
Volume 119, Issue 11, November 2009
ISSN 0024-3841
ELSEVIER
Lingua
International Review of General Linguistics
Special Issue Section
The Forests behind the Trees
Guest Edited by John Nerbonne
Lingua available on Science Direct from Volume 1, published 1949
Available online at
ScienceDirect
www.sciencedirect.com

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Lingua 119 (2009) 1589–1606
Lingua
www.elsevier.com/locate/lingua

Relevance of typology to minimalist inquiry

Jan-Wouter Zwart

Department of Linguistics, University of Groningen, P.O. Box 716, NL-9700 AS Groningen, The Netherlands

Received 1 October 2006; received in revised form 11 September 2008; accepted 11 September 2008
Available online 28 November 2008

Abstract

This article considers to what extent sample based typological research may shed light on questions of central concern within the minimalist program of generative grammar. Focusing on the nature of the structure building operation assumed within the minimalist program (Merge), it is argued that typological data may be relevant to the question whether the product of Merge, a pair of sisters, is symmetric or asymmetric. To this end, a case study is presented showing a fundamental and universal asymmetry between the members of a binary noun phrase coordination, such that the relation between the two conjuncts is invariably marked on the second conjunct. Assuming coordination to present the product of 'pure Merge' (i.e. undisturbed by further movement operations or rearrangements of any other kind), the result suggests that Merge creates an asymmetric sister pair, in which linear order and morphological dependency marking are consistently patterned. The article argues that this is generally the case in contexts of 'restricted syntax', which might equally well be studied using large typological samples. The article furthermore argues that the asymmetry between members of sister pairs involves both phonological and semantic dependence, suggesting that the asymmetry originates within the central syntactic component of grammar.

Keywords: Minimalism; Typology; Dependence; Sisterhood; Coordination; Linear order; Restricted constructions

1. Introduction

This article considers to what extent sample based typological research may shed light on questions of central concern within the minimalist program of generative grammar (Chomsky, 1995). Empirical research within generative grammar has traditionally been characterized by in depth study of data from a single language, or small groups of closely related languages. As a result, the question of the relevance of large-scale data collections has received relatively little attention. In recent years, linguistic typology has claimed a place in the generative research tradition through important studies by Kayne (1994), Baker (1996), Cinque (1999), and Julien (2000), among others. These studies use data of a typological nature to map out the universal lay-out of the clause (the 'cartographic' tradition), to discover the connections among morphosyntactic phenomena (macroparameters), or to contribute to our understanding of linguistic universals by bringing out phenomena that are conspicuous by their absence.

I would like to argue here that the latter strategy, that of looking for patterns that one might expect, but still do not occur, potentially yields information relevant to more abstract questions considered in the context of minimalist syntactic theory. In particular, assuming (as in Epstein, 1999) that syntactic relations between α and β are a function of

E-mail address: c.j.w.zwart@rug.nl.

0024-3841/$ – see front matter
doi:10.1016/j.lingua.2008.09.011

the operation merging α and β into a sister pair γ, a question that may be studied typologically is how the relation between α and β is morphologically marked on either α or β (or both). This article intends to elucidate that question. To that end, I (i) present relevant aspects of the minimalist program (section 2), (ii) discuss various pitfalls having to do with the complex nature of a syntactic derivation (section 3), (iii) present a case study on dependency marking in noun phrase coordination (section 4), and (iv) discuss further consequences, both for typology and syntactic theory (section 5). The sample used in the case study and subsequent work in progress is presented in the Appendix.

2. The minimalist program

With the term 'minimalist program' I refer to the stage of generative grammar inaugurated by class lectures by Noam Chomsky at MIT in the Fall of 1991 (Chomsky, 1993, 1995), and by Richard Kayne's keynote address at the 1992 Lisbon GLOW conference (Kayne, 1994). The minimalist program shares with earlier stages of generative grammar the idea that the object of grammatical inquiry is not (just) linguistic phenomena per se, but the human cognitive capacity generating these linguistic phenomena. It is distinguished from earlier stages of generative grammar in its attempt to reduce all properties of the syntactic component of grammar (referred to as 'narrow syntax') to requirements set by the interfaces between the computational system of human language (CHL) on the one hand and the components of the mind/brain dealing with sound (the articulatory–perceptual interface or A–P interface) and meaning (the conceptual–intentional interface or C–I interface) on the other.

I take the following to be key assumptions of the minimalist program dealing with the construction of sentences:

(1) a. syntactic structure is created via a recursive operation (called 'Merge') combining elements (from the Lexicon or created previously) into constituents
  b. linear order is established when the structure created by Merge is spelled out by the interface component dealing with sound (i.e. is not a property of syntax in the narrow sense)

In this article, we will mainly be concerned with these two assumptions. In particular, we will be concerned with the question how the operation referred to in (1a) determines the morphological properties of the elements merged, and how it might affect the linear ordering of these elements (cf. (1b)). This leaves many aspects of syntactic theory and syntactic analysis undiscussed, and this is perhaps as good a point as any to declare that no familiarity with the finer issues of the minimalist program is presupposed or required here.

In the two subsections to follow I briefly discuss relevant aspects of the structure building operation Merge and the linearization of syntactic structure.

2.1. The nature of Merge

A key assumption of the minimalist program is that CHL involves a system generating structure, called Merge. Merge takes two elements and combines them in a group (a constituent). It is recursive, since the output of Merge may be merged with other elements yielding a further constituent. As Chomsky states, Merge is an operation that comes 'free', in the sense that it is required in some form for any recursive system (Chomsky, 2001:4).

The significance of Merge as a property of the faculty of language is underscored by the contention of Hauser et al. (2002), who hypothesize that recursion is the only uniquely human component of the faculty of language. (This hypothesis has been criticized by Pinker and Jackendoff, 2005, who point out that human language is characterized by many other features not present in animal communication; we return to this issue in section 5). From an empirical point of view, if sentences in all languages are hierarchically organized and potentially infinite (as every sentence may be embedded in or coordinated with another sentence, or preceded by a quotative introduction), it appears that some operation like Merge is needed to adequately describe characteristic and universal properties of syntactic structure.

Accepting this, it is a natural question to ask how exactly Merge operates, or, more pointedly, what the simplest form is that Merge could take. I therefore take this to be a key question of the minimalist program, and in the context of the present article we may wonder whether typological investigations have anything to contribute here. I would like to suggest that the question of the nature of Merge becomes tractable once it is reformulated as a question concerning the properties of the *product* of Merge: the constituent it creates.

I will assume here without discussion that syntactic structures are invariably and universally binary branching (Kayne, 1984). A constituent, then, is a pair of sisters, each member of which is defined by the property of having merged with the other. Chomsky (2005) proposes that the sister pair created by Merge (in unmarked cases) forms an unordered set. An alternative, which Chomsky (2008) describes as 'more complex', would be that the operation Merge does not create a set but an ordered pair. The choice between Merge creating an unordered set or an ordered pair has consequences for the derivation of linear order (cf. (1b)). If Merge creates an unordered set, linear order will have to be imposed on the members of the set in some way. If Merge creates an ordered pair, we might hypothesize that linear order is a function of the ordering of the members of the ordered pair.

(The issue is complicated by factors we will not further take into account in this article, namely that an unordered set becomes an ordered pair as soon as one of the members is designated as the head of the pair, as pointed out in Langendoen (2003:310). I believe this does not resolve the issue, as there remains a difference: if Merge *creates* an ordered pair, the ordering may be taken to be a function of the operation Merge itself, for instance because Merge applies one element to another (asymmetrically); if the unordered set becomes an ordered pair by virtue of a difference in status between the two elements merged, one of them being a head and the other a phrase, the ordering is not an automatic function of the way Merge operates, and we may continue to assume that Merge takes two elements and combines them symmetrically.)

Assume now that Merge creates an ordered pair rather then an unordered set. This then leads to the following hypothesis:

(2) The order of the members of the ordered pair π that is the product of Merge is reflected in the morphosyntactic properties of the members of π.

Conversely, if Merge creates a set, we might expect morphosyntactic marking to be distributed in a way that is not dependent on the process of Merge, but on the various syntactic relations among the members of the sets created by Merge.

Notice that the hypothesis in (2) is in principle testable. For practical purposes, the hypothesis will have to be broken down into a range of more tractable research questions. For example, we may focus on the sister pairs consisting of a subject and a predicate, and then consider the question whether languages mark the predicate for agreement with the person/number features of the subject, or mark the subject for agreement with the tense features of the predicate. When formulated in a sufficiently precise way, such questions may receive answers from large-scale research of a typological nature. This may then lead to a rejection of the hypothesis in (2), suggesting that the initial assumption, that Merge creates an ordered pair, is false.

2.2. The question of linear order

It is commonly assumed by researchers working within the minimalist program that the CHL creates a *structure*, not a *word order*. Thus, a sequence of operations merging (i) a head α and its complement β, yielding a phrase γ, and (ii) γ and a specifier δ, creates a structure $[\ \delta\ [_{\gamma}\ \alpha\ \beta\ ]\ ]$, but the linear order in which α, β, and δ are pronounced is not determined by Merge. Hence, all orders of the sequence α–β–δ are possible, as long as the constituency of γ (i.e. the pair α–β, order irrelevant) is respected (cf. Maxwell, 1984:283 for an early representation of this view).

The question of how linear order (a temporally ordered sequence) is ultimately established may be answered in two ways. First, we may assume that linear order is a *function* of syntactic structure, in the sense that properties of the elements merged determine which is going to be spelled out first. The most prominent exponent of this approach is Kayne (1994), whose Linear Correspondence Axiom (LCA) states that precedence relations among terminals of constituents are a function of asymmetric c-command relations between the nonterminal elements of those constituents (where α c-commands β iff every γ dominating α dominates β). Second, we may assume that linear order is properly the domain of the A–P interface. Chomsky (2001) argues that this might be the case for several phenomena previously considered to be of a purely syntactic nature, such as verb movement and object shift.

Note that the question of how linear order is imposed on the members of a syntactic structure is in principle independent of the question of the nature of Merge. In particular, if we take Merge to yield an unordered set, it may still be the case that linear order is a function of syntactic structure, if the members of the set differ in the relevant properties (e.g. the way they enter into c-command relations with other elements in the structure). Nevertheless, the assumption that Merge yields ordered pairs allows one to formulate a maximally simple hypothesis about the relation between syntactic structure and linear order:

(3) The order of the members of the ordered pair π that is the product of Merge is reflected in the linear order of the members of π.

Like the hypothesis in (2), the hypothesis in (3) is in principle testable (but see section 3 for various pitfalls). Take the subject–predicate relation for example (here, as before, I use the term ‘predicate’ to refer to the subject’s sister, not as a synonym of ‘verb (phrase)’). On the assumption that the subject may be a raised internal argument of the verb (as with passive and unaccusative verbs), at least a subset of subject–predicate constructions involve a subject which is extracted from and hence merged to the predicate. On the assumption that the subject of an unergative verb is also generated internal to the verb phrase, as is now commonly assumed within the minimalist program (e.g. Chomsky, 1995:249–250), *all* subjects are merged to their predicates. There is, then, an inevitable asymmetry between the subject and the predicate (its sister), in the sense that the predicate is a target of movement for the subject, and the subject is not a target of movement for the predicate.

Now it is a well-established typological fact that the subject precedes the predicate in the great majority of the languages of the world (in neutral constructions; e.g. Tomlin, 1986 estimates 90% subject–predicate orders). This typological observation, then, is consistent with what is assumed about the derivation of subject–predicate pairs in the minimalist program, assuming (3). More concretely, it would suggest that α precedes β iff α is merged to β.

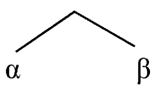
3. Methodological issues

Kayne’s intuition that linear order is a function of asymmetric structure (see section 2.2) was based on a range of empirical observations, mainly involving patterns that are significant by their absence (Kayne, 1992). For example, whereas a number of languages show verb second effects (with the verb appearing in the position after any first clausal constituent), no languages seem to exist which display ‘verb second to last’ effects (with the verb showing up in the position preceding any final clausal constituent). If verb second involves movement of the verb up to a functional position (the structural position for Tense features or for the Complementizer), the observation leads to the conclusion that the functional projections which may play host to a moving verb must all be head-initial. Observations of this kind led Kayne (1994:132) to the striking conclusion ‘‘that the specifier-head-complement order is the only order made available by U[niversal] G[rammar] and consequently that there can be no directionality parameter for word order.’’

Though some progress has been made in studying the empirical consequences of this conclusion, a fundamental difficulty remains that a particular linear order may be the result of direct Merge (also called ‘external merge’), or of a further operation, Move (also called ‘internal merge’). Move differs from Merge only in that one of the two elements combined is extracted out of the other (as in the example of subject raising discussed in section 2.2). To illustrate, the order in (4) may be a straightforward realization of the structure in (5a), resulting from the operation in (5b), or from the structure in (6a), resulting from the sequence of steps in (6b).

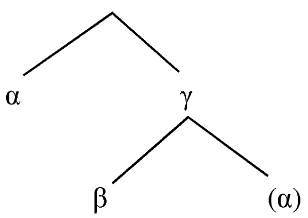
(4) / α β /

(5) a.



b. Merge α to β

(6) a.



b. Merge β to α (yielding γ) (external merge)
Merge α to γ (= Move) (internal merge)

This uncertainty (is a particular order the result of 'pure Merge' or of Merge + Move?) makes it imperative that empirical data be analyzed in considerable detail before their relevance to any theoretical issue can be fully ascertained. Consequently, it becomes difficult, if not impossible, to bring to bear on these theoretical issues the large-scale data collections underlying current typological research.

This illustrates the main difficulty of using large-scale data collections for fundamental syntactic research: the questions I am interested in (the nature of the structure building operation Merge and the properties of the structures it yields) can only be studied if we are certain about the structural analysis of the data, which requires an in depth study of the phenomena, practically impossible in sample based typological research.

A further problem in bringing large sets of empirical data to bear on more fundamental questions of abstract syntactic theory is that the hypotheses in (2) and (3) crucially involve the concept of asymmetric Merge, in the sense that one element is 'merged to' the other (instead of two elements 'merged with' each other). The problem here is that we can only be certain that one element is merged to another after a more or less sophisticated syntactic analysis, involving the thorny and unresolved question of what drives Merge/Move (as in the case of subject raising discussed in section 2.2). This suggests once again that an inherent conflict exists between questions of abstract syntax and answers provided by large-scale empirical research (given practical constraints).

Nevertheless, I believe it is possible to address the fundamental question of the nature of Merge using large language samples if the research domain is carefully circumscribed. In particular, the research should focus on types of constructions where we are relatively certain that movement played no role in their derivation. In the next section, we consider a type of construction, noun phrase coordination, where it is plausible that we see 'pure Merge' in action, i.e. the simple combination of two elements in a constituent, without movement (whether of a syntactic or a 'phonological' nature) coming into play. Section 5, then, considers in a very preliminary way how conclusions reached on the basis of the coordination facts may be extended to other 'restricted' construction types or even more involved syntactic structures.

4. Coordination

We now turn to a description of sample based research on coordination addressing the question of how the (two) members of a constituent created by Merge are distinguished by their morphological properties and by linear order. We will reach the conclusion that in a simple coordination construction combining two noun phrases, it is invariably the second noun phrase (in the sense of linear order) that is morphologically marked for being in a coordination relation with the first. This then leads to the conclusion that linear order and morphological marking are not arbitrary properties of members of pairs created by Merge.

4.1. Preliminaries

Coordination constructions may be described as symmetric or asymmetric based on the type of coherence relation between the members of the coordination (e.g. Kraak and Klooster, 1968:252f.; Kehler, 2000:539f.). In (7), the coherence is of a balanced type (symmetric), whereas in (8), the coordinated members are organized in a contiguous (8a) or even resultative (8b) association (asymmetric):

(7) John likes Mary and Bill hates Susan

(8) a. John came home and grabbed a sandwich
b. Federer won the US Open and became the year-end number one for the third year in succession

In the study reported on here, I looked only at coordinations that could be characterized as 'symmetric' in this sense. More particularly, while the typological survey covered coordinations of all categories, conclusions were drawn on the basis of noun phrase coordinations only.

Based on the number of conjunctions (not: conjuncts) N we can distinguish asyndetic ($N = 0$), monosyndetic ($N = 1$) and polysyndetic ($N > 1$) coordination types. We concentrate on the monosyndetic type here, which may in principle be expected to appear in one of three positions (where & is the conjunction and α, β are the coordinands):

(9) a. & α β
b. α & β
c. α β &

Since type (9a) does not occur at all (cf. Haspelmath, 2007), I use the terms ‘initial conjunction’ and ‘final conjunction’ for the types in (9b) and (9c), respectively.

It should be noted that certain textbook examples of final conjunctions on closer inspection turn out to be initial conjunctions, suffixed to the second coordinand’s first word. This is the case in Latin, where the bound morpheme *-que* ‘and’ might be taken as an example of a final conjunction:

(10)	arma	virum-que	canō	(Latin)
	weapon:PL.ACC	man:SG.ACC-and	sing:1SG.PRES	

‘I sing of war and of a man.’ (Vergil, *Aeneid* I.1)

However, where the second coordinand contains more than one word, we see that *-que* must appear as a second position element within the second coordinand:

(11)	ingenia	fecunda	totius-que	naturae	capacia	
	mind:PL-ACC	fertile:PL.ACC	all:SG.GEN-and	nature:SG.GEN	grasping:PL.ACC	(Latin)

‘minds that are fertile and able to grasp the entire universe.’ (Pliny the Elder, *Natural History* II.190)

This type of ‘second position initial conjunction’ is found in at least 13 languages in the sample (Bella Coola, Evenki, Fon, Hausa, Jacaltec, Kalasha-ala, Lezgian, Shipibo, Turkish, Wardaman, West Greenlandic, Zaghawa, Zay), mostly with clausal coordination, but sometimes with noun phrase coordination (where the type is potentially confusing) as well.

Coordinations may be formally marked by various types of elements (ignoring intonation, which Mithun, 1988:331f. identifies as a major coordination marking device). Next to *true conjunctions* of the type of English *and*, which are not currently used in any other function than coordination, we find *comitative* coordination markers (cf. Stassen, 2000), illustrated in (12), and *summary* coordination markers (cf. Haspelmath, 2007, section 6.3), illustrated in (13).

(12) *comitative strategy*

Péédoró-mútsi-kye	Jóáá-**ma**	ájtyúmɨ́ɨ́be	(Bora)
Pedro-DU-ACC	Juan-with	see:1SG	

‘I see Pedro and Juan.’ (Thiesen, 1996:75)

(13) *summary strategy*

Péédoro-o	Jóáa-á	Perípe-é	**éhdume**	péé	téhullévu	(Bora)
Pedro-RED	Juan-RED	Felipe-RED	this quantity	go	away	

‘Pedro, Juan, and Felipe went away.’ (Thiesen, 1996:75)

The comitative coordination marking strategy uses a marker otherwise used to express comitativity or instrumentality (‘with’), mostly an adposition or a particular case marker. As we will see, the comitative marker may develop into a true conjunction (cf. Mithun, 1988:339).

In the summary coordination marking strategy, coordinands are listed and then resumed by some element identifying the listed elements as a single participant in the event. This summarizing function may be performed by a range of elements, including pronouns, quantifiers, number markers (plural or dual), copulas, focus markers (such as *also*, *too*), adverbs of the *together* type, etc. As discussed in Mithun (1988:337), this coordination strategy may be regarded as a minimal elaboration of a more basic, asyndetic type, where the coordinands are merely juxtaposed.

4.2. Word order generalizations

What follows replicates the findings of Zwart (2005), now based on an enlarged sample of 214 languages (see Appendix). As the bottom line we may state that true conjunctions are invariably of the initial type (9b).

First of all, if we define a language as head-initial (or final) based on the position of a verb and/or adposition with respect to its complement (where a language is head-initial if a head precedes its complement), the languages in the sample are about evenly partitioned, as illustrated in Table 1.

Table 1
Head position in a 214 language sample.

Head-initial	96
Head-final	91
Split	10
Unclear	17

In contrast, initial conjunctions (of any type) are found much more frequently than final conjunctions, as Table 2 illustrates.

Table 2
Conjunction position in a 214 language sample.

Initial	135
Final	12
Mixed	26
Other	39

In the mixed group, 11 languages have both initial and final conjunctions and 15 languages show a preference for polysyndetic coordination; of these, 5 languages have final conjunctions as a monosyndetic alternative. In the 'other' group, 16 use polysyndetic coordination, and data from the remaining 23 are either missing or unclear.

It follows that a sizeable number of head-final languages use initial conjunctions. Conversely, no strict head-initial languages use final conjunctions (although some languages with split headedness do use final conjunctions, and some head-initial languages use final summary elements as a secondary strategy). Table 3 shows the distribution of initial and final conjunctions over head-initial and head-final languages.

Table 3
Conjunction position in head-initial vs. head-final languages.

Initial conjunction	
135	
Head-initial	85
Split	3
Head-final	47
Final conjunction	
12	
Head-initial	0
Split	2
Head-final	10

Table 4 shows that the 12 languages using final conjunctions exclusively invariably employ either the comitative or the summary strategy.

Table 4
Languages using final conjunctions exclusively.

Language	Summary	Comitative	True
35:3 Slave	X	X	
56:2 Yaqui		X	
58:1 Ika		X	
64:1 N Junin Quechua		X	
65:1 Jaqaru		X	
71:1 Yagua	X		
72:1 Bora	X	X	
75:1 Sanumá	X		
76:1 Barasano	X		
76:2 Retuarã	X		
82:1 Paumarí	X		
86:1 Trío	X		

To illustrate, the Ika (Aruak, Chibchan) coordination marker -*sin* in (14a) has a primary use as an instrumental or comitative element in (14b) (Frank, 1990:37–38).

(14) a. ribru rapi-sin pa ú (Ika)
book pencil-with put down AUX
'Put down the book and pencil!' (Frank 1990:38)

b. kʌnsia-sin si aʔsir-i ...
vine-with string tie-while
'He tied it with a vine ...' (Frank 1990:37)

Likewise, the Barasano (East Tucanoan) noun phrase conjunction *kẽde* (15a) has an independent use as a focus marker 'also' (15b) (Jones and Jones, 1991:30).

(15) a. ũbʉ-a rõbi-a dake-rã kẽde yã-ka-bã ĩdã (Barasano)
male-PL female-PL young-AN.PL also be-far:PAST-3PL 3PL
'There were men, women and children there.' (Jones & Jones 1991:133)

b. to bahi-ro yi-ya bũ kẽde
that be-NOM do-PRES 2SG also
'You will do that too.' (Jones & Jones 1991:174)

Also to consider are the 16 languages in the mixed group of Table 2 (11 using both initial and final conjunctions, 5 using final conjunctions next to a polysyndetic type; two of these, Baram Kayan and Wari', appear to be head-initial). Here, too, the true conjunction is exceedingly rare (Table 5).

Table 5
Languages using final conjunctions not exclusively.

Language	Summary	Comitative	True
3:1 Logbara	X	X	(X?)
7:6 Kalasha-ala	X		X[a]
10:1 Kolyma Yukaghir		X	
12:1 Ket		X	
15:2 Kham	X		
19:21 Baram Kayan	X		
33:4 W Desert Lg	X	X	
33:5 Kayardild	X		
35:2 Navaho		X	
46:1 Hualapai	X	(X)	
56:1 Shoshone		X	
64:2 Imbabura Quichua	X		
73:1 Pirahã	X		
83:1 Tariana		X	
85:1 Wari'	X		
94:1 Kwaza			X[a]

[a] Also used as initial conjunction.

If we look closer at the relevant cases of true conjunctions in final position, it turns out that none of them is above suspicion.

In Logbara (Central Sudanic, Nilo-Saharan), the final conjunction is the element *pie*, illustrated in (16).

(16) à mu ɛ̀ri pie àkú-a (Logbara)
we go he and home-to
'I and he go home.' (Crazzolara 1960:100)

In this example, the first coordinand is a so-called inclusory pronominal, semantically singular ('I') but formally marked plural (Haspelmath, 2007: section 6.2, Lichtenberk, 2000). The pronoun is called 'inclusory' because, as Ladusaw (1989) argues, the pronoun encodes a set which includes the referent of the second coordinand. The first coordinand, then, functions as the head of the construction, and the second as an adjunct. Not surprisingly, the inclusory strategy commonly employs a comitative marker to link up the two coordinands. A true conjunction, on the other hand, appears to be rare (cf. Haspelmath, 2007).

The element *pie* is not a comitative marker, however, but it may be related to the element *pi*, a verb meaning 'to complete', and the element *pi* (with mid-low tone) which, when following a noun phrase, means 'cum suis' (Crazzolara, 1960:101). Example (17) illustrates the use of *pi*, as well as the comitative marker used in Logbara coordination:

(17)	ɛ̀tɔ́ɔ̄	p`ỉ	mǜ	dȉ̃	a'ú-ȃ̃	bɛ	(Logbara)
	hare	c.s.	go	then	fowl-DIM	with	

'the hare and a small fowl went together.' (Crazzolara 1960:101)

If *pi* is a '*cum suis*' pluralizer (17) represents another example of the inclusory strategy, not with pronouns but with full noun phrases (similar to the case of Margi reported by Haspelmath, 2007). In all, there is reason to believe that *pie*/*pi* is not a true conjunction but a pluralizer of some kind, which would imply that (16) is not true coordination but coordination of the summary type.

In Kalasha-ala (Nuristani, Indo-Aryan), the final conjunction *ye* is apparently found with fixed expressions only. Whereas (18a) simply means 'a man and a woman', (18b), which may be further reduced, refers to a man-woman pair, such as 'husband and wife' (Degener, 1998):

(18)	a.	e	meši	ye	e	muša	(Kalasha-ala)
		a	woman	and	a	man	

'a woman and a man' (Degener 1998:166)

b.	meši	moša	ye	(> meši-moša-y)
	women	men	and	

'man and woman' (ibid.)

As far as I have been able to ascertain, the final conjunction only occurs in fixed expressions of this type, suggesting that we are dealing with a special case.

Finally, the final conjunction in Kwaza (unclassified, Brazil) is what Van der Voort (2004) terms a 'cosubordinator' (CSO):

(19)	si	xyi-a-'ta	oja-'nã-a-ki	Ba'hozo-nã	(Kwaza)
	I	you-1PL-CSO	go-FUT-1PL-DEC	Barroso-LOC	

'I and you we are going to Barroso.'
[lit: me, you, we being, we go to Barroso] (Van der Voort, 2004:707)

The cosubordinating suffix *'ta/tja* is analyzed as a medial clause same subject marker, i.e. as belonging to the switch reference marking system. Switch reference markers are used in coordination constructions in a number of languages in the sample, including Kiowa, Nateni, Shipibo, and Western Desert Language. Like in Kiowa (Watkins, 1980:288), the switch reference marker seems to have developed into a true (initial) conjunction, in examples like (20) (the declarative marker is a verbalizer, cf. Van der Voort, 2004:208).

(20)	'meza-na	kore'ja-tja	di'hu-tse	(Kwaza)
	table-LOC	knife-CSO	spoon-DEC	

On the table are a knife and a spoon.' (Van der Voort, 2004:706)

But in other constructions, such as (19), the switch reference marker seems to have retained the original function of linking a (clausal) constituent up with a matrix clause. In this respect the appearance of the 1PL suffix *-a* seems relevant, leading to the literal translation indicated, where the coordination acts like an adjunct clause. This case, too, then does not appear to be a case of a true conjunction in final position.

If the evaluation of the final conjunctions in Logbara, Kalasha-ala and Kwaza is correct, none of the 214 languages in the sample features true conjunctions in final position. Moreover, a substantial number of languages have developed initial true conjunctions out of final comitative markers (17 languages in the sample: Burmese, Dogon, Haida, Japanese, Ket, Kinnauri, Kokborok, Korean, Ladakhi, Lavukaleve, Mikir, Navaho, Northern Qiang, Shipibo, Slave, Suppire, Turkish), or out of other final elements such as the switch reference markers of Kiowa and Kwaza (cf. (19)–(20)). This is illustrated in (21) for Kinnauri (Western Himalayish, Sino-Tibetan) (Sharma, 1988).

(21) a. əṅ rəṅ doː chaṅ due (Kinnauri)
1SG:GEN with 3SG:GEN son be:3PAST
'His son was with me.' (Sharma 1988:91)

b. gə rəṅ ki bi-tič
1SG:DIR and you:HON go-FUT:1DU.INCL.HON
'I and you will go.' (Sharma 1988:182)

Finally, whereas conjunction borrowing is a common phenomenon (for reasons discussed in Mithun, 1988), no language seems to have come up with a final conjunction as the result of borrowing (based on 13 languages in the sample with borrowed conjunctions).

Mithun (1988), discussing the origin of formal coordination markers, notes that formally marked noun phrase conjunction is disfavored in spontaneous discourse, and suggests that the emergence of conjunctions may be a function of the development of literacy. She describes the summary and comitative strategies as minimal elaborations of the more basic type of asyndetic coordination, in fact as types "originally used to circumvent coordinate noun phrases" (Mithun, 1988:339). Indeed we regularly find the comitative or summary strategy described as a (marked) alternative to asyndetic juxtaposition (e.g. Nedjalkov, 1990:90 on Evenki, Koroma, 1994:76 on Gola, Enrico, 2003:1079 on Haida, Evans, 1995:394 on Kayardild, Davies, 1981:72 on Kobon, Smeets, 1989:180 on Mapadungu, De Smedt and Mostaert, 1964:163 on Monguor, Dayley, 1989:336 on Shoshone, Aikhenvald, 2004:484 on Tariana, Lewis, 1986:206 on Turkish, Glass and Hackett, 1970:64–65 on Western Desert Language, Payne, 1985:97 on Yagua, Dedrick and Casad, 1999:360 on Yaqui, Maslova, 2003:313f. on Kolyma Yukaghir, among others).

This suggests that a distinction be made between juxtaposition, comitative coordination and summary coordination on the one hand, and true coordination on the other. The generalization that arises from the languages in the sample, then, is that true coordination is invariably marked on the second conjunct, more precisely, at the second conjunct's left edge.

4.3. Discussion

The coordination phenomena reviewed in section 4.2 show up a clear asymmetry between the two members of a coordinated structure: a coordination marked by a true conjunction (i.e. not a comitative or summary element) invariably singles out the second conjunct for expressing the coordination.

As discussed by Munn (1993) and Haspelmath (2007, section 2.2), the English conjunction *and* forms a constituent with the second member, witness its position in discontinuous coordinations of the type in (22b,c). The reasoning here is that a constituent allows isolation or displacement.

(22) a. I saw [John and Bill] the other day
b. I saw [John] the other day [and Bill]
c. * I saw [John and] the other day [Bill]

As Haspelmath (2007) notes, initial conjunctions which may be grouped with a preceeding coordinand are typically cliticized or suffixed; with conjunctions that may be grouped with a following coordinand (e.g. on the basis of constituency tests like (22)), this does not appear to be the case. If we are allowed to consider cliticization a secondary, phonological effect, we may proceed on the assumption that syntactically the conjunction properly belongs to the second coordinand. Importantly, no data exist to suggest that the conjunction is equally closely linked to both coordinands. As Haspelmath (2007, section 2.2.) states: "Monosyndetic coordination seems to be universally asymmetric."

Munn's (1993) conclusion, adopted by Kayne (1994), is that the conjunction actually functions as a head, taking the second conjunct as its complement. If this is correct, the facts surveyed indicate that true coordination is invariably head-initial, suggesting that even clear head-final languages are not entirely without head-initial syntactic organization (Zwart, 2005).

Alternatively, we may describe the conjunction in more neutral terms as a linker, a morpheme appearing between (Maxwell, 1984:252) α and β and expressing that α and β are syntactically linked. Since the linker marks the second conjunct, this leads to the conclusion that in noun phrase coordination the second conjunct is invariably the dependent of the first conjunct. The findings then suggest an order-dependency correlation, such that the nondependent invariably preceeds the dependent.

If this is correct, and if coordination is established via the structure building operation Merge, we are allowed to conclude that Merge invariably creates a dependency between the elements merged, hence that the product of Merge is an asymmetric pair of sisters.

5. Consequences

The noun phrase coordination phenomena reviewed in section 4 suggest that two elements combined by Merge invariably realize a dependency, expressed by linear order and morphology (the conjunction functioning as a linking particle). The relevance of this finding to the issues raised in section 2 appears to be the following.

In section 2, we were interested in the nature of the structure building operation Merge, and, more concretely, in the nature of the product of Merge, a two-member constituent. We noted the problem that sister pairs may be created by Merge directly, or by a sequence of steps involving both Merge and Move (section 3). This problem makes it difficult to draw conclusions as to the nature of the structure building operation from empirical data that have not been thoroughly (and time consumingly) analyzed.

There is no reason to believe, however, that the coordination phenomena presented and discussed in section 4 involve any steps beyond pure Merge. I know of no arguments suggesting that the members of a pair of conjuncts must be taken to be reversed in the course of a derivation. Likewise, it seems implausible that a phonological rearrangement of the coordinands is called for. It follows that coordination data mined from a large language sample can be readily used, and that the analysis of the data need not progress beyond the elementary questions of the type of coordination strategy and the nature of the conjunction used.

If we then find that the second member of a pair of conjuncts is invariably marked (i) morphologically, via the conjunction, and (ii) syntactically, via linear order, we may conclude that the crosslinguistic study of coordination phenomena indicates that the product of Merge is invariably and consistently asymmetric. This conclusion is consistent with the hypotheses in (2) and (3), suggesting that a pair [α β] is derived by merging α to β, i.e. by creating an ordered pair rather than an unordered set.

In this final section before the conclusion, I would like to argue that the conclusion drawn on the basis of the study of coordination facts may be supported if we take a wider range of phenomena into account. Based on these phenomena, I argue that indicators of dependency are distributed over all components of the grammar, not restricted to the A–P interface, and that these indicators tend to identify the dependent and the nondependent category in a consistent way. This suggests that the asymmetry between members of sister pairs is not created at any of the two interfaces (A–P and C–I), but must be part of narrow syntax. The relevance to the topic of this article is that these dependency indicators can be readily inspected from standard reference grammars.

5.1. Dependency

In the structuralist tradition, dependency is generally defined in terms of phrase structure, capitalizing on the head/nonhead distinction (e.g. Tesnière, 1959; Nichols, 1986). Since a defining property of heads is that they may select nonheads within their projection, this approach views dependency as the function of a semantic selection relation.

I believe this definition is both too narrow and too much dependent on particular views of phrase structure. For example, the question of whether the subject is a dependent of the verb is answered differently depending on the hypothesized base position of the subject (inside the VP or not) and on the hypothesized structure of the functional domain (involving separate heads for functional categories like Tense or not).

A more basic approach to dependency would start from the core case in (23):

(23) δ is a dependent of α if δ expresses in its morphology a feature of α

On this view, a target element δ whose morphology is controlled by some other element α is by definition the dependent of α (cf. Corbett, 2003). More generally, we can say that in this situation δ is brought in opposition to α by some marked feature, in this case morphology. But other features may also be used to establish the intended oppositions, most notably pitch, linear order, and semantic interpretation.

For example, in the colloquial American English pejorative expression of the type illustrated in (24) (cf. Grohmann and Nevins, 2004), *schmoney* is brought in opposition to *money* in at least four ways:

(24) Money schmoney

(i) Morphologically, *schmoney* is dependent on *money* in that it reduplicates its root. (ii) Intonationally, *schmoney* is set off by a marked (higher) pitch on its first syllable. (iii) In linear order, the reduplicating (target) element invariably follows the reduplication source (the controller). And (iv) semantically, *schmoney* predicates a particular property of *money*, mainly the property of irrelevance.

I would therefore like to propose the following generalized definition of dependency:

(25) δ is a dependent of α if δ is brought in opposition to α by some marked property

This definition is intended to define dependency in an asymmetric fashion, since the element with the marked property is brought in opposition to an unmarked element. This is in line with the concept of opposition proposed by Jakobson (1935), where two elements α, β are in opposition if (i) β expresses a positive marking for a feature F, and (ii) α does not express any marking for F (i.e. neither positive nor negative). In the case of (24), *schmoney* is positively marked as [+reduplication], whereas *money* is not marked for the feature [reduplication] at all.

Notice that, in the example (24), the dependency indicators are distributed over the components of the grammar, some belonging to the A–P side of the grammar (morphology, intonation, linear order) and others belonging to the C–I side (semantic interpretation). In (24) the various dependency indicators express the dependency in a consistent way, in the sense that the 'phonological' dependency indicators and the 'semantic' dependency indicators mark the same element as being the dependent. We therefore would like to know whether this is generally the case. If so, the source of the dependency must not be associated with the interface components A–P or C–I, but with the component feeding both interface components, CHL (narrow syntax).

5.2. *Indicators of dependency*

It is clear that both within a single language and across languages, linear order is the least reliable indicator of dependency. Thus, whereas in the Dutch example (26a) all indicators of dependency converge, as detailed in (26b), (27) from the same language has inverted the linear order, but arguably not the dependency. (In the examples, small capitalization indicates high pitch.)

(26) a. Vogel-s VLIEG-en (Dutch)
bird-PL fly-PL
'Birds fly.'

b. morphology: [PLURAL] is an inherent feature of the noun phrase *vogels*, expressed redundantly on the verb *vliegen* (i.e *vogels* = α, *vliegen* = δ)
linear order: δ follows α
intonation: δ has high pitch, α has neutral pitch (in the unmarked intonational pattern)
semantics: δ is interpreted as a predicate of α (and not as another noun phrase; NB *vliegen* could also be interpreted as the plural noun 'flies', but not here)

(27) Toch VLIEG-en vogel-s (Dutch)
yet fly-PL bird-PL
'Yet birds fly.'

In (27), morphological, intonational and semantic dependency indicators still identify *vliegen* 'fly' as the dependent of *vogels* 'birds', in spite of the inverted linear order.

Examples illustrating the unreliable character of linear order as a dependency indicator can easily be multiplied. (28), for instance, shows examples of N-V compounds in French and Dutch:

(28) a. ouvre-bouTEILle (French)
open-bottle
b. FLES-open-er (Dutch)
bottle-open-NOM
both: 'bottle opener'

In both cases, the element translated by 'bottle' is the dependent of the element translated by 'open', as 'bottle' is interpreted as the element affected by 'open'. This dependency is expressed by pitch accent in both French and Dutch, but by linear order only in French. We find similar facts within a single language, in the expression of geographical terms:

(29) a. de berg HOreb (Dutch)
the mountain Horeb
'mount Horeb'

b. het ATlas-gebergte
the Atlas-range
'the Atlas range'

In (29a) we find the geographical term realized in a construct state syntax, with the name specifying the geographical object following it and set off via pitch accent. In (29b), a similar geographical object is described in a compound, with the specifying element preceding the element specified, but still carrying the pitch accent indicative of its dependent character.

The circumstance that linear order is an unreliable indicator of dependency is reflected in the more or less even distribution of head-initial and head-final languages in the sample (Table 1). Taking a complement to invariably be a dependent of the head it is merged with, the very existence of the head-initial/head-final typology indicates that linear order does not reflect dependency directly (see also Nichols, 1986). In terms of the structure building operation Merge, linear order alterations may be realized by remerging one of the elements merged to the derivation in a previous step (i.e. via Move; cf. (5)–(6)).

To abstract away from movement, we must study indicators of dependency in a range of constructions where we are certain that no remerge takes place, such as coordination (section 4). In addition to coordination, an entire range of such 'restricted' expressions may be identified (30).

(30)	a.	sports result	1-0	één-NUL	[one-nil]	(Dutch)
	b.	digit sequence	1, 2, 3	één-twee-DRIE	[one-two-three]	
	c.	numbers	21	een-en-TWINtig	[one and twenty]	
	d.	the time	1:30	half TWEE	[half two]	
	e.	the amount	2,50	twee-VIJFtig	[two-fifty]	
	f.	reduplication	zozo	zo-ZO	'so-so'	
	g.	titles		luitenant-koloNEL	'wing commander'	
	h.	acronyms	PvdA	payvaydayAH	[socialist party]	

These examples arguably involve mere juxtapositions, and the only dependency indicators which appear to be consistently present are linear order and intonation. In each case, the second element in the linear order (rightmost on paper, following in time) is set off by higher pitch.

Morphological dependency indicators appear to be generally absent in this type (except, possibly, in reduplications of the kind in (24)). Semantic dependency may be detected, however. Consider the example of sports results (30a).

In (30a), the leftmost term, *één* ‘one’, seems to have a simple cardinality reading ($n_1 = 1$), but the rightmost term appears to be ambiguous between the cardinality reading ($n_2 = 0$) and a reading that opposes the number against that of the leftmost term, such that n_2 is smaller than n_1. This reading is only possible if the rightmost term is intonationally set off from the leftmost term by high pitch. When both terms are pronounced equally low, the relative reading of the rightmost term disappears. This indicates that the interpretation of the second of the two terms is dependent on that of the first: what *nul* ‘nil’ in (30a) really means is not the cardinality of zero, but the fact that it is less than *één* ‘one’. (This semantic interpretation is irrespective of the actual score, so that a similar effect would be achieved with a result like *0-1*; here, *one* is ambiguous between the cardinality reading of ‘1’ and the relative reading of ‘more than zero’.)

Other types in which some semantic dependency may be identified are (30d,e,g), where the second term is interpreted as specifying the hour (30d), fraction (30e), or subrank (30g) of the first.

If restricted expressions of the type in (30) are created by Merge, it is interesting to hypothesize that the convergence of the dependency indicators that are available in this type reflect the basic asymmetric organization of pairs created by Merge. That this conclusion might be justified is suggested by the observation that slightly less restricted juxtaposition types (including coordination) appear to maintain the convergence of dependency indicators, most notably the correlation of intonation and linear order. This can be seen in the cases listed in (31) (Zwart, 2003).

(31)	a.	coordinations	john and MAry		(Dutch)
	b.	asyndetics	me TARZan		
	c.	construct state	huis USHer	[house (of) Usher]	

In these types, the semantic contribution of the second member is often that of predication or specification. In addition to the basic dependency indicators of linear order and intonation, we see the beginnings of morphological dependency indicators, via linking particles (the conjunction in (30a), and the functional preposition *of* in the English version of the construct state construction (30c)).

From here it appears to be a small step to full fledged predication and complementation structures, where the predicate/complement is intonationally set off from its sister by the Nuclear Stress Rule, which puts pitch accent on the complement (Chomsky and Halle, 1968; Cinque, 1993), or, in the spirit of Zubizarreta’s (1998:43) reformulation, on the dependent element in an asymmetric sister pair. In addition, we see the emergence of a range of morphological dependency indicators with predication and complementation, the exact distribution of which may very well be studied in large-scale language samples. The only complication with predication and complementation is that here (especially with complementation), the syntactic derivation may involve movement, eliminating linear order as a reliable dependency indicator.

I refer to Zwart (2006a, 2006b to appear) for more extensive discussion of dependency and dependency marking in (nonrestricted) syntactic constructions involving subject–predicate combinations, adjective–noun combinations, possessor–possessum combinations, and adposition–complement combinations. It is argued there that inflectional morphology (both agreement and case), with few exceptions, can readily be analyzed as markers of dependency as a function of Merge. For example, number agreement on the verb may be seen as a morphological device marking the predicate (the subject’s sister) as a dependent of the subject. Zwart (to appear) argues that structural dependent case (accusative) has the same function, i.e. not expressing dependency of the object to the verb, but dependency of the predicate containing the object to the subject (itself unmarked for case). Though many questions remain to be pursued, this approach to inflectional morphology presents a possible answer to some of the objections raised by Pinker and Jackendoff (2005) to the idea, expressed in Hauser et al. (2002), that the human faculty of language is defined by the ability to recursively perform the operation Merge: on the view expressed here, inflectional morphology is a *function* of Merge, and hence not a separate property distinguishing human language from animal communication.

The relevance of this discussion of dependency indicators to the general question at hand is the following. If we take the dependency indicators linear order, morphology, and intonation to properly belong to the A–P interface domain, and the semantic dependency indicator to the C–I interface domain, we observe that:

(32) a. dependency indicators tend to converge across interface domains,
b. divergence of dependency indicators takes place within the A–P interface domain, and
c. the cause of divergence is a syntactic operation belonging to neither of the interface domains

These observations strongly suggest that the source of the asymmetries underlying the dependency indicators is not contained within one of the interface domains, but is instead centrally located in the component feeding both interface domains, CHL (narrow syntax).

6. Conclusion

I have argued in this article that sample based typological research may be instrumental in addressing central issues in minimalist syntactic theory. Concretely, I have shown that there is reason to believe that the structure building operation Merge assumed within the Minimalist Program (Chomsky, 1995) generates ordered pairs rather than unordered sets. I have suggested that a fruitful area of investigation relevant to this claim is provided by restricted syntactic constructions, including juxtapositions of various types, and slightly more complicated structures, such as coordinations marked by a (true) conjunction, asyndetic predications and construct state constructions. The research might focus on the realization of various indicators of dependency in these construction types, both belonging to the A–P (phonological) and C–I (semantic) side of the grammar. In particular, the question whether these indicators of dependency are realized in a consistent way across languages and constructions may be fruitfully studied using the empirical data provided by standard reference grammars. As an example, I reviewed a sample based study of noun phrase coordination, showing that true coordination invariably involves morphological marking of the left edge of the second coordinand.

Acknowledgments

Thanks to Leon Stassen, John Nerbonne, and the anonymous reviewers of this contribution. The research reported here is part of the NWO-funded research program 'Dependency in Universal Grammar' (grant number 360-70-200), which is gratefully acknowledged.

Appendix A

The sample used in the research discussed here is a variety sample of 214 languages, created for the study of morphosyntactic variation across the languages of the world. Its aim is to arrive at maximal coverage of linguistic variation, not to estimate the probability of certain features (co)occurring. In the terminology of Bell (1978), we determine as the universe of the sample: the total of all possible human languages, or universal grammar; and as the frame of the sample: the total of actual extant human languages. The frame was partitioned in 71 language families, following the classification in 96 families of Gordon (2005) (the Ethnologue) (25 very small families could not be included; the number of families in the Ethnologue has since increased to 108, and the sample may be adjusted accordingly). The number of languages selected from each family was determined by size and estimated diversity, and the selection of individual languages was determined first by subgrouping and then by convenience (i.e. the availability of excellent reference grammars).

The following families and languages were included in the sample:

(1) 1. Khoisan, 2 languages (!Kung, Nama); 2. Niger-Congo, 27 lgs (Pulaar, Gola, Kalabari, Kikuyu, Nkore-Kiga, Ewondo, Tikar, Tiv, Limbum, Ibibio, Yoruba, Degema, Igbo, Duka, Jukun, Birom, Dogon, Grebo, Fon, Baule, Mundang, Ngbaka, Nateni, Suppire, tbd, Bobo Dioula, Soninke); 3. Nilo-Saharan, 5 lgs (Logbara, Dilling, Lango, Zaghawa, Songhai); 4. Afro-Asiatic, 8 lgs (Tamasheq, Margi, Lele, Hausa, Iraqw, Kafa, Gulf Arabic, Zay); 5. South Caucasian, 1 lg (Georgian); 6. North Caucasian, 2 lgs (Lezgian, Kabardian); 7. Indo-European, 9 lgs (Albanian, Lithuanian, Breton, Dutch, Gojri, Kalasha-ala, Dimli, Portuguese, Russian); 8. Uralic, 2 lgs. (Nenets, Hungarian); 9. Dravidian, 4 lgs. (Kolami, Brahui, Abujhmaria, Tamil); 10. Yukaghir, 1 lg. (Kolyma Yukaghir); 11. Altaic, 3 lgs. (Monguor, Evenki, Turkish); 12. Yenisei Ostyak, 1 lg. (Ket); 13. Japanese, 1 lg. (Japanese); 14. Chukotko-Kamchatkan, 2 lgs. (Chukchi, Itelmen); 15. Sino-Tibetan, 12 lgs. (Cantonese, Ladakhi, Kinnauri, Kham, Kokborok, Eastern Kayah Li, Mao Naga, Burmese, Meithei, Karbi, Digaru, Northern Qiang); 16. Miao-Yao, 1 lg. (Iu Mien); 17. Austroasiatic, 6 lgs. (Temiar, Chrau, Khmer, Khasi, Vietnamese); 18. Daic, 2 lgs. (Kam, Thai); 19. Austronesian, 31 lgs. (Sedeq, Paiwan, Tsou, Kambera, Paulohi, Roti, Loniu, Fijian, Samoan, Iaai, Ponapean, Tiri, North Efate, Sie, Kwaio, Hoava, Manam, Kilivila, East Makian, Malagasy, Baram Kayan, Ida'an,

Chamorro, Tagalog, Ivatan, Sama, West Bukidnon Manobo, Mongondow, Muna, Acehnese, Toba Batak); 21. Andamanese, 1 lg. (A-Pucikwar); 24. East Papuan, 1 lg. (Lavukaleve); 28. Sepik-Ramu, 3 lgs. (Yimas, Abelam, tbd); 30. Torricelli, 1 lg. (Arapesh); 31. Trans New Guinea, 9 lgs. (Tauya, Amele, Central Asmat, Kobon, Nabak, Marind, Daga, Eipo, Nimboran); 32. West Papuan, 1 lg. (Abun); 33. Australian, 6 lgs. (Gooniyandi, Wardaman, Dyirbal, Western Desert Language, Kayardild, Djingili); 34. Eskimo-Aleut, 2 lgs. (Aleut, West Greenlandic); 35. Na-Dene, 4 lgs. (Haida, Navaho, Slave, Tlingit); 36. Algic, 1 lg. (Eastern Ojibwe); 38. Wakashan, 1 lg. (Nootka); 39. Salish, 2 lgs (Bella Coola, Thompson); 41. Siouan, 1 lg. (Lakota); 42. Caddoan, 1 lg. (Wichita); 43. Iroquoian, 1 lg. (Tuscarora); 44. Penutian, 2 lgs. (Yokuts, Coos); 46. Hokan, 2 lgs. (Hualapai, Karok); 48. Kiowa-Tanoan, 1 lg. (Kiowa); 50. Muskogean, 1 lg. (Koasati); 52. Totonacan, 1 lg. (Misantla Totonac); 53. Mixe-Zoque, 1 lg. (Chimalapa Zoque); 54. Mayan, 3 lgs. (Tzotzil, Jacaltec, Cakchiquel); 55. Subtiaba-Tlapanec, 1 lg. (Tlapaneco); 56. Uto-Aztecan, 2 lgs. (Shoshone, Yaqui); 57. Oto-Manguean, 5 lgs. (Sochiapan Chinantec, Chalcatongo Mixtec, Mezquital Otomi, Southern Popoloca, Zoogocho Zapotec); 58. Chibchan, 1 lg. (Ika); 61. Choco, 1 lg. (Saija); 63. Zaparoan, 1 lg. (Zaparo); 64. Quechuan, 2 lgs. (North Junín Quechua, Imbabura Quichua); 65. Aymaran, 1 lg. (Jaqaru); 66. Aracaunian, 1 lg. (Mapudungu); 71. Peba-Yaguan, 1 lg. (Yagua); 72. Witotoan, 1 lg. (Bora); 73. Mura, 1 lg. (Pirahã); 75. Yanomam, 1 lg. (Sanumá); 76. Tucanoan, 2 lgs. (Barasano, Retuarã); 78. Nambiquaran, 1 lg. (Sabanês); 79. Maku, 1 lg. (Dâw); 81. Tupi, 3 lgs. (Kanoé, Guaraní, Yuqui); 82. Arauan, 1 lg. (Paumarí); 83. Arawakan, 1 lg. (Tariana); 85. Chapacura-Wanham, 1 lg. (Wari'); 86. Carib, 2 lgs. (Trió, Hixkaryána); 89. Mataco-Guaicuru, 1 lg. (Pilaga); 90. Panoan, 1 lg. (Shipibo); 91. Tacanan, 1 lg. (Araona); 92. Macro-Ge, 2 lgs. (Borôro, Canela); 93. Isolates, 6 lgs. (Ainu, Basque, Burushaski, Korean, Mosetén, Nivkh); 94. Unclassified, 1 lg. (Kwaza); 95. Creoles, 3 lgs. (Ndyuka, Mauricien, Cape Verdean Creole); 96. Pidgins, 1 lg. (Naga Pidgin); 97. Deaf Sign Languages, 1 lg. (American Sign Language).

If we set the total number of languages in the world, excluding isolated/unaffiliated languages, pidgins and creole languages and signed languages, at 6871 (Gordon, 2005), the sample as a whole represents 3.1% of that total number. The 25 groups not represented in the sample total 112 languages, i.e. 1.6%. The only sizeable group among the nonrepresented families is the Geelvink Bay group at 33 languages, a situation which needs to be remedied. The reason these groups are not represented is the current unavailability of good reference grammars.

Underrepresentation occurs where a language family's representation in the sample, calculated as the percentage of languages of that family included in the sample, is under 3.1%. This typically occurs with large language families, notably Niger-Congo (1514 lgs, repr 1.8%), Indo-European (445 lgs, repr 2.0%), Afro-Asiatic (375 lgs, repr 2.1%), Australian (263 lgs, 2.3%), and Austronesian (1268 lgs, repr 2.4%). In other cases, underrepresentation is the result of families or subgroups being poorly described, as with Arawakan (1.6%), Trans-New Guinea (1.6%), and Torricelli (1.9%). Underrepresentation may also occur with very small language groups with estimated low diversity (i.e. few or no subdivisions). Overrepresentation is not necessarily problematic in a variety sample, where no correlations are sought. In the sample, small language groups are typically overrepresented, with a view to covering the family internal diversity.

In what follows I briefly discuss the issue how the quality of a variety sample of this type might be assessed, considering questions of diversity and coverage.

Diversity is to some extent a function of size: a small family of four languages, divided over two branches, shows more diversity than an arbitrary set of four languages belonging to a larger subgroup. In the first case we want representation of two out of four languages, in the second we want the subgroup to be represented, for which a single language may suffice. Quantification of diversity is done by Bell (1978) and Rijkhoff et al. (1993). Bell's diversity assessment is based on the projected number of groups per phylum in separation at 3500 years before current time. We do not adopt this approach, for two reasons. First, for a number of phyla, the historical record does not allow for an informed estimate. Second, historical accident may cause deviations from the presupposed steady rate diversity increase. Rijkhoff et al.'s *diversity value* is a function of the complexity of genetic language trees (with weighting of higher versus lower branches), arrived at by calculating the average number of nonroot/nonterminal nodes at each intermediate level in the tree. This approach is more promising, we feel, but an additional factor that might be taken into account is the size of the terminal groups. We proceed from the idea that diversity arises when an opposition between (groups of) languages necessitates branching. It follows that two factors contribute significantly to diversity: a) the size of the terminal groups (where small = diverse), and b) the number of splits in the genetic language family tree. Note that the approach to estimating diversity assumes that branching reflects diversity in a consistent way, an assumption which is certainly false. This is an inherent weakness of any approach to diversity based on configurational properties of linguistic affiliation trees (see also Croft, 1990:22).

More important to a variety sample than genetic (or regional) bias is the effective coverage. This can be seen as a function of the proportion of the branches within each family represented in the sample. Sampling technique should be directed at gaining maximal coverage out of a limited number of languages. We propose to measure coverage by calculating for each phylum the average percentage of the branches of each layer of the genetic tree represented in the sample. If we number the layers of the tree from the bottom up from 1 to n, additional weighting can be done (in the spirit of Rijkhoff et al., 1993) by multiplying the percentage for each layer by its number, and dividing the sum of the percentages by $n!$. In the current sample, coverage values range from .34 (Austronesian) to 1.00 (Yukaghir). This method of estimating coverage provides an easy tool for improving the sample, as a low coverage value is typically caused by nonrepresentation of one or more high level branches. For instance, the relatively low coverage value of Nilo-Saharan (.44) is explained by the circumstance that the Nilo-Saharan family branches directly into a high number of subgroups, most of which are small and difficult to represent for lack of available descriptions. Incorporation of one or more languages from these branches would effect an immediate improvement. As it stands, certain coverage values are unacceptably low, but the expectation is that with the addition of another 30–40 languages, the situation might be significantly improved.

References

Aikhenvald, A., 2004. A Grammar of Tariana. Cambridge University Press, Cambridge.

Baker, M.C., 1996. The Polysynthesis Parameter. Oxford University Press, New York.

Bell, A., 1978. Language sampling. In: Greenberg, J.H. (Ed.), Universals of Human Language, vol. 1. Stanford University Press, Stanford, pp. 125–156.

Chomsky, N., 1993. A minimalist program for linguistic theory. In: Hale, K., Keyser, S.J. (Eds.), The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger. MIT Press, Cambridge, MA, pp. 1–52.

Chomsky, N., 1995. The Minimalist Program. MIT Press, Cambridge, MA.

Chomsky, N., 2001. Derivation by phase. In: Kenstowicz, M. (Ed.), Ken Hale: A Life in Language. MIT Press, Cambridge, MA, pp. 1–52.

Chomsky, N., 2005. Three factors in language design. Linguistic Inquiry 36, 1–22.

Chomsky, N., 2008. On phases. In: Freidin, R., Otero, C.P., Zubizarreta, M.L. (Eds.), Foundational issues in linguistic theory: essays in honor of Jean-Roger Vergnaud. MIT Press, Cambridge, MA, pp. 133–166.

Chomsky, N., Halle, M., 1968. The Sound Pattern of English. Harper & Row, New York.

Cinque, G., 1993. A null theory of phrase and compound stress. Linguistic Inquiry 24, 239–297.

Cinque, G., 1999. Adverbs and Functional Heads. Oxford University Press, New York.

Corbett, G., 2003. Agreement: terms and boundaries. In: Griffin, W.E. (Ed.), The Role of Agreement in Natural Language: TLS 5 Proceedings. University of Texas, Austin, pp. 109–121., http://uts.cc.utexas.edu/tls/.

Crazzolara, J.P., 1960. A Study of the Logbara (Ma'di) Language: Grammar and Vocabulary. Oxford University Press, London.

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

Davies, J., 1981. Kobon. North Holland Publishing Company, Amsterdam.

Dayley, J.P., 1989. Tümpisa (Panamint) Shoshone Grammar. University of California Press, Berkeley.

De Smedt, A., Mostaert, A., 1964. Le dialecte monguor parlé par les Mongols du Kansou occidental, 2ème partie: grammaire. Mouton, The Hague.

Dedrick, J.M., Casad, E.H., 1999. Sonora Yaqui Language Structures. University of Arizona Press, Tucson.

Degener, A., 1998. Die Sprache von Nisheygram im afghanischen Hindukusch. Harrassowitz, Wiesbaden.

Enrico, J., 2003. Haida Syntax. University of Nebraska Press, Lincoln.

Epstein, S.D., 1999. Un-principled syntax: the derivation of syntactic relations. In: Epstein, S.D., Hornstein, N. (Eds.), Working Minimalism. MIT Press, Cambridge, MA, pp. 317–345.

Evans, N., 1995. A Grammar of Kayardild with Historical-Comparative Notes on Tangkic. Mouton de Gruyter, Berlin.

Frank, P., 1990. Ika Syntax. Summer Institute of Linguistics, Dallas.

Glass, A., Hackett, A., 1970. Pitjantjatjara Grammar: A Tagmemic View of the Ngaayatjara (Warburton Ranges) Dialect. Australian Institute of Aboriginal Studies, Canberra.

Gordon Jr, R.G., 2005. Ethnologue: Languages of the World, 15th edition. SIL International, Dallas., http://www.ethnologue.com/.

Grohmann, K.K., Nevins, A., 2004. Echo-reduplication: when too-local movement requires PF-distinctness. University of Maryland Working Papers in Linguistics, vol. 13. pp. 84–108.

Haspelmath, M., 2008. Coordination. In: Shopen, T. (Ed.), Language typology and syntactic description, vol. 2: complex constructions. second edition. Cambridge University Press, Cambridge, pp. 1–51.

Hauser, M.D., Chomsky, N., Fitch, W.T., 2002. The faculty of language: what is it, who has it, and how did it evolve? Science 298, 1569–1579.

Jakobson, R., 1935. Beitrag zur allgemeinen Kasuslehre. In: Selected Writings II (1971), Mouton, The Hague, pp. 23–71.

Jones, W., Jones, P., 1991. Barasano Syntax. Summer Institute of Linguistics, Dallas.

Julien, M., 2000. Syntactic heads and word formation: a study of verbal inflection. Doctoral Dissertation. University of Tromsø.

Kayne, R.S., 1984. Connectedness and Binary Branching. Foris, Dordrecht.

Kayne, R.S., 1992. Word Order. Keynote Lecture at the GLOW Colloquium, Lisbon, April 14.

Kayne, R.S., 1994. The Antisymmetry of Syntax. MIT Press, Cambridge, MA.

Kehler, A., 2000. Coherence and the resolution of ellipsis. Linguistics and Philosophy 23, 533–575.

Koroma, R., 1994. Die Morphosyntax des Gola. Institut für Afrikanistik, Cologne.

Kraak, A., Klooster, W., 1968. Syntaxis. Stam-Robijns, Culemborg.

Ladusaw, W., 1989. Group reference and the plural pronoun construction. Papers on the Plural Pronoun Construction and Comitative Coordination. Syntax Research Center, University of California at Santa Cruz, Santa Cruz, pp. 1–7.

Langendoen, D.T., 2003. Merge. In: Carnie, A., Harley, H., Willie, M. (Eds.), Formal Approaches to Function in Grammar: in Honor of Eloise Jelinek. John Benjamins, Amsterdam, pp. 307–318.

Lewis, G.L., 1986. Turkish Grammar. Oxford University Press, Oxford.

Lichtenberk, F., 2000. Inclusory pronominals. Oceanic Linguistics 39, 1–32.

Maslova, E., 2003. A Grammar of Kolyma Yukaghir. Mouton de Gruyter, Berlin.

Maxwell, D.N., 1984. A typologically based principle of linearization. Language 60, 251–285.

Mithun, M., 1988. The grammaticization of coordination. In: Haiman, J., Thompson, S.A. (Eds.), Clause Combining in Grammar and Discourse. John Benjamins, Amsterdam, pp. 331–359.

Munn, A., 1993. Topics in the syntax and semantics of coordinate structures. Doctoral Dissertation. University of Maryland.

Nedjalkov, I., 1990. Evenki. Routledge, London.

Nichols, J., 1986. Head-marking and dependent-marking grammar. Language 62, 56–119.

Payne, D.L., 1985. Aspects of the grammar of Yagua: a typological perspective. UCLA Dissertation.

Pinker, S., Jackendoff, R., 2005. The faculty of language: what's special about it? Cognition 95, 201–236.

Rijkhoff, J., Bakker, D., Hengeveld, K., Kahrel, P., 1993. A method of language sampling. Studies in Language 17, 169–203.

Sharma, D.D., 1988. A Descriptive Grammar of Kinnauri. Mittal Publications, Delhi.

Smeets, I., 1989. A Mapuche grammar. Doctoral Dissertation. Leiden University.

Stassen, L., 2000. AND-languages and WITH-languages. Linguistic Typology 4, 1–55.

Tesnière, L., 1959. Éléments de syntaxe structurale. Klincksieck, Paris.

Thiesen, W., 1996. Gramática del idioma bora. Instituto Linguïstico de Verano, Yarinacocha.

Tomlin, R., 1986. Basic Word Order: Functional Principles. Croom Helm, London.

Van der Voort, H., 2004. A Grammar of Kwaza. Mouton de Gruyter, Berlin.

Watkins, L.J., 1980. A grammar of Kiowa. Doctoral dissertation. University of Kansas.

Zubizarreta, M.L., 1998. Prosody, Focus and Word Order. MIT Press, Cambridge, MA.

Zwart, J.W., 2003. What's in a name? Syntactic and asyntactic accentuation in Dutch. In: Delsing, L.-O., Falk, C., Josefsson, G., Sigurðsson, H.Á. (Eds.), Grammatik i focus/Grammar in focus: Festschrift for Christer Platzack, vol. II. Department of Linguistics, Lund, pp. 395–401.

Zwart, J.W., 2005. Some notes on coordination in head-final languages. In: Doetjes, J., van de Weijer, J. (Eds.),Linguistics in the Netherlands 2005. John Benjamins, Amsterdam, pp. 232–241.

Zwart, J.W., 2006a. Complementizer agreement and dependency marking typology. Leiden Working Papers in Linguistics, vol. 3. 2. pp. 53–72 http://www.let.leidenuniv.nl/pdf/lucl/lwpl/3.2/zwart.pdf.

Zwart, J.W., 2006b. Local agreement. In: Boeckx, C. (Ed.), Agreement systems. John Benjamins, Amsterdam, pp. 317–339.

Zwart, J.W., to appear. Structural case and dependency marking: a neo-Jakobsonian view. Proceedings of the Chicago Linguistic Society 42. http://www.let.rug.nl/zwart/docs/cls42zwart.pdf.