoined structures in Belhare (in contrast to the converbs in the related language Puma, which imposes disjunct scope, as in (12)):

(14) Belhare (Bickel 1993)

a. *taw-a=lo kam n-cokg-att-u-n.* ('COM')
[3sS-]come-PST.SBJV=COM work[NOM] NEG[-3sA-]do-PST-3sP-NEG
'He didn't keep working until he came.' (i.e. 'he worked but not until he came.')

b. *yaŋ- his-sa la- ŋŋ-um-ʔ-ni.* ('CVB')
around- look-CVB walk- [3nsS]NEG-walk-NPST-NEG
'He didn't look around while walking.' (i.e. 'he walked without looking around')

In this case, the negation marker has scope over the dependent clause; the main clause must be interpreted as affirmative. This type of 'transported' scope is so far unattested with other operator categories. Indeed, with regard to illocutionary operators, Belhare *lo(k)*-constructions have local scope, as illustrated by the following examples:

(15) Belhare

a. *tupt-u-ŋ=lo khem-t-u-ŋ.* ('COM')
understand[SBJV]-3sP-1sA=COM listen-NPST-3sP-1sA
'I'll listen so that I understand.'

b. *tupt-u-ŋ=lo prʌsta ka-lur-a!* ('COM')
understand[SBJV]-3sP-1sA=COM clear 1sP-tell-sIMP
'Talk to me clearly so that I understand!'

Here, the difference between the illocutionary markers in the main clause has no impact on the interpretation of the dependent clause. The fact that the same constructions show different scope behavior under negation and under illocutionary operators confirms that these are independent typological variables.

Illocutionary force and negation are the best known operators whose scope behavior differentiates between clause linkage types. Other operators are not well-studied in this regard. One exception is tense and status (realis/irrealis) markers. Especially for Papuan languages, the scope behavior of these markers has often been noted to differ across clause linkage structures (Foley & Van Valin 1984, Foley 1986). Amele *ʔe*-structures illustrate conjunct tense scope (cf. (1), which shows conjunct illocutionary scope of the same construction):

(16) Amele (Roberts 1988)

a. *ho busale-ʔe-b dana age gbo-ig-a.* ('CHAIN')
pig run.out-DS-3s man 3p hit-3p-T.PST
'The pig ran out and the men killed it.'

b. *ho busale-ʔe-b dana age gbo-gbag-an.* ('CHAIN')
pig run.out-DS-3s man 3p hit-3p-FUT
'The pig will run out (*not:* ran out) and the men will kill it.'

Here, the temporal interpretation of the dependent clause strictly depends on the tense choice in the main clause. This contrasts with Belhare *ki*-clauses, where the dependent clause is not necessarily within the scope of the main clause tense marker. The scope is primarily limited to the main clause, but can optionally be extended into the dependent clause. Accordingly, a sentence like the following can have different interpretations, depending on context:

(17) Belhare
khimm-e n-ta-ch-u ki mun-n-dhup-chi. ('CHAIN')
house-LOC 3nsA-reach[SBJV]-d-3sP SEQ chat-3nsS-chat[NPST]-d
'They will reach home and chat.'

or 'When they reach home, they'll chat.'
or 'They reached home and now they will chat.'

For a similar construction in another Sino-Tibetan language of Nepal, Chantyal, Noonan (1999:412) notes: "Because of the context in which this was said, we understand the converb to be referring to a time future relative to the moment of speaking; however, had the sentence been uttered after he had arrived in his home village, but before he met his father, we would understand the clause as referring to an event past relative to the moment of speaking:"

(18) Chantyal (Sino-Tibetan: Tamangic; Noonan 1999)
na thɦem-əŋ ɦya-sirə bəw-ra dhɦo-wa ɦin. ('CHAIN')
I house-LOC go-SEQ father-DAT meet-NOM be.NPST
'I will go home and see my father.'
or 'Having gone home, I am to see my father.'

Tense markers whose scope is not conjunct but extensible are also characteristic of some Indo-European participle constructions. Participles in Ancient Greek, for example, are attested both within and outside the scope of main clause tenses:

(19) Ancient Greek
a. *pollakhoû dé me epéskhe légo-nt-a*
often PTCL 1sACC stop.3sIMPERFECT talk-IPFV.ACT.PTCP-ACC.s
metaxú. ('PART.CONI.')
in.the.middle
'[The oracle] has often stopped me when I was in the middle of talking.' (Plat. *Apol.* 40b)
b. *egṑ eréō hōs eû epistá-men-os.* ('PART.CONI.')
1sNOM speak.1sFUT PTCL well understand-IPFV.MED.PTCP-NOM.s
'I will speak out because I understand it well.' (Herod. *Hist.* IX 42)

In (19a), the time reference of the participial clause coincides with that of the main clause; in (19b), by contrast, the participial clause makes a present tense assertion while the main clause refers to the future.

Tense markers with local scope, i.e. scope that is restricted to the main clause, are also attested. This is mostly the case when both the dependent and the main clause are marked for their own tense. The other scope types observed with illocutionary force and negation, viz. disjunct or transported scope, never seem to occur with tense markers.

3.3 Finiteness and marking possibilities

At first sight, it seems plausible to analyze every kind of non-local scope constraint as resulting from the fact that a dependent clause is not fully finite, i.e. lacks some operator specification of its own. However, as we saw in the preceding, nonfinite dependent clauses do not always lead to scope constraints (cf. the data in (18) and (19) above). And, finiteness is determined not only by operator categories (i.e. categories with a certain scope behavior) but also by referential categories like person

marking (which does not define a specific scope over other linguistic objects). Thus, the variable of finiteness is logically independent from variables of scope.

Finiteness is traditionally defined by the range of categories that can be expressed by inflectional forms in those main clause types that allow the maximal number of categories. Dependent clauses can deviate from this by allowing more or less categories. If there are less categories, the verb form heading the dependent clause is called 'nonfinite'. The case where a dependent clause expresses more categories than a main clause does not fit any traditional term (but see the Wambule data in (30) below for an example).

Finiteness is relevant for clause linkage on its own, but there are two additional aspects: first, dependent clauses vary as to exactly which categories can be marked and which categories are not allowed. To the extent that categories are realized by inflectional morphology, this follows from finiteness; however, many such categories, e.g. tense or illocutionary force, can also be realized by independent elements (words, clitics, particles) and therefore, possibilities here can be independent of finiteness. Second, some linkage types require that two clauses contain the same degree of finiteness and the same range of expressible categories, while other linkage types do not require this. In the following, I concentrate on the first issue, i.e. on marking possibilities of dependent clauses in their own right. The second issue will be taken up in Section 3.4.

3.3.1 Illocutionary force marking

An issue that is often critical is whether or not dependent clauses allow the marking of illocutionary force operators. In many languages conjunct illocutionary scope entails that illocutionary markers cannot themselves appear on dependent clauses. This is the case for example in the Amele constructions that were first illustrated by (1) and (16):

(20) Amele (Roberts 1988)

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

But such a constraint on illocutionary marking is not universal. Turkish dependent clauses in *-Ip* show conjunct scope when the illocutionary marker appears on the finite final clause (as in 21a), but nothing bars such markers from appearing on the dependent clause (as in 21b):

(21) Turkish (Bickel 1991, Johanson 1995)

a. *gel-ip sana bir şey söyle-di mi?* ('CHAIN')
come-CVB 2sDAT one thing say-PST[3s] Q
'Did he come and say something?

b. *otur-up mu konuş-tu-lar?* ('CHAIN')
sit-CVB Q speak-PST-[3]p
'Did they speak (while they were) sitting?' (asking about the circumstance of speaking)

And similar possibilities are known from the Papuan language Fore:

(22) Fore (Trans New Guinea: Kainantu-Gorokan; Scott 1978)

a. *kana-a:-kí-tá* *a-ka-us=ó.* ('CHAIN')
come-3sPRES.DS-DEP-1dAS 3sP-see-1dA=Q
'Is he coming and we see it?'

b. *na-m-e-g-ánt=ó* *na-ku-w-e.* ('CHAIN')
1sP-give-2sFUT.DS-DEP-1sAS=IMP eat-FUT-1s-DECL
'Give me something and then I will eat it.' (*i.e.*, 'Give me something to eat!')

The sentence in (22a) illustrates the fact that main clause question markers have conjunct scope, while (22b) shows that this basic structure does not ban the occurrence of at least imperative markers on the dependent clause.

A number of languages allow illocutionary marking on dependent clauses only if it matches the marking in the main clause. In Belhare, for example, *ki*-clauses (of the kind illustrated before by (3) and (17)) can be marked by an imperative, but only if the same mood is also marked on the main clause:

(23) Belhare
caw-a ki khar-a! ('CHAIN')
eat-IMP SEQ go-IMP
'Eat and go!'

In Chechen, illocutionary marking on the dependent clause is possible in some but not other structures. For example, the marking is blocked in clauses marked by the sequential converb in *-na*, or the concessive converb in *-na-shehw* that is derived from this. In each case, the interrogative clitic (*=ii*) must occur on the main clause (24a,c) and cannot be attached to the dependent clause (24b,d):

(24) Chechen (Good 2003, Molochieva 2008)

a. *Maliika tyka-na='a j-ax-na c'a-j-e'a-r=ii?* ('CHAIN')
M.(J).NOM store-DAT=SS J-go-CVB home-J-come-W.PST=Q
'Did Malika come home, having gone to the store?' (presupposing 'having gone to the store')
or 'Did Malika go to the store and come home?'
but not: 'Did Malika go to the store, having come home?'

b. * *Maliika tyka-na='a j-ax-na=ii c'a-j-e'a-r* ('CHAIN')
M.(J).NOM store-DAT=SS J-go-CVB=Q home-J-come-W.PST
Intended: 'Did Malika come home having gone to the store?' (presupposing 'Malika came home')

c. *ahw toex-na-shehw j-axa-r=ii Zaara tyka-na?* ('CHAIN')
2SG.ERG hit-CVB-CONC J-go-W.PST=Q Z.(J).NOM store-DAT
'Did Zara go to the store, even though you hit her?'
not 'Did Zara go to the store, and is it even the case that you hit her?'

d. * *ahw toex-na-shehw=ii j-axa-ra Zaara tyka-na?* ('CHAIN')
2SG.ERG hit-CVB-CONC=Q J-go-W.PST Z.(J).NOM store-DAT
Intended: 'Did Zara go to the store, even though you hit her?'

But illocutionary marking is allowed in clauses headed by the temporal converb in *-cha*:

(25) Chechen (Molochieva 2008)

a. *Muusa c'a v-e'a-cha, naan-na xaza xiiti-r=ii?* ('CVB.TEMP')
M.(V).NOM home V-come-WHEN mother-DAT beautiful seem-W.PST=Q
'Was (his) mother happy, when Musa came home?' (presupposing either 'mother was happy' or 'Musa came home')
not 'Did Musa come and was his mother happy?'

b. *Muusa c'a v-e'a-ch=ii, naan-na xaza xiiti-ra?* ('CVB.TEMP')
M.(V).NOM home V-come-WHEN=Q mother-DAT beautiful seem-W.PST
'Was (his) mother happy, when Musa came home? (presupposing 'mother was happy')'

The possibility of illocutionary marking is again independent of scope behavior: both the concessive and the temporal converb are limited to structures with disjunct scope, as shown by (24c) and (25a). Yet the concessive converb bans illocutionary marking (24d), while the temporal converb doesn't (25b). This shows that one cannot define a universally applicable notion of 'subordination' which would entail both disjunct scope and a ban on illocutionary marking on dependent clauses. Instead, scope behavior and the possibility of illocutionary marking are independent variables.

3.3.2 Tense and status marking

Another type of operator that is often relevant for distinguishing clause linkage types is tense or status (realis/irrealis) markers. When discussing tense scope in Section 3.2, we noted that a lack of tense markers in the dependent clause does not necessarily entail conjunct scope. In other languages, such as Amele, however, a ban on tense marking does go together with conjunct tense scope (Roberts 1988). This suggests that, like in the case of illocutionary force, scope and marking possibilities are independent variables in the typology of tense as well.

In some languages, tense marking in dependent clauses is possible, but at the same time subject to constraints on tense choices across clauses (in the spirit of Latin *consecutio temporum* rules). One example is Belhare *lo(k)*-constructions of the kind illustrated by (14a) and (15). The dependent clause must be in the subjunctive, but within this, it can be either in the past (*-a*) or nonpast (zero-marked). The choice must match the tense choice in the main clause, however (Bickel 1996:92f):

(26) Belhare (Bickel 1996)

a. *nam kus-a=lo kam cog-he-ŋa (*cou-ʔ-ŋa)* ('COM')
sun [3sS]set-SBJV.PST=COM work do-PST-1sS (do-NPST-1sS)
'I worked till the sun set.'

b. *nam ku=lo kam cou-ʔ-ŋa (*cog-he-ŋa)* ('COM')
sun [3sS]set[SBJV]=COM work do-NPST-1sS (do-PST-1sS)
'I will work till the sun sets.'

Another example comes from Hua, a Trans-New Guinea language from the Papuan highlands (and further Papuan examples are discussed by Foley in this volume):

(27) Hua (Trans-New Guinea: Kainantu-Gorokan; Haiman 1980:421)

a. *fu=mo d-mi-sa-ga-da u-gu-e.* ('DS-CHAIN')
pig=TOP 1sP-give-FUT-3pDS-1sAS go-FUT-1sDECL
'They will give me pork and then I will go.'

b. * *fu=mo d-mi-sa-ga-da u-e.* ('DS-CHAIN')
pig=TOP 1sP-give-FUT-3pDS-1sAS go[NFUT]-1sDECL
Intended: 'They will give me pork and so I went', *i.e.*, 'I went because they will give me pork.'

In one type of clause linkage, marked by an agreement paradigm that also signals switch-reference (here in the form of the third person, different subject marker *-ga*), tense marking is allowed in the dependent clause only if it matches the tense marking of the main clause. This is the case in (27a), but not in (27b). In another type of clause linkage in Hua, marked by an agreement paradigm without switch-reference coding (here in form of the third person plural marker *-ma*), no such constraint is imposed. Dependent clauses freely allow tense marking, regardless of main clause tense choices:

(28) Hua
fu=mo d-mi-ga-ma-da u-e. ('TOPIC')
pig=TOP 1sP-give-FUT-3pSUB-1sAS go[NFUT]-1sDECL
'They will give me pork and so I went', *i.e.*, 'I went because they will give me pork.'

Unconstrained tense marking in dependent clauses is traditionally associated with either 'co-ordination' and 'subordination', while the presence of constraints has sometimes been suggested as diagnostic of 'chaining' and 'converb' constructions (e.g. Foley 1986). However, in a number of languages, including languages from the Trans-New Guinea family, structures traditionally classified as 'chaining' freely allow tense marking. A case in point is Korafe, where the different subject marker in the following examples establish past tense reference in the dependent clause. The main clause can be independently marked as present or past tense:

(29) Korafe (Trans-New Guinea: Binanderean, Papua New Guinea; Farr 1999)

a. *mut-eno er-ira-re.* ('CHAIN')
give.1s-SEQ.REALIS.1sDS IPFV-go.DUR.PRS.3sIND-CURRENT.RELEVANCE
'I gave it and he is currently going.'

b. *mut-eno i-sira.* ('CHAIN')
give.1s-SEQ.REALIS.1sDS go.DUR-F.PST.3sIND
'I gave it and he went (two or more days ago)'

In Wambule, a language from the same Sino-Tibetan branch as Belhare, tense marking is allowed exclusively on chained, dependent clauses. Main clauses, by contrast, cannot be marked for tense.[5]

[5] This is likely to result from a regular process of nominalizing main clauses, following a general trend characteristic of the Sino-Tibetan family at large (Bickel 1999b). The nominalizer was eventually reanalyzed as an illocutionary marker, with some kind of 'assertive' or 'affirmative' function.

(30) Wambule (Sino-Tibetan: Kiranti; Nepal; Opgenort 2004)

a. *saiso am kam pa-si tum-nu-ma-kho lwa-nu-mei.* ('CHAIN')
yesterday DEM work do-INF finish-2s-PST-SEQ go-2s-ASS
'You finished doing this work yesterday and you went.'

b. *na hep ja:-ma-k tyaŋ iskul di-ŋ-m.* ('CHAIN')
previously cooked.grain eat[1s]-PST-SEQ from.now school move-1s-ASS
'I ate cooked grain before, and now I will go to school'

Both sentences in (30) contain dependent clauses marked as past tense, but in (30a), this combines with past tense reference in the main clause, while in (30b), the adverb *tyaŋ* 'from now on' suggests future tense reference in the main clause.

3.4 Categorical symmetry

Issues of finiteness and marking possibilities are not only relevant with regard to the dependent clause on its own, but also in relation to the marking possibilities of the main clause. The core issue is wether or not the dependent and the main clause allow the same range of categories, i.e. whether the construction is symmetrical or asymmetrical in this regard. The difference between symmetrical and asymmetrical clause linkage was first emphasized by Haiman (1980, 1985a,b) in work on Hua and other Papuan languages; a related concept of 'balanced' vs. 'deranked' linkage was developed by Stassen (1985) and Cristofaro (1998, 2003).

Hua constructions like those in (27), for example, are asymmetrical in that the dependent clause allows anticipatory subject marking, while the main clause does not allow this, but instead allows marking of illocutionary force. This contrasts with symmetrical constructions such as those in *-rohí* and *-rehí*, which signal two iterative events in alternation:

(31) Hua (Haiman 1980)
p-mi-rohí de-rehí. ('ALTER.ITER')
3pO-give-1sALTER.ITER eat-3pALTER.ITER
'I gave them some, they ate; I gave them some more: they ate; and so on.'

Note that the verb forms in this example are not fully finite in Hua: they lack tense and illocutionary force marking, and the agreement paradigm is greatly reduced (Haiman 1980). What matters for symmetrical linkage is that conjuncts allow the same range of categories to be marked, not whether they are finite or not.

The contrast between symmetrical and asymmetrical construction is often associated with clause linkage types based on conjunctions vs. affixes. In Amele, for example, constructions involving dependent verb morphology and switch-reference of the kind illustrated by (1), (16), or (20) is asymmetrical. Conjunctions like *fo* 'or' or *gba* 'but', by contrast, require symmetry in the range of categories expressed (also cf. (7)):

(32) Amele (Roberts 1987, 1988)

a. *Fred ʔum ho-i-an fo gbee gbila ho-i-a?* ('OR')
F. yesterday come-3s-Y.PST or NEG today come-3s-T.PST
'Did Fred come yesterday or today?'

b. *ʔud weg-i-me-ig fo ʔeed weg-i-me-ig fo fal-do-ig-a!* ('CHAIN')
sago weave-PRED-SS-2p or bamboo weave-PRED-SS-2p or fence-3sP-2sA-IMP
'Weave some sago or bamboo and fence it!.'

c. * *ija ja hudo-ʔo-min gba ugba sab mane-i-a.* ('BUT')
1s fire open-DS-1s but 3s food cook-1s-T.PST

d. *ija ja hud-ig-a gba ugba sab mane-i-a.* ('BUT')
1s fire open-1s-T.PST but 3s food cook-1s-T.PST
'I lit the fire but she cooked the food.'

In (32a), both conjuncts are finite and tense-marked, in (32b) both conjuncts are nonfinite and marked for 'same subject', enclosed by a repetition of *fo*. The ungrammatical sentence in (32c) shows that the conjunction *gba* is not compatible with an asymmetrical choice of categories, e.g. switch-reference marking in one conjunct but tense marking in the other. Only a symmetrical choice, as in (32d), is accepted.

While such an association of symmetry with conjunctions is common, it is not universal. In Belhare, for example, some conjunctions license asymmetrical, while others require symmetrical category choices. The conjunction *ki* allows the subjunctive category, as shown by (25b) above — a possibility that is not given for main clauses. This contrasts with the (enclitic) conjunction *cha*, which is used in a symmetrical constructions for expressing alternating events. In this case, both conjuncts must include verb forms from the same paradigm:

(33) Belhare
ŋka=na ten-he-ŋ=cha phenɟ-he-ŋ=cha. ('AND')
1sNOM=TOP hit-PST-1sA=ADD drive.off-PST-1sA=ADD
'I hit him and drove him off.'

Asymmetrical clause linkage allows the expression of different ranges of categories. What is still required, however, is that conjuncts are of the same type, e.g. both verb-headed clauses. This is different in what I call here 'constraint-free' linkage, where it is not required that the part of speech of conjuncts be matched. This is typical for sentential topic clauses (Marchese 1977, Haiman 1978, Bickel 1993, 1998, 1999a), such as those of Usan:

(34) Usan (Reesink 1987: 283ff)

a. *munon iyau wârâm-or eng um-orei ʔiyo?* ('TOPIC')
man dog 3sP.hit-3sF.PST TOP die-3sF.PST Q
'Given that the man hit the dog, did it die?'

b. *yâr-âb eng ye-nipat ʔur big-âr.* ('TOPIC')
come-SS TOP 1sP-step.over.SS money put-pIMP
'If you come, step over me and put your money (in the basket).'

c. *munon eng, wonou man soau is-orei.*
man TOP 3sPOSS garden landslide go.down-3sF.PST
'As for the man, his garden went down in the landslide.'

In (34a), *eng* marks a finite clause as providing the background against which the main clause question is to be understood. In (34b), the same conjunction marks a same-subject converb as

providing the relevant background, leading, as is often the case with topic clauses (Haiman 1978), to an interpretation as a conditional. In (34c), finally, the framework is provided by a bare noun (*munon* 'man'), in a typical topic construction. There is no reason to assume that the semantic contribution of the conjunction varies across these three cases; in all instances the conjunction defines the framework within which the rest of the clause is to be interpreted, fairly close to Chafe's (1976) definition of a topic as "a spatial, temporal, or individual framework within which the main predication holds."

3.5 Focus or questions in linked clauses and extraction possibilities

In many languages some types of linked clauses ban the occurrence of question words. It is sometimes claimed that such constraints can be derived from general extraction constraints (Ross 1967, Chomsky 1973) or from information structure constraints (Foley & Van Valin 1984, Van Valin 1995), and that on either account, the constraints differentiate between universally defined types of clause linkage (coordination vs. subordination). However, it seems best to treat question and focus possibilities as a variable that is independent of extraction possibilities, and not to tie any of these to other variables of clause linkage such as illocutionary scope behavior. The cross-linguistic variation is just too big for any reductionist approach. There are three pieces of evidence for keeping the variables distinct.

First, as argued by Foley & Van Valin (1984) and Van Valin (1995), a ban on questions in dependent clauses cannot be reduced to extraction constraints because such a ban may obtain even without extraction, i.e. where question words remain *in situ*. Foley & Van Valin (1984) and Van Valin (1995) demonstrate this with data from the Siouan language Lakhota, and a similar situation obtains in Tauya:

(35) Tauya (MacDonald 1990)

a. *ne-ni we tu-a-te yau-i=ne?* ('CHAIN')
3s-ERG who[NOM] [3sP]give-3s-DS [3sP]see-3p=PARAMETRICAL.Q
'Who did he give it to? and they saw him' ('Who did he give it to when they saw him?')

b. * *we mei fofe-a-nani=ra …* ('TOPIC')
who[NOM] here come-3s-ASS=TOP
Intended: '*if who did come here...'

Example (35a) is based on the same 'chaining' construction that was illustrated earlier in (5). In these constructions, *in situ* question words are possible inside the dependent clause (headed by *tuate* 'he gave it to someone and...'). This is in minimal opposition to 'topic' clause constructions of the kind observed earlier in (8) and illustrated here by (35b): in these constructions, the dependent clause does not allow question words. This constraint is imposed by the type of clause linkage, and does not result from any extraction constraint since question words are never extracted in the language.

Second, the domain of possible extraction sites may be different from the domain that defines where question words can appear, and this suggests that these are independent variables. In Chechen, for example, converbs in *-na* and *-alc* allow *in situ* question formation, just like their

Tauya counterparts in (35a):[6]

(36) Chechen (Good 2003)

a. *Maliika hu iec-na c'a j-e'a-ra?* ('CHAIN')
M.(J).NOM what buy-CVB house J-come-W.PST
'What did Malika buy and came home?'

b. *mila c'a j-all-alc Ahwmad irs d-ol-ush v-a-ra?* ('CVB')
who.NOM home J-come-UNTIL A.NOM happy D-be-CVB V-AUX-W.PST
'Ahmed was happy until who came home?'

But the same constructions block extraction in the form of relative clauses:

(37) Chechen (Molochieva 2008)

a. *Zaara koch ec-na c'a j-e'a-ra.* ('CHAIN')
Z.(J).NOM dress.NOM buy-CVB home J-come-W.PST
'Zara bought a dress and came home.'

b. * *Zaara(-s) ec-na c'a j-e'a-cha j-olu koch* ('CHAIN')
Z.(J).(-ERG) buy-CVB home J-come-CVB J-AUX.PTCP dress
Intended: 'the dress that Zara bought and came home.'

This is in direct contrast to Tauya, where the switch-reference marked dependent clauses illustrated by (35a) above do allow extraction in the form of relative clauses:

(38) Tauya (MacDonald 1990)

a. *ʔe fenaʔa-ni ya-tu-a-te te-pa yate-e-ʔa.* ('CHAIN')
DEM woman-ERG 1sP-give-3s-DS get-SS go-1s-DECL
'That woman gave it to me and I took it away.'

b. *ya-tu-a-te te-pa yate-e-na fenaʔa* ('CHAIN')
1sP-give-3s-DS get-SS go-1s-NMLZ woman
'the woman who gave it to me and I took it away'

Example (38b) is a relative construction based on (38a), where the sole argument of the dependent clause (*fenaʔa* 'the woman') is extracted as the head.

Third, the possibilities of *in situ* questions or focus formation are independent of other properties of clause linkage. For example, while Tauya bans *in situ* questions in conditional clauses (as noted in (35b) above), other languages allow this in otherwise very similar clause linkage constructions. This is so, for example, in Amele and Belhare conditional clauses as well as in 'when' and 'while' clauses:

[6] and since they also allow conjunct scope of illocutionary force operators, as shown in (10), the construction qualifies as 'cosubordinate' in Role and Reference Grammar (cf. Good 2003).

(39) Amele (Roberts 1987)
ugba sab man-i j-igi-an ja in hude-ʔe-b fi. ('COND')
s/he food fry-PRED eat-3s-FUT fire who open-DS-3s if
'She will cook and eat if who will light the fire?'

(40) Belhare (Bickel 1993)

a. *sa-a ya=m-phekt-a-k=naa ŋ-khatd-at-ni-gak=phe?* ('SUB')
who-[s]ERG call=3nsA-call-SBJV.PST-2=TOP NEG-go-PST-NEG-2=IRR
'You hadn't gone if who had called you?'

b. *sa-a lap-yukt-u=lo m-phig-he?* ('COM')
who-[s]ERG [3sA-]grab-hold-3sP=COM 3nsA-pick-PST[-3sP]
'They picked the fruit when who was holding down [the branch]?'

c. *hene yuŋ-sa n-lur-he-ga?* ('CVB')
where sit-CVB 3sA-tell-PST-2sP
'While where sitting did he tell you?'

Direct translations into English typically fail because a question word in English 'if' and 'when' clauses is interpreted as an echo question. But in the original, these are genuine questions.

Structures with no illocutionary scope constraint often freely allow question and focus words in dependent clauses as well. This is shown by the earlier example from Tauya in (35a) and the following data from various languages:

(41) Kâte (Trans-New Guinea: Huon, PNG; Schneucker 1962)
ŋohe wena yu-ha-pire goŋgoŋ dâŋe-yeʔ. ('CHAIN')
2d where be-SIM-2dDS bell sound-3sN.PST
'Where were you when the bell rang?'

(42) Belhare (Bickel 1993)

a. *laitar hene lept-he-ga ki sʌlai am-t-u-ga?* ('CHAIN')
lighter where throw-PST-2sA SEQ matches light-NPST-3sP-2sA
'Where did you throw the lighter so that you have to use matches?'

b. *han-na=cha tha n-tog-u-n-ga ki emgari mun dhup-ka?* ('CHAIN')
2s-ERG=ADD know NEG-know-3sP-NEG-2sA SEQ how talk talk-2sNPST
'Since YOU don't know it either, how can you talk [about it]?'

(43) Nepali (Clark 1963:169)
tyo alchī-le ke garī-kana kamā-era khāncha? ('CHAIN')
DEM lazy.fellow-ERG what do-CVB earn-CVB eat-3sNPST
'How does that lazy fellow earn a living?'

In some languages, there are also structures with extensible illocutionary scope that allow question words. This is the case in Chechen *na*-constructions, as noted in (37) above.

While these data showed that *in situ* question (and, thereby, focus) formation is independent of other clause linkage variables, the following show that the same is true of the possibilities for

question word extraction. Some languages, like English, block such extraction from 'adverbial clauses', while other languages allow it. The following example from Latin shows extraction of an object argument (*permulta* 'many') from a conditional clause (marked by *nisi* 'if not') into the pre-clausal position that is generally used in the language for forming relative clauses:

(44) Latin
permulta$_i$ [[*quae*$_i$ *orator a natura nisi haberet*] *non*
many.N.NOMp REL.N.ACC orator.NOMs by nature.ABLs if.not have.3sPST.SBJV NEG
multum a magistro adiuvaretur] (Cic., *de Or.* 1, 126)
much.N.ACCs by teacher.ABLs help.3sPST.SBJV.PASS
'many [properties] such that if the orator didn't have them by nature, he couldn't be much helped by a teacher.' (*Literally:* '*many properties [which$_i$ couldn't be helped by a teacher [if the orator didn't have t_i by nature]])'

Other languages distribute possibilities in a different way. In Tauya, for example, we noted in (38) that extraction is possible in switch-reference-marked chaining constructions, but it is not possible with topic clauses (of the kind illustrated in (8) and (35b) above):

(45) Tauya (MacDonald 1990:273,298)
a. *ʔe fanaʔa yau-e-nani=ra ne-ni wawiya ya-tu-a=ʔa.* ('TOPIC')
DEM woman [3sP-]see-1-ASS=TOP 3s-ERG mango 1sP-give-3s=DECL
'I saw the woman and she gave me a mango.'
b. * *yau-e-nani=ra ne-ni wawiya ya-tu-a-na fenaʔa* ('TOPIC')
[3sP-]see-1-ASS=TOP 3s-ERG mango 1sP-give-3s-NMLZ woman
Intended: 'the woman who I saw and who gave me a mango.'

The preceding examples all involved extraction in the form of relative clauses, but also the possibilities for afterthought extraction (i.e. a special type of right-dislocation) are sometimes subject to variation across types of clause linkage. In Belhare, for example, *ki*-clauses (as exemplified in 3, 17 and 42) regularly support the extraction of arguments into post-verbal positions, whereas topic clauses marked by *naa* (cf. 6 and 40) do not:

(46) Belhare
a. *chokt-he ki, n-celi-ŋa,*
[3sA-]point.with.finger-PST[3sP] SEQ 2sPOSS-unmarried.female.agnatic-ERG
doko-ep=phu chaŋ-da-he. ('CHAIN')
basket-LOC=REP [3sS-]curl.up-come-PST
'When she pointed with her finger at him, your celi, that is, [Rainbow] came and curled up in [her] basket.'
b. *u-tak-ŋa ya phekt-u=naa ŋ-khatd-at-ni.* ('SUB')
3sPOSS-friend-ERG call [3sA-]call-[SBJV]-3sP=TOP NEG-[3sS-]go-PST-NEG
'When his friend called him, he didn't go.'
c. * *ya m-phekt-u=naa, u-tak-ŋa, ŋ-khatdatni.* ('SUB')
call [3sA-]call[-SBJV]-3sP=TOP 3sPOSS-friend-ERG NEG-[3sS-]go-PST-NEG
'When they called him, his friend, that is, he didn't go.'

As we have seen in (40) above, Belhare *naa*-clauses allow *in situ* question formation. This confirms again that the possibilities of question formation are a variable independent of the variable regulating extraction constraints.

3.6 Focus marking on dependent clauses

In the preceding we concentrated on question and focus formation inside dependent clauses, but focus markers can of course also occur attached to the edge of such clauses, with a scope that typically includes the entire clause but does not extend beyond it The typical functions of such markers lie in restricting, adding, or contrasting propositions. Especially when tightly integrated into the main clause syntax, dependent clauses allow such focus markers. This is well attested in many Eurasian converbs, as amply illustrated by Haspelmath (1995), but it also extends to some structures that are less integrated into main clauses. This is exemplified by Belhare *ki*-constructions (cf. the example in (3) above):

(47) Belhare
cama ca-he ki=cha rʌksi uŋ-he. ('CHAIN')
food [3sS]eat-PST SEQ=ADD liquor [3sS]drink-PST
'He drinks liquor even after the meal.'

The possibility of focus marking on a dependent clause seems to be independent of the possibility of interrogative or focused constituents within the dependent clause. Belhare *ki*-clauses allow both focus markers on the clause, as just shown in (47), as well as question words inside the clause, as shown earlier in (42). But this is not universal: Chechen *na*-clauses allow question words inside the clause, as shown in (36a), but the same clauses do not allow focus markers attached to their edges, regardless of whether the clause contains a question word or not:

(48) Chechen (Zarina Molochieva, p.c.)
a. *Maliika hu iec-na(*='a) c'a j-e'a-ra?* ('CHAIN')
M.(J).NOM what buy-CVB(=FOC) house J-come-W.PST
'What did Malika buy and came home?'
b. *Maliika tyka-na='a j-agh-na(*='a) c'a j-e'a-ra.* ('CHAIN')
M.NOM store-DAT=SS J-go-CVB(=FOC) home J-come-W.PST
'Malika went to the store and then came back home.'

Whether or not focus markers are allowed on dependent clauses therefore seems to be again an independent variable. The possibility of focus marking perhaps depends on the exact semantics of the relevant focus markers, but in some cases it seems to depend on the form of the dependent clause. While, as just noted, Chechen *na*-clauses ban focus markers, *-cha*-clauses allow them, although the difference in meaning between these converbs is minimal:

(49) Chechen (Good 2003)
Maliika hu iec-cha='a c'a j-e'a-ra? ('CVB.TEMP')
M.(J).NOM what buy-WHEN=FOC house J-come-W.PST
'What did Malika buy and came home?'

The possibility of purely formal constraints is further illustrated by Burúshaski, a language for which Tikkanen (1995:514f) notes that its converbs cannot host any focus particle. This is in contrast with Indo-Aryan languages of the same region, like Hindi or Nepali, where such particles freely attach to converb clauses.

Especially in work on European languages, it has been suggested that focus assignment and topic-comment articulation in general constrains the scope of illocutionary force and similar operators (e.g. Jacobs 1984, 1991, and many others since). Preliminary evidence in Bickel (1993) and Schackow et al. (in press) suggests that this may not be universal. For example, from all we know, the presence or absence of the restrictive focus clitic *=ŋa* does not resolve the scope ambiguity of the illocutionary force operator in Puma:

(50) Puma (Schackow et al. in press)
risiwa=cha *mu-so(=ŋa)* *mʌ-ta-a=ku,* *bura-ci?*
shamanic.rhythm[NOM]=ADD do-CVB(=FOC) 3pS-come-PST=NMLZ old.man-ns[NOM]
'Did they come and play the drum?' (conjunct)
or: 'Did they play the drum while coming?' (only converbal clause in scope of question)
or: 'Playing the drum, did they come?' (only main clause clause in scope of question)

While more research is needed to firmly establish the nature of *=ŋa* as a focus marker, it seems best to assume for now that choices in focus marking and in illocutionary scope are in principle independent of each other. If focus and illocutionary scope are structurally linked in a language, this may necessitate positing distinct structures — for example one structure with focus on the main clause and local illocutionary or negation scope, as opposed to a construction with focus on both clauses and conjunct illocutionary or negation scope.[7]

3.7 Clause position

The position of dependent clauses and main clauses has traditionally been taken as a key variable in clause linkage types. In many types, the position is fixed, as either initial or final. An example is the Amele construction that was already illustrated in (1), (16) or (20). The dependent clause in this construction must always precede the main clause:

(51) Amele (Roberts 1988)
a. *ho busale-ʔe-b* *dana age gbo-ig-a.* ('CHAIN')
pig run.out-DS-3s man 3p hit-3p-T.PST
b. * *dana age gbo-ig-a* *ho busale-ʔe-b.* ('CHAIN')
man 3p hit-3p-T.PST pig run.out-DS-3s
'The pig ran out and the men kill it.'

The same condition obtains for Usan clauses marked by the conjunction *eng* (cf. 34 above):

[7] Thanks to Volker Gast for drawing my attention to this possibility.

(52) Usan (Reesink 1987)
wau e-âb igo-iner eng unor mâni u-t-i-b-â. ('TOPIC')
child cry-SS be-3sUNCERTAIN.FUT TOP mother yam 3sP-give-s-FUT-3s
'If the child is crying, his mother will give him yam.'

Dependent clauses marked by *end* 'because', by contrast, can appear both before or after the main clause, *salva veritate*:

(53) Usan (Reesink 1987)
a. *ya itum der igâm-a igo-i urigerm-a end irumban sig boru.* ('CAUSE')
rain night come.down be-3sDS be-CESS light-3sDS BECAUSE mud very bad
'Because it has been raining all night until daybreak, it is very muddy.'
b. *irumban sig boru, ya âib dâr-a end.* ('CAUSE')
mud very bad rain big come.down-3sDS BECAUSE
'It is very muddy; that's because a big rain came down.'

While such flexibility as with Usan *end*-constructions are traditionally associated with 'subordination', it can also be observed in some languages with constructions that are functionally closer to narrative 'chaining' uses. An example is Belhare *ki*-constructions. In most cases, *ki*-clauses appear before their main clause (cf. 3, 17, 42, 46a, and 47), but they can also follow the main clause, as in the following example:

(54) Belhare
a. *ca-ma=na ca-yau-t-u, tʌrʌ he-lleŋ leŋ kina?* ('CHAIN')
eat-INF=TOP [3sA-]eat-IPFV-NPST-3sP PTCL where-DIR [3sS-]direct SEQ
'It (the cow) is eating, but after having turned towards which direction?'
b. *pʌscim i-baŋ khat-ke, lotthi tekap cou-se ki* ('CHAIN')
west one-HUM [3sS]go-TEMP stick hold.on [3sA]do-PRF[3sP] SEQ
'He is going west, after having taken hold of the (walking) stick.'

For many constructions, positional flexibility is limited, however, in that the dependent clause must still be adjacent to the main clause. This is so in Belhare. By contrast, Chechen converbs in *-na*, which cover a similar range of narrative sequential uses as Belhare *ki*-clauses, are not subject to an adjacency constraint. The following data show some possible *salva veritate* permutations of two converb clauses and one main clause:

(55) Chechen (Good 2003)

a. *Maliika tyka-na='a j-agh-na zhejna='a iec-na c'a j-e'a-ra.* ('CHAIN')
M.NOM store-DAT=SS J-go-CVB book.NOM=SS buy-CVB home J-come-W.PST

b. *Maliika tyka-na='a j-agh-na c'a j-e'a-ra zhejna='a iec-na.* ('CHAIN')
M.NOM store-DAT=SS J-go-CVB home J-come-W.PST book.NOM=SS buy-CVB

c. *Maliika c'a j-e'a-ra tyka-na='a j-agh-na zhejna='a iec-na.* ('CHAIN')
M.NOM home J-come-W.PST store-DAT=SS J-go-CVB book.NOM=SS buy-CVB
'Malika went to the store, bought a book, and came back home.'

The example in (55c) shows that a converb clause need not be adjacent to the main clause that it refers to: here, the sequential converb suffix *-na* situates the event of book buying into a direct relation to the event expressed in the main clause ('coming home', here initial), not in the immediately adjacent clause ('going to the store', here the second clause). Such behavior is traditionally taken to be associated with 'subordination' and can often be observed with what translates English 'adverbial clauses'.

3.8 Layer of attachment

Adjoined clauses are not embedded in the sense of being arguments of main clause predicates. But they can attach to various levels inside a main clause. While more fine-grained distinctions are certainly possible, what seems to be the most important difference is between (a) clauses adjoining to a predicate or verb ('ad-V'), (b) clauses adjoining to entire clauses ('ad-S'), or (c), clauses that are adjoined to some higher level and appear as 'detached' from main clauses in the sense of König & van der Auwera (1988, 1990), König (1995) and others working on European clause linkage patterns.[8] A further possibility is for clauses to be adjoined not so much to a sentence but to an utterance, as in the case of speech-act modifying clauses like *frankly speaking*.

Of all these distinctions, the one that seems to be of widespread relevance is that between ad-V and ad-S constructions. The traditional term 'adverbial subordination' conflates the two, but many languages make sharp distinctions (Bickel 1991, 1993, 1998). Ad-V clauses perform the function of adverbial modifiers; ad-S clauses, by contrast, provide general frameworks for the main clause (as in 'if', 'when' or general 'topic' clauses) or sequences in an event chain. Ad-V clauses typically behave like ordinary adverbial constituents, often case-marked like NP constituents and entirely transparent to the assignment of case to other constituents of the main clause. The critical result of all this is that ad-V clauses can be center-embedded:

(56) Belhare

a. *Dhankuta him-yakt-a-lok=to khar-e.* ('COM')
D.[LOC] [3sS-]stumble-IPFV-PST.SBJV-COM=FOC [3sS]go-PST
'He went to Dhankuta stumbling.'

[8] also cf. the notion of 'left-detached' and 'right-detached' position in Role and Reference Grammar (Van Valin & LaPolla 1997, Van Valin 2005).

b. *pit-chi-lo ap-khar-ket.*
cow-ns-COM [3sS]come.on.the.same.level-go-TEMP
'She is passing by with the cows.'

The dependent clause in (56a) (also cf. 14a, 15, and 26) is marked by the same comitative case suffix as the NP in (56b). The NP *Dhankuta* in (56a) appears with a zero allomorph of the locative case rather than the regular overt locative in *-e*. This version of the locative is licensed by the main clause verb (and cannot be licensed by the dependent verb *him-* 'stumble'), which suggests that the dependent clause does not interfere with case and semantic role assignment. This is different with ad-S constructions, which block case assignments and therefore center-embedding:

(57) Belhare

a. *u-chom pok=naa Dhankuta kha?-yu.* ('SUB')
3sPOSS-desire [3sS]rise[SBJV]=TOP D.[LOC] [3sS]go-NPST

b. * *Dhankuta u-chom pok=naa kha?-yu.* ('SUB')
D.[LOC] 3sPOSS-desire [3sS]come.up[SBJV]=TOP [3sS]go-NPST
'If he wants, he will go to Dhankuta.'

Because of the intervening ad-S clause *u-chom pok=naa* 'if he wants', the main clause verb cannot assign case to the first NP (*Dhankuta*) in (57b). Here, the ad-S clause is a topic clause, but the same blocking effect obtains if the ad-S clause is a sequential chain:

(58) Belhare

a. * *a-tak [sʌppʌi mai-mat-pir-he ki] khar-e.* ('CHAIN')
1sPOSS-friend[NOM] all[NOM] 1s-[3sA]narrate-BENEF-PT SEQ [3sS]go-PST

b. *[sʌppʌi mai-mat-pir-he ki] a-tak khar-e.* ('CHAIN')
all[NOM] 1s-[3sA]narrate-BENEF-PT SEQ 1sPOSS-friend[NOM] [3sS]go-PST
'My friend told me everything and went off.'

Nominative case on *atak* 'my friend' in (58a) can only be assigned the verb form *khare* 'went', but this is impossible because of the intervening *ki*-clause. If this clause is to be included, it must precede the entire main clause (*atak khare*), as in (58b).

In the absence of case, patterns of semantic role assignment may be the only reflex of ad-S attachment. This is illustrated by Amele constructions of the kind exemplified earlier in (1), (16) and (51):

(59) Amele

a. * *dana age [ho busale-ʔe-b] gbo-i-ga.* ('CHAIN')
man 3p pig run.out-DS-3s hit-3p-T.PST
'The men, the pig having run out, killed it.'

b. *[ho busale-ʔe-b] [dana age gbo-ig-a].* ('CHAIN')
pig run.out-DS-3s man 3p hit-3p-T.PST
'The pig ran out and the men killed it.'

Sentence (59a) is ungrammatical because *dana age* 'the men' cannot be assigned a semantic role: the NP belongs to the main verb (*gboiga* 'they killed it'), but role assignment by this verb is blocked by the intervening dependent clause (*ho busale?eb* 'the pig ran out and'), which itself has no extra role to assign. The sentence becomes grammatical in (59b) where the dependent clause is not center-embedded but precedes the main clause containing both *dana age* and the main verb.

3.9 Other variables

The preceding discussion suggested a number of variables that are logically independent of each other. There are other variables. Some of these, such as the difference between adjoined and subcategorized (argumental) dependent clauses are outside the scope of this survey. Another variable that has received great prominence is syndesis, i.e. the question of whether clauses are linked by affixes, independent conjunctions and/or by intonational means. This is clearly an important issue, but it derives by and large from the general morphological and phonological structure of the language and is perhaps not as relevant to syntactic issues as has traditionally been assumed.

An important class of variables relates to issues of cross-clausal coreference. This includes constraints on backward anaphora (which has often been taken to be diagnostic of clause linkage types), control and raising patterns (which is critical in distinguishing various complement clause types), and the presence of reference-tracking devices such as switch-reference morphology, cross-clausal reflexivization or logophoric pronouns. A survey of the relevant variables would take us far beyond what can be covered in this chapter.

Another issue that goes beyond the scope of this survey is the nature of the interpropositional relation that is realized by a clause linkage construction. Developing variables for this is a tall order because especially temporal relationships are tightly connected with the aspectual system of a language, and reference grammars typically do not provide sufficient detail in order to understand these connections.[9]

4 Typological patterns

A set of typological variables like the one discussed here raises two questions: (i) Are some structures more similar to each other than to others so that they define cross-linguistic type clusters? (ii) Which variables are correlated, forming statistical implicational universals? In the following, I discuss methods that allow answering these questions, based on standards in other disciplines. To illustrate the methods, I apply them to a pilot database of adjoined clause linkage structures.

4.1 A pilot database

I sampled languages based on what analyses I had readily available from an earlier survey (Bickel 1991), and this leads to a bias towards Eurasia, Papua New Guinea and Africa. Therefore, any

[9] Another topic that is is difficult to survey in reference grammars is verb gapping, i.e. whether pro-verbs are allowed or not, as in *John cooked dinner, and so did Harry*.

results are preliminary and will have to be tested against a larger and better balanced database. The pilot database contains data on 69 constructions from 24 languages and codes each construction for eleven variables, plus one variable registering the set of markers used. For ease of reference, each construction is identified by a short label. The full database is reproduced in the Appendix.

Within each language, constructions are differentiated by unique combinations of markers and the levels across all variables. The set of variables largely follows the discussion in Section 3, but I exclude negation scope because relevant data was not available with sufficient analysis. I also use a simplified version of the finiteness variable. The following definitions of levels were applied, listing only levels that are actually attested in the database:

ILL-scope: The scope of illocutionary operators in the main clause is
- CONJUNCT: extends to the main clause and the dependent clause
- DISJUNCT: extends to either the main or the dependent clause but never to both
- LOCAL: is limited to the main clause
- EXTENSIBLE: extends to either the main clause alone or to both the main clause and the dependent clause, but never to the dependent clause alone
- CONSTRAINT-FREE: is not regulated by the clause linkage type

T-scope: The scope of tense or status operators in the main clause is
- CONJUNCT: extends to the main clause and the dependent clause
- LOCAL: is limited to the main clause
- EXTENSIBLE: extends to either the main clause alone or to both the main clause and the dependent clause, but never to the dependent clause alone

Finiteness: The dependent clause is headed by a verb form that is
- FINITE: at least as many categories must be marked as in main clauses
- NONFINITE: only fewer categories are allowed
- ANY: either the same range or less categories can be marked

ILL-mark: Marking of illocutionary force operators in the dependent clause is
- OK: allowed
- BANNED: not allowed
- HARMONIC: allowed but only if it matches the marking on the main clause

T-mark: Marking of tense or status operators in the dependent clause is
- OK: allowed
- BANNED: not allowed
- HARMONIC: allowed but subject to constraints based on the tense or status choice in the main clause

Symmetry: The range of categories that can be expressed on linked clauses is
- SYMMETRICAL: must match
- ASYMMETRICAL: can be different
- FREE: can be different and can even include elements of different type (different parts of speech, clauses and NPs, etc.)

WH: Question words and constituent focus inside dependent clauses are
- OK: allowed
- BANNED: not allowed

Extraction: Extraction of elements of dependent clauses is

OK: allowed
BANNED: not allowed

FOC: Focus marking on dependent clauses is
OK: allowed
BANNED: not allowed

Position: The position of the dependent clause vis-à-vis the main clause with which it enters a dependency relation is
FIXED:POST-MAIN: is fixed and is always after the main clause
FIXED:PRE-MAIN: is fixed and is always before the main clause
FLEXIBLE-ADJACENT: can be before or after the main clause but must be adjacent to it
FLEXIBLE-RELATIONAL: can be before or after the main clause and can be separated from the main clause by other dependent clauses

Layer: The dependent clause adjoins
AD-V: to the predicate and can be center-embedded
AD-S: to the clause and cannot be center-embedded
DETACHED: to the clause but is separated syntactically and intonationally

For current purposes I treat all these variables as unstructured multinomial variables. This is a simplification since in some cases, there may be an underlying structure — for example, 'local' and 'extensible' scope are arguably more similar to each other than either is to 'disjunct' or 'conjunct' scope; or the levels 'banned', 'harmonic', and 'ok' could be modeled in terms of degrees of restrictiveness, i.e. as a rank variable.[10] I leave exploration of such possibilities to future research.

4.2 Prototypes and Scales

A key question is to what extent there are cross-linguistic clusters or 'prototypes' of clause linkage patterns — that is, types of structures to which language-specific constructions bear similarity to a quantifiable degree. To find out, we need methods for measuring similarity between individual constructions and for aggregating the similarities of all constructions at once.

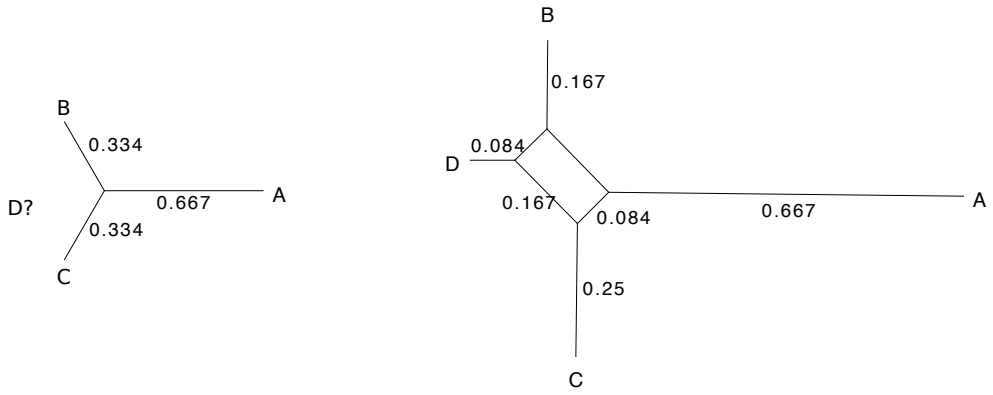
4.2.1 Methods

A standard measure of similarity used in many fields is the inverse relative Hamming distance (also known as the Gower coefficient; see Kaufman & Rousseeuw 1990): the distance or dissimilarity between two constructions is equal to the proportion of different levels in all non-empty variables. For example, according the pilot database in the Appendix, the difference between the Amele 'but'-construction and 'chain'-construction is 5/8 = .625 since there are eight nonempty (non-NA) variables and five of them have different levels. The distance between the Amele 'but'-construction and the Belhare 'and'-construction is smaller, viz. 3/8 = .375, and this captures the intuition that these constructions are relatively similar to each other. Some structures may even be identical. For

[10] Thanks to Jeff Good and Volker Gast for reminding me of this.

example, from all we know, Chantyal and Burúshaski 'chain'-constructions have identical properties in all non-empty variables; this results in a relative Hamming distance of 0.

Since the database contains 69 constructions, computing Hamming distances of all pairs of constructions results in a matrix of $\frac{69\cdot(69-1)}{2}$=2,346 pairs, similar in kind to distance charts found in road maps (except that there can be 0-length distances). The challenge is now to find patterns in these 2,346 pairs by searching for clusters of elements that are closer together across all pairwise distances.[11] One way of going about this is by plotting all distances at once in a graph. However, distances in a 69 × 69 matrix can be faithfully plotted only in 68-dimensional space, which is not visually accessible. All reductions to two- or three-dimensional space are non-trivial and necessarily impose some degree of distortion. For example, if we have three elements with the distances DIST(A,B)=1, DIST(A,C)=1, and DIST(B,C)=2/3, this can be faithfully plotted in two-dimensional space (cf. Figure 1a). If we now add an element D with distances DIST(A,D)=1, DIST(B,D)=1/3, and DIST(C,D)=1/2, this immediately becomes impossible. For example, if we try to respect the distance of D to C (which is 1/2) and B (which is 1/3), we necessarily violate the distance to A, which is 1. One popular class of solution in typology and psychology, Multidimensional Scaling, seeks a geometrical arrangement with an optimal balance between minimal distortions and a minimal number of dimensions. Often, however, no good balance can be found and all reasonably low-dimensional solutions (two- or three-dimensional ones) are heavily distorted (cf. Cysouw 2007).



(a) Simple graph *(b)* Split graph

Figure 1: Representing all pairwise distances between four items in two-dimensional space.

An alternative solution, which I adopt in the following, consists in simultaneously plotting competing distances in what is known as a split graph (Bandelt & Dress 1992). This is shown in Figure 1b, where the lines show the possible paths one can take while respecting the true distances:

[11] Except where noted, all computations here and in the following were done in R (R Development Core Team 2009), with the additional packages `cluster` (Maechler et al. 2005), `vcd` (Meyer et al. 2009a), `entropy` (Hausser & Strimmer 2009b), and `minet` (Meyer et al. 2009b).

each true distance between elements is the sum of the indicated (rounded) lengths on the shortest connecting path. As a result, the less distances (i.e. dissimilarities) are in conflict, the less there are splits and elements cluster into clear groups. If there is more conflict, there are more splits, such as the parallelogram in Figure 1b, and the evidence for clusters is weaker: in Figure 1b, for example, there is no particular cluster except for a trend towards separating B and D from all else (since DIST(B,D) is smaller than any other distance in the example). Split graphs of this kind can be efficiently computed by an algorithm known as the NeighborNet algorithm (and all split graphs here and in the following are produced with the program SplitsTree4 implementing this algorithm; see Bryant & Moulton 2004, Huson & Bryant 2006).[12]

A general problem with similarity analysis (whether done computationally or by hand) is that it works best if data are complete, with no missing values: as soon as one variables lacks a value (marked as 'NA' in the Appendix) for a given construction, the construction can no longer be compared to these other constructions with regard to this variable, even if the other constructions have a value. As a result, a considerable part of the information in the database cannot be exploited in the analysis. Dissimilarity methods are known to be fairly robust and not very sensitive to the amount of information used. But it is clear that they can only be used as a heuristic and need to be backed up by other methods (cf. Section 4.3).

4.2.2 Results

Traditional conceptions of clause linkage would lead one to expect two or three clearly distinct clusters, representing 'coordination', 'adverbial subordination', and possibly 'cosubordination' or 'chaining'. The split graph in Figure 2 casts doubt on the this.[13] There is some degree of clustering of structures in the top left region of the graph that one might loosely associate with 'adverbial subordination', including some 'topic' clauses (e.g. in Godié or Nepali) as well English or German 'when' and 'if' clauses (labeled as 'SUB' in the graph and the Appendix). But this excludes ad-V structures, which are dissimilar in many regards and are therefore placed in a different region, the lower right.

One cluster that emerges as somewhat more pronounced is a group of 'and'-like structures on the left side of the graph. Interestingly, these structures share a significant proportion of properties with 'topic'-constructions in Tauya and also with detached finite clauses in German. Moreover, they are closer to the 'when', 'if' and 'topic' constructions in the top left region than to the chaining-like structures that are placed at the other end of the graph on the right. The positioning of 'and'-constructions closer to 'topic', 'when' and 'if' clauses than to chaining-like structures seems to be mostly caused by the fact that these structures all have flexible position, symmetrical category

[12] `http://www.splitstree.org/`. Split graphs are usually applied to genetic data, both biological and linguistic. Distances then represent language change. For another application to typology, see Cysouw (2008).

[13] The graph follows exactly the same principles as the 'toy' graph in Figure 1b, but now applied to the whole dataset in the Appendix. For readability, the numerical distances are left off the edges of the graph, but the relative length of the edges is proportional to these distances.

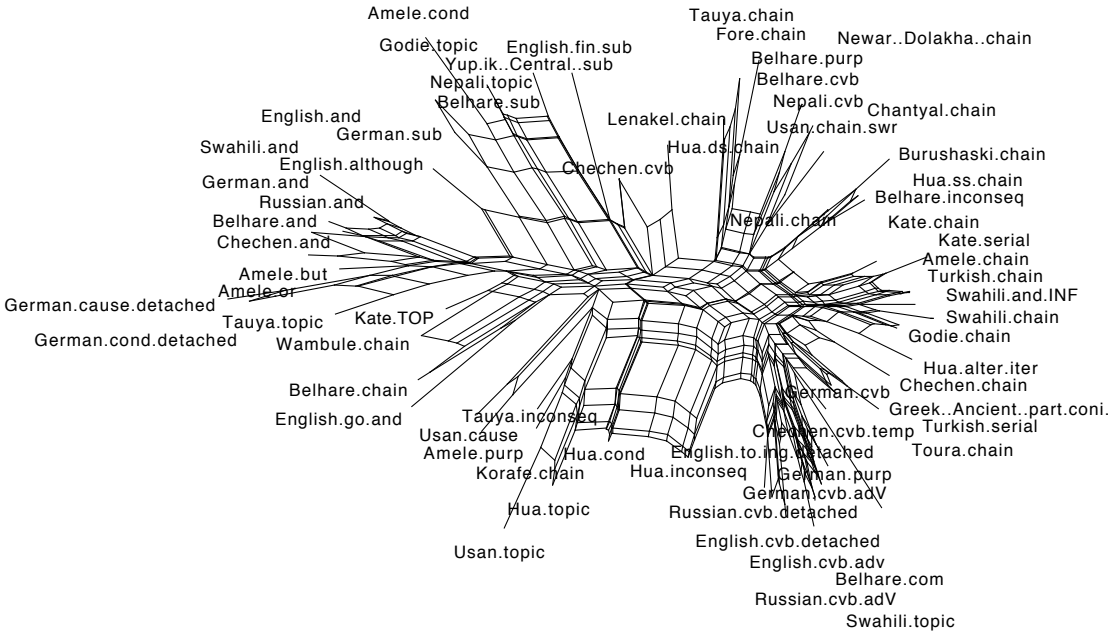


Figure 2: A NeighborNet representation of clause linkage patterns based on the relative Hamming distances between the constructions in the pilot database. Nodes represent individual constructions and are identified by the language name and the constructional label given in the Appendix.

marking, and involve finite clauses, whereas chaining-like constructions are mostly nonfinite and asymmetrical.

The right side of the graph contains various converb and chaining structures. The way these are arranged suggests a possible continuum between two sub-clusters, arranged by the extent to which illocutionary force scope is constrained: the cluster at the top right of the graph contains structures with less such constraints (e.g. Fore or Hua different subject chains) than the cluster in the mid-right region (e.g. Amele, Swahili or Turkish chains). Interestingly, Belhare and Wambule chains (both Kiranti), but no others, are closer to 'and'-coordination, in line with their finiteness.

4.3 Correlations

Clusters such as those emerging from the similarity analysis in the preceding result from co-varying variables, where particular levels on one variable (i.e. specific constructional properties) are associated with particular levels on other variables. Indeed, traditional cross-linguistic notions can in fact be thought of as property bundles in which properties are strictly correlated and fully entail each other. For example, a notion like 'subordination' entails (at least) disjunct illocutionary scope, the availability of restrictive focus, and a ban on extraction and question formation. The evidence discussed in Section 3 suggests that the variables surveyed are all logically independent of each other. Therefore, there is no reason to expect exceptionless correlations (also cf. Lehmann 1988). However, there may be statistical trends, and it is such trends that underly the possible clusters that we observed in the similarity analysis in Figure 2. Thus, to the extent that there are probabilistic

clusters based on specific kinds of subordination and coordination, as suggested by Figure 2, we can expect corresponding correlations of the properties defining these clusters.

The difference between traditional property bundles and probabilistic correlations is the same as the difference between absolute and statistical (or 'empirical') implicational universals. As property bundles, traditional cross-linguistic notions have the form 'A $\leftrightarrow$ B' (where A and B are properties), and they suffer from the same validation problem as absolute universals (Bickel in press): no language sample can guarantee that the universal is without exceptions because we cannot survey all languages that have ever been or will ever been spoken. The only available route to justification is logical deduction. Apparent counter-examples (e.g. constructions that show disjunct illocutionary scope but allow questions, such in Amele (39) or Belhare (40)), then need to be re-analyzed (e.g. by positing suitable underlying structures with different scope properties), or the definition of the notion (the property bundle) needs to be revised. Both options bring us back to the discussion in the introductory section and the problems associated with finding universal definitions of property bundles. Probabilistic correlations have the same structure as statistical universals, and so they can be validated in the same way, by statistical evidence and significance testing.

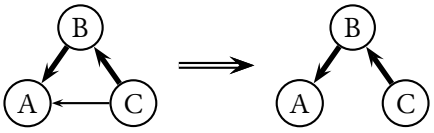
4.3.1 Methods

With eleven variables there are $\frac{11\cdot(11-1)}{2}$=55 possible pairs of variables. Given these possibilities, we need a heuristic technique to find those pairs with substantial correlations. One method for finding these would simply consist in performing statistical tests for each pair of correlations, but this leads to well-established problems with spurious success rates that arise from multiple testing (known as 'familywise errors' in the statistics literature). An alternative heuristic is based on information theory and has been developed for exploring associations between genes.[14] The key idea is to estimate what is called the mutual information between two (categorical) variables, which can be informally thought of as the extent to which knowledge of the frequency distribution in one variable (i.e. knowledge of the actual counts of each level defined by the variable) allows prediction of the frequency distribution in the other variable. Estimates of mutual information can be normalized by the frequency distribution in one of the variables, and this yields an estimate of the extent to which the frequency distributions in one variable can be predicted by the frequency distribution of the other variable. This can be represented as $\pi(A|B)$, i.e. the probability of the frequency distribution of A given the frequency distribution of B. π(A|B) captures the intuition behind a classical implicational universal of the form 'B $\rightarrow$ A' and, conversely, $\pi(B|A)$ is the equivalent of what is traditionally written as 'A $\rightarrow$ B', and if $\pi(A|B)$ is about the same as $\pi(B|A)$, we find what is traditionally written as 'A $\leftrightarrow$ B'.[15]

[14] I am grateful to Korbinian Strimmer for drawing my attention to this technique as well as to Margolin et al. (2006) and Hausser & Strimmer (2009a).

[15] Formally, the mutual information (MI) of variables A and B is defined in terms of the Shannon entropy (H) of A and B and the entropy of the joint probability distributions of A and B: *MI(A;B) = H(A) + H(B) − H(A,B)*. The probability $\pi(A|B)$ is then defined as $\frac{MI(A;B)}{H(A)}$. The entropy H itself describes the extent to which the probabilities of the levels of a variable are skewed. For a variable A with levels $a_i \in \{a_{1\ldots k}\}$ and associated level probabilities p_{a_i},

When computing implications between many pairs of variables in this way, one risks detecting artifacts caused by chains of associations. For example, if C predicts B and B predicts A, then C predicts A redundantly. What we are interested in is the strongest of these associations, i.e. the ones with the highest probability $\pi(X|Y)$. A powerful method to achieve this is suggested by Margolin et al. (2006) for bioinformatic purposes and consists in removing all implications with the lowest values on $\pi(X|Y)$ given an association of X and Y via some other term. Assume that the implication $B \rightarrow A$ is associated with $\pi(A|B)$ and that for the implications $C \rightarrow A$ and $C \rightarrow B$ we observe that $\pi(A|C) < \pi(B|C)$. Then, $C \rightarrow A$ is an artifact of the stronger implications and we set $\pi(A|C) = 0$, i.e. we effectively remove the link:

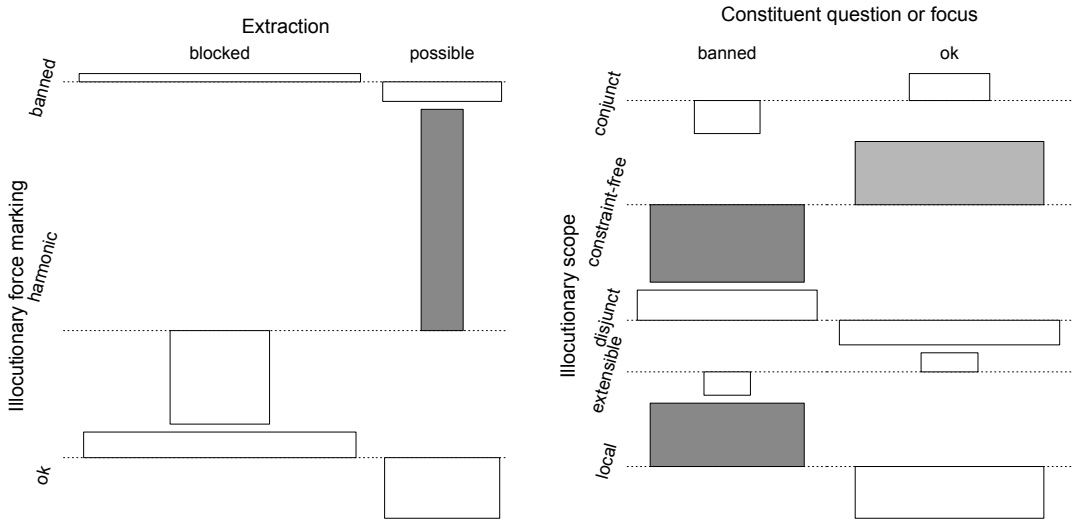


Once implications and their strengths are established, the question arises as to which particular levels cause them. The mutual information method only shows which variables are likely to be correlated (e.g. 'ILL-scope → WH'), but for analyzing clusters, it is more important to know which exact levels (properties) are behind this (e.g. 'local illocutionary scope → question words banned'). A standard solution rests on the analysis of the Pearson residuals of each cell in a contingency table defined by the variables of interest (e.g. the cell at the intersection of 'local' and 'ok' in a contingency table defined by the variables 'ILL-scope' and 'WH'). Informally, the Pearson residual describes the extent to which the observed frequency of a given cell in a contingency table deviates (positively or negatively) from what can be expected under the null hypothesis of no association, i.e. what one would expect in this cell if the proportions of level counts were constant across rows and columns (e.g. so that the overall .4 vs. .6 proportion of banned vs. allowed question words in the data is the same for all levels of the 'ILL-scope' variable, i.e. for local scope in the same way as for conjunct scope and all others).[16] Adopting methods developed by Zeileis et al. (2007) and Meyer et al. (2006), I use a permutation test to establish which residuals exceed what one would find under random reshuffling of cell counts. If one finds the observed residuals in less than 5% (or 1%) reshufflings, they point to statistically significant associations of specific levels (at what is called a 5% or 1% significance level). Note that it is possible that while the overall table shows a significant association (because the total sum of residuals is high), no individual residual may be significant or several (or all) are significant (because the residuals spread uniformly over the cells). In this case, it

$H(A) = -\sum p_{a_i} \log(p_{a_i})$. *H(A)* is zero if there is a total bias towards a single level, e.g. with p_{a_1}=1, p_{a_2}=0, and p_{a_3}=0; it reaches its maximum in uniform distributions, e.g. with p_{a_1}=.33, p_{a_2}=.33, and p_{a_3}=.33. The entropy of the joint probability distribution of *A* and *B* is $H(A,B) = -\sum p_{a_i,b_i} \log(p_{a_i,b_i})$. The probabilities p_{a_i} and p_{a_i,b_i} can be estimated by the relative level counts ($p_{a_1} = \frac{N(a_1)}{N(a_{1\ldots k})}$ etc.) in the database, but this method (the 'Maximum Likelihood' method) is known to be unreliable when individual level frequencies are relatively low, as is often the case in typological data with small sample size. Alternative methods are discussed by Hausser & Strimmer (2009a), and in this paper I use the James-Stein-type estimator developed there and implemented in the `R` package `entropy`.

[16] Formally, the Pearson residual is the relative contribution of a cell in a contingency table to the χ^2 sum of the table and is defined as $\frac{n-\hat{\mu}}{\sqrt{\hat{\mu}}}$, where n is the observed count and $\hat{\mu}$ the expected count.

is impossible to establish which specific properties are responsible for the overall correlation without expanding the dataset considerably, or by collapsing distinctions in the definitions of the variables.



(a) ILL-mark → extraction *(b)* ILL-scope → WH

Figure 3: Residual analysis: box heights are proportional to the residuals and box width to the number of datapoints, the direction of boxes from the baseline signals their sign (raising boxes for positive, falling boxes for negative residuals), and shading indicates that residuals are individually significant under a permutation test, with light grey at a 5% and dark grey at a 1% significance level (see Meyer et al. 2006). The order of levels in the plot is arbitrary (alphabetical).

The plots in Figure 3 illustrate the method for two of the implications found, 'ILL-mark → extraction' and 'ILL-scope → WH'. Figure 3a suggests that if a construction allows illocutionary marking on dependent clauses only if the marking matches that in the main clause ('harmonic ILL-mark'), then it is likely (at a 1% significance level, dark shading) that constituents can be extracted. In other words, extracting constituents is significantly facilitated by harmonic illocutionary force marking. Structures illustrating this include English 'go and'-constructions, where extraction is possible and which allow, say, imperatives if they apply to both conjuncts (cf. *the beer I went to the store and bought* and *Go and buy beer!*; Stassen 1985, Lakoff 1986). Another example is Belhare *ki*-constructions (cf. (23) for illocutionary harmony and (46a) for extraction). In Figure 3a, no other residuals are significant. Figure 3b shows three significant associations. First, there is a highly significant association (at a 1% significance level, dark shading) between local illocutionary scope and a ban on constituent question formation, i.e. question formation is banned more often when scope is local than when it is conjunct, extensible, or constraint-free. The same trend can be observed if scope is disjunct: although not statistically significant on its own, the positive residuals suggest that structures with disjunct scope appear to ban question formation more often than what one would expect under the null hypothesis of no association. Second, constraint-free scope is negatively associated with a ban on question formation, i.e. if there is no constraint on illocutionary

scope it is unlikely that question formation is banned (as is the case for example in Belhare *ki*-construction, as shown by by (42) and (54)). This is mirrored in the third association, which is the positive association between constraint-free scope and allowed question formation. This association is individually significant only at a lower (5%) significance level (light shading), which suggests that there is somewhat less evidence for clause linkage with constraint-free scope explicitly allowing question words than there is evidence for the finding that this type of clause linkage tends to lack an explicit ban on question words.[17]

4.3.2 Results

Table 1 summarizes the findings on all correlations which have an MI-based probability $\pi(X|Y)$ that is not set to zero after removing all artifacts. Where there is evidence from unequal probabilities, the implications are formulated in a directed way ('$\rightarrow$'); where the probabilities are identical or near-identical, the correlations are formulated as two-way implications ('$\leftrightarrow$'). The table also reports the results from the residual analysis, where '*' indicates significance at a 5% and '**' at a 1% level. No entry means that none of the residuals was significant, i.e. the pilot database does not allow identifying which specific properties are responsible for the overall association of the variables. In those cases where an additional (but non-significant) trend in the same direction as that of a significant association can be discerned (as was the case with disjunct scope in Figure 3b), this is added in brackets.

Most of the findings in Table 1 receive straightforward theoretical interpretations and are good candidates for genuine universals that deserve testing against larger and less biased datasets.

One set of associations concerns scope and marking possibilities. First, there are direct links between these two properties (ILL-scope $\leftrightarrow$ ILL-marking, conjunct T-scope $\leftrightarrow$ banned T-marking, local or extensible T-scope $\leftrightarrow$ allowed T-marking). Although these links are not logically necessary, many languages seem to conventionalize the pragmatic principle that a category value is maintained as long as there is no explicit marking against this, and, vice-versa, that the explicit marking of a category blocks the scope of other markers of the same category. Supporting evidence for this comes from the association of conjunct tense or status scope with ad-V, but not ad-S structures (conjunct T-scope $\leftrightarrow$ ad-V): ad-V structures are often center-embedded within the main clause and thus a category value needs to be maintained only for a shorter time (in processing) than in ad-S structures, which cannot be center-embedded (by definition).

T-marking, but not ILL-marking, possibilities are also associated with finiteness (banned T-marking $\leftrightarrow$ nonfinite, freely or harmonically allowed T-marking $\leftrightarrow$ finite). This may reflect a universal trend to the effect that tense and status have a higher probability of being realized through verb morphology than illocutionary force, which is more commonly realized through clitics and particles on the sentence level. While many theories assume that illocutionary force is on a more external (peripheral) syntactic 'shell' than tense and status (e.g. Foley & Van Valin 1984, Van Valin

[17] This reflects a fundamental descriptive problem: if there are no constraint on question formation in a language, linguists are naturally tempted not to discuss question formation at all in that language – but then we don't know whether question formation is allowed or simply not studied.

Implication	$\hat{\pi}(X\|Y)$	$\hat{\pi}(Y\|X)$	*Statistically significant associations of properties*
ILL-scope ↔ ILL-mark	.09	.05	
ILL-scope → WH	.18	.08	local (or disjunct) → banned**
			constraint-free → ok*
ILL-scope → FOC	.22	.06	
ILL-scope ↔ position	.23	.20	constraint-free ↔ flexible-adjacent**
			disjunct ↔ flexible-relational**
T-scope ↔ T-mark	.29	.26	conjunct ↔ banned**
			local (extensible) ↔ok**
T-scope ↔ symmetry	.21	.17	local ↔ symmetrical (or constraint-free)**
			extensible ↔ constraint-free*
T-scope → FOC	.11	.05	
T-scope ↔ layer	.09	.06	conjunct ↔ ad-V*
ILL-mark → extraction	.22	.12	harmonic → possible**
T-mark ↔ finiteness	.31	.28	banned ↔ nonfinite*
			ok (or harmonic) ↔ finite*
symmetry ↔ finiteness	.28	.28	constraint-free ↔ any (or finite)*
			symmetrical ↔ finite**
symmetry ↔ ILL-mark	.13	.12	symmetrical ↔ ok (or harmonic)*
position → layer	.12	.06	

Table 1: Implications detected by analysis of the relative mutual information between variables in the pilot database (see Appendix) and by permutation tests of the Pearson residuals.

2005), there is in fact no clear motivation for this because just like illocutionary force, tense and status are properties of entire propositions or even entire texts, and there is no sense in which they are properties of predicates. Perhaps this an arbitrary fact of our language faculty, but at any rate, the issue calls for further research if the trend observed here is robust.

Marking possibilities and finiteness are furthermore connected with regard to symmetry: if a clause linkage type requires the same amount of marking, this involves typically finite structures and allows independent or harmonic ILL-marking, and vice-versa (symmetrical ↔ finite, symmetrical ↔ freely or harmonically allowed ILL-marking). If a clause linkage type does not select a specific type of dependent (e.g. allows NPs as much as clauses), it imposes no constraints on the kind of morphology (e.g. finite) when the dependent is a clause (constraint-free symmetry ↔ 'any' or finite).

Scope properties of illocutionary force are moreover linked to question formation (local or disjunct ILL-scope ↔ banned question formation, constraint-free ILL-scope ↔ allowed question formation). Structures with local ILL-scope often seem to block dependent clauses from any modulation of illocutionary force, assimilating them to presupposed, 'taken for granted' topic expressions, as Foley (1986) has noted for Papuan languages. In turn, the flexible kind of illocutionary scope behavior found in chaining-like constructions significantly often allows question formation (cf. Section 3.5).

Some of the correlations in Table 1 are responsible for the clusters that emerged from the similarity analysis in Figure 2: the cluster around 'topic', 'when', and 'if' clauses noted in the upper left of the figure is characterized by the association of disjunct ILL-scope, local T-scope, finite dependent clauses, symmetry, and flexible-relational position (disjunct ILL-scope ↔ flexible-relational position, local T-scope ↔ symmetrical, symmetrical ↔ finite). There is in addition a statistically non-significant trend for disjunct ILL-scope to be associated with a ban on question formation or constituent focusing (disjunct ILL-scope → banned WH, but see e.g. (39) or (40) for counter-examples). The properties of symmetry and finiteness are shared by the adjacent cluster of 'and'-constructions, but in addition, these structures are characterized by flexible (constraint-free) illocutionary scope, flexible-adjacent position and allowing ILL-marking and question formation (constraint-free ILL-scope ↔ flexible-adjacent, symmetrical ↔ freely or harmonically allowed ILL-marking, constraint-free ILL-scope ↔ allowed question formation). Harmonic ILL-marking characterizes specifically Belhare *ki*-constructions ('chaining') and English *go and*-constructions, which show up as closely related in Figure 2 (lower left). This type of 'and'-like structure also allows extraction (harmonic ILL-marking ↔ allowing extraction).

5 Conclusions

In this chapter I have argued that cross-linguistic diversity in clause linkage is far too big as to be amenable to traditional typologies which seek to define a couple of universal notions that entail sets of properties without exception (cf. Lehmann 1988). What is needed instead is an approach similar to what is standard in other disciplines studying diversity: sets of variables capturing the variation and probabilistic assessment of clusters and correlations. Methods for doing this are readily available and have a solid and well-tested mathematical foundation.

Applied to a pilot study of 69 constructions from 24 languages, this approach has revealed a set of possible implicational universals that now need to be tested against larger datasets, with systematic control of genealogical inheritance and areal spread factors, before they can be fully explained. The approach proposed here also allows examining the evidence for probabilistic clusters, or typological 'prototypes' of constructions. The pilot database casts doubt on universal cluster concepts close to traditionally expected structures like 'coordination', 'subordination', or perhaps 'cosubordination'. Instead, there is tentative evidence for a specific prototype of 'subordination', which tends to associate disjunct illocutionary scope, local tense scope, finite and symmetrical clauses, and flexible position (and, with slightly less probability, also a ban on question formation or focusing inside the dependent clause). This cluster is in close neighborhood not to chaining structures but to 'and'-like constructions, with which it shares symmetry and finiteness, but differs by having more constraints on illocutionary scope, illocutionary marking and question formation or constituent focusing inside the dependent clause. Chaining constructions do not form a tight cluster but are spread out in a continuum of structures with more vs. less tightly constrained scope properties.

The advantage of a multivariate approach as suggested here is that it brings typology closer to language-specific analysis. Instead of being separated enterprises (as proposed by Lazard (2006) or Haspelmath (2007)), typological survey work is identical to language-specific analysis and consists in detailed descriptions of properties. To the extent that we learn more about language-specific properties, the better can we define fine-grained typological variables. And such fine-grained vari-

ables provide in return the ideal questionnaires for fieldwork. No fieldwork can proceed out of a theoretical vacuum. But what this chapter has shown is that none of the received theoretical concepts in clause linkage fills this vacuum with the right questions: it makes little sense to debate whether a particular construction is subordinate or not because such a notion is far too coarse to be universally applicable. But it makes sense to ask whether the construction has disjunct illocutionary force scope, or whether it allows question formation. Such questions are indeed much closer to the level of granularity that is needed in field linguistics. In return, only by answering such questions in great detail can we develop full datasets, with a minimum of missing data. And this is a precondition for establishing and understanding implicational universals, cross-linguistic prototypes and continua.

References

Bandelt, Hans-Jürgen & Andreas Dress, 1992. A canonical decomposition theory for metrics on a finite set. *Advances in Mathematics* 92, 47 – 105.

Bearth, Thomas, 1986. *L'articulation du temps et de l'aspect dans le discours Toura*. Bern: Lang.

Bickel, Balthasar, 1991. *Typologische Grundlagen der Satzverkettung*. Zürich: ASAS.

Bickel, Balthasar, 1993. Belhare subordination and the theory of topic. In Ebert, Karen H. (ed.) *Studies in clause linkage*, 23 – 55. Zürich: ASAS.

Bickel, Balthasar, 1996. *Aspect, mood, and time in Belhare*. Zürich: ASAS.

Bickel, Balthasar, 1998. Converbs in cross-linguistic perspective [review article of Haspelmath and König, eds., *Converbs*, Berlin: Mouton de Gruyter 1995]. *Linguistic Typology* 2, 381 – 397.

Bickel, Balthasar, 1999a. From *ergativus absolutus* to topic marking in Kiranti: a typological perspective. *Proceedings of the 25th Annual Meeting of the Berkeley Linguistics Society* 38 – 49.

Bickel, Balthasar, 1999b. Nominalization and focus constructions in some Kiranti languages. In Yadava, Yogendra P. & Warren W. Glover (eds.) *Topics in Nepalese linguistics*, 271 – 296. Kathmandu: Royal Nepal Academy.

Bickel, Balthasar, 2007. Typology in the 21st century: major current developments. *Linguistic Typology* 11, 239 – 251.

Bickel, Balthasar, in press. Absolute and statistical universals. In Hogan, Patrick Colm (ed.) *The Cambridge Encyclopedia of the Language Sciences*. Cambridge: Cambridge University Press [pre-print available at `http://www.uni-leipzig.de/~bickel/research/papers/universals_cels_bb.pdf`].

Bickel, Balthasar & Johanna Nichols, 2002. Autotypologizing databases and their use in fieldwork. In Austin, Peter, Helen Dry, & Peter Wittenburg (eds.) *Proceedings of the International LREC Workshop on Resources and Tools in Field Linguistics, Las Palmas, 26 - 27 May 2002*. Nijmegen: MPI for Psycholinguistics.

Bryant, David & Vincent Moulton, 2004. Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution* 21, 255 – 265.

Chafe, Wallace L., 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Li, Charles N. (ed.) *Subject and topic*, 27 – 55. New York.

Chomsky, Noam, 1973. Conditions on transformations. In Anderson, Stephen R. & Paul Kiparsky (eds.) *A Festschrift for Morris Halle*. New York: Holt, Rinehart and Winston.

Clark, T.W., 1963. *Introduction to Nepali*. Kathmandu: Ratna Pustak Bhandar (Reprint).

Cristofaro, Sonia, 1998. Deranking and balancing in different subordination relations: a typological study. *Sprachtypologie und Universalienforschung* 51, 3–42.

Cristofaro, Sonia, 2003. *Subordination*. Oxford: Oxford University Press.

Croft, William, 2001. *Radical construction grammar: syntactic theory in typological perspective*. Oxford: Oxford University Press.

Cysouw, Michael, 2007. Building semantic maps: the case of person-marking. In Miestamo, Matti & Bernhard Wälchli (eds.) *New challenges in typology: broadening the horizons and redefining the foundations.*, 225–248. Berlin: Mouton de Gruyter.

Cysouw, Michael, 2008. Generalizing scales. In Malchukov, Andrej & Marc Richards (eds.) *Scales (Linguistische ArbeitsBerichte 86)*, 379 – 396. Leipzig: Institut für Linguistik.

Farr, Cynthia J., 1999. *The interface between syntax and discourse in Korafe, a Papuan language of Papua New Guinea*. Canberra: Pacific Linguistics.

Foley, William A., 1986. *The Papuan languages of New Guinea*. Cambridge: Cambridge University Press.

Foley, William A. & Robert D. Van Valin, Jr., 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

Forker, Diana, 2009. Converbs in Hinuq and some remarks on linguistic concepts. Ms. Max-Planck-Institute for Evolutionary Anthropology, Leipzig.

Genetti, Carol, 2005. The participial construction in Dolakha Newar: syntactic implications of an Asian converb. *Studies in Language* 29, 35 – 87.

Good, Jeff, 2003. Clause combining in Chechen. *Studies in Language* 27, 113 – 170.

Green, Georgia M., 1976. Main clause phenomena in subordinate clauses. *Language* 52, 382 – 397.

Haiman, John, 1978. Conditionals are topics. *Language* 54, 564 – 589.

Haiman, John, 1980. *Hua: a Papuan language of the Eastern Highlands of New Guinea*. Amsterdam: Benjamins.

Haiman, John, 1985a. *Natural syntax*. Cambridge: Cambridge University Press.

Haiman, John, 1985b. Symmetry. In Haiman, John (ed.) *Iconicity in syntax*, 73 – 95. Amsterdam: Benjamins.

Hale, Austin & Kedār P. Shrestha, 2006. *Newār: Nepāl bhāsā*. Muenchen: LINCOM.

Haspelmath, Martin, 1995. The converb as a cross-linguistically valid category. In Haspelmath, Martin & Ekkehard König (eds.) *Converbs in cross-linguistic perspective*, 1 – 55. Berlin: Mouton de Gruyter.

Haspelmath, Martin, 2007. Pre-established categories don't exist: consequences for language description and typology. *Linguistic Typology* 11, 119 – 132.

Hausser, Jean & Korbinian Strimmer, 2009a. Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks. *Journal of Machine Learning Research* 10, xx–xx.

Hausser, Jean & Korbinian Strimmer, 2009b. `entropy`: entropy and mutual information estimation. `R` package, `http://www.R-project.org/`.

Horn, Laurence R., 1989. *A natural history of negation*. Chicago: The University of Chicago Press.

Huson, D. H. & D. Bryant, 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 32, 254 – 267.

Jacobs, Joachim, 1984. Funktionale Satzperspektive und Illokutionssemantik. *Linguistische Berichte* 91, 25–58.

Jacobs, Joachim, 1991. Focus ambiguities. *Journal of Semantics* 8, 1–36.

Johanson, Lars, 1995. On Turkic converb clauses. In Haspelmath, Martin & Ekkehard König (eds.) *Converbs in cross-linguistic perspective*, 313 – 347. Berlin: Mouton de Gruyter.

Kaufman, Leonard & Peter J. Rousseeuw, 1990. *Finding groups in data: an introduction to cluster analysis*. New York: Wiley.

König, Ekkehard, 1995. The meaning of converb constructions. In Haspelmath, Martin & Ekkehard König (eds.) *Converbs*, 57 – 95. Berlin: Mouton de Gruyter.

König, Ekkehard & Johan van der Auwera, 1988. Clause integration in German and Dutch conditionals, concessive conditionals, and concessives. In Haiman, John & Sandra A. Thompson (eds.) *Clause combining in grammar and discourse*. Amsterdam: Benjamins.

König, Ekkehard & Johan van der Auwera, 1990. Adverbial participles, gerunds and absolute constructions in the languages of Europe. In Bechert, Johannes, Giuliano Bernini, & Claude Buridant (eds.) *Toward a typology of European languages*, 337 – 355. Berlin: Mouton de Gruyter.

Lakoff, George, 1984. Performative Subordinate Clauses. *Proceedings of the 10th Annual Meeting of the Berkeley Linguistics Society* 472–480.

Lakoff, George, 1986. Frame-semantic control of the Coordinate Structure Constraint. *Proceedings of the 20th Annual Meeting of the Chicago Linguistics Society* 152–167.

Lazard, Gilbert, 2006. *La quête des invariants interlangues: la linguistique est-elle une science?* Paris: Champion.

Lehmann, Christian, 1988. Towards a typology of clause linkage. In Haiman, John & Sandra A. Thompson (eds.) *Clause combining in grammar and discourse*, 181 – 226. Amsterdam: Benjamins.

Lynch, John, 1978. *A grammar of Lenakel*. Canberra: Pacific Linguistics (PL - B55).

MacDonald, Lorna, 1988. Subordination in Tauya. In Haiman, John & Sandra A. Thompson (eds.) *Clause combining in grammar and discourse*, 227 – 246. Amsterdam: Benjamins.

MacDonald, Lorna, 1990. *A grammar of Tauya*. Berlin: Mouton de Gruyter.

Maechler, Martin, Peter Rousseeuw, Anja Struyf, & Mia Hubert, 2005. `cluster`: cluster analysis basics and extensions. `R` package, `http://www.R-project.org/`.

Marchese, Lynell, 1977. Subordinate clauses as topics in Godié. In Mould, Martin & Thomas J. Hinnebusch (eds.) *Papers from the 8th Conference on African linguistics*, 157 – 164. Los Angelese: University of California.

Margolin, Adam, Ilya Nemenman, Katia Basso, Chris Wiggins, Gustavo Stolovitzky, Riccardo Favera, & Andrea Califano, 2006. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a Mammalian cellular context. *BMC Bioinformatics* 7, S7.

Meyer, D., A. Zeileis, & K. Hornik, 2006. The `strucplot` framework: visualizing multi-way contingency tables with `vcd`. *Journal of Statistical Software* 17, 1 – 48.

Meyer, David, Achim Zeileis, & Kurt Hornik, 2009a. `vcd`: visualizing categorical data. `R` package, `http://www.R-project.org/`.

Meyer, Patrick E., Frederic Lafitte, & Gianluca Bontempi, 2009b. `minet`: mutual information network inference. `R` package, `http://www.R-project.org/`.

Molochieva, Zarina, 2008. Scope properties of Chechen converbs. Handout of paper given at the Syntax of the World's Languages III conference, Berlin, September 25-28, 2008.

Noonan, Michael, 1999. Converbal constructions in Chantyal. In Yadava, Yogendra P. & Warren G. Glover (eds.) *Topics in Nepalese linguistics*, 401–420. Kathmandu: Royal Nepal Academy.

Olson, Michael L., 1981. *Barai clause juncture: toward a functional theory of inter-clausal relations*. Ph.D. thesis, Australian National University.

Opgenort, Jean Robert, 2004. *A Grammar of Wambule*. Leiden: Brill.

Pilhofer, G., 1933. *Grammatik der Kâte-Sprache in Neuguinea*. Berlin: Reimer.

R Development Core Team, 2009. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, `http://www.r-project.org`.

Rappaport, Gilbert C., 1984. *Grammatical function and syntactic structure: the adverbial participle of Russian*. Columbus, Ohio: Slavica.

Reed, Irene, Osakito Miyako, Steven Jacobsen, Paschal Afcan, & Michael Krauss, 1977. *Yup'ik Eskimo grammar*. Fairbanks: Alaska Native Language Center, University of Alaska.

Reesink, Ger P., 1987. *Structures and their functions in Usan*. Amsterdam: Benjamins.

Roberts, John R., 1987. *Amele*. London: Croom Helm.

Roberts, John R., 1988. Amele switch-reference and the theory of grammar. *Linguistic Inquiry* 19, 45 – 64.

Ross, John R., 1967. *Constraints on variables in syntax*. Ph.D. thesis, MIT, Cambridge, Mass.

Schackow, Diana, Balthasar Bickel, Shree Kumar Rai, Narayan P. Sharma (Gautam), Arjun Rai, & Martin Gaenszle, in press. Morphosyntactic properties and scope behavior of 'subordinate' clauses in Puma (Kiranti). In Gast, Volker & Holger Diessel (eds.) *Clause-combining in cross-linguistic perspective*. Berlin: Mouton de Gruyter [pre-print available at `http://www.uni-leipzig.de/~autotyp/download/schackowetal2009puma.pdf`].

Schneucker, Carl L., 1962. *Kâte language handbook*. Madang: Lutheran Mission.

Scott, Graham, 1978. *The Fore language of Papua New Guinea*. Canberra: Pacific Linguistics (PL-B47).

Stassen, Leon, 1985. *Comparison and universal grammar*. Oxford: Blackwell.

Suter, Edgar, 1992. *Satzverbindung im Kâte*. Master's thesis, University of Zürich.

Takahashi, Hidemitsu, 2008. Imperatives in concessive clauses: compatibility between constructions. *Constructions* 2.

Tikkanen, Bertil, 1987. *The Sanskrit gerund: a synchronic, diachronic, and typological analysis*. Helsinki: The Finnish Oriental Society.

Tikkanen, Bertil, 1995. Burushaski converbs in their South and Central Asian areal context. In Haspelmath, Martin & Ekkehard König (eds.) *Converbs in cross-linguistic perspective*, 487 – 528. Berlin: Mouton de Gruyter.

Van Valin, Robert D., Jr., 1995. Towards a functionalist account of so-called extraction constraints. In Devriendt, Betty, Louis Goossens, & Johan van der Auwera (eds.) *Complex structures: a functionalist perspective*, 29 – 60. Berlin: Mouton de Gruyter.

Van Valin, Robert D., Jr., 2005. *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press.

Van Valin, Robert D., Jr. & Randy J. LaPolla, 1997. *Syntax: structure, meaning, and function*. Cambridge: Cambridge University Press.

Zeileis, Achim, David Meyer, & Kurt Hornik, 2007. Residual-based shadings for visualizing (conditional) independence. *Journal of Computational and Graphical Statistics* 16, 507–525.

Appendix: Pilot database on clause adjoining (NA = information not available)

Language	Label	Marker	ILL-scope	T-scope	Finiteness	ILL-mark	T-mark	Symmetry	WH	Extraction	FOC	Position	Layer	References and examples
Amele	BUT	*gba*	local	local	finite	banned	ok	symmetrical	NA	NA	NA	fixed:pre-main	ad-S	Roberts (1987, 1988), (7), (34c-d)
Amele	CHAIN	*-me, -ʔV*	conjunct	conjunct	nonfinite	banned	banned	asymmetrical	NA	NA	NA	fixed:pre-main	ad-S	Roberts (1987, 1988), (1), (16), (20), (32b), (51), (59)
Amele	OR	*fo ~ o*	conjunct	local	finite	ok	ok	symmetrical	NA	NA	NA	fixed:pre-main	ad-S	Roberts (1987, 1988), (32a)
Amele	COND	*fi*	disjunct	extensible	any	ok	ok	constraint-free	ok	NA	NA	flexible-relational	ad-S	Roberts (1987, 1988), (39)
Amele	PURP	*nu*	disjunct	extensible	any	banned	ok	asymmetrical	NA	NA	NA	fixed:pre-main	ad-S	Roberts (1987, 1988)
Belhare	AND	*=cha...=cha*	constraint-free	local	finite	ok	ok	symmetrical	NA	banned	banned	flexible-adjacent	ad-S	Bickel (1993), (32)
Belhare	CHAIN	*ki(na)(huŋ)*	constraint-free	extensible	finite	harmonic	ok	asymmetrical	ok	possible	ok	flexible-adjacent	ad-S	Bickel (1993), (3a), (17), (23), (42), (46a), (47), (54), (58)
Belhare	INCONSEQ	*-kone*	local	extensible	nonfinite	ok	banned	asymmetrical	ok	banned	ok	flexible-adjacent	ad-S	Bickel (1993)
Belhare	SUB	*-naa, -huŋ(go/do/cha), =cha*	disjunct	extensible	finite	banned	ok	constraint-free	ok	banned	ok	flexible-relational	ad-S	Bickel (1993), (6), (40a), (46b-c), (57)
Belhare	CVB	*-sa*	constraint-free	conjunct	nonfinite	ok	banned	asymmetrical	ok	banned	ok	flexible-adjacent	ad-V	Bickel (1993), (3b), (14b), (40c)
Belhare	PURP	*-si*	constraint-free	conjunct	nonfinite	ok	banned	asymmetrical	ok	banned	ok	flexible-adjacent	ad-V	Bickel (1993)
Belhare	COM	*-lo(k)*	local	conjunct	finite	banned	harmonic	asymmetrical	ok	banned	ok	flexible-relational	ad-V	Bickel (1993), (14a), (26), (40b), (56)
Burúshaski	CHAIN	*n(V)-ˊ-Σ-(i)n*	constraint-free	extensible	nonfinite	NA	banned	asymmetrical	NA	NA	banned	flexible-adjacent	ad-S	Tikkanen (1995), (13)
Chantyal	CHAIN	*-si, -rə*	NA	extensible	nonfinite	NA	banned	asymmetrical	NA	NA	NA	flexible-adjacent	ad-S	Noonan (1999), (18)
Chechen	CHAIN	*-na, -(u)sh, -i*	extensible	conjunct	nonfinite	banned	banned	asymmetrical	ok	banned	banned	flexible-relational	ad-S	Molochieva (2008), (10), (24), (36a), (37), (48), (55)
Chechen	CVB	*-nach, -alc* etc.	disjunct	local	nonfinite	banned	ok	asymmetrical	ok	banned	ok	flexible-relational	ad-S	Molochieva (2008), (36b)
Chechen	CVB.TEMP	*-lie, -cha*	disjunct	conjunct	nonfinite	ok	banned	asymmetrical	ok	banned	ok	flexible-relational	ad-S	Molochieva (2008), (25), (49)
Chechen	AND	*t'q'a*	local	local	finite	ok	ok	symmetrical	ok	banned	NA	flexible-adjacent	ad-S	Molochieva (2008)
English	AND	*and, but*	constraint-free	local	finite	ok	ok	symmetrical	ok	banned	ok	flexible-adjacent	ad-S	Ross (1967)
English	GO-AND	*and*	conjunct	conjunct	finite	harmonic	ok	symmetrical	ok	possible	ok	flexible-adjacent	ad-S	Lakoff (1986), etc.
English	FIN.SUB	*if, when* etc.	disjunct	conjunct	finite	banned	harmonic	symmetrical	banned	banned	ok	flexible-relational	ad-S	
English	CVB-DETACHED	*-ing*	local	conjunct	nonfinite	banned	banned	asymmetrical	banned	banned	ok	fixed:post-main	detached	
English	ALTHOUGH	*although*	disjunct	local	nonfinite	ok	ok	symmetrical	banned	banned	ok	flexible-relational	ad-S	
English	TO.ING.DETACHED	*to, -ing*	local	conjunct	nonfinite	banned	banned	asymmetrical	banned	banned	ok	fixed:pre-main	detached	
English	CVB-ADV	*-ing*	disjunct	conjunct	nonfinite	banned	banned	asymmetrical	banned	possible	ok	fixed:post-main	ad-V	
Fore	CHAIN	*-ki*	constraint-free	conjunct	finite	ok	harmonic	asymmetrical	NA	NA	NA	flexible-adjacent	ad-S	Scott (1978), (22)
German	SUB	*weil, wenn*	disjunct	local	finite	banned	ok	constraint-free	banned	banned	ok	flexible-relational	ad-S	
German	CAUSE.DETACHED	*weil*	local	local	finite	ok	ok	symmetrical	banned	banned	banned	fixed:post-main	detached	
German	PURP	*um zu, ohne zu*	disjunct	conjunct	nonfinite	banned	banned	asymmetrical	ok	banned	ok	flexible-relational	ad-V	
German	CVB-ADV	*-nd*	disjunct	conjunct	nonfinite	banned	banned	asymmetrical	banned	banned	ok	fixed:post-main	ad-V	
German	AND	*und*	constraint-free	local	finite	ok	ok	symmetrical	ok	banned	ok	flexible-adjacent	detached	
German	COND.DETACHED	*wenn*	local	local	finite	ok	ok	symmetrical	banned	banned	banned	fixed:pre-main	detached	
German	CVB	*-nd*	disjunct	conjunct	nonfinite	banned	banned	asymmetrical	banned	banned	ok	flexible-relational	ad-S	
Godié	CHAIN	*yi*	conjunct	conjunct	nonfinite	banned	banned	asymmetrical	NA	NA	NA	flexible-adjacent	ad-S	Marchese (1977)
Godié	TOPIC	*nʌ*	disjunct	local	finite	banned	ok	constraint-free	NA	NA	NA	flexible-relational	ad-S	Marchese (1977)
Greek (Ancient)	PART.CONJ.	*-men, -nt*	disjunct	extensible	nonfinite	banned	banned	asymmetrical	NA	NA	NA	flexible-relational	ad-S	(19)
Hua	ALTER.ITER	*-rohI, -rehI*	conjunct	conjunct	nonfinite	banned	banned	symmetrical	NA	banned	NA	flexible-adjacent	ad-S	Haiman (1980), (31)
Hua	DS-CHAIN	*-ga*	constraint-free	conjunct	nonfinite	banned	harmonic	asymmetrical	ok	NA	NA	fixed:pre-main	ad-S	Haiman (1980), (27)
Hua	TOPIC	*-ma*	local	local	nonfinite	banned	ok	constraint-free	banned	NA	NA	fixed:pre-main	ad-S	Haiman (1980), (28)
Hua	INCONSEQ	*-mana*	local	local	nonfinite	banned	banned	asymmetrical	banned	NA	NA	fixed:pre-main	ad-S	Haiman (1980)
Hua	COND	*-si*	local	local	nonfinite	banned	ok	asymmetrical	banned	NA	NA	fixed:pre-main	ad-S	Haiman (1980)
Hua	SS-CHAIN	∅	conjunct	conjunct	nonfinite	banned	banned	asymmetrical	ok	NA	NA	fixed:pre-main	ad-S	Haiman (1980)
Korafe	CHAIN	*-o, -i,* ∅	constraint-free	local	NA	NA	ok	asymmetrical	NA	NA	NA	fixed:pre-main	ad-S	Farr (1999), (29)
Kâte	CHAIN	*-ra, -me, -ku, ...*	constraint-free	conjunct	nonfinite	banned	banned	asymmetrical	ok	NA	NA	fixed:pre-main	ad-S	Pilhofer (1933), Suter (1992), (41)
Kâte	SERIAL	∅	conjunct	extensible	nonfinite	banned	banned	asymmetrical	NA	NA	NA	fixed:pre-main	ad-S	Pilhofer (1933), Suter (1992)
Kâte	TOPIC	*mutsaʔ*	disjunct	local	finite	NA	ok	asymmetrical	ok	NA	NA	NA	ad-S	Pilhofer (1933), Suter (1992)
Lenakel	CHAIN	*m-*	conjunct	local	nonfinite	banned	ok	asymmetrical	NA	NA	NA	flexible-adjacent	ad-S	Lynch (1978)
Nepali	CHAIN	*-era*	constraint-free	conjunct	nonfinite	banned	ok	asymmetrical	ok	NA	ok	flexible-adjacent	ad-S	fieldnotes, (4)
Nepali	TOPIC	*bhane*	disjunct	extensible	finite	banned	ok	constraint-free	ok	NA	ok	flexible-relational	ad-S	fieldnotes
Nepali	CVB	*-era*	constraint-free	conjunct	nonfinite	banned	ok	asymmetrical	ok	NA	ok	flexible-adjacent	ad-V	fieldnotes, (43)
Newar (Dolakha)	CHAIN	*-en*	constraint-free	NA	nonfinite	NA	banned	asymmetrical	NA	NA	ok	flexible-adjacent	ad-S	Genetti (2005)
Russian	AND	*i*	constraint-free	local	finite	ok	ok	symmetrical	ok	banned	ok	flexible-adjacent	ad-S	
Russian	CVB-ADV	*-a, -v*	disjunct	conjunct	nonfinite	banned	banned	asymmetrical	banned	banned	ok	flexible-relational	ad-V	
Russian	CVB.DETACHED	*-a, -v*	disjunct	conjunct	nonfinite	banned	banned	asymmetrical	banned	banned	ok	fixed:pre-main	detached	Rappaport (1984)

language	Label	Marker	ILL-scope	T-scope	Finiteness	ILL-mark	T-mark	Symmetry	WH	Extraction	FOC	Position	Layer	References and examples
Swahili	CHAIN	*ka-*	conjunct	conjunct	nonfinite	banned	banned	asymmetrical	NA	NA	NA	fixed:post-main	ad-S	fieldnotes, (2)
Swahili	AND.INF	*na*	conjunct	conjunct	nonfinite	banned	banned	asymmetrical	NA	NA	NA	fixed:post-main	ad-S	fieldnotes
Swahili	AND	*na*	conjunct	local	finite	ok	ok	symmetrical	NA	NA	NA	flexible-adjacent	ad-S	fieldnotes
Swahili	TOPIC	*ki*	disjunct	conjunct	nonfinite	banned	banned	asymmetrical	NA	NA	ok	flexible-relational	ad-S	fieldnotes
Tauya	CHAIN	*-pa, -te/-fe/-tefe*	constraint-free	conjunct	nonfinite	ok	harmonic	asymmetrical	ok	possible	NA	fixed:pre-main	ad-S	MacDonald (1990), (5), (37a), (40)
Tauya	INCONSEQ	*-na*	local	local	nonfinite	ok	ok	asymmetrical	NA	banned	NA	fixed:pre-main	ad-S	MacDonald (1990)
Tauya	TOPIC	*=ra*	local	NA	finite	ok	NA	constraint-free	banned	banned	NA	fixed:pre-main	ad-S	MacDonald (1990), (8), (35b), (45)
Toura	CHAIN	*le*	conjunct	conjunct	nonfinite	banned	banned	asymmetrical	NA	NA	NA	flexible-adjacent	ad-S	Bearth (1986)
Turkish	SERIAL	*∅*	conjunct	conjunct	nonfinite	banned	banned	asymmetrical	NA	NA	NA	fixed:pre-main	ad-S	fieldnotes
Turkish	CHAIN	*-Ip*	conjunct	conjunct	nonfinite	ok	banned	asymmetrical	NA	NA	ok	fixed:pre-main	ad-S	Johanson (1995), (11a), (21)
Usan	CHAIN.SWR	NA	extensible	conjunct	nonfinite	NA	ok	asymmetrical	NA	NA	NA	flexible-adjacent	ad-S	Reesink (1987), (9)
Usan	TOPIC	*eng*	local	local	any	banned	ok	constraint-free	NA	possible	NA	fixed:pre-main	ad-S	Reesink (1987), (34a-b), (52)
Usan	CAUSE	*end*	disjunct	NA	any	NA	NA	asymmetrical	NA	NA	NA	flexible-adjacent	ad-S	Reesink (1987), (53)
Wambule	CHAIN	*-kho*	constraint-free	local	finite	NA	ok	asymmetrical	NA	NA	NA	flexible-adjacent	ad-S	Opgenort (2004), (30)
Yup'ik (Central)	SUB	NA	disjunct	extensible	nonfinite	banned	ok	constraint-free	NA	NA	NA	flexible-relational	ad-S	Reed et al. (1977)