1 *What breaks the symmetry in syntactic structuring*

1.1 The asymmetry of syntax

Let us refer to syntactic structures as *symmetric* if for a class of (sub-) trees [A B], there is a corresponding class of (sub-)trees that differ only with respect to the order of their immediate sub-constituents, i.e. [B A], within a given language, or cross-linguistically. In this sense, the attested language *structures* are asymmetric.[1] This means that the *inverse order* of a well-formed sequence of phrases in a complex phrase structure is in general not a well-formed sequence, neither cross-linguistically nor within a single language. In terms of phrase-structure trees, a symmetric organization of syntactic structuring as a cross-linguistic property would entail that there is a *mirror image* of a given *structure* as a well-formed structure at least in some other language. Compare, for instance, a German VP and its English counterpart, as in (1a,b):

(1) a. *give* the reader a hint
b. dem Leser einen Hinweis *geben*
the reader$_{DAT}$ a hint$_{ACC}$ give
c. $[_{VP}\,[_{V'}\,V°\ DP]\ DP]$ (Chomsky 1981: 171)
d. $[DP\,[DP\ V°]_{V'}]_{VP}$

Until the mid-eighties, before Barss and Lasnik (1986), (1a,b) and (1c,d), respectively, were assumed to betray mirror-image structures. The two nominal complements in (1a,b) are first combined with the head and then with the resulting intermediate projection. Head-*initial* structures were deemed to integrate the complements on positions *following* the head (1c), while head-*final* structures were seen as the result of integrating the complements in positions *preceding* the head.

[1] For a non-empty set M and a relation $R \subseteq M \times M$: R is *asymmetric* $=_{def.}$ $\forall x,y \in M$: $xRy \Rightarrow \neg(yRx)$.

Later, starting with Barss and Lasnik (1986) and Larson (1988) and subsequent work, and contrary to Baker (2003: 350), syntacticians came to know that (1c) is *empirically* inadequate for English double-object structures in particular, and for double-object constructions with head-initial VPs in general. Unfortunately, this has not spread to typologists yet.[2] But, why is left-branching (1c) excluded on *theoretical* grounds? What is it that makes the symmetric siblings unavailable for the grammars of human language? In other words, the theory has to provide for *symmetry breaking*.

Particularly clear cases of a basic asymmetry are found at the *sentence periphery*. There are quite a few languages – e.g. Germanic languages (except English), Kashmiri, Romansh in Switzerland; see Holmberg (2012) for an up-to-date report – with a *V-second* property (= V-movement + phrasal movement to a clause-initial functional projection), but there are no languages with a *V-penultimate* property. This would be the mirror image of the V2-structure: the finite verb would move to a clause-final functional head, and a single phrase would move to the clause-final spec position of this very head.

The clause-initial phrase in the V2-structure is the result of A′-movement to a clause-initial spec position. This position provides alternative accommodation for various kinds of elements, neither of which has a mirror-image counterpart in other languages.

There are languages that move *question phrases* to the clause-initial position, both in main and in embedded clauses. There are no languages, however, that move these items to a right-hand *spec* position.[3] Analogously, neither *relative* nor *comparative* clauses show movement of phrases to the right. And finally, there is no language with a *clause-final expletive*, as a counterpart to the (Germanic) clause-initial expletives in V2-declaratives. This is a clear case of asymmetry, even if it has not been fully appreciated yet.

As for apparent rightward movement, there are numerous proposals for *extraposition* as an instance of A′-movement to the right, but proponents of these accounts remain silent on the grammatical causality of the profound syntactic differences between the alleged A′-movement to a clause-final position and the well-established cases of A′-movement to the clause-initial position

[2] Until now, typologists have tended (see, for instance, Dryer 2009: 185) to equivocate *head-initial/final* with *left-/right-branching*, without presenting any substantive syntactic evidence at all for this alleged correlation between word order and structure. The correlation between word order and structure is merely presupposed or stipulated and not substantiated by appropriate syntactic analyses that are sensitive for and predictive of structural differences.

[3] There are languages with clause-final wh-positions (see Chapter 3). But these are clause-final focus positions in a cleft construction, not the result of movement.

(see Haider 2010a: ch.5). Extraposition as A′-movement would violate several core constraints on A′-movement.[4]

The overall asymmetry puzzle struck me first in 1991 when I was asked why anaphor binding within a German NP apparently violates the c-command constraint under the assumption – until then unquestioned – that a head-initial phrase like the German NP is left-branching.[5] If the PP that contains the reflexive in (2a), or the bound pronoun in (2b), were indeed higher in the structure than their respective binders, the result should be a weak-crossover violation. But there is no detectable violation, and moreover the 'correctly' inverted order is deviant, contrary to expectations:

(2) a. die Wut [des Mannes^{i} [auf sich^{i}]]
the fury the man's at himself
'the man's^{i} fury against himself^{i}'
b. die Zerlegung [jeglicher^{i} Substanz [in ihre^{i} Bestandteile]]
the dismantling (of) each^{i} substance$_{GEN}$ into its^{i} components
'the dismantling (of) each^{i} substance$_{GEN}$ into its^{i} components'

The same considerations apply to English VPs, as discussed by Barss and Lasnik (1986). And the solution is the same, too: both head-initial and head-final projections are structured in the same way, namely right-branching only. This universal property of phrase structures has been dubbed the BBC (*basic branching constraint*) in Haider (1992/2000).

(3) BBC $=_{def.}$ The structural build-up (*merger*) of phrases (and their functional extensions) is universally *right-branching*.[6]

If a phrase α is merged to a phrase β,[7] the resulting structure is $[_{\beta}{}^{n}\ \alpha\ \beta]$. Hence, merger produces *right-branching structures only*. Left-branching merger structures $*[_{\beta}{}^{n}\ \beta\ \alpha]$ are universally ruled out.[8] The result is an *asymmetry* property for phrase

[4] Extraposition patently violates the core constraints of A′-movement (namely: no extraction out of a subject phrase, no extraction out of an NP inside a definite DP, no extraction out of an adverbial phrase).

[5] Thanks to Werner Frey, my colleague at the University of Stuttgart at this time, for this stimulating question.

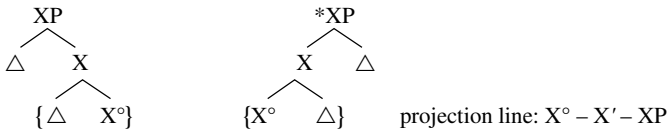
[6] 'Right-branching' $=_{def.}$ a structure is right-branching iff the node on the projection line follows the node attached to the projection line. In other words, the branching node on the projection line is on the right-hand side.

[7] 'Merge α with β' $=_{def.}$ combine α with β into a phrase structure $[_{\gamma}\ \alpha\ \beta]$, where γ is a projection of either α or β.

[8] This premise applies to (internal) merger. The BBC remains silent on the question as to whether there could be a transformational source of left-branching structures, as, for instance, adjunction to the right by movement to the right. Extraposition might appear as a prima facie candidate, but its properties (see Haider 2010a: ch.5) do not support this conjecture.

structures (4). The curly brackets in the bottom line of the structures in (4) are to signify that the branching restriction is independent of the linearization restriction of head and complement, that is, head-final or head-initial order. The linearization follows from the directionality parameter for identification by a head.

(4) a. **right**-branching b. **left**-branching



The BBC in combination with an identification requirement (5) between the head and its dependants produces a rich set of implications that cover a significant portion of the syntactic differences between head-initial and head-final phrases and clauses. The identification requirement is characterized as follows:

(5) *Directional Identification*: merged positions in the projection of the head of a phrase need to be properly identified by the head under *canonical directionality*.

Canonical directionality is the basic parametric factor that produces head-final or head-initial structures, respectively.[9] The constraint is formulated under (6):

(6) *Principle of Directional Identification (PDI)*: a merged phrase P must be properly identified.
Def.: A merged phrase P is *properly identified* by the *head* of the host phrase h° iff
(i) P is in the canonical directionality domain of h°, and
(ii) P and (an extension of) h° *minimally, mutually c-command* each other.[10]
(Extension of h° $=_{def.}$ projection of h° [-maximal])

The BBC and the PDI in combination immediately cover and explain a wide range of properties of head-initial vs. head-final structures. For details see Chapter 3.[11] Here is a summary of the main consequences:

[9] The idea that directionality is a relevant parametric factor is not a new one. It has been discussed since the advent of the Principles and Parameters (P&P) Model. (e.g. Koopman 1984; Travis 1984, 1989; Frazier 1985) and has been revived in the Minimalist Program (e.g. Saito and Fukui 1998).

[10] This subsumes any submaximal projection of a head, including the trivial projection, that is, the head itself.

[11] The *binary branching* property of phrase structures immediately follows from (6ii), once the relation is strengthened to *uniqueness*, that is, if identification is strengthened to biuniqueness: the head or its projection *uniquely* identifies the dependent element and for any dependent

Corollary 1: OV languages (i.e. phrases with a head-final organization) are the straightforward instantiations of the BBC in combination with the PDI, because the canonical directionality is in harmony with the branching direction for head-final phrases.

Corollary 2: Head-*initial* organization (VO) is necessarily more complex because of shell structures, but it has an advantage for processing. It presents the head early (preferred for bottom-up processing).

Corollary 3: Both OV and VO are each *necessarily* not fully optimal for the conditions of usage, that is, for the processing of grammatical structures when parsing them (see also Chapter 2):

(i) OV has the *simpler* structure, but the head is presented *late*;

(ii) VO has a more *complex* structure, but the head is presented *early*.

From the processing point of view, each of the two options has its unavoidable price. Hence it must not come as a surprise that we find them implemented in approximately the same percentage of languages. There is no unique way of organizing phrase structure such that it is optimal for the *simultaneous* application of *bottom-up* and of *top-down* strategies. OV and VO are the best approximations.

In addition, OV and VO are *not complementary*. There is a *third* option (*Type III*), namely the option of *underspecified* canonical directionality. The diachronically attested stages of Indo-European languages (and many present-day languages, as e.g. Slavic ones) are of Type III (see Chapter 3, section 3.5, and Chapter 5). The diachronic Germanic word-order split (North Germanic languages became VO; most West Germanic languages became OV) finds its explanation as a change from underspecified (i.e. Type III) to specified directionality. This change had two possible alternative outcomes, namely OV and VO. These options found their implementations.[12]

1.2 Symmetry breaking – adaptation for *time* and *dimension* management

The exclusion of left-branching structures is the result of symmetry breaking. This contrasts with the fact that there are *apparently symmetric* linearization

element there is not more than a single identifier. In an n-ary projection, the identifier would minimally c-command – and be c-commanded (by) – more than a single item.

[12] The fact that this happened only in the Germanic family is not accidental. Romance languages started from the Type III Latin and each language ended up as a VO language. It was the V2-property of the Germanic languages that became the key property for opening alternative channels for the change from underspecified to specified directionality (see Chapter 5).

patterns. The German examples in (7) *appear* to be symmetric in their linear order, since either the head precedes the complement, or the complement precedes the head.

(7)

a.	$[\text{P DP}]_{\text{PP}}$	$\text{wegen}_{\text{P}^\circ}$ des $\text{Geldes}_{\text{GEN}}$	– (for the) sake (of) the money
b.	$[\text{DP P}]_{\text{PP}}$	des $\text{Geldes}_{\text{GEN}}$ $\text{wegen}_{\text{P}^\circ}$	– (for) the money's sake
c.	$[\text{PP V}]_{\text{VP}}$	auf $\text{jemanden}_{\text{ACC}}$ $\text{warten}_{\text{V}^\circ}$	– for someone wait
d.	$[\text{V PP}]_{\text{VP?}}$	$\text{warten}_{\text{V}^\circ}$ auf $\text{jemanden}_{\text{ACC}}$	– wait for someone
e.	$[\text{PP A}]_{\text{AP}}$	mit $\text{jemandem}_{\text{DAT}}$ $\text{unzufrieden}_{\text{A}^\circ}$	– with someone incontent
f.	$[\text{A PP}]_{\text{AP?}}$	$\text{unzufrieden}_{\text{A}^\circ}$ mit $\text{jemandem}_{\text{DAT}}$	– incontent with someone

However, a little knowledge about German syntax suffices for destroying the symmetry suspicion: a pre- and postposition such as *wegen* is the rare exception in German, and the other examples are secondary patterns. The phrase-final PP in (7d,f) is in an extraposed position. Symmetric patterns are rare for a given *language*, and so they are *cross-linguistically*, apparent mirror-image examples of the kind in (8a,b) notwithstanding.

(8)

a.	walk along the river	
b.	den Fluss entlang gehen the river along walk	German
c.	be content with this	
d.	damit zufrieden sein this-with content be	German

Again, this cross-linguistic symmetry – English is serializing in one direction, German in the other – is merely incidental but not representative. A representative case is illustrated in (9). The order in complex head-initial phrases compared to the order in complex head-final ones is identical, modulo the opposite positioning of the *head*. There is no mirror-image order otherwise (Haider 1992). An exemplary piece of evidence for the shared c-command property is the binding effect between quantifier and the bound variable in the form of a pronoun.

(9)

a.	dass sie *jedem*^{i} *ein Paket an seine*^{i} *Privatadresse* schicken werden that they everybody a parcel to his private address send will	German
b.	omdat *ze iedereen*^{i} *een pakje naar zijn*^{i} *privaatadres* zullen opsturen that they everybody a parcel to his private address will send	Dutch
c.	that they will send *everybody*^{i} *a parcel to his*^{i} *home address*	
d.	at de forklarede *hver deltager*^{i} *problemet på hans*^{i} *eget sprog* that they explained every participant problem-DEF in his own language	Danish

This invariance property follows immediately from the BBC and the PDI. The BBC accounts for the universal right-branching structure and the PDI accounts for the positioning of the head on the phrasal periphery (preceding or following the dependent elements, respectively). The argument structure of the verbs guarantees that the ranking of the arguments determined by the lexical argument structure of the head is mapped onto the syntactic structure. In sum, this accounts for the invariant cross-linguistic order of the dependent elements, with the conclusion that the source of asymmetry is the BBC.

Why should natural languages embody the BBC as a restriction on their core grammar? The answer must be sought in the *cognitive evolution* of grammars. Languages, just like species, are subject to evolution. In the case of grammars it is *cognitive* evolution rather than *biological* evolution.[13] Evolution is adaptation by selection out of a pool of variants. For languages, the selector is the (child's) processing brain that has to acquire the grammar of a given language by merely being exposed to the language. Like in biological evolution, the winner is the variant that 'infests' more brains than the other competing variants. The winning variant multiplies itself more often. And just like in biological evolution, the emergent result is an *adaptive* quality. See also Haider (forthcoming).

What is the adaptive quality of a right-branching architecture that is not embodied in a left-branching architecture? The adaptive quality, as in biological evolution, is a quality that becomes operative in the environment of the system/organism. The environment of grammars is *language processing*. Grammar is the system that the brain employs for language processing, either in reception or in production. What is the adaptive advantage of *right*-branching over *left*-branching, then? It is the greater ease for time and dimension management.

Processing time is a limited resource (Haider 1997e) and right-branching structures offer a significant advantage for *dimension management* under *time* limitations. Grammar is a dimension management system. It is the algorithm that we employ when we map one-dimensional arrays (PF) onto at least two-dimensional box-in-box structures (SF), and vice versa.

Syntax is (in part) an algorithm that projects at least two-dimensional structures onto one-dimensional arrays (i.e. the phonetic/phonological representation) of terminals (in reception) and compresses two-dimensional structures to one-dimensional strings of terminals (in production). It thereby bridges a

[13] A grammar is a self-reproducing cognitive system. It is a cognitive virus. The language learning brain gets 'infected' by a given grammar variant in the course of language acquisition. The acquired grammar then determines the language which is the basis for spreading this grammar to other brains.

dimension gap. It enables the mapping of the one-dimensional representations (strings) of phonetic/phonological structure to the at least two-dimensional hierarchical box-in-box structure of semantic representations, back and forth. The dimension mismatch is an unavoidable consequence of the respective interfaces. Sound structures are organized along the time axis (linear organization), conceptual representations are timeless, hierarchically organized complex structures (hierarchical, box-in-box organization).

The receptive side is the crucial filter. It is the input side for language acquisition and it is the side with the limited resource in language use, namely the processing time span. This is the selector for cognitive evolution. Any data structure that can be processed more effectively will have an advantage over a structure that incurs higher processing costs. The production side, on the other hand, is under no time pressure, but the perception side is. As a speaker, I may consume as much time as I need for the production of an expression. But, as a listener, I am dependent. I have to finish the processing of a given expression while the next utterance is coming in. Otherwise my processing channels will get jammed.

Given the resource limitation, this amounts to a constant selection mechanism for an organization of data structures that saves processing time. The data-to-parser fit is optimal if the parser – a left-corner parser – can *simultaneously* operate bottom up and top down, that is, with immediate and continuous data processing (bottom up) plus immediate *grammar guidance* (top-down information on possible structures). This clearly favours right-branching structures (10b) over left-branching ones (10a):

(10) a. $[[[[h^\circ\ X]\ Y]\ Z]_{HP}\ \ldots]$ left-branching
b. $[\ldots\ [Z\ [Y\ [X\ h^\circ]]]_{HP}]$ right-branching

The bracket notation in (10) already demonstrates the disadvantage of (10a): the parser has to *guess* how many brackets need to be opened. Without look-ahead, the parser cannot possibly know how deeply embedded the phrase is. So, the cost is frequent backtracking.

(10b) does not involve this disadvantage at all. For every phrase under construction, its mother node is necessarily the *immediately* dominating node. This is the optimal architecture for simultaneously combining top-down information (associated with the dominating phrasal nodes) and bottom-up information, provided by the terminals that are parsed at any given moment.

On top of this, there is involved a third facet. This is the lexically stored syntactic information associated with the head of the phrase. Early availability of this information is advantageous for predictions in parsing. (10b), however, is

head-final, because of the BBC (as a reflex of the advantage of right-branching) and the fact that phrases are endocentric: a phrase is a head plus its dependants, and the structural build-up starts with the head. Hence the head is in the most deeply embedded position. Improving this suboptimal situation is the source of VO structures, but this improvement is costly to a certain extent.

Presenting the head of a phrase earlier incurs costs in a BBC-constrained system. The cost is the more complex structure as a result of the mismatch of the canonical directionality and right-branching. Head-initial structures induce a shell structure. Consequently, a head-final structure on the one hand, and a head-initial organization on the other, are equivalent, that is, equally costly or equally cheap outcomes of *cognitive selection* for parser friendliness, otherwise one of the two types would have vastly outnumbered the other in the course of grammar change over the past millennia.

1.3 Head first or head last – costs and gains

SOV and SVO are the most widely attested basic serialization types for natural languages. This means that a grammar variant that produces one of these two serialization types is more often superior to its competitors and therefore it is reproduced more often in language acquisition. Why does one of these two types not outperform the other, or, even stronger, why are not *all* languages of only *one* of these types? A satisfactory answer must be able to locate cost and gain characteristics that each of these types fulfils equally well. Here is a sketch of the answer.

OV is harmonic with the BBC. The canonical directionality of OV is identical with the directionality imposed by the BBC. As a consequence, phrases are head-final. The result is a strict OV language. But, if the early presentation of the head is an advantage for the parser, the head-final property of the OV type is a disadvantage at the same time. Compensation for this disadvantage has its costs, however. Early presentation of the head requires a *more complicated* structure, namely a shell structure (11).

(11) [… [Z [$\dot{V}_i$ [Y [e_i X]]]]]

The head position is in the deepest position with the required directionality in (11). This is the only position that obeys the BBC and the PDI in each of the two directionality settings. The position of the two other dependants, namely Y and Z, is on a left branch according to the BBC, and therefore not in the directionality domain of the lower head, if its canonical directionality is to the right. Hence, the PDI is not fulfilled. The solution is the re-instantiation of the

head. This produces the shell structure in (11). A simple count of the closing brackets in (11) and (10b) shows that (11) is more complex than (10b). This is the cost factor: either a simpler structure, with head *last* as 'disadvantage', or head *first*, but with a complex structure as 'disadvantage'. Structural complexity is the price for the advantage of an early presentation of the head, and late presentation of the head is the unavoidable disadvantage for the structurally simple head-final architecture.

If the verb in (11) moves even across Z, the result is the VSO order. In SOV, the subject is merged as the highest local argument in the VP, but in this position, it is not within the directionality domain of the head, and hence it is not directionally identified. It needs a c-commanding head for its identification. This is the source of the *obligatory functional subject* position in VO (viz. Chomsky's EPP property). The functional head that selects the VP identifies the position of Z in (12a). It may attract the finite verb, and its spec accommodates the raised subject. SOV, in comparison, does not need any of these measures. The objects and the subject are merged *within the same* directionality domain (12b). Each dependant is on the canonical side, identified by the head or one of its projections.

(12) a. $[_{\text{Spec-F}^\circ} \ldots [\text{F}^\circ\rightarrow [\text{Z} [\text{V}_\text{i} \rightarrow [\text{Y} [\text{e}_\text{i} \rightarrow \text{X}]]]]]]$
b. $[\text{Z} \leftarrow [_{\text{V}'} \text{Y} \leftarrow [_{\text{V}'} \text{X} \leftarrow \text{V}^\circ]]]_{\text{VP}}$

The special treatment of the subject introduces a specific set of asymmetries into VO that are alien to OV, namely the structural asymmetry between the VP-internal objects and the VP-external subject. In the VO-biased literature on grammar theory, this property has mistakenly been elevated to the status of a universal (cf. Chomsky's EPP). It is a universal, but only of the VO clause structure.

1.4 The OV/VO syndrome

The interaction between the BBC (3) and the PDI (6), with its parametric option for the canonical directionality, is the source of an intricate set of syntactic properties that predict and cover systematic differences between head-final (OV) and head-initial (VO) phrases and clauses. Here is a non-exhaustive list of ten syntactic properties that are differentiated by the head-initial vs. head-final property of syntactic phrases in general and VO vs. OV in particular:

(13) a. Complex head-initial structures require a *shell structure*; head-final structures do not.
b. Head-initial structures are *compact*, head-final ones are not.