BBC or LCA? – Fact finding and evaluation

0. Background

The agenda of the final chapter is the comparative evaluation of the role of asymmetry and linearization in two competing approaches: BBC (Haider 1992) vs. LCA (Kayne 1994). It is a comparison of competing conceptions of the relation between an *input/output property* at the PF-interface, namely *linearity*, and a *computational property*, namely the *structural organization* of syntactic representations. And moreover, these are *complementary* conceptions as for the basic syntactic process that relates the cross-linguistically apparently diverse structural architectures (viz. head-*final* and head-*initial* structures):

In the BBC model, the basic quality that relates and distinguishes OV and VO[1] is *head positioning* (viz. the make-up of the shell structures of complex head-initial phrases, or the partial shell formation in the type III structures. In the LCA model, however, the basic tool is *phrasal* movement. SVO is taken as axiomatic and OV is derived from the axiomatic VO structure by means of massive movement of *non-head* constituents.

Given the fact that the two models employ complementary grammatical tools, viz. *head*-chaining vs. *phrasal* movement, it should not be overly difficult to identify and empirically falsify or 'verify' relevant predictions of each model. Since the two models cannot both be fully right, at least one of them must be eventually identifiable as empirically more adequate and consequently also theoretically superior.

In the BBC perspective, grammar is a cognitive algorithm for mapping *linear arrays* (strings) onto hierarchically organized *box-in-box structures* (= phrase structures) and vice versa; in other words, grammar is conceived of as a *string-to-structure* or *structure-to-string mapping algorithm*. The fact that utterances look like *linear concatenations* of terminals is primarily an interface property (the phonetic channel provides only a single, viz. linear, dimension, and its only degrees of freedom are *precede* and *follow*).[2]

From the perspective of the mapping algorithm, the basic asymmetry of phrase structures – *No left-branching projection*![3] – is understandable as the UG-determined parser friendliness of grammars (see chapter 2), which is a result of adaptation by cognitive selection in the cognitive co-evolution of grammars and grammar processing. The crucial gain of this type of structuring is the simultaneous applicability of bottom up (= data driven) and top down (= grammar driven) processes, since the earlier (= left) element in a phrase is always the higher one. 'OV' and 'VO' are just the effects of the alternatively admissible directionality factors, given the string property of linguistic expressions and the universal phrase structure constraints. Technically, these data structures are optimal for a *left corner parser* (see Phillips 1995, 2003).

[1] Here, ‚VO' is meant to cover both SVO and VSO, abstracting away from the final destination of the verbal heads that start out in a head-initial V-projection, and also the V-movement in the 'third type'.

[2] This characterization applies to the organization of morpho-lexical terminals. This does not exclude that there are other tiers, for other information qualities (e.g. intonation contours). But, crucially, for the morpho-lexical terminals, there is only a single dimension, namely the dimension of the time course of the utterance. This is a linear dimension.

[3] In *merger* terminology of Chomsky's Minimalist Program, this reads as follows: merger in a given projection is to the left.

The mainstream stance in grammar theory of the generative camp favours a different, opposed view, namely a *structure-on-structure mapping algorithm*. Grammar is regarded as an algorithm for *generating* structures, successively mapping them onto other structures in the course of the derivation and linearizing them ('spell-out'). Grammar is not primarily seen as a *direct* PF-to-structure-to-SF ('semantic form') mapping algorithm, but rather as a *generation algorithm* that takes lexical items as input ('numerations') and generates a series of representations with specific PF and SF properties. This approach, despite its popularity, is neglecting and therefore missing a crucial objective, both on empirical as well as on theoretical grounds. It neglects the undeniable fact that a grammar is put to use as a mapping algorithm, and that the conditions of use are part of the cognitive selection environment that shaped the grammars in the course of their cognitive evolution.[4]

It is reasonable to take for granted that grammars have the core properties they have because the brain provides effective processing routines (as the neuro-cognitive substrate that is reflected by and compiled in UG) for the unique properties of linguistic expressions characterized by these grammars. In other words, it is entirely *un*reasonable to maintain that grammars might primarily hamper the processing routines by insisting on structures and relations that are hard to process or even intractable for the human brain's processing capacities.

Let me very briefly point out two aspects of language processing that clearly favour a *string-to-structure mapping* perspective over a *generation & linearization* view.[5]

i) *Language acquisition* (feasibility): L1 acquisition presupposes a 'mapping algorithm', not a 'generation algorithm': What the child is confronted with are linear arrays, whose structures and regularities the child has to detect. UG ought to support the child in identifying the core grammar properties of the string-to-structure mapping properties of the language the child is confronted with, otherwise the task will be beyond the child's capacities.

   A mapping algorithm is tractable for the child, a production algorithm is not. The latter is not tractable because it *starts* with *a hidden step* (viz. with the generation process on the speaker's side) and produces the overt product (linearization). The mapping algorithm, however, starts with the *public part* (the linearization) and yields the hidden quality (the phrase structure). Since in a production grammar, the start is hidden, the non-deterministic derivational history of expressions in a production grammar makes is impossible for the child to identify the language specific initial state. S/he could do so only by constant analytic step-by-step backtracking (*reverse engineering*), which is intractable (even with strong UG support), however, since a UG compatible core grammar is not narrowly deterministic. UG, as is well-known, must allow for considerable cross-linguistic variation which – see the OV-VO debate – is characterized as the effect of complex derivational relations. How could the child find out the *language specific* settings, if they are hidden in a non-deterministic derivational algorithm? By testing & retesting?

[4] Here are examples of structural parsing theories with principles that favor right branching structures: late closure (Frazier 1987), recency (Gibson 1991), connect bottom-up (Stabler 1994), low attachment (Abney 1989).

[5] Much of the attractiveness of the generation & linearization approach, in my opinion, rests on its fascinating system complexity. It is much more entertaining than a simple string-to-structure mapping algorithm because it allows devising and solving intriguing puzzles of possible derivational interactions of rules under various conditions of application. What I doubt is that the human language faculty is algebraic, that is, a *derivational* puzzle-solving algorithm. I prefer the conjecture that is geometric, that is, a *gestalt mapping* algorithm.

It is highly unlikely for a child to proceed in a *testing mode* that continuously fixes and re-fixes several interacting parameters in order to determine what is the momentarily optimal fine-tuning for a given input language. This would presuppose a highly effective, domain-general capacity of executive control and crypto-analytic supervision, which is unavailable for a child's brain. Even professional linguists are struggling with this task since decades. A child is unaware of the problem and accomplishes the task within months or few years.

Generative (viz. structure-generation) models give rise to a *learning paradox* which does not arise in a mapping model (Haider 1999): in order to know what the structure of a given input is, you have to generate it (and compare its spell-out with the input), but in order to be able to generate it, you must know what the structure is, that is, you must have learned how to fix the various open parameters that apply to the given structure. This in turn presupposes that the learning process crucially involves a comparison process: generate the alternative structures and compare them and find out which one is the adequate one. If, as it is the case, the structures (as in the case of phrase structures) involve more than one parameter, the multi-dimensional parametric system space of alternatives soon becomes too complex and intractable for any human learner.[6] The direct mapping model (BBC-based) does not suffer from this paradox, evidently. The learner locally compares direct mapping alternatives rather than complex alternative derivations that produce the mapping as their output.

ii) *Processing efficiency (economy of usage):* Efficiency involves the economic employment of limited resources. What is the limited resource in language processing? It is the processing time in the *comprehension* process, that is, the time for processing the structure of a given string when parsing it. It is not the time for the *production* (generation) of a string. In principle, I could allow myself as much time as I need for generating an utterance. But, when I am engaged in processing the on-going discourse, I have to be finished with an utterance when the following utterance reaches my brain. So the efficient system is one with *mapping* efficiency rather than *production* efficiency[7] (see Haider 1997e). The generation-algorithm approach has no direct way of optimizing the receptive side. This side is not a respected part of the system. The string-to-structure mapping, on the other hand, provides a natural system-property. Effectiveness is enhanced by gestalt properties of pattern matching procedures. Patterns are compiled into *complex gestalt properties*. For the generation approach, compiling would amount to a packaging of derivations. A packaged derivation is cognitively intractable; a complex gestalt property, however, still is in a direct and continuous relation to the string properties of terminals.

[6] Note that the appeal to the 'invisible guiding hand' of UG would not help. It is precisely the variation space admitted by UG that is the problem for the learner. The learner cannot know in advance that (s)he is exposed to a VO or OV or third type language, with or without V_{fin}-movement, for instance.

[7] A production system like the Minimalist Program is of no advantage for the production process either, ironically: It starts with a lexical choice ('numeration'), but when it is going to assemble the pieces, it takes off from the 'wrong' side, namely the bottom. Imagine me as a speaker of German, who is planning a medium length utterance. I cannot utter the first word until the whole sentence has been generated bottom up, if production is grammar based, and the grammar is organized like the MP wants to have it. The first word is the top item, but generation starts at the bottom. The inevitable but wrong prediction is this: every speaker knows the very last word of the utterance already when he starts with the very first word.

Once one abandons the generation-algorithm bias, the LCA-based system loses its major attractive quality of imposing structural restrictions on the derivation of a structure suitable for the given linearization. In fact, the LCA merely serves as a *benchmark condition* for production grammars. The benchmark condition is that a grammar must guarantee a particular function, namely the unique mapping of structures onto linearizations. The benchmark is checked at the PF interface: Does the given representation meet the LCA-requirement? LCA-compatible grammars are grammars that are organized such that their outputs meet the LCA. Assuming LCA as a principle of UG is like postulating "I expect a grammar to organize the data structures in such a way that this guarantees the property I require to hold, namely a unique mapping of structure to linearization, based on asymmetric c-command."

Note, that the LCA as a benchmark condition is a *global* well-formedness condition rather than a principle of phrase structure generation ('transderivational constraint'): If a given representation cannot obey the LCA (e.g. a head that merges with several complements simultaneously), a movement operation has to be applied and the grammar has to enable this by providing a functional projection to accommodate the moved item (see Kayne 1994:30, on the small-clause structure of double object constructions).

From the string-to-structure mapping perspective, it is not evident why one should adopt the LCA as an axiom from the beginning. The mapping function[8] it is meant to cover is crucial only in a generation perspective, but not in a string-to-structure mapping perspective, since the inverse of the LCA – *Generate a structure, starting from a linearization*! – remains just one of the many one-to-many mappings[9] embodied by grammars.

That the LCA allows deriving restrictions for generating phrase structures that previously had to be simply postulated is not enough justification, though. It must be shown that the restrictions plus the implications of the model are empirically adequate. In section 3, the focus will be set on empirical issues that play an essential role in evaluating the relative success of the LCA approach in comparison with a BBC-based account.

1. Verification versus falsification

An obvious and crucial testing ground for the empirical adequacy of an LCA- compared to a BBC-based account of the OV-VO distinction is the empirical success of the given model in explaining the contrasting aspects of grammars. The crucial question that calls for an answer is this: How straightforward or how recalcitrant do the numerous, apparently independent structure properties of closely related languages fit into the respective theoretic model?

It is not unfair to underscore that the applicability of the LCA-approach benefits to a significant extent from its high degree of derivational *flexibility*.[10] In system-internal terms, it is free & costless to move to the left as many phrases as one deems necessary and thereby introduce as many functional layers as one needs as landing sites, as long as the result obeys the LCA, -

[8] For every structure there is a unique linearization the structure is mapped on.

[9] For a given linearization there may be more than one LCA-compatible structure (simply because it may involve empty positions that are not visible/audible for linearization.

[10] Flexibility is greatly enhanced when – as it is the case – the LCA environment is combined with the Minimalist Program's abundant tool kit: *overt & covert movement, covert lowering, remnant movement, roll-up movement, strong or weak features, overt or covert checking of features,* etc.

if one does not hesitate to flatly disrespect Occam's razor as a maxim of scientific work (and the obligation for providing independent evidence for each of these hypothetic derivational steps).

From the point of view of theory evaluation, most applications of the LCA merely pursue a *verification* approach. You simply show that the system is able to produce a reasonable output after a series of derivation steps. But it is well-known that the fact that a model is flexible enough to capture a given phenomenon (or a subset of phenomena) does not prove the model to be empirically adequate and correct. It is a truism of the theory of science that verification is not the decisive step. The decisive test for a theory is demonstrating its merits under thorough falsification trials.

What is crucially underrepresented in the LCA literature is this scientific obedience to thorough *falsification testing*: What are the main empirical predictions and how successful does the proposed system deal with them? The problems are less conspicuous for VO languages, the favorite demonstration area for the LCA-model. For OV languages, however, the problem areas are numerous and easy to identify (see section 3, on comparative counterevidence). The literature has proliferated many affirmative contributions (e.g. DiSciullo ed. 2003, Haegeman 2001, Zwart 1996), but no systematic accounts of the potentially falsifying empirical evidence, with few exceptions, as for instance Haider (2000). Answers are still missing in Kayne (2005), where essential LCA-saving derivational procedures are extrapolated for Japanese, but the paper stops short of justifying the adequacy of these postulated measures on the basis of independent evidence from Japanese. It is not enough to merely point out a way through a maze if one cannot plausibly make sure that it eventually leads out of the maze rather than deeper into it.

2. Theoretical shortfalls

This section lists obvious areas of actually or potentially falsifying evidence for an LCA approach. A prominent problem for the LCA system is the *missing trigger theory*. There is no general account for a reasonable *trigger* in the derivational continuation from VO to OV, that is, a trigger for turning an otherwise well-formed VO base structure into an OV structure. In the LCA perspective, an OV language appears to be an unexpected and puzzling distortion of a well-behaved VO system: If the OV structure is indeed derived from an underling VO organization, this is the primary question: Why should a well-formed VO-structure be further transformed into an OV arrangement by starting a machinery of massive derivational dislocations and evacuations? Why are *not all* languages VX languages, if UG allegedly is so strongly VO-biased? Why did non-VO languages not change into VO languages in the preceding millennia of diachronic changes, given the prevailing system pressure towards VO?

The lack of a trigger-*theory* is a serious shortfall of the LCA model.[11] In the absence of a trigger theory for movement, the LCA theory simply postulates that natural languages are head-initial languages. But there are OV languages. So, if the OV order is the result of derivational-

[11] Note that a *trigger theory* is not identical with a set of guesses as to what particular features might be useful candidates for the unknown trigger of the respective individual movement process. A trigger theory is an essential part of the theoretical account. It must follow from principles of the theory as an integral part and it must be empirically adequate, of course.

ly transformed head-initial projections, this transformation must be triggered. Moreover, the trigger in the LCA system must be a *global* one because it indiscriminately affects all elements in a superficially head-final structure, independent of their category and grammatical function: arguments, attributes (including relative clauses), adverbials, particles, (secondary) predicates, and so on.

Holmberg (2000) proposes *c-selection* (i.e. strict subcategorization) as the parametric locus for the global trigger of the respective movement types: in his view, the basic difference between OV and VO languages is reducible to a difference in terms of checking c-selection features: In OV languages, c-selection features are to be checked by phrasal movement, but in VO-language they are checked by head movement. As he himself acknowledges, this fails to account for the serialization of non-subcategorized elements, as PP adverbials, adverbial clauses, secondary predicates, extraposed relative clauses, and so on, without help of additional non-standard assumptions.

What is equally problematic for Holmberg's proposal is the existence of mixed language types, such as for instance German and Dutch: N°-, most P°- and all functional projections are head initial; V°- and A°-projections are head-final. If the trigger is framed as a computational property of the different implementation of a feature checking mechanism, parameterization in terms of category subclasses would amount to a language-internal parameterization on the computational level, and not on the level of lexical properties. This raises additional questions as to how plausible and insightful this account would be. In addition, Holmberg's proposal remains open on all the particular empirical issues that are raised against an account that takes head-initial structures as a starting point for deriving head-final ones (see below).

The really tough task for devising a triggering scenario is the *mixed bag characteristics* of the elements that have to be made to move when turning a basic VO structure into an OV structure. The set of elements to be moved out of the VP is completely inhomogeneous. The LCA simply requires moving *any* elements whatsoever out of the head-initial phrase that is to be turned into an apparent head-final structure. This leaves hardly any room for a grammatical feature that could be made responsible for the distribution of the elements that are allegedly moved in order to arrive at a consistent OV linearization. But, it would be just begging the question if one invented a 'kick out' feature as a feature of the head that makes the phrase internal elements flee their place of origin and move to the left just in order to produce the impression that the head is a phrase-final head. Presently, the model operates as if there was an *expulsion feature*, without admitting it, though.

The mixed bag problem could be circumvented, as proposed in Kayne (1998:134), if the clause final position of the verb was the result of V-movement out of the VP, followed by remnant VP-fronting across the moved V. This, however, is equally problematic, for different reasons, however. As pointed out in Haider (2010a, ch.2), there is diverse, robust and recalcitrant counterevidence for a proposal that assigns the (clause final finite) verb to a functional head position. Second, invoking clause internal remnant VP fronting would over-generate excessively (see the discussion on the distribution of particles and on opacity of extraction, below).

3. Comparative (counter)evidence

Let me repeat: in the LCA model, OV is derived from a structure that is basically VO. Post-verbal material is fronted to preverbal positions and even the emptied VP may be moved subsequently, if necessary. Hence, if OV contains VO as a derivationally transformed core, ceteris paribus, OV is predicted to embody VO properties (that is, at least the subset that is unaffected by the derivational changes) plus a derivational fringe benefit.

For the sake of concreteness, let me take German and Icelandic as representatives of Germanic VO and OV languages, respectively. A comparison of Icelandic and German is a most advantageous testing ground for competing grammar theoretic accounts on the relations between 'VO' and 'OV'. Icelandic is consistently head-initial ('VO'), whereas German is head-final in the [+v]-projections, that is, for VP and AP. The two languages differ with respect to headedness (OV vs. VO in the narrow construal of OV/VO) but their morpho-syntactic systems of verbal and nominal inflection are similar to an extent that guarantees a broad enough common background for identifying the specific effects of the different headedness options. Moreover, German and Icelandic are historically closely related members of the same language family, and they share the Germanic V2-property.

In the LCA account, German would turn out as a particular derivational continuation of the generation of Icelandic structures. So, German is predicted to be derivationally more complex than Icelandic. In the BBC approach, however, Icelandic and German structures are the result of implementing opposite directionality values for identification. Here, Icelandic is predicted to be the structurally more complex language since it not only requires VP-shells, but also a functional projection for the subject, because of its head-initial V-projection. Let us therefore look out for straightforward predictions based on contrasts and similarities between Icelandic and German in particular,[12] and the *mixed-bag problem* in general (see also Haider 2000b). Consider the following German example. The four underlined items are obligatorily post-verbal in VO (1b), and hence they would have to be fronted somehow, given the LCA premises on deriving OV from VO.

(1) a. dass ja keiner diese Dinge heutzutage Syntaktikern auf ihr Brot schmieren muß
that nobody$_{\text{-Nom}}$ PRT these things$_{\text{-Acc}}$ these-days syntacticians$_{\text{-Dat}}$ on their bread butter must
b. that nobody must butter syntactians these things on their bread these days

The four items in (1a) consist of two objects, a PP interpreted as a directional predicate, and the main verb. The objects may come in any order, relative to each other or to the subject. This fact makes is unlikely that they target special licensing positions for nominal arguments, e.g. for the purpose of functional checking of case. PP arguments are optionally pre- or post-verbal in German, that is, they appear in the midfield or in an extraposed position. Case licensing as a trigger for fronting would fail in this case as an explanation for the preverbal PP positioning since PPs are not assigned case. Analogously, preverbal infinitival argument clauses in German are a puzzle for a case-based (or nominal feature based) trigger theory.

The PP in (1) is a directional PP. This is a sub-instance of result predicates. Result predicates are obligatorily preverbal in OV, regardless of their syntactic category. This separates result PPs from PP objects. Only the latter are extraposable in German and in Dutch.

[12] This section is largely based on a more detailed account of the pertinent phenomena presented in Haider (2005).

Finally, the verb (or the emptied VP containing only the verb) must have been inverted. Three grammatical categories (DP, PP, VP), three grammatical functions (argument, secondary predicate, main verb (phrase)), three different triggers for a uniform result? Technically, this is no problem. We could simply postulate an '*expulsion feature*' with the property that it cannot be checked within the licensing phrase. So, the DPs, the PP, and the VP must leave their original home position.

Obviously, this would be entirely ad hoc, even if it worked somehow. However, grammar theory is not practical engineering. It strives to become an empirical science. For science it does not matter whether an idea 'works' somehow. The idea must be shown to be empirically adequate and superior to all presently known alternative accounts.

For an *engineering* problem there may exist many diverse and mutually incompatible solutions. Each one may be a viable solution in itself, though. However, if there appear to exist several incompatible solutions for a problem in *science*, at most one of them has the chance of being (close to) the 'right' one. If there is a competing theory, you have to show why yours fares better, and if there is none, you have to investigate and check possible alternatives. There are empirical obstacles for any interesting theory, but your particular solution will be uninteresting and ad hoc, as long as you merely propose ancillary assumptions but do not present solid independent evidence for the assumptions that are needed for protecting your central assumptions. It is not prejudiced if one states that thoroughgoing, uncompromising empirical testing of its consequences has not been a key part of the biography of the by now already teen-aging LCA theory.

Under the BBC approach, the difference between (1a) and (1b) is predicted and follows from the opposite licensing directionality: In OV, the arguments have to precede. In (S)VO, the subject precedes, the objects follow, due to the different licensing relations (lexical, functional). The main verb precedes the auxiliary in OV, unless the auxiliary is fronted. Finally, result predicates are selected by a governing verb (see chapter 7) and thus they precede in OV, and they follow in VO. A simple, single fact of grammar covers the distribution of the four elements in (1a) and in (1b), respectively, without any ad hoc excuses and hand waving.

Let us now enter the German-Icelandic fact finding mission. The following non-exhaustive list of OV/VO-dependent properties will be discussed in the given order:

(1) functional subject position and EPP-effects
(2) quirky subjects
(3) opacity for extraction
(4) scrambling versus object shift
(5) particle-verb order (even in the case of verbs that do not move)
(6) stacked VPs vs. V clustering
(7) participle agreement
(8) word structure
(9) diachronic history of VO and OV
(10) verb clustering and verb order variation

First, the *obligatory functional* subject position is the hallmark of SVO clause structures. In OV, the subject is VP internal and it remains VP internal. Since it is within the directionality

domain of the head, just like any other argument of the head, it does not need an external functional licenser. As a consequence, there are no robust subject-object asymmetries in German (or any other OV language) that would compare to grammatical subject-object asymmetry contrasts in English or in other Germanic VO languages.

Generative Grammar has not put forth an interesting answer for the fact that VO clauses have obligatory functional subject positions. Instead, an EPP-feature is postulated that allegedly needs to be checked by a functional head and something in its spec. This is merely begging the question. In plain English, the answer reads as follows: The grammatical property is as it is because there is a grammatical feature that produces this property. This merely shifts the problem to feature theory and urges paraphrasing the question: Why should grammars contain such a particular feature? The theory itself does not provide an explanation. This situation would be acceptable as long as the property covered by the assumed feature is empirically valid. Its validity is improbable at least, given the cross-linguistic counterevidence.

OV languages do not have obligatory functional subject positions. The uninteresting answer would be that they do not possess the EPP-feature. Again, this ignores a crucial question, namely: Is the presence/absence of an EPP-feature an independent property? If it is so, we ought to see OV languages with and without the EPP-property and we should see VO languages with and without the EPP-property.[13] This is not the case, and therefore the feature story loses explanatory force because of the weakness of its narrowly circular formulation.

OV languages do not implement obligatory functional subject positions and therefore they do not need obligatory expletives[14] in case the obligatory functional subject position could not be lexicalized. The fact that VO needs a functional licenser for the subject is a VO property. The OV situation is the unmarked situation: all arguments are structurally equal. They are projected into phrase structure within their VP. They are licensed there and therefore they stay there. In SVO, the subject is the only VP-internal argument that is not in the directionality domain of the verbal head. Hence, it needs an external, functional licenser. This is the causality of the otherwise unexplained EPP property.

Second, under the LCA approach, German makes available the grammatical 'ingredients' for the *quirky subject* phenomenon. However, there is no quirky subject phenomenon in German. Icelandic admits quirky subjects, German does not. Given the parallel situation in German (viz. availability of VP internal nominative licensing and the availability of inherent morphological case marking), German should/could have quirky subjects too, says the LCA account. Both languages allow VP-internal nominatives and both languages have morphological datives. The functional subject position is not reserved for checking nominative and therefore it is able to accommodate the highest-ranking DP of the VP. If the highest ranking argument happens to be an inherently cased argument (e.g. in the passive of intransitive verbs with a dative object, or in a clause with an unaccusative dative-nominative verb), the result should be

[13] See Biberauer & Roberts (2005) for an EPP-based discussion of diachronic developments in English.

[14] In Scandinavian languages, *intransitive* verbs may be passivized, and there is an obligatory expletive. Modern English, for instance, does not have a suitable candidate for a dummy subject and therefore intransitive verbs cannot be passivized: 'There' must be associated with a postverbal DP, and 'it' is either a lexically selected quasi-argument or it is associated with an extraposed clause. There is no independent dummy element available.

a quirky subject construction, with the oblique DP in the subject position and the nominative in the VP-internal position, as in Icelandic.

German and Icelandic would just differ with respect to the postverbal occurrence of arguments in Icelandic. In German, these would be fronted to preverbal positions. But the candidate for the functional subject position would be the same, given that German and Icelandic both have verb classes with an oblique argument as the highest ranking argument.

The BBC theory predicts a different outcome: Icelandic employs a functional subject position, because the highest VP-internal argument – the argument in the preverbal VP-spec position – is not identified by the verbal head. So it needs a functional identifier, which in turn is associated with a functional spec position. This position determines the structural subject qualities.

In a head-final VP, all arguments are within the directionally restricted domain of the verbal head, since all positions homogenously precede the verb. The corresponding construction in German is not a quirky subject constructions but a VP with an oblique argument as the leftmost argument in the neutral word order, with a nominative argument following (if there is one). There is no structural room for a quirky subject construction in an OV language.

Third, for OV, the LCA approach inevitably and wrongly predicts the same kind of structure-triggered opacity effects as in VO. The conditions on extraction domains make a difference for pre- vs. postverbal arguments in VO. The preverbal ones are opaque for extraction in English and other VO languages (see also Haider 2010a, ch.1, observation 8). What constrains extraction out of the subject as the only preverbal argument in VO is bound to constrain extraction from any preverbal argument in OV. This empirical generalization is the robust result of more than a decade of research in conditions of extraction domains at the end of the past century. The prediction is false for OV, obviously:

(2) a. He said that [eating eels] is fun
b. * What$_i$ did he say that [eating e$_i$] is fun?
c. * What$_i$ did he say that [eating e$_i$] he dislikes?
d. What$_i$ did he say that he dislikes eating e$_i$?
f. * Whom$_i$ would [to have dinner with e$_i$] please you more?
e. Mit wem$_i$ hätte denn [e$_i$ speisen zu dürfen] dich mehr gefreut?
with whom had PRT [dine to be-allowed] you-ACC more pleased
g. Wen$_i$ hat [damit e$_i$ zu konfrontieren] keiner versucht?
whom has [it-with to confront] nobody tried
h. Was$_i$ hat sich [ihr e$_i$ zu schenken] Fritz denn vorgenommen?
what has himself [her to present] Fritz PRT decided
'What did Fritz decide to present her?'

In English, preverbal positions are opaque for extractions. In (3b,f), it is the subject position, in (3c) is a fronted object phrase. In German, none of these preverbal sites are opaque. This is a flat contradiction for the LCA account, since in this account, the preverbal positions in OV are structurally of the same kind as in VO. In the BBC model, however, the prediction is clear and supported by the facts. In OV, preverbal positions are in the directionality domain of the

verb; in VO, preverbal positions are *not* in the directionality domain. This accounts for the opacity contrast.

Opacity conditions had been the primary concern of syntactic research in the eighties and these investigations produced reliable descriptive generalizations. Today, this area has fallen victim to a kind of neglect syndrome. Neither the LCA system not the Minimalist Program has been shown to be capable of subsuming these findings in a straightforward way.

But this is not the real problem.[15] The crucial problem is that the facts are as they are, and that the LCA system predicts opacity effects where there are none. In an OV language, any preverbal argument phrase is predicted to be as opaque as any preverbal argument phrase in VO. This is an unavoidable prediction since the positions targeted by the fronted argument phrases in OV are the same kind of positions as the subject position in VO and positions preceding the subject position.

Fourth, *scrambling* is problematic for the LCA account. It is predicted for OV and for German in particular by the BBC-based account, however. And, scrambling is ruled out for Icelandic (and for VO in general) in the BBC system. Both, Icelandic and German, show word order variation for arguments. In German this is referred to as *scrambling*, the Icelandic phenomenon is referred to as *object shift*. The crucial differentiating properties are these: scrambling *permutes* arguments whereas object shift strictly *conserves* the relative order of arguments, and object shift presupposes an empty left edge of the VP (Holmberg 1999; Haider 2010a:159), but scrambling is independent of this restriction. If German was just a derivational continuation of Icelandic, the contrast seems to be accidental. In other words, the contrast could equally well be even the inverse, with Icelandic as a scrambling language, and German as an object-shifting one. From the BBC point of view, the prediction is straightforward: Scrambling is possible only in a head-final environment. It is internal merger-by-adjunction to the VP *within* its domain of identification. In a head-initial VP, adjunction to the left would leave the domain of identification and thus produce a different outcome, and adjunction to the right is banned by the BBC anyway.

Moreover, the LCA theory predicts scrambled phrases to become opaque for extraction, which is contrary to the facts. In an LCA perspective, scrambled phrases are phrases moved to preverbal spec positions. Hence, they are expected to display the behaviour of phrases in preverbal spec positions just like an English subject. In particular, these phrases are predicted to be as opaque for extraction just like preverbal subjects are. This is not the case, however. It is uncontroversial that scrambled clauses do not become opaque in German.

Fifth, *particle-verb order* (even in the case of immoveable verbs) is an excellent testing ground for the opposing theories since the LCA approach needs to invoke particle movement for deriving the preverbal particle position in OV. Normally, if the particle of a particle verb is not V-adjacent, this indicates that the particle is stranded. Languages like German and

[15] It is not the sign of a development towards a mature field of science, though. Major theoretical changes in the mature sciences conserve the findings of previous models as special cases or successfully reinterpret them. Proponents must not flatly ignore central and highly relevant data areas of the empirical range of data if they want the new model to be regarded as improvement on the predecessor models.

Dutch clearly show that functional particles (i.e. particles without truth-conditional semantic content) are inert.[16] They are only found in V-adjacent positions.

Vikner (2001:37) formulates a descriptive generalization: "*In the Germanic OV languages, particle verbs whose particles are postverbal under V2 (separate) nevertheless always have preverbal particles in non-V2 contexts, whereas in the Germanic VO-languages, particle verbs whose particles have to be stranded under V2 never have preverbal particles in non-V2 contexts."* If particles could move, that is, move to the left, they ought to occur in the midfield in OV languages or in preverbal positions in the VO languages. This is not the case, however.

Hence, if particles do not move, how do they end up preverbally? And why is there no OV language with postverbal particles? They could surface in preverbal position only if the verb could move to the right, but this is excluded, both in the LCA as well as under the BBC, since there is no head-final functional projection to accommodate the verb. So, in the LCA account a syntactic process of *particle fronting* must be invoked, or else a particle would never appear in a preverbal position (Kayne 1998:136). But, once you assume particle fronting for deriving the OV order, you allow this rule in principle also for VO. It becomes part of the grammatical tool kit and you could not discriminate VO languages. Of course there is no VO language with preverbal particles, but the LCA theory would admit it.

BBC predicts the preverbal position of particles in OV (and for word structure both in VO and OV): If the lexical head of a C-projection is head final, the dependents of this verbal head must precede; if it is head-initial, the dependents must follow. The LCA system, on the other hand, does not provide a straightforward account. It is bound to assume an ad hoc measure, namely particle movement, and this measure overgenerates.

A second aspect of particle constructions is the fact that VO languages may differ as to whether they permit stranding. Danish obligatorily strands, Swedish does not strand at all; English, Icelandic and Norwegian allow both, stranding or pied piping of the particle. However, crucially, there is no Germanic OV language that allows a particle position that is not V-adjacent, and the particles are invariably preverbal. If OV results from VO, there should exist a stranding OV language with postverbal particles. Note that this argument rests on the assumption that OV is the result of VP evacuation. But, an account in terms of *remnant VP fronting* would not work either.

Remnant VP-fronting as a vehicle of particle movement is doomed to fail too. It is a fact that remnant VP-topicalization must not pied-pipe a stranded particle. This is easy to check with a fronted remnant VP and a stranded particle in German. Take the particle verb 'einbilden' - *imagine* (lit. in-build). If the LCA account of preverbal particles in VO is based on remnant VP movement, it would incorrectly allow (3a). This structure shares the basic configuration with the alleged remnant VP fronting that is necessary for producing preverbal particles: First, the verb is fronts, stranding the particle. Then the VP with the particle is fronted.

(3) a.*[Mir auf-e_i]$_{VP}$ fiel$_i$ das nicht
(to) me up fell this not ['up fall' = come to one's mind]

[16] This is true also for ‚cluster creepers', that is, stranded particles within the verbal complex in Dutch. This is the result of V movement to the right (V-cliticization), and not the result of particle movement (see Haider 2003).

‚It did not come to my mind' (= I did not notice it)

b. [Mir aufgefallen]$_{VP}$ ist das nicht
(to) me up-fallen] is this not

As (3a) shows, particle stranding in a fronted VP is not allowed. This is the consequence of a crossing violation. A phrase that contains a trace is fronted across the antecedent of the trace. Thus, the trace of the verb in the fronted remnant VP is not properly bound. This is true for (3a) and analogously for (4), that is, for the hypothetical derivational vehicle that is to produce particle fronting.

(4) dass mir etwas [$_{VP}$ auf-e_i ……]$_j$ fiel$_i$ e_j [beware: this structure is hypothetical only]

The ***sixth*** region of contrasting data is *verb clustering* in OV (see Haider 2010a, ch.7). In languages with head-*initial* VPs, verbs that select a VP are projected/merged in stacked VPs (5a). In languages with head-*final* VPs, they are merged in clusters (5b). Stacked head-*initial* VPs are right-branching structures. Stacked head-*final* VPs would be left-branching (5c). This violates the BBC. OV languages stack VPs, VO language cluster the verbs. Why should this be so?

The bracketing in (5c) already reveals the crucial factor. For (5c), the parser would have to guess how many brackets to open; in (5a), the parser encounters the head of the topmost VP first, with the maximal projection of the head as the immediately dominating node. Clustering (5b) reduces the indeterminacy for the parser to the *local* environment of a single V-projection. Clustering is UG's contribution to parser friendliness of complex head final structures.[17]

(5) a. [$_{VP}$ V_1 [$_{VP}$ V_2 [$_{VP}$ V_3 …..]]] head-initial
b. [$_{VP}$ ..... [$_{V°}$ [$_{V°}$ V_3 V_2] V_1]] head-final
c.* [$_{VP}$ [$_{VP}$ [$_{VP}$ ….. V_3] V_2] V_1] head-final

Icelandic verbs, like in English and the other Scandinavian languages do not cluster, neither with auxiliaries nor with other verbs. Why do German and Dutch verbs cluster? The answer has to be sought in the headedness difference again: The constraint against left-branching (extended) projections does not only rule out structures with head-*final functional* heads (see appendix 1 in ch. 7) but also *lexically* extended[18] head-final V-projections, that is, V° selecting VPs as its *left-hand* complement.

For the LCA, the mirror image order of auxiliary sequences in German and the clustering property in general are unexpected. Moreover, the core problem is the *compactness* of the cluster: The verbs in the cluster do not tolerate non-verbal, intervening material (except particles; these I consider as part of a complex verbal lexical entry). This property is not captured by movement analyses (e.g. remnant VP fronting, as suggested by Koopman & Szabolcsi

[17] It is presupposed here that the structures provided by the core grammar (as a function of UG) are parser friendly, given that UG and parsing developed in a process of cognitive co-evolution (see Haider 2003). 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 continuous data processing (bottom up) plus grammar guidance (top down information on possible structures). This is implemented best with right-branching projection structures.

[18] 'Lexical extension of a VP' refers to verbal elements that contribute to the aspectual, modal, epistemic or other semantic modifications of a VP by selecting the respective VP as complement. In German and Dutch, these verbs are obligatorily clustering. In addition, a large sub-class of control verb is clustering optionally.

(2000), except for a head-to-head adjunction analysis (Zwart 1993). He primarily analyzed Dutch and advocates a V-raising approach. The auxiliary order is derived by iterative head-to-head adjunction (cf. Zwart 1993 sect. 2.4). Even if it is possible to derive the order by means of the devices suggested by Zwart, the theoretical and empirical soundness remains to be demonstrated.

First, it is unclear why clustering is found only in OV, but not in VO languages. Zwarts's (1993) movement account would work in VO as wells as in OV. The BBC provides a principled answer, the LCA apparently not. Second, the BBC provides an answer to the specific variations within the cluster, and in particular the contrast between German and Dutch in this respect. In the BBC framework, which starts from a VO organization of auxiliary verb sequences, the inversion of this order is a disturbing factor, and the variation in the inverted sequences is even more disturbing. In the absence of a reasonable trigger theory, the head-to-head adjunction proposal (actually a revival of Arnold Evers' original verb raising idea) remains an ad hoc feature of the model.

Second, *clause union phenomena* provide robust evidence that the crucial aspect of clustering is not so much the formation of a verb cluster by means of head-to-head adjunction, but, as already acknowledged in Evers (1975), the crucial aspect is the lack of the VP- or functional complement structure, which provoked Evers to resort to a pruning operation. Head movement would not destroy the stacked VP or IP structure. Nevertheless, the available evidence points to the absence of these remnants of an infinitival clause structure (see the list of 16 properties in Haider 2010a: 311-313). Here, the LCA model is at a loss, on principled grounds, because it predicts the conservation of clause internal structures on the one hand and has no answer at hand for the question as to why the this structure is absent on the other hand, giving rise to numerous clause union phenomena. Movement would not destroy these structures, but the diverse evidence is incompatible with them.

The clustering analysis is extensively explicated in Haider (2010a, ch.7). Here, a single[19] piece of evidence may suffice as representative, namely *rigorous compactness* as a sign of clustering. Another piece of evidence has been discussed in chapter 4, namely cluster nominalization.

(6) a. dass er was zu sehen war [gesehen *wird* haben]$_{cluster}$
that he what to seen was seen shall have
'that he shall have seen what was to be seen'
b. [Gesehen, was zu sehen war]$_{VP}$ *wird* er haben
c.* dass er [*gesehen* was zu sehen war] *wird haben*
that he [seen what to see was] shall have
d. dass er gesehen wird haben [was zu sehen war]

The specific cluster variant (6a) has the finite verb in between the non-finite verbs. In a derivational analysis, the participle in (6a) would be the right edge of the deepest VP. This VP is a possible target of extraposition, as illustrated by the topicalized VP with extraposition (6b). Hence, (6c) should be a well-formed order if the participle is projecting a VP, but it is not. In

[19] In chapter 1 and 8, the availability of cluster nominalization has already been mentioned as a specific piece of evidence for the V°-category of the cluster, consisting of head-to-head adjoined verbs.

the clustering analysis, the three verbs in (6a) are elements of a verbal cluster. This immediately rules out a word order as in (6c). In a derivational analysis, a clause with three verbs would start with three VPs and these VPs could not be done away with. Koopman & Szabolcsi (2000) are fully aware of the problem of over-generating. The set of filters they propose are a last resort measure for taming the derivational machinery (see also the discussion of verb order variation in the cluster below, under 'tenth').

Seventh, predicative adjectives agree in VO, but not in OV. The OV languages German, Frisian, Dutch, and Yiddish have inflected *attributive* adjectives, but *un*inflected *predicative* adjectives, whereas those VO languages that have inflected attributive adjectives (that is, all the Scandinavian languages and all the Romance ones) *also* have inflected predicative adjectives.

The fact that none of the present day Germanic OV languages has adjective agreement for predicative adjectives is surprising in the LCA model, but not in the BBC system. Only if OV was the derivational continuation of VO, this would entail the agreement relations like in VO. Neither A-movement (e.g. NP-movement in passive) nor A'-movement (e.g. topicalization or wh-movement) destroys previously established agreement relations.

The crucial difference that is responsible for the contrast seems to be the following: In VO, the subject of the predicative adjective needs an external licenser since its pre-adjectival, AP-internal subject position is not in the canonical directionality domain of the head. In OV, however, the subject argument in the predicative AP is in the canonical domain of the head. In VO, licensing is implemented in terms of a functional head and this head is lexicalized with the copula. The copula enters an agreement relation with the subject, and selects the AP as complement.

(7) a. VO: [DP_i [Copula [$_{AP}$ e_i A°]]] (as in: *It$_i$ is* [*e$_i$ green*] vs. in: *paint* [*it green*])
b. OV: [[$_{AP}$ DP [A° Copula]]]

In OV, the copula and the adjectival head (optionally) form a head-head-cluster. This is predicted by the BBC. The copula should not select the AP just like an auxiliary cannot select a VP in OV. It is functional selection, and the phrase would be a functional extension of the AP. Hence the AP would end up as a left sister in an extended projection. This is avoided.

In VO, however, the AP is a full-fledged complement. It is this difference in structure that prevents agreement in the OV type. In this type, the adjective is part of the cluster, and the cluster is the complex head of the VP. So, there is only a single candidate for agreement, and this is the copula. In VO, the subject agrees with the copula, and moreover, it is in a relation to the AP-subject position. Therefore, if the adjective is able to agree, it may agree as the head of a separate projection.

The situation for participle agreement in Romance languages like French, Italian, or Old Spanish is instructive in this respect. In active sentences, the perfect participle is in an agreement relation with the object if the object is fronted (by wh-movement or by cliticization). The resulting configuration is analogous to (4a): A DP is fronted to a functional position and thereby leaves the phrase it was born in. Consequently, in a derivational analysis of OV, predicative adjectives should agree in OV languages just like in VO, given the massive move-

ment of phrases to the left. The antecedent of the predicative adjective would be in a higher spec-position.

What these contrasts tell us is that the LCA owes us an explanation, whereas the BBC model is compatible with this set of differences. In fact, they support the principled difference between an OV and a VO set up of clause structure.

Eighth, *word structure* contradicts the LCA and confirms the BBC. The latter predicts merger to the left (see chapter 8). The former predicts a left-headed structure. It is an undeniable fact that word structure in the vast majority of languages is right-headed, both morphologically as well as semantically. Suffixes determine the category, prefixes don't, as in the German examples (8a,b):

(8) a. $\text{Fett}_{N^\circ} \rightarrow [\text{fett-}\textit{en}_{V^\circ}]_{V^\circ} \rightarrow [\textit{ent}\text{-fetten}]_{V^\circ} \rightarrow [\text{Entfett-ung}_{N^\circ}]_{N^\circ} \rightarrow [\text{Voll } [\text{Entfett-ung}_{N^\circ}]_{N^\circ}]_{N^\circ}$
fat fatt*en* *de*-grease de-greas*ing* full de-greasing
b. $\text{white}_{A^\circ} \rightarrow \text{whit-}\textit{en}_{V^\circ} \rightarrow \text{re-whiten}_{V^\circ} \rightarrow \text{re-whiten-}\textit{ing}_{N^\circ}$

The BBC is a UG property and constrains the make-up of recursively built structures. Hence, it will apply both to phrase structure as well as to word structure. It requires merger to the left. Merger to the left produces head-final structures. This is what we see in word structure (see chapter 8). Word structure is the pure reflex of the BBC since in English or German word structuring there is no directionality parameter involved.

The LCA account predicts a universal spec-head-complement structure.[20] This covers all kinds of structures, and presumably it is to cover word structure, too. Hence (9a) should have the order (9b), since *radar* is clearly the object of *detecting,* nominalized in the form *detector.*

(9) a. [radar-[detector]]
b. detector-radar (with 'detector' as the semantic base)

Roeper & Snyder (2005) accept the implications of the LCA account and consequently postulate movement operations as in (10a,b), without independent evidence, though. What they seem to overlook, for instance, is the fact that in word formation, children do not produce the kind of variation they produce in phrase structuring, namely a variation between OV and VO patterns. This apparent variation is in fact an independent, third-type option, resulting from an unfixed directionality value (see also the ninth property, below).[21] From a BBC point of view this is predicted: when there is no movement, there is no variation. Word structure is not a domain of movement. Roeper & Snyder (2005), however, accept the LCA dogma that the head-final organization of word structure is a derivation result of a head-initial base structure.

(10) a. hold pen $\rightarrow$ $[\text{pen}_i\ [\text{hold } e_i]$
b. [er $[\text{pen}_i\ [\text{hold } e_i]]$ $\rightarrow$ $[\ [\text{pen}_i\ [\text{hold } e_i]_j$ er $[\ e_j]]$

[20] Recall that Kayne (1994) and Chomsky (1995) have advanced the following claims:
a. Any movement rule is universally left-bound.
b. Complex structures are represented in the form of Specifier-Head-Complement.

[21] Karl (2010) studied the acquisition of nominal compounds $[N_1\ N_2]$, with N_2 being the base, as deverbal noun, and an N1 in an object relation (e.g. *nut cracker*). It turned out that children at the age of 3-4 already use the correct order. The percentage of mistakes with the inverted order was very small and correlated significantly with X-V-O order in the elicitation sentence.

Although it is not immediately evident that a simple 'penholder' has as rich a structure and derivational history like the one suggested in (10b), Roeper & Snyder (2005) do not tell what kind of evidence might have convinced them that (10b) is really more than a forced ad hoc analysis for those who want to maintain the LCA against the evidence of word structure. For a BBC account of the word structure, see the discussion of compounding in chapter 4.

Ninth, the LCA and the BBC model differ in their prediction of *diachronic changes* within the Germanic (and also in the Romance) language family in particular. This is the topic of chapter 5.

The present discussions of OV and VO tend to downplay the existence of languages that exhibit both VO and OV surface orders simultaneously, plus orders that do not occur in VO or in OV (see chapter 3, section 4 and chapter 5). Examples of this type of languages are historic variants of Germanic languages, like Old Icelandic (e.g. Rögnvaldsson 1996, Hróarsdóttir 2000) and Old English (e.g. Pintzuk 1991), as well as a present day Germanic specimen, namely Yiddish.[22]

(11) a.	[XP [YP [ZP V°]]]	OV
b.	[XP [V_i° [YP [V_i° ZP]]]]	VO
c.	{ [XP [YP [ZP V°]]], [XP [YP [V° ZP]], [XP [V_i° [YP [V_i° ZP]]]] }	type III

The OV and the VO option are the result of merger with either the value *left* or *right*, just like in (11a) and (11b). The third possibility is the result of *adjustable* (or underspecified) directionality, that is, the possibility of switching the directionality in the course of merging. The second variant in (11c) starts out with *directionality value* = *right*,[23] licensing a complement to the right, and then the value switches to *left*, proceeding in an OV manner.

The implication for diachronic syntax is this: If the historic variants of Germanic languages are viewed from this perspective, the development of the Germanic languages is much easier to understand. The basic change has been one from an *adjustable* (i.e. underspecified) headedness value to a *rigid* one. The rigid directionality allows either of two values. The choice of the value is in principle free. One branch of Germanic languages fixed it in the OV way (West Germanic), the other in the VO way (North Germanic). So, two very similar languages in terms of their morpho-syntactic make-up, namely Icelandic and German, ended up in different systems by accident. The accident is the choice of the directionality value when giving up the adjustability option.

The BBC model provides the right kind of tool for capturing these developments, namely the directionality parameter of licensing by a lexical head. In the LCA model, the West-Germanic OV languages would have to be filed as continuations of a VO system that is still conserved in the North-Germanic VO languages.

This is not what diachronic studies tell us. There is no indication that in the attested history of Germanic languages the predecessors of OV languages had been VO before. The known facts

[22] Slavic languages should be listed in this group as well and Hungarian also. Since the descriptive tools of typologists are not accurate enough for capturing the differences, the list of type III languages is likely to be approximately as long as the list of OV- or VO-languages.

[23] Note that this is the only available possibility. Obviously, structure building could not start in the OV-style and switch to VO.

support the BBC based account: The modern Germanic languages with their OV-VO split are continuations of a predecessor language. This language resembled the only surviving Germanic 'type three' language, namely Yiddish, in a crucial respect, that is, the apparently variable pre- and postverbal order of nominal arguments. In the BBC model, this is the option you get if the directionality parameter is underspecified. Fixing the parameter value yields two possible outcomes, namely strict head final or strict head initial projections. And, moreover, the direction of the attested drift towards OV and VO is in accordance with the degrees of freedom the model offers for diachronic change.

Tenth, and finally, OV and VO robustly differ with respect to V-V(P) dependencies. In VO languages, the relative order of auxiliaries, quasi-auxiliaries and verbs is *invariant*. In VO languages with V-movement, verb order variation is common and typical. In all Germanic OV languages there is verb order variation, and this variation does not contribute anything to semantics or pragmatics. It is entirely optional:

(12) a. dass er es *lesen müssen* $wird_{fin}$ */lesen* $wird_{fin}$ *müssen / * $wird_{fin}$ *lesen müssen*
that he it read must shall /read shall must /shall read must
b. dat hij niets *gezien* kan_{fin} *hebben / * kan_{fin} *gezien hebben / * kan_{fin} *hebben gezien* Dutch
that he nothing seen can have / can seen have / can have seen (*ANS* 1984: 1069)

The first variant in (12a), namely ‚*lesen müssen wird*' has a fully inverted counterpart in Dutch, namely ‚*zal moeten lezen*' whose order is congruent with English ‚*shall have-to read*'. Koopman and Szabolcsi (2000) suggested a simple account: VO as in English is the base configuration (13a). Then the object is fronted (13b). The variant (13c) is the result of VP-fronting (13c) to a spec position in between the two auxiliary positions or in front of the two auxiliaries (13d).

(13) a. dat hij [kan_{fin} [*hebben* [*gezien* niets]]] LCA base order
b. dat hij niets$_i$ kan_{fin} *hebben gezien* e_i object shift
c. dat hij niets kann [$_{VP}$ gezien e_i]$_j$ hebben e_j VP shift
d. dat hij niets [$_{VP}$ gezien e_i]$_j$ kann hebben e_j VP shift
e.*dat hij niets [[$_{VP}$ gezien e_i]$_j$ hebben e_j]$_k$ kann e_k VP shift

The *first* remarkable aspect of this analysis is that it massively over-generates. The authors admit this, but they are content with proposing a set of filters for eliminating the overgenerated variants. The reason for massive over-generating is this: First, the verbal cluster in Dutch is compact. No nonverbal material must intervene between the verbs in the cluster, except a verbal particle. The VP however may pied-pipe (extraposed) material. This would produce interveners. Second, the very VP fronting option needed for (13d) produces the order (13e), which is ungrammatical in Dutch. It is a possible German order, however, and therefore it could not be excluded on principled grounds.

The *second* remarkable aspect is the complete absence of independent evidence. There is not the slightest evidence that a VP has moved in these constructions, and there is no reason why it should have moved. The analysis is ad hoc and it makes the order variation properties appear ad hoc too. It is motivated by theory-internal considerations only, wilfully elevating the VO order property to the level of an unquestioned but unfounded axiom.

In the BBC-based approach, clustering in OV and its absence in VO is predictable: Head-*final* complementation produces centre-embedded V-projections (14a). A grammar with clustering in OV is a parser-friendly grammar. Clustering reduces centre-embedding and confines the domain of left-branching to the local domain of the verbal cluster (14b). The variations of word order in the cluster are partial (or full) 'repairs' of the left-branching cluster structure (see Haider 2010a, ch. 7.5):

(14) a. ….. $[_{VP}\,[_{VP}\,[_{VP}$ ….. $V_1]\ V_2]\ V_3]$
b. ….. $[_{VP}$ ….. $[[[V_1]V_2]\ V_3]$ German cluster variant
c. ….. $[_{VP}$ ….. $[V_3\,[V_2\,[V_1]]]$ Dutch cluster variant

In the BBC-based approach, clustering is not a bizarre property but the immediate outcome of the head-final option for V-projections. Clustering is not a Germanic peculiarity but holds for OV languages in general. In languages without V-movement, as Japanese or Korean, it has no linearization effect and hence it tends to be overlooked (see Haider 2010a: 324 and the literature cited there).

4. Ockham's Razor,[24] epicycles and (im)perfect systems

Is the LCA- or the BBC-based system closer to the ideal of a minimal, necessary and sufficient set of conditions for an at least descriptively adequate account of the relation between head-initial and head-final structures?

Empirical success or failure is only one side of the coin. As every scientist implicitly knows, and as Duhem and Quine[25] established on philosophical grounds, you always can mend, bend and expand your system in order to cover or block off problematic evidence. So, there is no final and definite way of falsifying a theory. All one can do is pointing out what kind of data are problematic for a given version. The more diverse and widespread the varieties of counter-evidence are or become, the greater is the likelihood that the theoretical premises a system is based on are inappropriate. Having discussed empirical issues, let us turn now briefly to theoretical considerations.

A distinctive area of theory comparison is a comparison in terms of the overall system properties. Is the system *elegant*, *economic*, and *heuristically* productive? Here, Ockham's razor as a principle of theoretical economy has its place. It applies to several aspects:

- economy of the inventory
- economy of the combinatorics (i.e. structuring and derivational processes)
- relation to the null hypothesis

The merits of the LCA and the BBC model clearly differ with respect to the various parameters of evaluation. Let me start with the inventory. The LCA system claims to work without a *di-*

[24] William of Ockham (1285-1347): „Non sunt multiplicanda entia praeter necessitatem" [You must not introduce a (theoretical) entity without (having shown its) necessity]. Or, in the words of Georg Christoph Lịchtenberg (1742-1799), who inverted Hamlet's sigh ("There are more things in heaven and earth, Horatio, than are dreamt of in your philosophy"): "There are even more things in our philosophy that do not occur anywhere between heaven and earth."

[25] According to the "Duhem-Quine thesis", the holistic nature of scientific theories makes the rejection of a *particular* hypothesis as a consequence of counterevidence virtually inconclusive: One can always retain the hypothesis at hand by making *appropriate adjustments* elsewhere in the system. Theories tend to be rescued by modifications (of their *periphery*) in order to protect the *core* against the very counterevidence (Lakatos 1977).

rectionality parameter (Kayne 1994:132). This is true in a literal sense, since directionality is not used as a primitive feature of the system. It seems, however, that this is merely a terminological difference. In fact, *directionality* is a primitive element of the LCA system, too. It is necessary for symmetry breaking: the LCA (Kayne 1994:5-6) merely requires that the grammar provide a means of mapping asymmetric c-command relations into linearization relations. This mapping can be grammatically implemented in various ways, however. The structures in (15) are obviously different structures, but they share the same asymmetric c-command relations and therefore they are each mapped on the same linearization relation: the item that asymmetrically c-commands is linearized in a relation R with respect to the item that it c-commands. R could be 'precede', as Kayne postulates (15), or 'follows'. So, a decision on directionality is at stake on the *linearization* side, and on the *structure* side, too. The LCA system is *inherently* parametric. The *hidden* parameter is the linearization relation (*precede* or *follow*).

(15)	a. [A [B [C [D]]]]	→	A-B-C-D	(or: D-C-B-A)
	b. [[[[D [C [B [A]	→	A-B-C-D	(or: D-C-B-A)
	c. [A [B [[D] C]]]	→	A-B-C-D	(or: D-C-B-A)
	d. [A [B [[D] C]]]	→	A-B-C-D	(or: D-C-B-A)
	e. etc.			

Directionality covers the essential degree of freedom for the PF representations: PF representations are one-dimensional (i.e. with the single ordering relation: precede vs. follow). This is a necessary consequence of the coding of complex linguistic representations as *sequences*. Since complex expressions cannot be produced instantly, they must be sequenced on the time axis, either as a stream of sounds or a stream of gestures (in sign languages).

The degree of freedom for the single dimension of a string is *before* and *after*. Algorithmic parameterization, on the other hand, would become intractable if it starts from the PF side. But it must start from the PF side when it unfolds the syntactic structure of a string. So, there is no continuous relation between the organization of the string in terms of *before* and *after* and the algorithmic feature that is responsible for deriving a head final organization from a head initial one by means of derivational processes.

Therefore, the question "Should the theory contain a directionality feature as a primitive" should be answered with *yes*, if one is pleased with the idea that the way the grammars manage linguistic data structures is in a continuous relation with the properties of the data structures at the interfaces. In other words, if one thinks that our brain employs effective and efficient processing algorithms for transforming objects with one specific property (namely strings with precede/follow, as their inherent order relation) into objects with another specific property (namely phrases as hierarchically organized box-in-box structures), one should not be surprised to see that a core property of one set of objects (namely precede/follow) maps directly on a core property of the other set (directional identification in a structure). Here comes the essential difference:

If, on the other hand, the mapping of structures onto PF is a *stepwise derivational process* (= algebraic model), it would not matter at all whether a structure has a linear order or not. In the latter case, each one of the different bracketings in (15a-d) is an instantiation of one and the

same structure, with its notational variants. If, however, the mapping is a *direct pattern matching* process (= geometric model), the linearization of the structure matters. Only (15a) can be mapped directly on 'A<B<C<D'.

The BBC model is a *'geometric'* one, the LCA model is an *'algebraic'* one. This is a neuro- and psycho-linguistically testable difference, and it is a relevant aspect for the comparative evaluation: The BBC model embodies a *direct* matching between structure and linearization; the LCA model defines a global constraint for grammars: A grammar is licit only if it 'somehow' manages to guarantee that phrase structure is anti-symmetric, in order to get it uniquely mapped on linearizations. The direct approach by the BBC, on the other hand, is evidently the stronger theory as it provides fewer degrees of freedom for the bridging relations between syntactic structure and linearization.

Let us now turn to the algorithmic quality of the theory in the characterization of OV vs. VO. The LCA model is a system that characterizes the OV properties as the result of a derivationally *more complex* grammar. For the BBC system, on the other hand, OV structures are *less complex* than VO structures (because the latter, but not the former, need shells plus a functional projection for the subject). Given the massive derivational efforts needed for turning VO into OV,[26] Occam's razor simply says: Give me good reasons for the extra derivational apparatus and show me that it really is at work. The extra machinery is necessary if you cannot plausibly do without it, and it is at work if you can show us several immediate and independent effects of this apparatus in the data properties.

A system that achieves the same result (namely providing insight into the principal relations that differentiate OV and VO, and characterizes them as alternative options within the same system potential) without an excessive and hitherto largely unmotivated (see: lack of a trigger theory) derivational machinery is the BBC model.

What is a reasonable *null hypothesis* with respect to OV versus VO, you should ask yourself. 'OV' and 'VO' refer to different phrasal organizations, but the difference is not haphazard. 'OV' and 'VO' are dependent variables and the dependency rests on factors that correspondingly partition the system space provided by UG for the phrasal architecture.

At least two possibilities have to be acknowledged and checked: The dependency can be an *opposition*, with OV and VO as alternative outcomes, or, and this is the more restricted hypothesis, the dependency is a case of an *inclusion* relation, with one state as the more specific variant of the other. The LCA proposes the specific inclusion relation, the BBC holds an opposition relation. In the light of the arguments developed in this volume, the inclusion hypothesis appears to be insufficiently grounded and unlikely to be appropriate. The more adequate model is one that characterizes OV and VO as alternative (parametric) options within the *same* system space of UG.

[26] Heavy syntactic traffic to the left is not limited to the derivation of an OV structure. Even for VO grammars, Kayne (1998) employs Neg phrase preposing (p.135); Neg° raising (p.152), raising of *not* (p. 153), VP-preposing (p.135), particle preposing (p.136, 143), predicate raising (p.138), *only*-phrase preposing (p.146), raising of *only* (p.150), direct object preposing (p.162).

Let me finally briefly highlight a well-known problem of dealing with *economy and perfection* on the level of theory design.[27] It is the desire of most theorists to reduce the observed complexities to complexities in the output of an elegant and parsimonious system (of *'virtual necessity*'). Perfection, however, does not guarantee explanatory success. The history of science offers numerous examples of elegant and complex systems whose only defect was that they turned out to be empirically wrong. Unfortunately, nature had not bothered instantiating them.

A case in point readily comes to mind from the history of astronomy. It has been possible to maintain a simple, easy to understand, wrong idea for quite some time (approximately from the time of Apollonius of Perga at the end of the 3rd century BC until Kepler's days). The leading idea, which has turned out to be empirically wrong after all, was the idea that celestial bodies move on geometrically minimal and perfect trajectories, namely circles. This restricts the degree of freedom for movement to a single parameter. It is the strongest constraint on possible movements.

Only 'planets' (Greek for 'rovers') appeared to be exceptional, but their exceptionality was explained away as a *complexity effect*: Like other celestial bodies, planets were thought to move on circles, but unlike other celestial bodies, the centre of the respective circle moves on a circle (the so-called *deferent*), too (whose centre may again move on a circle, and so on). In other words, this system was a system with a highly restricted theory of movement (movement on circles only), and with a simple derivational component for integrating the actual movement of planets (namely *epicycles*, that is, circles on circles on circles …). How can we assure ourselves that the massive movements in the LCA-based accounts are not epicycles? The answer is obvious. Provide positive and independent empirical results for the numerous implications, like Copernicus, Kepler and others did in Astronomy.[28]

Without sufficient independent evidence for the derivational complexity invoked by the LCA system for OV, the need for massive derivational re-arrangements on the way to OV raises the suspicion that this is just the epicycle part of an allegedly minimalist & elegant grammar theory. Its design may be elegant, minimal, and sophisticated, but the way it models the facts may nevertheless be inadequate, simply because the linguistic products of a domain-specific mental capacity do not fit into this very model.

With the idea of perfection, linguists in the vicinity of the Minimalist Program have revived a scholastic kind of reasoning that had been given up with the idealist concepts in the 18th century. It is the *argument from perfection*: a property is postulated and declared as existing simply because its absence would make the system less perfect than one would like to have it. This is an argument known from mediaeval debates on metaphysical matters but its value in empirical domains is nil. Inferences from an alleged quality of *'perfection'* (Roberts 2000) are

[27] "Universal grammar is perfectly designed, that is, it contains nothing more than what follows from our best guesses regarding conceptual, biological, physical necessity." (Boeckx 2006:4). Pinker & Jackendoff (2005: 227) justly emphasize: "The overall claim that language is 'perfect' or 'optimal' is a personal view of how language ought to be characterized rather than an empirical discovery about the way language is."

[28] The social biotope of Generative Grammar apparently instigates ambitions for becoming a linguistic Copernicus if not Newton, more than the desire to come close to a respected linguistic Regiomontanus or Brahe. The average syntactician's desire for innovating or revolutionizing a present variant of a theory usually prevails over the apparently less impressive attempt of carefully assessing a generalization over a relevant data basis.

simply irrelevant in empirical domains. *Perfection is a property of a platonic object like a calculus, but not of grammar*[29] *as a mental capacity embodied in a neuro-biologically evolved organ.* Even if a grammar might resemble a calculus in important aspects, it is obvious that it is an indirect product of *biological* and *cognitive* evolution. Evolution is tinkering and by no means perfect. "*In reality, evolution is a blind process, with absolutely no guarantee of perfection*" (Kinsella & Marcus 2009:187). What is true for biological evolution is equally true for cognitive evolution. It is extremely unlikely that a grammar could be a perfect system on the one hand and on the other hand a product of an organ, namely the human brain, whose capacities and operative qualities are the result of biological and cognitive evolution (see Haider (2011), for examples of grammatical inconsistencies).

A grammar may be a highly adaptive system for the purposes it is employed, but this is a property of selection (on the biological, neuro-cognitive and cognitive level), and not of thriving for perfection. The argument from perfection uses 'perfection' as a *teleological* quality: In order to be/become *perfect* a grammar must have the properties $P_{1\text{-}n}$, hence UG must guarantee the properties $P_{1\text{-}n.}$ This metaphysical conception of grammar theory I do not want to adopt. It is clearly idealist in all but name; it is the contemporary version of linguistic idealism of the 18th century. Science is orthogonal to idealism, and I prefer linguistics to develop into a science.

Appendices

A*ppendix* 1: On hyper-derivational syntax and the scientific conduct of grammar theory

The Minimalist Program and the LCA program are strongly (if not hyper)-derivational. Their major syntactic tool is massive overt as well as covert 'movement',[30] plus the devices for taming these hyper-active 'animals'. The evidence, however, comes only from properties of the *output* of the assumed mental counterpart of the machinery and in particular from introspective judgements on, and eclectic informant consent to, these expressions. The aim of these theories apparently is not so much a model of a domain-specific capacity of human brains but rather a demonstration of the algorithmic properties of a system that produces these properties. The relation between the properties of the algorithm and the properties of the processing brain is one of *weak* generative *equivalence* between the derivational apparatus on the one hand and whatever mental system there is that is operative in the human brain on the other hand.

It cannot be more than *weakly* equivalent because these approaches are linguistic ones, that is, they investigate the properties of the *output* but not of the system itself that is embodied in a

[29] Those who claim that a grammar is the result of a perfect UG must firmly close their eyes & ears when being confronted with English as a specimen of an evidently imperfect instantiation: no passive of intransitive verbs, although all Germanic languages have it; movement of finite auxiliaries, but not of finite main verbs (in V1 and V2 patterns), which lead to a patch-up construction ('do-support'), and no other (Germanic) language enjoys this exception; no morpho-syntactic marker for infinitive and as a consequence an exceptional class of verbs that cannot be used in infinitival constructions (i.e. modals). These are but a few examples for properties that are much closer to tinkering than to perfection (see Haider 2010a: 9).

[30] 'Overt' vs. 'covert' movement is a weird and wonderful concept. It suggests a kind of eyewitness-like direct perception ('overt') contrasting with a highly indirect perception. 'Covert movement' means that phrases are not where they appear to be, and you cannot *perceive* where they have gone to. It is only the *smile* of a ph(r)ase that you see (cf. Alice's Cheshire cat), but 'in reality' the ph(r)ase has disappeared and gone away. Entertaining, but difficult to prove conclusively as an empirically testable reality.

human brain as a cognitive and neuro-cognitive entity. This objection has already been raised by W.V.O.Quine (1960, 1970) in the early days of the Generative program (see also Stich 1979, Fraser 2000). Chomsky's (1968) answer to this challenge was suggestive but not entirely appropriate: you can find out many things about the sun without having visited the star physically. But, this is clearly not the essential point. A scientist who wants to test a hypothesis on what is going on in the interior of the sun is of course unable to run an experiment on or inside this star. But the analogy with solar research is inappropriate. The crucial difference is this: the scientist has a model at his disposal whose *components* have been tested and experimentally assessed under immediate control in the lab. Linguistics does neither have an animal model nor a reliable background theory from the (cognitive) neuro-sciences.[31]

Theoreticians of these two related camps (MP, LCA) do not investigate (or tell how to investigate) whether there is any empirical evidence for the psycholinguistic or neuro-cognitive reality of some (implications) of the general postulates (e.g. for the cascades of processes for overt or covert structure-on-structure mapping operations). It is speculative, in the best sense. It is hypothetical. But this is not enough for the modelling of an empirical object. It is not enough to model the output somehow, but forget about the fact that the capacities you study are put to use under real time conditions. They have a history in cognitive evolution. They are embodied in a brain with a biological evolution that vastly outdates the time span since humans exist (appr. $6x10^6$ years have passed since the split into the *homo* and the *pan* branch). How can you be sure that the human mind uses a specific *derivational* machinery, as long as you have not shown that there is some piece of conclusive experimental evidence? The probability that your initial speculative bold guess on such a complex matter is right is indistinguishably near or equal to zero and not near or equal to 1, of course.

Until the nineties, grammar theory had to treat the language faculty as a black box that could not be accessed directly. But, since then, experimental psycho- and neuro-linguistics have built viable access paths for observing brain activities. It is not completely inaccessible anymore. Here is a conclusion from Haider (2009:96-97):

> "The self-understanding of linguists as members of the scientific community of the cognitive sciences sharply contrasts with the established working traditions. If linguistics is to be grounded as a branch of science, grammar theory will have to give up its 'splendid isolation' with respect to the experimental camps in psycho- and neuro-linguistics and in psychology. The theoreticians must acknowledge that the practice that proved successful in the pioneering phase of the past decades, namely introspection and eclectic feed-back from informants, has reached its limits.
>
> If they do not actively seek the cooperation with experimental research, it is at least their responsibility to clearly point out what are empirically testable implications of their model. It is not enough to present novel and bold hypotheses and embed them into a set of assumptions that is complex enough to make an experimenting linguist immediately resign."

Note, importantly, that the preceding paragraphs are not meant as a *revival* of the '*psychological reality*' debate of the seventies (as in Devitt 2006, countered by Slezak 2009). It is just a

[31] We do not have the lightest idea of what kind of 'sofware' is runing on the brain's 'wetware' and what kind is excluded.

plea for improving the quality of data assessment (beyond introspection and eclectic informant consent) with means provided by today's psycholinguistic methods of experimenting.

It is beyond reasonable doubt that the old fashioned reliance on introspection and informant inquiries produces unreliable results (see Haider (2011) and (2009); note in particular the discussion of the exemplary case of the incoherent judgements for Dutch wh-in-situ on p.79).[32] The problem is simple: 'crystal clear', run of the mill data are handled by every theory variant somehow. If you want to discriminate among the competing variants, you have to resort to more *peripheral* issues for which the variants diverge. But more often than not, the judgements become peripheral, too. Data assessment is indispensable, like in any empirical discipline. What is crucially missing in linguistics is a mutually agreed and obeyed standard of data evaluation, as for instance in psychology. Data assessment by self-observation (introspection) and eclectic informant consent would be far away from meeting these standards in the neighbouring disciplines in cognitive science. Psychology has made its way since Wundt's (1888, 1901) initiatives, but linguistics still goes with Delbrück's time (1901) in its predilection for mere introspection as the primary source of data.[33]

But even if we disregard the problems on the data side, the *theory testing* side is underdeveloped as well. The LCA model (and the Minimalist Program) is a highly flexible tool,[34] and in fact it invites overly flexible, and highly complex, implementations of syntactic analyses. Let me demonstrate this with an example: Assume, counterfactually, just for the sake of demonstration, that English is basically OV, and that this would be so because languages are universally OV.[35] With this premise, English would leap into its derivational maze with an initial structure like Japanese (1a).
Next, LCA enters the grammatical saloon and insists on asymmetric c-command for all items that need to be serialized. So, (1a) must undergo some derivational procedures that guarantee this outcome, namely movement of at least one of the two sister items in (1a). One of the two sister constituents has to move out, either the verb, or the object, or maybe both.

(1)	a. $[O\ V°]_{VP}$	starting point
	b. $[V°_i\ [O\ e_i]_{FP}$	the verb moves to a head[36] position, *or*
	b. $[O_i\ [e_i\ V°]]_{FP}$	the object moves to spec, *and*
	d. $[[e_i\ V°]\ [[O_i\ e_j]]]_{FP}$	the emptied VP is fronted to a spec-position ('roll up')

Having exemplified the ingredients for a simple case, we move on to a more entertaining case, viz. a verb with two objects:

[32] Incoherent set of judgements of twenty-two native Dutch syntacticans on wh-in-situ data from Dutch.
[33] 1901was the year of the very first debate on psycholinguistics and grammar theory: Wundt (1901:17-18) had argued that linguistics would greatly benefit from psychological insights, but Delbrück (1901) refused this offer: It would be of no help for a linguist who is struggling to reconstruct grammar change if he had to consult Herbart's associative psychology or Wundt's volitionistic theory for explanatory assistance.
[34] You may, if you wish, derive *any order* of elements, even if you strictly obey the rules of the game: Move everything to the left, and make sure that your landing position is always in a spec-head-complement configuration. For the rest you are in principle free.
[35] Here is a plausible-sounding reason for this hypothetical axiom: They are OV because the construction process starts with the head that is merged stepwise with its argument phrases. Fukui &Takano (1998) assume the order complement-before-head as basic. As for English and Dutch, the idea of the universality of OV has found a (part time) advocate in Koster (1988).
[36] *FP* is a cover variable for any kind of functional projection.

(2) a.	$[IO\ [O\ V°]]_{VP}$	starting point
b.	$[V°_i\ [IO\ [O\ e_i\ ]]]_{FP}$	the verb moves to a head position, *or*
c.	$[IO_j\ [O_i\ [\ e_j\ e_i\ V°]]]_{FP}$	the objects move to specs, *and*
d.	$[[e_j\ e_i\ V°]_k\ [[IO_j\ [O_i\ e_k]]]_{FP}$	the emptied VP is fronted to a spec-position ('roll up')

Note that the OV/VO surface order would depend only on fronting the VP in (1d) and (2d) versus (1b) and (1c), in the case of a surface OV order. This scenario gets close to the results of the BBC approach, with an essential difference, though. In the latter approach, the object positions in a head-final VP and in a head-initial VP are VP-*internal*. In the derivations sketched above, any argument position is VP-external at the derivational end point.

This implementation of VO produces the same surface orders as the implementation of VO in Kayne's original conception. Which one is right? This is an empirical question, of course. The answer does not come from the inspection of the output. The output is the same, as far as linearization is concerned. The answer comes from testing the various implication of each of the two competing *structural* analyses. This would also be the answer to the central question of this chapter, namely: What is the better model for understanding the properties of an OV language, an LCA- or a BBC-based one? Derive the implications and see which is right!

Moreover, the onus of proof for the empirical adequacy of an LCA-based massive movement approach towards OV is on the side of the proponents. One of the clear but clearly wrong predictions is the opacity prediction:

Preverbal arguments in OV are claimed to be in positions of the same kind as preverbal arguments in VO, viz. in spec-positions of functional heads preceding the VP. The same kind of conditions that make subject phrases and phrases that precede the subject opaque for extraction in VO, will make any preverbal argument phrase in OV opaque for extractions, too.

This is a grand prediction, and languages could easily be this way. Unfortunately, natural languages are not this way. Hence the prediction is wrong and the theory that produced this prediction is wrong, too. It clearly cannot be fully right.

In sum, there is still an unsound imbalance between theoretical creativity and experimental rigidity in grammar research. It is time for a change. Linguistics prides itself as a discipline with scientific standards but disregards the need for, and the outcomes of, an experimental discipline (experimental psycho- and neuro-linguistics) that would plainly put to test the numerous platonic objects of the theoreticians on the one hand, and confront the theoretician with its own findings that call for a theoretical modelling.

Every mature branch of science has a theory camp and an experimenting camp. Linguistics has an experimental camp but theory developing people largely ignore and marginalize the results of the experimenting groups. Enthrallingly, this has not been recognized as a serious deficit yet. For psychology, Wundt's (1888) plea for thorough experimental checks of theoretical claims (based on introspection) has been the starting point of present day psychology. Linguistics still suffers from the introspectionist heritage that was given up in psychology already at the end of the nineteenth century. If linguistics does not seriously take up the experimental program, it will sooner or later and deservedly get marginalized in the family of cognitive sciences.

Appendix 2: Linearization is not a (series of) spell-out snapshot(s)

In the Minimalist Program, linearization has become a piecemeal by-product of structure generation. In the Chomskyan view, linear order is characterized as a snapshot through a language-specific, cross-linguistically adjustable inspection window on the order of terminals at a given stage of *structure generation*. For a more recent modification, Fox & Pesetsky (2005) suggested regarding spell-out not as a single snapshot but as a series of snapshots taken through a sequence of windows, each looking at a different spell-out domain, e.g. VP, DP and CP (roughly corresponding to *phases*).[37]

They claim a principle of o*rder preservation* that is to constrain the mapping between syntax and phonology when movements take place: Once established at the end of a given spell-out domain, this very linear order is not changed any more in a derivation. When the derivation reaches a spell-out domain S, spell-out applies and the terminals get linearized. When a new spell-out domain S' is completed, spell-out linearizes the new material while preserving the information produced by previous applications of spell-out. Thus, spell-out is strictly incremental. It *adds* new material to the linearization already established at the previous spell-out domains and thereby constrains movement:

As a consequence, only movement of an element X from the *edge* position of a spell-out domain to a position in a higher spell-out domain is licit, since only in this cases the moved item continues to precede the elements that it precedes in D (and conversely for rightward movement).

Movement from a phase-internal position that skips the edge is illicit since the ordering statements established in the phase are inconsistent with the ones established after movement. Suppose the derivation has created the Spell-out domain S in (1), with the given linearization:

(1) $[_S$ X Y Z]

Subsequently, Y moves to the left, up into the next Spell-out domain, as in (2). As a consequence, the linearization information in (1) and the linearization information in (2) are contradicting: Y is linearized after X in (1), but in (9), Y precedes D, which contains X. So, Y must move to the edge of S before undergoing further movement.

(2) *$[_{S'} \ldots Y_i \ldots [_S$ X e_i Z]

Movement from a non-edge position is licit only if the previously established linearization is not disrupted. Y can move from the non-edge position in (1) only if X moves as well and X and Y preserve their original order in the higher spell-out domain.

The data base for the proposal is small, as usual: Scandinavian Object shift, Scandinavian preverbal negated quantifiers, and (a non-problem with) Scrambling, exemplified with Korean Data. In Korean, like in German, Scrambling may separate a DP and a quantifier associated with it, as in (3a). The counterpart of (2b) is not well-formed in Korean. For Korean a stranding analysis for the dissociated quantifier has been motivated.

[37] It is not only a series of snapshots; they must match, in addition. How this is to be guaranteed, they do not tell. Apparently, grammar theory must provide a snapshot comparison file and a checking mechanism that certifies what they call order preservation.

(2) a. dass *die Bilder*$_i$ keiner e_i *alle vier* gesehen hat
that the pictures nobody all four seen has
‚that nobody saw each of the four the pictures'
b. dass *die Kinder*$_i$ das Bild (e_i) *alle* vier angefasst haben
that the children the picture all four touched have

The null-hypothesis for (2b) is clearly this: The order of the arguments in (2b) is congruent with the base order. Hence a base structure is projected. If in the base structure, the DP and the quantifier must be adjacent, then (2b) is simply ill-formed because of the distant quantifier. The distant quantifier itself does not trigger an analysis for (2b) in which both DPs are recognized as scrambled, since scrambling is non-string-vacuous; (see the argumentation in Haider 2010a:185-187). In German, the quantifier expression and the DP do not necessarily form a constituent; otherwise they would make a well-formed Spec-C phrase (3b). Hence, it must not come as a surprise that (2b) is well-formed in German, but not in Korean, if in this language, the quantifier is a constituent of the nominal expression.

(3) a. *Die Männer* haben *alle zwei* den Raum verlassen
the men have all two the room left
b.*/??[*Die Männer alle zwei*] haben den Raum verlassen
[the men all two] have the room left

A more challenging case for the order conservation principle is extraposition in a fronted constituent in V2-languages that would be ill-formed in the base position. The phenomenon has been first brought to attention in Haider (1990). A VP is a site for extraposition. This can be clearly demonstrated with a topicalized VP. In (4a), the PP is extraposed. In the non-extraposed version, the PP precedes the verbal head ‚gesprochen' (*spoken*).

(4) a. [Gesprochen mit dem Zimmermädchen]$_i$ würde er nicht e_i haben
[spoken with the room maid] would he not have
b.*dass er nicht [gesprochen *mit dem Zimmermädchen*] haben würde
that he not [spoke with the room maid] have would
c. dass er nicht *mit dem Zimmermädchen* gesprochen haben würde
that he not with the room maid spoken have would
d. dass er nicht gesprochen haben würde *mit dem Zimmermädchen*
that he not spoken have would with the room maid

The crucial point is this: The VP in (4a) cannot and must not be reconstructed, since in the base order, extraposition targeting an embedded VP is ill-formed, because of obligatory cluster formation. The well-formed extraposition version is (4d). (4a) is well-formed, but it is obviously ruled out by cyclic spell-out: In the base structure, extraposition is impossible (4b). The correctly extraposed version (4d) cannot be the base for VP-fronting with the extraposed PP. So, there is no way for deriving (4a):

Assume you start with the base order of the lowest VP ‚*mit dem Zimmermädchen gesprochen haben*' as in (4c). In this phase, you must not extrapose; otherwise you derive (4b). But, if you topicalize the VP, you cannot extrapose anymore. The PP in (4a) is representative for any extraposable constituent, as for instance a relative clause (5a) or a wh-clause (5b):

(5) a. [Das Buch lesen, das ich ihr empfohlen habe]$_i$ würde$_j$ sie nicht e$_i$ wollen e$_j$
[The book read, that I her recommended have] would she not want
b. [Gefragt, wo sie wohne]$_i$ würde$_j$ er nicht e$_i$ haben e$_j$
[asked, where she lived] would he not have

Note that (4b) and (5) are disastrous for a *copy theory* of movement, too. In each case, the copy is ill-formed because of the illicit extraposition site in the would-be base position. The well-formed order would not match the fronted phrase.

German and other V2-languages with VP topicalization are not the only source for this kind of noncompliant data. English offers exactly the same kind of constellation, with preverbal adverbial phrases in combination with the edge effect:

(6) a. Everyone has much faster [(**than his competitors*) [reached his destination]$_{VP}$]$_{VP}$
b. [How much faster *than his competitors*]$_i$ has everyone [e$_i$ [reached his destination]]?

(6a) illustrates the edge effect for adjuncts of head-initial phrases. In this case it is an adjunct of the VP. The edge effect is absent in the *spec*-position (6b), of course. What is the point of departure for the fronted adverbial in (6b)? It is ungrammatical in the pre-VP position, because of the edge effect. So, it could be generated at the end of the VP (7), but then it would have to move to the left edge of the VP before it could leave this VP, according to F&P's order preservation restriction. But, the left edge is accessible only for the adverbial phrase *without* its post-head material, because of the restriction perceived as the edge effect in (6a):

(7) Everyone has *much faster* reached his destination *than anyone else*

Note that the puzzle is identical with the puzzle for extraposition in a topicalized VP. For a sketch of a solution of these puzzles consult chapter 6, sect. 3.4., and be aware that the puzzle is a puzzle only for derivational accounts, but not for representational approaches.

A *representational* account for this kind of derivational problem is sketched in Haider (2010a, sect. 7.7.1): The topicalized constituent is a (not necessarily maximal) V projection. It is related to an empty category in the base-position. The empty category is not structured or layered, but atomic. It gets assigned the semantic content of its antecedent (but not its shape, since it remains atomic) and it's not yet discharged part of the argument structure. Its syntactic shape is that of an atomic empty element of category V. So, it matches the cluster requirements. Any element in the verbal cluster is an atomic element of the category V.

The phrase in the derived position (i.e. the topicalized V projection) must be a well-formed phrase. Crucially however, this does not imply that it can be embedded in any other position in its given, well-formed shape. Since an OV clause structure does not allow stacked VPs (but only V-clustering), the topicalized VP could not be reconstructed into the position of the empty category (i.e. the so-called 'trace') in the verbal cluster. From a representational point of view, this is unproblematic; from a derivational point of view, it is detrimental.

The very same considerations apply to (6a). The fronted complex adverbial phrase is related to its atomic trace marking the base position. This position 're-imports' the semantic interpretation of the fronted phrase (basically as a *'how'*, with semantic modifications), but syntactically it is atomic.

Let me finish by emphasizing once more the surprisingly underestimated role allotted to *linearization* in the Minimalist Program. Imagine the nontrivial problems for the L1 learning brain. If the MP was right, linearizations would not be predictive of the structure of a given expression until you have managed to determine which window position the given spell-out snapshot belongs to. I have not the slightest cue how a learner could possibly solve this intractable problem. The problem is intractable because of the obvious indeterminacy of the linearization-to-structure relation for an L1 learner in this situation:

Before you can determine the structure of an expression, you have to find out the language-specific ‚window-positions', but before you can find out the window-positions, you have to determine the given structure in order to be able to determine the corresponding window position. An innate tractability device would have to be really prophetic. It would have to anticipate the specific syntactic properties of the language that the brain is going to acquire.

This is a nasty paradox for L1 acquisition, if the world is like the MP wants to have it. It seems that cracking the Enigma-code was as demanding for Turing as cracking the ‚window code' would be for a child. Fortunately, the child does not have to rely on 'Turingeries' and is able to solve its task without crypto-analytical methods. The task becomes achievable once you accept that there are no windows, but only a single orifice, namely the mouth that spells out the linearization at what is customarily called the surface structure level.