

Phrase-structure based typology

Hubert Haider

Univ. Salzburg, Dept. of Linguistics & Centre for Cognitive Neuroscience

Abstract

The predictive power of word order typologies in the Greenbergian tradition is avoidably weak. The set of syntactic properties that can be *unfailingly* predicted from the mere fact that a language is categorized as SVO rather than SOV or VSO, is undersized. Presently, word order patterns are unreliable predictors, for at least two reasons. First, VX patterns are structurally ambiguous. Second, the SVO pattern invites misclassifications to a large extent.

The predictive power of a taxonomy, that is, the assignment of a given language to a type, can be notably enhanced if word order patterns are matched with the properties of phrase structures that are causal for the respective serialization patterns. A more reliable predictor of a wide range of syntactic properties is the structural position of the head of a phrase within the phrase: head-*final*, head-*initial*, and crucially, *variable* positioning of the head within its phrase, as a hitherto underestimated property. If the latter type fails to be recognized as a type of its own, misclassifications are unavoidable.

Misclassifications frequently result from contaminating SVO languages and languages with variable head positioning since the subject-verb-object serialization is a frequent pattern of simple clauses in languages of the latter type. Since the two types – head-initial vs. variable – differ in essential syntactic respects, the aggregate of genuine and misclassified SVO languages is syntactically inconsistent, with almost zero predictive power.

This paper lists and analyses ten syntactic properties. These properties directly correlate with the grammatically determined, canonical positioning of the head within its phrase. Hence they serve as diagnostics for more accurate type assignments, with SOV, SVO, VSO, and VHP (variably head-positioning) as major syntactic clause types. A taxonomy based on phrase structure gains a much higher predictive power than the mere inspection of frequent linearization patterns.

1. Introduction

In word order typology, taxonomic accuracy as well as predictive power need to be enhanced. Current typological investigations usually understand the Greenbergian word order denominators (e.g. SOV, SVO, VSO) as immediate diagnostic *markers*¹ for the presence or absence of particular syntactic properties. However, the reliability of correlations anchored on such elementary word-order patterns is limited, the main source of limitations being the following.

Word order is a *secondary* property. It is the correlate of a primary property, namely the syntactic structure of linguistic expressions. Structure determines word order but word order does not fully predict structure. There is no one-to-one relation between word order and structure. The very same word order may be compatible with different but incompatible structures.

¹ A linguistic "*marker*" – in analogy to the concept of bio-markers in biology – is a linguistic trait such as a particular word order pattern that serves as an identifier of a particular language type. The types are correlated with the presence or absence of specific grammatical properties.

Grammars determine the mapping from structures to word orders, that is, from structure to linear arrays, but the inverse is a one-to-many relation. In other words, a given linearization of lexemes may be structurally ambiguous, even within the same language.

Nonetheless, it is understandable that typologists prefer watching out for word-order based correlations. Word order is directly accessible information, structure is not. Structures are determined by the grammar of a language, but for many languages in typological samples, there is a shortage of reliable structural information. Relevant information may be missing and the grammatical reality may be inadequately presented in crucial aspects by descriptive grammarians or informants.

Fortunately, this impasse does not lead into a take-it-or-leave-it dilemma. There are some crucial structural properties that closely correlate with word order cross-linguistically. Such a property is the positioning of the head of a phrase within its phrase. This paper shall focus on syntactic properties that correlate with the position of the heads of major phrases, such as verb phrases, noun phrases, adjective phrases and particle phrases. Since the verb phrase is a basic constituent of a clause, the properties of verb phrases determine core properties of clauses as well.

1.1 Head positions within phrases - final, initial, or variable

It is commonplace that linguistic expressions are *structured*, that is, they are not mere linear concatenations of lexical units. For the majority of languages, a level of organization that is customarily referred to as the level of *phrases* is easily detectable, even without full knowledge of all relevant grammatical details of a language.

Phrases are endocentric, that is, they consist of a "head" that is associated with other phrases by various grammatical means. The syntactic category of the head determines the category of the phrase.² However, the items that constitute a given phrase are not always contiguous. Grammars of most languages permit displacements. Typically, a displaced item is a phrasal constituent, but in quite a few languages, heads may be displaced as well.³ Typically, a constituent of a phrase in a position outside of the phrase it belongs to is in a filler-gap relation. The outside phrase is the filler for something absent inside the phrase it belongs to.

In many languages, the positioning of the head within a phrase follows a uniform pattern. The head position is peripheral. It either follows or precedes the dependent phrases⁴ within its own phrase. When this property is uniform across all head categories in a given language, this language is a strictly head-initial or a strictly head-final language. English, for instance, is strictly head initial. Japanese, on the other hand, is strictly head final.

² There seems to exist a minority group of languages, as for instance the Salish languages (Jelinek & Demers 1994:698), that do not provide a partitioning of lexical items into distinct lexical categories. In such languages, the only way of phrasal organization is the combination of a functor with its argument. Complex utterances are composed of multiple elementary functor-argument propositions: „*Salish languages are as close to 1st order predicate logic as natural languages get.*“ (Cable 2008:1). Such languages are outside the scope of this paper.

³ A well-known case is "V-second" in Germanic languages. The finite verb is in a fronted position. In typology, this should be common place at least since Mallinson & Blake (1981:129).

⁴ Modifying phrases – attributes of noun phrases or adverbials of verb phrases – do not count as dependent phrases. Their serialization is not strictly tied to the serialization of the dependent elements.

In addition, the relative order of the nominal arguments in (2) may vary, too. This amounts to twenty-four (= 4!) possible different serializations of the four items following the complementizer in (2).

- | | |
|---|-----------------------|
| (2)a. (że) Marek Ewie kwiaty <i>dał</i> . | Polish |
| (that) Marek _{Nom} Ewie _{Dat} flowers _{Acc} gave | (Leszkowicz 2015:121) |
| b. (że) Marek Ewie <i>dał</i> kwiaty. | |
| c. (że) Marek <i>dał</i> Ewie kwiaty. | |
| d. (że) <i>dał</i> Marek Ewie kwiaty. | |

All of these variants are grammatical but of course not equivalent with respect to effects that depend on pragmatic side effects, viz. information structure. Thus, some of these serializations are compatible with more contexts of utterance than others.

An essential ingredient of such a word order freedom on the clause level is the positioning of the verbal head within the VP. Variable positioning of a verbal head within its phrase is a general property not only of Slavic languages but in fact of an underestimated number of languages worldwide. Current descriptions of languages such a Polish – in typological as well as in theoretical schools of linguistics – take (2c) to be the order closest to an assumed 'base' order or 'dominant' order. As a consequence, Polish is labelled SVO. Generative grammarians would derive the other orders by shifting back and forth the subject and the objects, respectively. In any of these accounts, Polish appears as a perplexing specimen of an alleged SVO-type language.

However, as will be explicated in detail in Section 3, the syntactic properties of Slavic languages are not exceptional properties of SVO languages. These languages are not exceptional at all. They are regular languages of a type that has not been acknowledged by grammar theory, namely the type of languages with phrase structures that are not constrained by a directionality constraint for the position of the lexical head of a phrase, that is, the *variably-head-positioning* (VHP) type.

As for the other major phrases, the head-complement order of noun phrases or adjective phrases is rarely investigated and documented in typological surveys. Direct information on the phrase-internal position of adjectives as heads of complex adjective phrases is usually missing completely. However, this information is crucial since it interacts with constraints on adjective phrases as modifier phrases of noun phrases. Here is an example.

In German, adjective phrases are head final, as illustrated by the AP in (1b) and (3d). In uniformly head-initial languages such as English or Romance languages, APs are head initial as well. This has an immediate effect on the positioning of APs as adnominal attributes, since 'left' adjuncts of 'left-headed' phrases are subject to an adjacency requirement (see sect. 3.2). The head of a modifier phrase that is pre-adjoined to a head-*initial* phrase must be *adjacent* to the phrase it is adjoined to (Haider 2018). Consequently, head-initial adjective phrases with complements are ruled out as prenominal attributes (3a,b).

- | | |
|--|--------|
| (3) a. a [proud (*of herself)] _{AP} woman | |
| b. une [fière (*de soi)] _{AP} femme | French |
| a [proud _{Ag} of herself] woman | |

- c. une femme [fière_{A°} de soi]_{AP}
- d. eine [auf sich stolze_{A°}] Frau German
 an [of herself proud_{A_{gr}}] woman
- d. a woman, proud of herself
- a. eine Frau, stolz auf sich und unbeugsam
 a woman, proud of herself and unrelenting

In languages such as French, in which an attribute may precede or follow a noun phrase, complex AP attributes invariably follow (3c). In English (3e) as well as in German (3f), complex AP get postponed and treated as appositions to a preceding NP. Their status as apposition shows in the lack of agreement morphology in German in the postnominal position.

German with its head-final adjective phrase (3d) is able to meet the head-adjacency restriction for adjuncts of head-initial noun phrases even if the adjective phrase is complex since the head-final adjectival phrase meets the adjacency requirement in every case. Cross-linguistic surveys of the order of adjectival attributes relative to their target nouns suffer both from neglecting the head-final vs. head-initial property of the adjective phrases in the given languages as well as from neglecting the head-positioning type of the NP, that is, head-initial, head-final or unconstrained.

Surveys such as Dryer's (1992:95; Table 17), which displays correlations between OV/VO and adjective noun order, would benefit. If the noun phrase is head-initial, as in French (3b), *complex* adjective phrases will *always* follow, even if adjectival attributes may precede or follow the noun phrase. If the adjective phrase is head-final as in German, even complex adjectival attributes may, and in fact have to, precede a head-initial noun phrase. This indicates that the positioning of adjective phrases is sensitive to factors that relate to the head position of the adjective phrase *and* the head-positioning of the noun-phrase. As long as these factors are not differentiated, overall correlations derived from such mixed samples will remain less informative than they could be.⁷

As for noun phrases, typological surveys typically focus on the positioning of nominal heads in an NP relative to the serialization of "genitives". In the typological tradition, this term unfortunately subsumes different and partially incompatible phrasal realizations (see Dryer 2013c, 2007). In typological terminology, a "genitive" expression may be a noun phrase (4a) or a prepositional phrase (4b). Moreover, it may be a complement of the noun, as in (4a), or an item that precedes the head-initial noun phrase, such as in (4c).

- (4) a. die [Beschreibung_N [solcher Phänomene]_{NP}] German
 the description such phenomena_{Gen}
- b. die Beschreibung [von solchen Phänomenen_{Dat}]_{PP}
 the description of such phenomena
- c. Dryers/deren/*solcher Phänomene/ Beschreibung
 Dryer's_{Gen}/their_{Gen}/such phenomena_{Gen}/ description

⁷ Even if Dryer (1992: 96) is right – "*I conclude that <noun, adjective> is not a correlation pair*" – regularities such as the adjacency requirement for modifiers preceding head-initial noun phrases will be recognizable only when the head-positions of the involved phrases are taken into consideration.

The terminological umbrella "genitive" that subsumes noun phrases, particle phrases, prenominal possessor phrases or genuine nominal genitives (or equivalent cases) – gives rise to misleading classifications, as for instance in English, which is classified as a language with no dominant order for nouns relative to 'genitives' (Dryer 2013a). But English clearly has strict and dominant orders. PPs always follow the head-noun of the NP (5a) and never precede (5b), while a noun phrase with a possessive marker "-s" never follows (5c) but always precedes (5d). What precedes and what follows is clearly different and should not be confounded. It is just the fuzzy usage of "genitive", which subsumes structurally disjoint items under a functional cover term, that permits misleading descriptions of "genitive" items that may precede or follow a noun. In English, nominal phrases are *never* complements of a preceding noun.

- (5) a. the description of the suspect
 b. *(the) of the suspect description
 c. *the description the suspect's
 d. the suspect's description

Because of the taxonomic fuzziness underlying such classifications, their predictive accuracy is weakened and it is difficult to extract precise cross-linguistic information on the head positioning within noun phrases from typological surveys. A prenominal noun phrase with a possessive marker must not be confounded with a postnominal prepositional phrase that may also denote a possessor. What is usually missing completely in typological surveys is sufficiently reliable information on the position of complements of nouns in an NP, on a par with complements of verbs in a VP, and information on possible order variations within complex NPs. Here is an instructive illustration from German. Of course, not every language provides such minimal-pair contexts for the different word order properties of verb phrases and noun phrases.

In German, and in Germanic as well as Romance languages in general (see Sleeman [2010], Sleeman & Perridon [2011]), verbs are convertible. The infinitival form can be used as a nominal form. In German, verbs are phrase-*final* heads (6a,b) while nouns are phrase-*initial* (6c,d), as in all Germanic languages. Verbs assign accusative to their *preceding* direct object; verbs converted to nouns assign genitive to the very same item that *follows*, like all nouns do. Head-final phrases permit order variation (6a,b), head-initial phrases don't (6c,d). Consequently, a German VP admits order variation (6a,b) but head-initial phrases such as NPs don't (6d). This property, by the way, is one of the prime candidates for cross-linguistically relevant correlations of head-positioning and order variation (see Section 3.3).

- (6)a. [einen Konflikt mit Gewalt lösen]_{VP} German
 a conflict_{Acc} by violence solve
 b. [mit Gewalt einen Konflikt lösen]_{VP}
 by violence a conflict solve
 c. das [Lösen eines Konflikts mit Gewalt]_{NP}
 the solve_{Inf} a conflict_{Gen} by force ('the solving of the conflict by force')
 d. *das [Lösen mit Gewalt eines Konflikts]_{NP}
 the solve by violence a conflict

In sum, two kinds of highly desirable information should be part of typological surveys, namely first, information on the position of the head in the noun phrases relative to a complement of the noun in a given language, and second, information on the order regularities *within* complex noun phrases. Head-final noun-phrases are predicted to admit order variation since head-final phrases permit this. Strictly head-initial noun phrases, on the other hand, are expected to be strictly ordered, like any other strictly head-initial phrase. This correlation between head-positioning and order restrictions is a general correlate of the structural make up of head-initial phrases in contrast to head-final ones (see Haider 2013: 62-64). Head-final phrases admit phrase internal variation, head-initial ones don't.

1.2 Verb positioning in clauses

When Greenberg (1963:45) decided to utilize the relative order of S, O, and V as simple and easily accessible markers for distinguishing clausal word order patterns, he could not know that one of his three major types, namely SVO, viz. his Type II, is difficult to reliably identify by the surface order of the three items. Cross-linguistically, the mapping of this order onto a structure is a one-to-many mapping. The very order S-V-O is compatible with different but incompatible clause structures, all of which give rise to a linear ordering in which in a simple clause, a subject precedes the verb, which precedes its object. It is a consequence of this difficulty that in typological surveys, numerous languages are arguably misclassified. Presently, "SVO" is not a sufficiently reliable predictor for correlating grammatical properties.

For Greenberg and many typologists since then, SVO, SOV, and VSO have merely been serialization *patterns*. However, a serialization pattern is a reflex of structural properties of the involved phrases in a given language. The structural conditions provide the relevant level of explaining the correlation of patterns within and across syntactical types. Therefore, it is essential to understand SVO and the other types as type *labels* rather than phenomenological *markers* of specific types of clause structure. The label is the name of a distinct set of syntactic properties. "Same serialization" in simple clauses cannot not be taken at face value and equated with "same type" especially in the case of S-V-O.

If, as many typological surveys do, the order patterns of minimal declarative clauses are taken as primary grammatical markers, it is easy to arrive at inaccurate type assignment decisions. The mere inspection of the sequence of subject, object and verb in a simple clause of a given language is not informative enough. An illustrative case is the variety of alternative, but incompatible, type assignments in current literature even for linguistically very well-known languages such as the continental West-Germanic languages, that is, Dutch, Frisian, German, and Afrikaans.

For some contemporary typologists, these count as SVO languages (Gell-Mann & Ruhlen 2011; appendix⁸), for others they have "*no dominant word order*"⁹ (Dryer 2013a), and for yet

⁸ In particular, German and Afrikaans are classified as SVO, Dutch as SVO/SOV, and Frisian is missing. However, Afrikaans, Dutch, Frisian and German do not differ with respect to the positioning of the finite verbs. They are SOV languages with a superimposed verb-second requirement: "*The order used for a stylistically unmarked version of John saw Mary in German would be SVO, too, but to simply call German an SVO language would disguise the verb-second nature of its word order.*" (Mallinson & Blake (1981:129).

⁹ Actually, the definition of '*no dominant order*' is infelicitous since – in Plank's (2009:61) words – "*There is not a bit of flexibility about where the German verb is.*" For Dryer (2013d), "*Dominant order [...] means that it is*

others they are OV languages (Hawkins 2014:140), which by the way is the structurally adequate categorization.

Such a state of affairs should be a warning signal. If the basic linguistic classification of well-studied languages suffers from such a high degree of discrepancy, the chances that the reliability of type assignments for less well-known languages are of a better quality are proportionally low. The problem lies not so much in the lack of information but in the choice of the classification parameters.

Dryer (2013a) acknowledges the verb-second property of German and refines the classification. "*In German and Dutch, the dominant order is SVO in main clauses lacking an auxiliary and SOV in subordinate clauses and clauses containing an auxiliary.*" This characterization is appropriate only for the positioning of the finite verb. However, any claim that SVO in the sense of subject-before-verb is a *grammatically* determined order of main clauses in Dutch or German would be wrong. The clause-initial position preceding the finite verb is open for *any single* constituent of the clause, that is, not only for subjects,¹⁰ but alternatively for objects, adverbials, or nonfinite verbal constituents. In the typological literature, this has been explicitly acknowledged first by Mallinson & Blake (1981:129); see Fn. 8.

For the appropriate assignment of the label SVO to a given language, it is by no means sufficient to present well-formed, simple, minimal clauses with an S-V-O serialization. Such a criterion will produce a large amount of so-called *false positives*. Languages will be classified as SVO although they aren't "SVO languages". The very word order pattern is compatible with different and mutually incompatible clause structures. Here is an inexhaustive sample (7).

- | | | | |
|--------|---|--------------------------------|-----------------------------|
| (7) a. | [S [VO] _{VP}] _{clause} | "SVO" | |
| b. | [S V _{fin} [-- O --] _{VP}] _{clause} ¹¹ | "SOV" + V2 | (e.g. German) |
| c. | [S [V -- O] _{clause}] _{clause} | "VSO" + subject fronting | (e.g. Syrian) ¹² |
| d. | [{S V O}] _{VP&clause} | "VHP" – variably V positioning | (e.g. Slavic) |

A language justly filed as SVO is a language with an obligatorily head-initial VP and an obligatory, VP-external subject position (7a). In (7b-d), the very same surface order results from different conditions. In (7b), the verb second property turns the word order of a simple transitive clause of an OV language either into S-V-O or O-V-S, but neither order would warrant typing this language as SVO or OVS by itself. (7c) is familiar from VSO languages with the option of fronting the subject. (7d), eventually, refers to clause structures based on a VP with variable head-positioning. In such languages, the position of the head in a phrase is not re-

either the only order possible or the order that is more frequently used." "*The expression dominant order is used here, rather than the more common expression basic order, to emphasize that priority is given here to the criterion of what is more frequent in language use, as reflected in texts.*"

¹⁰ More than a century ago, this has been understood and stated explicitly by Oskar Erdmann (1886:183): "*Durchaus unrichtig ist es, wenn einige Grammatiker hier dem Subjektsnominativ besonderen Anspruch auf die erste Stelle einräumen wollen.*" (It is entirely incorrect if some grammarians concede a special privilege to the subject nominative for the first position).

¹¹ "--" marks the gap position of a displaced item. Here, the subject and the verb are in displaced positions.

¹² The same order variation could be the result of verb fronting, on top of a SVO structure. Such an analysis predicts SVO properties. The analysis indicated in (7c) predicts VSO properties. These theoretical options can be distinguished, for instance with respect to the grammatical restrictions on subject positions (see sect. 3.4).

stricted to the peripheral position. It may appear in any linear order relative to its dependent items in the phrase. Slavic languages are well-known for the property in question, as already exemplified by Polish in (2). The order S-V-O is a frequent order in these languages but their clause structure is not the clause structure of an SVO language.

Languages such as the Slavic languages in general, and Polish in particular, do not neatly fit into the Greenbergian scheme with SOV, SVO, and VSO as the major syndromes. In these types, there is a "dominant", that is, grammatically determined order, with the verb in the indicated position. Any typology that assigns Polish and English to the same syntactic type loses most of its predictive power because of the conspicuous syntactic differences between genuine SVO languages on the one hand, and languages that resemble Slavic languages, on the other hand. The majority of word order alternations in a simple Polish clause is ungrammatical in English and in languages similar to English.

1.3 Types as predictors

Adequately defined types are bundles of syntactic properties that characterize core properties of the members of the given type. Being a member of a given type means sharing the properties of this type. So, assignment of a given language to a type predicts that the given language matches the core properties of the type. There may be variation between the members of a type in peripheral properties but there must be an invariant core of properties if type assignment is of any practical use and theoretical significance. The invariant core amounts to a set of predictions for grammatical properties to be met by any language assigned to a given type.

If it were legitimate to file Slavic languages as SVO languages, they ought to share a substantive set of the defining properties of SVO languages, that is, the properties that follow directly from the particular clausal architecture of an SVO language. Slavic languages, however, systematically differ from uncontroversial SVO languages in numerous properties that are core properties of SVO languages (see Haider Hubert & Szucsich 2018).

As Dixon (2011:183) emphasizes, Slavic languages are not exceptional in this respect: “*More of the world’s languages are like Russian than are like English.*” If taxonomic considerations coerced the grouping of Slavic languages together with English, North-Germanic or Romance under the same type, then Slavic languages would have to be acknowledged as highly exceptional when comparing their syntactic profile with that of uncontroversial SVO languages. Clearly, this would be an undesirable taxonomic decision since it entails the weakening of the definition of the "SVO" type to such an extent that it loses most if not all of its predictive power.

In the following section, ten grammatical properties will be reviewed that correlate with the head positioning in phrases, and in particular in verb phrases. The first five properties (3.1 to 3.5) can be tested in virtually any language with phrases based on lexical categories. The final three properties (3.8 to 3.10) are limited to those languages that displace phrases in interrogative clauses and syntactically related constructions, such as relative clauses. The properties 3.6 and 3.7 are limited to SOV languages. They serve as evidence of a particular property of SOV languages, namely the syntactic clustering of clause-final sequences of verbs.

It will be argued that the properties discussed in the following sections provide a more adequate basis for re-defining the major Greenbergian types – SOV, SVO, VSO – once this set is completed with a type that Greenberg's scheme did not explicitly foresee, namely the type with variable V positioning, that is, the VHP type. The three major Greenbergian types (8a-c) are types with a fixed verb position within the verb phrase. In the missing type (8d), grammar does not narrow down the positioning of the verbal head in its phrase to a single position. Cross-linguistically, the type (8d) seems to be at least as sizeable as the two major types SOV and SVO since many languages that belong to (8d) are currently classified as SVO, as for instance the Slavic language and many other languages with parallel properties.

- (8) a. SOV: *fixed V-positioning*: phrase-final
 b. VSO: *fixed V-positioning*: phrase-initial
 c. SVO: *fixed V-positioning*: phrase-initial; obligatory preverbal subject position
 d. {S,V,O}: *variable V-positioning (VHP)*

2. Head-positioning as a predictor of grammatical properties

In this section, ten candidates for correlates of the positioning of the head within a phrase, most notably the verb phrase, are reviewed. Table 1 itemizes the properties for the three grammatical options of head-positioning in a phrase, namely initial, final, and variable. The "*head-initial*" column lists the properties of languages with head-initial phrases *and* an obligatory, pre-verbal subject position, that is, of genuine SVO languages. The properties 3.1 to 3.5 can – in principle – be tested with any language that differentiates phrases by the lexical category of the head. 3.8 to 3.10 are limited to languages with filler-gap constellations due to phrasal displacement. 3.6 and 3.7 merely serve as sources of independent evidence for 3.5.

Table 1 – Grammatical properties in correlation with head-positioning

	<i>head-initial</i>	<i>head-final</i>	<i>variable</i>	<i>section</i>
i. compact head-complement order	<input checked="" type="checkbox"/>	no	no	3.1
ii. head-adjacent preverbal adjuncts	<input checked="" type="checkbox"/>	no	no	3.2
iii. variable order of nom. arguments	no	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	3.3
iv. obligatory subject position	<input checked="" type="checkbox"/>	no	no	3.4
v. auxiliary-and-verb order variation	no	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	3.5
vi. compactness of Aux-V orders	no	<input checked="" type="checkbox"/>	no	3.6
vii. V+V(+V) nominalization	no	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	3.7
viii. positional filler-gap restrictions	<input checked="" type="checkbox"/>	no	no	3.8
ix. interrogative subject left behind	no	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	3.9
x. preverbal interrogatives left behind	no	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	3.10

3.1 Compact head-complement order

In strictly head-initial phrases (9a,e), the complement immediately follows the head. Interveners such as adverbial phrases are not admitted. The situation is illustrated in (9). There is no room for intervening items such as a prepositional phrase (9b,f). In English verb phrases (9a, b), but also in German noun phrases (9e,f), which are head-initial phrases, compactness is obligatory. In head-final phrases, as for instance (9c,d), or in phrases of the type with variable head-positioning (see Slavic below), on the other hand, no such restriction holds (Haider 2010:12).

- (9) a. He [cut the retaining rope *with his knife*]
 b.* He [cut *with his knife* the retaining rope]
 c. Er hat [das Absperrseil *mit einem Messer* durchgeschnitten]_{VP} German
 he has the retaining-rope with a knife cut-through
 d. Er hat [*mit einem Messer* das Absperrseil durchgeschnitten]
 e. das [Durchschneiden des Absperrseils *mit einem Messer*]_{NP}
 the cut-through the_{Gen} retaining-rope_{Gen} with a knife
 f.* das [Durchschneiden *mit einem Messer* des Absperrseils]

In Dryer's (2013b) survey on the distribution of 'oblique phrases', that is, phrases that function most frequently as adjuncts of verbs, the pattern "VXO", with X as the oblique phrase, is not attested as a dominant word order. XO_V, OX_V, and OV_X, on the other hand, are well represented in the sample. Noun phrases have not been studied in this respect.

Even if the criterion "*dominant word order*" is too vague,¹³ the survey provides two relevant pieces of cross-linguistic information. First, there is no language attested in which VXO is a dominant order, which is in clear contrast to OX_V. Second, there are languages, in particular Slavic languages, in which VXO is nevertheless an acceptable order. In Dryer's survey, Belorussian is listed under '*no dominant word order*'. This entails, that VXO is a licit serialization in this language in alternation with VO_X or XVO. However, Belorussian is representative of Slavic languages in this respect, as Polish (see Zabrocki 2016:135), Russian (Kallestinova¹⁴ 2007:81), and BCS (Bosnian, Croatian, and Serbian) confirm, which happen to be missing in the survey. In strictly head-initial languages, the VXO pattern is absent, but in Slavic languages, although they are usually listed under "SVO", it is an attested pattern.

- (10)a. Nikola piše *često* smiješna pisma. BCS (Bosnian, Croatian, Serbian)
 Nikola writes *often* funny letters
 b. Marek pisał *często* dziwne listy. Polish
 Marek wrote *often* funny letters

In sum, compactness of head-complement structures is a criterion for identifying strictly head-initial phrases as well as for distinguishing them from optionally head-initial ones. Since genuine SVO languages are strictly head-initial, they will feature the compactness property. Languages with variable head-positioning are optionally head-initial languages. In such lan-

¹³ "This means that it is either the only order possible or the order that is more frequently used." (Dryer 2013d).

If a specific order is characterized as 'dominant', it may be the only admissible order or the more frequent one.

¹⁴ Kallestinova tested 120 native speakers of Russian on order variants. The S-V-Adv-O order has not been rated unacceptable. However, the English control group clearly dismissed the V-Adv-O order in English.

The difference between (12a,c) and (12b,d), respectively, is one in terms of form. Semantically, these adverbial expressions are roughly equivalent. Phonologically and morphologically, the deviant version is less complex than the acceptable form. They differ syntactically. In (12a,c), the head of the adverbial is adjacent to the noun phrase; in (12b,d) it is not, since the adverbial phrase is a prepositional phrase and therefore head-initial.

In Dryer's (2013b) survey, (12b,d) would count as "XVO". He notes that "*the XVO type [...] is quite rare; the only known instances are varieties of Chinese.*" In head-final languages, adverbial phrases typically precede the verb but no such constraint is at work, as Dryer reports: "*All three types of OV languages are widely attested. Among them, XOV is the most frequent on the map (46 languages), OVX next most frequent (44 languages), and OXV least frequent (27 languages).*"

The crucial cases are those in which the "X" itself is not a head-final phrase. In strictly head-final languages, any phrase is head-final. Hence any adverbial phrase will be head-adjacent to the phrase it precedes. But, it is easy to allocate languages showing that the adjacency constraint is absent for adjuncts of head-final phrases, as for instance Dutch or German, or any other OV language that admits postponed phrases:

- (13) a. dat beslissingen [meer *dan werd gedacht*] door emoties worden gedreven Dutch
 that decisions [more *than was thought*] by emotions are triggered
 b. dass Entscheidungen [mehr *als man dachte*] durch Emotionen gesteuert werden
 that decisions [more *than was thought*] by emotions triggered are German

In (13), adverbial phrases as well as passive *by*-phrases precede, but the head of these phrases is not adjacent to the verb phrase. In VO, this is excluded, in OV it is normal. Slavic languages, which are testimonies of VHP languages, pattern like OV and not like VO, that is, they admit non-adjacent, preverbal adjuncts (Haider & Szucsich 2018).

- (14) a. V prošlom godu [gorazdo *bol'se* (čem Igor')] vyigrala tol'ko Maša Russian
 in last year [much more (than Igor)] *won* only Maša
 b. Prošle godine je [mnogo *više* (od Želimira)] radila samo Branka]] BCS
 last year has [much more (than Želimir)] *worked* only Branka
 c. W zeszłym roku [dużo *więcej* (niż Jarek)] pracowała tylko Katarzyna Polish
 in last year much more (than Jarek) *worked* only Katarzyna

Standardly, these languages are filed as SVO. However, they do not pattern like SVO languages in this respect and others. These are VHP languages, and S-V-O is just one out of a set of admissible orders. The adjacency property is a property that identifies strictly head-initial languages in general, and in particular, languages with a strictly head-initial verb phrase.

Prenominal attributes of optionally head-initial noun-phrases behave alike. The head of the attribute must be adjacent to the noun phrase, if the attribute precedes and the noun phrase is strictly head-initial. In Russian (15a), adjacency does not matter. This shows that the noun phrase is not strictly head-initial. In German (15b-d), the noun phrase is strictly head-initial, and the attribute must be head-adjacent (15b,c)

- (15) a. [vernij svoej žene] muž Russian
 faithful his wife_{DAT} man
 ‘a man faithful to his wife’

- | | |
|---|---------------|
| <p>b. ein [seiner Frau <i>treuer</i>] Mann
 a [his wife_{DAT} <i>faithful</i>] man
 c. ein [auf seine Frau <i>stolzer</i>] Mann
 a [of his wife <i>proud</i>] man
 d. *ein [<i>stolzer</i> auf seine Frau] Mann</p> | <p>German</p> |
|---|---------------|

The adjacency requirement produces collateral effects that have been mentioned already in section 1. In strictly head-initial languages, prenominal attributes cannot contain complements (16a,c) since these phrases would make the head of the attribute non-adjacent, as illustrated in (16). French, a language with prenominal as well as postnominal placement of adjectival attributes, complex attributes are postnominal therefore (16d). In English, such adjective phrases are replaced by postnominal appositions (16d).

- | | |
|---|---------------|
| <p>(16)a. un [curieux (*de tout)] homme
 a curious_{AGR} (about everything) man
 b. un homme [curieux_{AGR} de tout]
 a man curious about everything
 c. some [well-known (*to anyone)]_{AP} facts
 d. some facts, [well-known to anyone]</p> | <p>French</p> |
|---|---------------|

In the typological literature, information on the head position of adjectives within complex adjective phrases is usually missing. If adjectives are discussed in connection with the relative order *adjective-noun* or *noun-adjective*, they are not studied as phrases but only as words accompanying nouns. This state of imperfect knowledge is aggravated by the equally imperfect knowledge of the head position of the noun within the noun phrase. The relevant knowledge would be available but it is not standardly gathered in typological descriptions. Consequently, cross-linguistic statements about correlations of adnominal adjectives and nouns in relation to OV and VO remain indeterminate. Correlations can be found, however, once one zooms in into the involved phrases. Here is a correlation: In strictly head-initial languages, *complex* adnominal attributes are postnominal or they cannot be complex.

The reason is the following: If a language is strictly head-initial, noun phrases and adjective phrases are head initial. Hence, the adjacency requirement for adjuncts preceding a head-initial phrase rules out complex, prenominal adjective phrases, since the complement of the adjective would follow the adjective, which precludes adjacency.

3.3 Variable order of arguments

Variable versus strict order within a phrase is immediately correlated with the compactness property discussed in section 3.1. Compact phrases are strictly ordered. In other words, in compact phrases there is no room for order variation. The property that precludes a "VXO" order holds not only when X is an adjunct but also when X would be a reordered object. Phrases which are not compact feature variable word order for objects as well as for adverbials. This kind of word order variation is also known as '*scrambling*'. This term suggests that a given base order gets rearranged. Here, the term is merely used as a label for order variation within and among phrases. Scrambling is absent in head-initial phrases and admitted in head-final phrases and in phrase with variable head-positioning.

Languages in which head-positioning is differentiated by the lexical category of the head, such as German, provide immediately contrasting pairs of head-initial and head-final phrases. Head-final phrases display variable order; head-initial phrases are strictly ordered. These facts clearly indicate that scrambling is a property of phrase structure and not a holistic property of languages. German scrambles within the verb phrase, but not within the noun phrase, since only the former is head-final while the latter is head-initial. Note that the phrases involved in the order variation are morphologically clearly distinguishable. Nevertheless, the pattern (17d) is unacceptable. This contrast between verb phrases and noun phrases is the general contrast between head-final and head-initial phrases.

- (17) a. [an den Vorsitzenden eine Aufgabe übertragen]_{VP} German
to the chairman a task_{Akk} assign
b. eine Aufgabe an den Vorsitzenden übertragen
c. das [Übertragen einer Aufgabe an den Vorsitzenden]_{NP}
the assigning a task_{Gen} to the chairman
d. *das Übertragen an den Vorsitzenden einer Aufgabe
the assigning to the chairman a task_{Gen}

SVO languages are strictly head-initial languages. Hence, in these languages, word order is strict, even if morphology would clearly identify the arguments of a verb. Icelandic is an appropriate example. It is an SVO language with rich case inflection but the word order is strict. In (18), dative and accusative are distinctively marked on the nouns. This notwithstanding, Dehé (2004: 94) reports that "*the inverted order was rejected*", i.e. (18b), and it was rejected by all her informants, without exception.

- (18) a. Þau sýndu foreldrunum krakkana. Icelandic
They showed parents-DEF-DAT kids-DEF-ACC
b. *Þau sýndu krakkana foreldrunum
They showed kids-DEF-ACC parents-DEF-DAT

This situation is in striking contrast with the situation in Slavic languages with their word order freedom on the one hand and their classification as SVO languages on the other hand. So, either the correlation between SVO and strict word order or the classification of Slavic languages is wrong. The criteria discussed in this section converge on the latter. Slavic languages are VHP languages, hence the constraints responsible for the strict order in head-initial phrases do not apply. Bulgarian is particularly instructive in the following respect.

Although Bulgarian lacks morphological case marking, the variability of word order in Bulgarian is as free as in any other Slavic language. The subject and the objects may be serialized freely, with the familiar, concomitant effects on information structuring. (19a-c) are just three variants (see Avgustinova 1997:127-136) out of the set of grammatically admissible twenty-four permutations of the three arguments and the verb. They are semantically equivalent but differ in information structure, that is, they are felicitous answers to different questions.

- (19) a. Ivan *izprati* kuklata na decata Bulgarian
Ivan sent doll_{Def.} to children_{Def.}
b. Kuklata Ivan na decata *izprati*
Ivan sent to children_{Def.} doll_{Def.}

- c. *Izprati* na decata kuklata Ivan
 sent to children_{Def} doll_{Def}.Ivan

In sum, phrase-internal word-order variability is a property of head-final phrases and phrases of the VHP type. Head-initial phrases are strictly ordered. Consequently, SVO languages, understood as strictly head-initial languages with an obligatory subject preceding the head-initial verb phrase, are languages with strict word order. Apparent exceptions come from languages that are misclassified, such as Slavic languages. "Free" word order is not a holistic property of a language. It is a property of phrases. In languages with category-dependent differences in head-positioning, there are phrases *without* word order variation, namely the head-initial ones, and phrase *with* variation, namely the head-final ones. German is a representative instance of this class of languages.

3.4 Obligatory subject position

In SVO languages, unlike SOV and VSO languages, the subject is an obligatorily instantiated grammatical relation in a clause. There is an obligatory *structural* position reserved for a subject and this position must not remain vacant. In SVO, the obligatory subject position precedes the head position of the verb. The objects follow (20a). In the other types, the subject plus the objects uniformly precede (SOV) or follow the verb (VSO). Genuinely subjectless¹⁸ clauses are common in SOV (20b) and in VSO languages such as Syrian Arabic (20c) or Irish (McCloskey 1996). In SVO languages, the subject position must not remain genuinely empty (20d).

- (20) a. Here, *the subject* precedes the verb.
 b. Aus diesem Glas wurde nicht getrunken German
 out-of this glass was not drunk
 c. ma nšarab b ha l-kaseh Syrian Arabic
 not drink_{3sg.Pass} out-of this the-glass (Farhat 1991:178)
 d. *Out of this glass was not drunk.

Subject expletives are a reliable indicators of obligatory subject positions. In the absence of a subject candidate, the position reserved for the subject in an SVO-type clause must be filled with an item whose only grammatical function is to serve as a dummy for a missing subject. A subject expletive is typically a personal pronoun, such as French ,il‘ (21a) or Norwegian ,det‘ (21b), or a locative adverbial, such as English 'there' (21c) or Danish ,der‘ (21d).

- (21) a. *Il* a été dormi dans ce lit. French (Rivière 1981: 42)
 it has been slept in this bed
 b. Ofte vart *det* telefonert. Norwegian (Åfarli 1992:85)
 often was EXPL telephoned
 c. Since then, *there* has been a long decline.
 d. (at) *der* blevet danset Danish (Vikner 1995: 209)
 (that) *there* was danced

¹⁸ "Genuinely subjectless" must not be equivocated with "without a lexical subject", since in null-subject languages, pronominal subjects are not lexicalized but a subject is (morpho-)syntactically identifiable (see below).

German and Dutch are particularly instructive. On the one hand, they are SOV and consequently there is no obligatory subject position. On the other hand, they are V2 languages, with an obligatory clause-initial position in declarative clauses. In declarative clauses, this position must be filled with an expletive (22a,c) if no other item is placed there. However, these expletive items do not figure as subject expletives in otherwise subjectless clauses (22b,d).

- (22) a. *Es* wird an einer Lösung gearbeitet. German
 EXPL is on a solution worked
 b. dass (**es*) an einer Lösung gearbeitet wird
 that (EXPL) on a solution worked is
 c. *Er* wordt aan een snelle oplossing gewerkt. Dutch
 EXPL is on a quick solution worked
 d. dat aan een snelle oplossing gewerkt wordt
 that on a quick solution worked is

Another instructive combination of two typological traits is the combination of SVO and the null-subject property.¹⁹ The latter property eliminates pronouns as candidates for subject expletives, since in null-subject languages, unstressed subject pronouns are obligatorily omitted and expletives are always unstressed. Romance languages provide a good example.

The majority of Romance languages are null-subject SVO languages. As a consequence, the standard passive is not applicable to intransitive verbs since this would result in a subjectless clause (23a,b). French, however, does not share the null-subject property and therefore a pronoun can serve as an expletive (23c). In at least one null-subject Romance language, viz. in Venetian, a regional vernacular spoken in the northeast of Italy, a locative pronoun has been recruited as an expletive subject and consequently, intransitive verbs may get passivized in this language (23d).

- (23) a. **Fue* trabajado duro aquí. Spanish
 was worked hard here
 b. **È* stato dormito bene in questo letto. Italian
 has been slept well in this bed
 c. *Il* a été dormi dans ce lit. French
 EXPL has been slept in this bed (Rivière 1981:42)
 d. *Z'è* stà parlà de ti. Venetian
 there has been spoken about you

In sum, expletive subjects in otherwise subjectless constructions are reliable identifiers of SVO languages. In this type of languages, genuinely subjectless sentences do not exist. If a language allows genuinely subjectless sentences, it cannot be an SVO language.

Slavic languages, once more, are an instructive set of languages. They have been classified as SVO. However, the appropriate classification is not SVO but variable head-positioning (VHP). Slavic languages allow subjectless clauses without expletive subjects. This is incom-

¹⁹ In languages with this property, unstressed pronominal subjects are omitted. Superficially, clauses that otherwise would have a pronominal subject appear to be subjectless. Syntactically, they are not subjectless. The subject is null subject pronoun.

patible with SVO but expected for languages with flexible head-positioning. Here are pertinent examples from Russian (24).

- (24) a. Za granicej ne bylo opublikovano takix statej.²⁰ Russian
 beyond border NEG were_{NEUT} published_{NEUT} such articles_{GEN}
 ‘There weren’t any such articles published abroad.’
 b. Borisu bylo ne istratit' tak mnogo deneg na sebja.
 Boris_{DAT} was_{N.SG} NEG spend_{INF} so much money_{ACC} on self
 ‘It was not (in the cards) for Boris to spend so much money on himself.’

Norwegian, on the other hand, demonstrates how the full grammatical potential is tapped when it comes to filling the subject position in an SVO language (Taraldsen 1979:49; Lødrup 1991:127). The filler for the obligatory preverbal subject position may be the direct object (25a), turned into a derived subject. In (25b), the so-called pseudo-passive, the complement of the prepositional object is turned into a subject. Eventually, in (25c) the subject position is filled with a dummy subject in spite of there being available candidates for the role of a syntactic subject. (25d) is unacceptable since the subject position would not be filled. This is true for main clauses as well as embedded ones.

- (25) a. (at) frimerker ble klistret på brevet. Norwegian
 (that) stamps were pasted on letter_{DEF}
 b. (at) brevet ble klistret frimerker på.
 (that) letter_{DEF} was pasted stamps on
 c. (at) *det* ble klistret frimerker på brevet.
 (that) EXPL was pasted stamps on letter_{DEF}.
 d. *(at) ble klistret frimerker på brevet.

In sum, the obligatory presence of an unequivocal expletive subject is an SVO identifier. On the other hand, a language cannot be of the SVO type if it admits genuinely subjectless finite clauses. The standard passive applied to an intransitive verb produces a subjectless verb. If the result is grammatical in the absence of a subject expletive, the language is not an SVO language. Note that this criterion is not affected by the null-subject property. In either case, a subjectless SVO clause is unacceptable since a null-subject cannot serve as an expletive. This is exemplified by Romance languages. This kind of fine-grained structural constraints on passive constructions easily escape typological surveys.

3.5 Order variation between auxiliaries and the main verb in a clause

In VO languages, auxiliaries typically precede the main verb while in OV they follow. As for the grammatical source of this correlation, Dryer (2009: 204) frankly admits: “*The primary conclusion is that there is no obvious explanation for why auxiliary verbs tend to precede the main verb in VO languages but follow in OV languages.*” However, there is a source and the source is the canonical directionality of verbs in a given language.

Dryer has been confused by the fact that typologists tend to classify auxiliaries as "modifiers" and therefore they expect them to pattern like modifiers, what they don't. They pattern the

²⁰ Alternatively, there is a variant with a subject. In this variant, the genitive of (23a) is a nominative, with the concomitant agreement relations.

way *governing* verbs pattern. Syntactically, auxiliaries are governing items since they determine the grammatical form of the verbs they combine with. English is a convenient example. It is the auxiliary that determines whether the dependent verb is an infinitive, a participle, or suffixed by "-ing".

That auxiliaries are governors is Gunnar Bech's (1955) original insight, who explicitly emphasizes the parallel between case government and "status government". The "status" of the dependent verb (e.g. participial or infinitival or aspectual form) is determined by the governing (auxiliary) verb. The order patterns reflect the directionality property of verbal government. Consequently, OV correlates with V-before-Aux and VO correlates with Aux-before-V. The canonical government direction of verbal government, that is, the selection of the grammatical form of the head of the complement, is uniform in each case. What is different is the governee.

A typologically relevant issue is the variability of these order patterns since variability immediately correlates with the respective word order type. If a language allows order variation for auxiliaries, the language cannot be an SVO language. In other words, in SVO languages, the order of an auxiliary verb relative to the main verb is invariant. The auxiliary precedes the main verb. Order variation is found in OV languages (26) and in VHP languages (27). For this reason, the order variability may serve as a feature for distinguishing SVO languages from apparent SVO language, that is, languages with variable V-positioning for which S-V-O happens to be a frequent serialization pattern.

Variable positioning of auxiliaries is well-attested for Indo-European SOV languages (26). Alternative positions may be available for the very same auxiliary, as in Dutch and German (26a,b), or positions may vary with the kind of auxiliary, as in Persian (26c,d). While the auxiliary for passive obligatorily follows (26c), the future tense auxiliary (26d) obligatorily precedes (Goldberg 2002, §6.1).

- | | |
|--|---------|
| (26) a. dat hij niets gezien <i>heeft</i> / <i>heeft</i> gezien
that he nothing seen <i>has</i> / <i>has</i> seen | Dutch |
| b. dass ich es nie beantworten <i>würde</i> können / beantworten können <i>würde</i>
that I it never answer <i>would</i> be-able-to/ answer be-able-to <i>would</i> | German |
| c. ānhā gošude <i>šodænd</i>
they opened _{Past-Partic.} became _{3P} ('They were opened') | Persian |
| d. ānhā gošude <i>xāhænd</i> šod
they open _{Past-Partic.} FUT-3 rd become ('They will be opened') | |

Uncontroversial SVO languages, that is, languages that meet all the other criteria of the *head-initial* column in Table 1, do not admit order variation for auxiliaries. If there is order variation in an alleged SVO language, this language arguably is a VHP language. A case in point is the Slavic language family (27).

- | | |
|--|--------|
| (27) a. We wtorek <i>poukladać musisz</i> w szafie.
on Tuesday <i>tidy-up must</i> _{2nd.sg.} in wardrobe | Polish |
| b. We wtorek <i>musisz poukladać</i> w szafie. | |
| c. Sutra <i>pospremiti moramo</i> samo našu sobu.
tomorrow <i>tidy-up must</i> _{1st.pl.} only our room | B/C/S |

- d. Sutra *moramo pospremiti* samo našu sobu.
- e. Zavtra *ubirat' budem* v Izmajlovskom parke. Russian
tomorrow *tidy-up shall*_{1st.pl.} in Izmajlovo Park
- f. Zavtra *budem ubirat' v* Izmajlovskom parke.

The positioning of the auxiliary relative to the dependent verb parallels the positioning of the verbal head relative to its objects, in so far as the relative order is variable or not.

3.6 Compactness of V-aux orders

In languages with head-initial VPs, adverbs may be interspersed between stacked VPs. The English example (28a) by Quirk et als. (1986:495, §8.20) presents an adverb in front of each of the four verbal heads of the four stacked verb phrases. In an SOV clause structure such as in German (28b), these adverbs precede the whole sequence of verbs. Any interpolation of adverbs into the canonically ordered sequence of verbs in (28b) would render the clause strongly deviant. This compactness property of the sequence of verbs is a property of head-final languages only. In head-initial languages, each verb is the head of a verbal phrase, and each verb phrase may be modified by preceding adverbials.

- (28) a. The new theory *certainly* may *possibly* have *indeed* been *badly* formulated
 b. ob es *unter Umständen vielleicht tatsächlich schlecht* [formuliert worden sein könnte]
 whether it *under circumstances perhaps indeed badly* [formulated been be could]
 'whether it possibly could perhaps have been badly formulated'

In head-final languages, the verbs of a single clause obligatorily 'cluster' and these clusters do not leave any room for intervening non-verbal items, except for particles of particle verbs. This is the grammatical source of the compactness property.

Clustering, in brief and without further explication (for an explicit analysis see Haider 2015:87), is a grammatical way of avoiding stacked, centre-embedded verbal phrases in head-final languages. Such a structure, unwelcome for the parser, would be unavoidable if each verb was the head of a separate, head-final verb phrase. Instead, there is a single VP with clustered verbs. As discussed in section 3.1, a head-final VP is not compact with respect to object-verb sequence. Therefore, adverbs may intervene between objects and the verb cluster, but not within a cluster. The cluster is compact, but the head-final verb phrase is not.

3.7 V+V(+V) nominalization

This property is a sequel of the verb clustering property. It is listed here mainly because it provides independent evidence for the clustering property of the head-final verbs in SOV clauses. Verb clusters may be the input of word formation processes. Word formation processes operate on word-level items and a cluster is a complex unit of word-level items, namely verbs. Since the cluster is the result of adjoining verbs to verbs, that is, word-level categories to word-level categories, the category of a verb cluster is a word-level category, too, as illustrated by a 3-verb cluster in (29). Crucially, clusters are not morphological compounds but syntactic units.²¹ Compounds cannot be separated by syntactic displacement rules, but

²¹ The same type of head+head clustering structure is known from separable particle verbs, as syntactic units. Particle verbs with separable particles are complex lexical items, consisting of a verb and a particle that precedes

clusters can. For instance, in V2-languages, the position of the finite verb alternates between the clause-final position in the cluster and the displaced position in the clause-initial region. Particles of particle verbs are stranded in the clause-final base position of a displaced finite verb in the Germanic OV languages.

(29) $[[[v^\circ v^\circ]_{v^\circ} v^\circ]_{v^\circ}]_{v^\circ}$

If a language provides a conversion rule for converting items of the category verb into items of the category noun, this type of nominalization-by-category-conversion will not only be applicable to single verbs but also to verbal clusters:

- (30) a. *singen*<sub>V[°] - *das Singen*_{N[°]} German
to *sing*_{Inf} - the singing
b. *das* $[[$ *Befolgen müssen*_{N[°]} $]$ *von Anweisungen*_{NP}
the $[[$ *comply*_{Inf} *must*_{Inf.} $]$ of instructions
'the need to comply with instructions'
c. *zingen*<sub>V[°] - *het zingen*_{N[°]} Dutch
to *sing*_{Inf.} - the singing
d. *het* $[[$ *moeten laten zien*_{N[°]} $]$ *van het witte doekje*_{NP}
the $[[$ *must*_{Inf.} *let*_{Inf.} *see*_{Inf.} $]$ of the white piece-of-cloth
'the obligation to let (someone) inspect the white little piece of cloth'</sub></sub>

Although such a phenomenon is not expressly studied in typological surveys, for obvious reasons, it is not difficult to find testimonies in OV languages and in VHP languages. Japanese is a strictly head-final language, therefore the verbs of a simple clause cluster. The following example (32b) is an instance of clause union²² with a nominalized cluster.

- (31)a. $[$ *tegami-wo* $[$ *kak-i* - *wasure* $]-$ *ru/ta* $]$ _{VP}²³ Japanese
letter-_{Acc} $[$ *write-inf* - *forget* $]$ -Present/Past
b. $[$ *tegami-no/*wo* $[$ *kak-i* - *wasure* $]]-$ *wa*_{Nom} $]$ _{NP}
letter-_{Gen/*Acc} write forget ('the forgetting of the writing of a letter')

As for languages with variable V-positioning, the head-final position is one of the various options. The selection of the head-final position for verbs in a simple clause is the trigger for clustering. Hence a clustering option is expected in these languages, and as a consequence cluster nominalization. Hungarian, a language with variable verb positioning (Kenesei et. als. 1998) confirms this prediction, as Bartos (2004) shows, who was the first to study the phenomenon of nominalized verb clusters in detail in Hungarian.

3.8 Positional restrictions on filler-gap relations

Restrictions show for phrases fronted to the clause-initial position in languages that employ phrasal displacement in interrogative, comparative or relative clauses, and in some languages

in OV-languages such as English *put-by/on/off/up* or follows in VO-languages such as German *an/aus/ein/mit/-zu/machen* (on/out/in/with/at put).

²² In German, clause union in combination with clustering is an option for many verbs that otherwise combine with a clausal infinitival complement. Here is an example with the verb 'forget':

i. *Abzuschließen vergessen wurden die Türen*_{Nom} *nicht*.
to-lock forgotten were the doors not

²³ These examples have been kindly provided to me by Masayuki Oishi (p.c.).

also in declarative clauses. Germanic languages, for instance, are V2-languages and so, they front phrases in declarative (32a) as well as in interrogative clauses (32b). In each case, there is a filler-gap relation between the fronted item as the filler and its canonical position as the gap, indicated by "--" in (32).

- (32) a. *Ingenting* ska jag göra -- imorgon Swedish
 nothing shall I do tomorrow
 b. *Vad* ska du göra -- imorgon?
What will you do tomorrow

The filler-gap relation is grammatically constrained. One of these constraints is a good indicator of an SVO clause structure. Since it is absent in SOV and in VHP languages, it discriminates between genuine SVO languages and putative SVO language, such as the VHP languages.

The constraint is as follows. The gap must be within the directionality domain of a (verbal) head. In SVO, this domain excludes the pre-verbal subject position, since the dependent items *follow* the verb. In SOV, VSO, and in languages with variable positioning, subjects as well as objects are within the respective domains of the verbs. In SOV, subjects and objects precede, in VSO they follow, and in the VHP type, the directionality domain includes preceding as well as following items. As a consequence, fronting an item out of a subject constituent is unacceptable in SVO, but not in languages of the other types.

- (33) a. *What* should I avoid [saying --]?
 b. **What* should [saying --] be avoided?
 c. I invited more people than she had asked me to [invite --].
 d. *I invited more people than [to invite --] was reasonable.

In (33a,b), the clause-initial 'what' is related to a gap position. This relation is blocked whenever the gap position is inside the preverbal subject phrase (33b). In (33c,d), the filler of the gap is the target of comparison²⁴ and the relation is blocked in (33d), for the same reason as in (33c). In OV languages such as German, such a subject-object discrepancy does not exist.

- (34) a. *Wen* hätte [-- dazu zu überreden] sie zu viel Zeit gekostet? German
 who would-have [to-it to persuade] her too much time costed
 b. Er hat schließlich mehr Leute eingeladen als [-- dazu einzuladen] sinnvoll war.
 he has finally more people invited than [-- at-it to-invite] reasonable was

The following well-known property of Slavic languages is instructive for more than one reason. First, it confirms that Slavic languages do not pattern like SVO languages with respect to filler-gap relations, and second, it shows that the gap may even be within a prenominal constituent of a noun phrase, which is also known as 'left-branch extraction'.

Whoever files Slavic languages as SVO languages, falsely predicts the typical pre- vs- post-verbal asymmetry for filler-gap relations. Left-branch extractions are predicted to be acceptable only for gap-phrases in *postverbal* positions but excluded when a gap-phrase is in a *preverbal* position.

²⁴ In English, it is a null pronoun, like in some relative clauses:

i. the people [~~who~~ I talked [to --]]

- (35) a. *Kakuju* Ivan [-- mašinu] *kupil* svoej žene? Russian
 which Ivan [-- car] *bought* his wife
 ('Which car did Ivan buy for his wife?')
- b. *Japonskiju* Ivan [-- mašinu] *kupil* svoej žene.
 Japanese Ivan [-- car] *bought* his wife
- c. *Koju* Petar [-- knjigu] *daje* svojoj ženi? B/C/S
 which Petar [-- book] gives his wife
- d. *Jaki* Jarek [-- samochód] *kupił* swojej żonie. Polish
 which Jarek [-- car] bought his wife

In each example in (35), the gap-containing phrase is preverbal. Nevertheless, each construction is acceptable, given an appropriate context for the information structure effect of the particular word order with a preverbal object. The respective grammars of these languages do not rule out such a construction. These structures are well-formed since both, the verb phrase as well as the noun phrase is VHP. Hence the gap is in the directionality domain of the head and the filler-gap relation is well-formed.

3.9 Interrogative subjects left behind

This phenomenon is directly related with the issue of the preceding section since it singles out the preverbal subject position in SVO. In SVO languages that front only a single interrogative phrase in question constructions and leaves all other interrogatives in their positions, an interrogative subject must not be 'left behind'. English (36) is representative. This restriction is absent in SOV and in languages with variable verb positioning, that is in VHP languages.

- (36) a. *Who* experienced what? – It is unclear [*who* experienced what]
 b. *What did *who* experience? – *It is unclear [what *who* experienced]

In a language with a head-final verb phrase (37) or one with variable head positioning (38), the subject has no priority for the clause initial position in interrogative clauses. Both patterns are attested in corpora.

- (37) a. *Was* hat wen schockiert? – Es ist unklar, *was* wen schockiert hat. German
 what has who shocked it is unclear what who shocked has
- a. *Wen* hat was schockiert? – Es ist unklar, *wen* was schockiert hat.
 who has what shocked it is unclear who what shocked has
- (38)a. *Kdo ho* kde viděl je nejasné? Czech
 who him_{CLITIC} where saw is unclear (Toman 1981: 298)
- b. *Kde ho* kdo viděl je nejasné?
 where him_{CLITIC} who saw is unclear
- c. *Ko* je koga vidio? Bosnian/Croatian/Serbian
 who is whom seen ('Who has seen whom?')
- d. *Koga* je ko vidio?
 whom is who seen

In the unacceptable patterns in (36b), an interrogative subject in the canonical position for a subject is preceded by another interrogative item. The ensuing deviance is a characteristic property of SVO languages. In SOV (37b) and in VHP languages (38), interrogative subjects

are not privileged over non-subjects. An interrogative subject may precede or it may follow another interrogative phrase which is placed in a fronted position.

3.10 Preverbal interrogative adverbial phrases left behind

This property is a correlate of the adjacency property of adjuncts discussed in section 3.2, in the context of multiple-interrogative constructions. In SVO languages, there is no room for an interrogative item in the position of a preverbal adjunct. The interrogative variant of the preverbal adjunct of (39a) is unacceptable in this position in a multiple question (39b). In fact, there exists no acceptable alternative serialization of (39b) at all²⁵ since fronting the adjunct as in (39c) would leave the interrogative subject behind, which would turn the result deviant for the reason discussed in the preceding section.

- (39) a. This has *therefore/very often* proven to be a good strategy.
 b. *What has *why/how often* proven to be a good strategy?
 c. *Why/How *often* has *what* proven to be a good strategy?

A further restriction applies to a subclass of adverbial interrogatives. This order restriction holds for "how" and "how x" (with x as a variable other items such as "often", "many", "much", or dimension adjectives), as well as for 'why'. These adverbials have in common that what they ask for is semantically of a higher type than the semantic type of interrogative pronouns like 'who', 'what', 'when' or 'where'.²⁶ Higher type interrogatives must precede the canonical position of the verb. Consequently, such an expression cannot follow the verb in SVO or VHP languages. (40a,b) illustrates this property for English.

- (40) a. *(I don't know) who saw this film *how often*.
 b. *(I don't know) who quitted his job *why*.

In SOV languages and in VHP languages, the restriction on the preverbal positioning (39b) is absent. So, this property may be adduced for distinguishing SVO languages from languages with variable verb positioning (provided these languages employ fronting in interrogative constructions). If an interrogative adjunct may occur between the subject and the finite verb, the language cannot be an SVO language.

As a consequence, in multiple questions, a subject interrogative plus a higher-order adverbial interrogative cannot co-occur in the same simple clause in SVO languages. If the adverbial interrogative is fronted, the subject is left behind (40c), which is unacceptable. If, on the other hand, the subject interrogative comes first, there is no well-formed serialization available for the adverbial interrogative. It is unacceptable in the postverbal position (40) as well as in the preverbal position (39b).

If Slavic languages were SVO languages, they ought to pattern like uncontroversial SVO languages in this respect. Russian is an apt test case since this Slavic language does not *obligatorily* front all question items in multiply interrogative clauses. (41a) shows that the grammatically well-formed position is the preverbal position. Sentences with higher-order interrogative

²⁵ An acceptable paraphrase of the intended utterance is a coordination of two simple questions: "What has proven a good strategy, and how often?"

²⁶ "Who", "what", "when", "where" quantify over *elementary* entities, that is, individuals, points of time and place. "Why" and "how" quantify over sets (of sets). 'How often', for instance, asks for the cardinality of a set of events.

adverbials in the postverbal position (41b) are ill-formed, as expected. In this case, as in (40), the finite verb would not be in the scope domain. Unlike in strictly head-initial languages, higher-order interrogative adverbials occur in the position immediately preceding the verb phrase (41). This separates VHP language from SVO languages.

(41) a. Mne interesno, kakoj fil'm Boris *kak často* smotrel Russian
 me interests which film Boris *how often* saw

b. *Mne interesno, kakoj fil'm Boris smotrel *kak často*
 me interests which film Boris saw *how often* (40b) is instructive for another reason.

(40b) also indicates that the verb position in (40b) is a canonical verb position and not a displaced position for a finite verb. In verb second languages, the fronted finite verb is in a displaced position and therefore it may precede this kind of adverbial interrogatives²⁷ without affecting the acceptability of the construction.

4. Outcomes

With respect to the positioning of the head of a phrase, the type of variable head-positioning should be acknowledged as a type of phrasal organization in addition to the two standardly assumed types:

- head-final (= right peripheral)
- head-initial (= left peripheral)
- variable (= peripheral or non-peripheral)

Heads have their canonical position either in a fixed *peripheral* position or their positioning is not constrained and therefore variable (variably head positioning = VHP). Phrases with a fixed head position are either head final or head initial. These three possibilities of head-positioning amount to three different types of phrase structures.

In many languages, head-positioning within phrases is uniform across all phrasal heads. Consequently, these languages are uniformly head initial, uniformly head final, or uniformly variable. In some languages, head positioning differs along the categories of heads. A well-known case is that of languages with a head-final verb phrase and head-initial noun phrases. Equally well-known are languages in which the head-positioning of verb phrases and particle phrases (PPs) diverge, such as SOV languages with prepositional phrases. In general, the position of the head within its phrase correlates with a wide range of syntactic properties and is therefore a good predictor for the entailed properties.

- The order VXO, with X as an adverbial modifier and V in its canonical position, is absent in SVO languages but not in VHP languages (3.1).
- Adjuncts preceding a head-initial verb phrase or a head-initial noun phrase are obligatorily *head-adjacent* in SVO. In VHP or in strictly head-final SOV (3.2) there is no adjacency requirement.

²⁷ i. Wer wird *wie oft* applaudieren? (Who shall how often applaud?) German
 ii. Wer applaudierte *wie oft* -- ? (Who applauded how often?)

- The relative serialization of subject and objects is *invariable* in SVO but variable in VHP and in SOV, especially when the grammatical functions are distinctively marked (3.3).
- Subjectless clauses obligatorily contain a *subject expletive* in SVO but not in VHP or SOV (3.4).
- Filler-gap constructions cannot relate the filler to a gap in a preverbal phrase in SVO, but they can in VHP and SOV (3.8).
- If interrogative phrases are fronted to the clause initial position, a subject interrogative must not be left behind in SVO, but it may in VHP and SOV (3.9).
- If interrogative phrases are fronted to the clause initial position, an adverbial interrogative of a semantically higher type must not be left behind in the position preceding the canonical position of the verb in SVO, but it may in VHP and SOV (3.9).

It is important to precisely define and differentiate syntactic types in terms of their core syntactic properties. It is equally important to avoid misclassifications of languages. Misclassified languages pollute the data base and weaken the predictive power of type assignments. A major source of misclassifications is the commingling of SVO-languages and languages with variable head positioning.

In many syntactic respects, SVO languages are more tightly constrained than VHP or SOV languages. If VHP languages, such as for instance the Slavic languages, are mistaken for SVO languages, the clear-cut syntactic profile of the SVO type suffers and loses much of its predictive power since VHP languages lack most of the core properties of SVO languages.

Bibliography

- Avgustinova, Tania 1997. *Word order and clitics in Bulgarian*. (Saarbrücken dissertations in computational linguistics and language technology 5). Saarbrücken: Universität des Saarlandes.
- Bartos, Huba 2004. Verbal complexes and morpho-syntactic merger. In Katalin É. Kiss and Henk van Riemsdijk (eds.) *Verb Clusters: A study of Hungarian, German, and Dutch*. 395-415. Amsterdam/Philadelphia: John Benjamins.
- Bech, Gunnar 1955. *Studien über das deutsche verbum infinitum*, Copenhagen: Munksgaard. (Second unchanged edition in 1983, introduced by Cathrine Fabricius-Hansen, Tübingen: Niemeyer)
- Börjars, Kersti; Engdahl, E; Andréasson, M. 2003. *Subject and object positions in Swedish*. LFG03 Conference. Miriam Butt; Tracy Holloway King (eds.). 43-58. Stanford: CA: CSLI Publications.
- Cable, Seth 2008. *Configurationality and the Salish language*. Ms. U. Mass at Amherst. (<http://people.umass.edu/scable/PNWSeminar/handouts/Config/Configurationality-Salish.pdf>; accessed June 1st 2018).
- Dehé, Nicole 2004. On the order of objects in Icelandic double object constructions. *UCL Working Papers in Linguistics* 16:85-108.
- Dixon, Robert M.W. 2011. *I am a linguist*. Leiden: Brill.
- Dryer, Matthew S. 2013a. Order of Subject, Object and Verb. In: Dryer & Haspelmath (eds.) (<http://wals.info/chapter/81>, Accessed 2018-03-27)

- Dryer, Matthew S. (with Orin D. Gensler). 2013b. Order of Object, Oblique, and Verb. In: Dryer & Haspelmath (eds.) (<http://wals.info/chapter/84>; accessed 2018-03-27)
- Dryer, Matthew S. 2013c. Order of Genitive and Noun. In: Dryer & Haspelmath (eds.) (<http://wals.info/chapter/86>. Accessed 2018-03-27)
- Dryer, Matthew S. 2013d. Determining Dominant Word Order. In: Dryer & Haspelmath (eds.) (Accessed 2018-04-01)
- Dryer, Matthew S. (2009) The branching direction theory of word order correlations revisited. In Sergio Scalise, Elisabetta Magni, Antonietta Bisetto (eds.), *Universals of Language Today*. 185-207. Berlin: Springer.
- Dryer, Matthew S. 2007. "Word order". In *Clause structure, language typology and syntactic description*, Vol. 1, Timothy Shopen (ed.), 61-131. 2nd Edition. Cambridge University Press.
- Dryer, Matthew S. (1992) The Greenbergian word order correlations. *Language* 68(1): 81-138
- Dryer, Matthew S. & Haspelmath, Martin (eds.) The World Atlas of Language Structures Online. (<http://wals.info>).
- Erdmann, Oskar. 1886. *Grundzüge der deutschen Syntax nach ihrer geschichtlichen Entwicklung*. Erste Abt. Stuttgart: Cotta. (Reprint 1985. Hildesheim: Georg Olms Verlag)
- Farhat, Haytham (1991). *Empty categories in Syrian Arabic*. Unpubl. PhD-Dissertation. Bangor: Univ. of Wales.
- Goldberg, Adele. 2002. Persian Auxiliaries. Ms. Princeton University. <https://www.princeton.edu/~adele/papers/Papers> (accessed 2018-04-27)
- Greenberg, Joseph. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph Greenberg (ed.) *Universals of Language*. 73-113. Cambridge, MA.: MIT Press.
- Haider, Hubert (2018). The Left-Left Constraint – a structural constraint on adjuncts. (to appear in) Ulrike Freywald & Horst Simon (eds.) *Headedness and/or Grammatical Anarchy?* Berlin: Language Science Press. (accessed via Academia.edu; June 20, 2018)
- Haider, Hubert 2015. Head directionality – in syntax and morphology. In: Antonio Fábregas, Jaume Mateu, Mike Putnam (eds.) *Contemporary linguistic parameters*. 73-97. London: Bloomsbury Academic.
- Haider, Hubert. 2013. *Symmetry breaking in Syntax*. Cambridge: Cambridge Univ. Press.
- Haider, Hubert 2010. *The Syntax of German*. Cambridge: Cambridge University Press. (Cambridge Syntax Guides).
- Haider Hubert and Luka Szucsich (2018). *Slavic languages – "SVO" languages without SVO qualities?* Ms. Univ. Salzburg (accessed via Academia.edu; June 20, 2018)
- Hawkins, John A. 2014. *Cross-linguistic variation and efficiency*. Oxford University Press
- Jelinek, Eloise, and Richard Demers 1994. Predicates and pronominal arguments in Straits Salish. *Language* 70: 697–736.
- Kallestinova, Elena Dmitrievna 2007. *Aspects of word order in Russian*. University of Iowa PhD Dissertation. [<http://ir.uiowa.edu/etd/165>]
- Kenesei, István and Vago, Robert M. and Fenyvesi, Anna. 1998. *Hungarian*. London / New York: Routledge.
- Leszkowicz, Joanna 2015. Scrambling im Polnischen als A-bar-Bewegung. In Elena Dieser (ed.) *Linguistische Beiträge zur Slavistik: XX. JungslavistInnen-Treffen in Würzburg, 22.-24. September 2011*. p. 117-133. München: Verlag Otto Sagner.
- Lødrup, Helge 1991. The Norwegian pseudopassive in lexical theory. *Working Papers in Scandinavian Syntax* 47:118-29.

- Mallinson, Graham & Barry J. Blake, 1981. *Language typology: Cross-linguistic studies in syntax*. (North-Holland Linguistic Series 46). Amsterdam: North-Holland.
- McCloskey, James 1996. Subjects and subject positions in Irish. In Robert D. Borsley & Ian G. Roberts (eds.), *The syntax of the Celtic languages: a comparative perspective*. 241-283. Cambridge University Press.
- Moinzadeh, Ahmad 2005. A new approach to the Ezafe phrase in Persian. *Ottawa Papers in Linguistics*. 33: 43-64.
- Payne, Doris L. 1990. *The pragmatics of word order: typological dimensions of verb initial languages*. Berlin: Mouton de Gruyter.
- Plank, Frans. 2009. WAL values evaluated. *Linguistic Typology* 13:41-75.
- Polański, Kazimierz 2003. *Encyklopedia Językoznawstwa Ogólnego*. Wrocław: Ossolineum.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey N. Leech and Jan Svartvik (1986⁴) *A comprehensive grammar of the English language*, London, Longman.
- Rivière, Nicole 1981. *La construction impersonnelle en français contemporain*. Documents de Linguistique Quantitative no. 41, St. Sulpice-de-Favière: Éditions Jean-Favard.
- Sleeman, Petra 2010. The nominalized infinitive in French: structure and change. In *Linguística. Revista de Estudos Linguísticos da Universidade do Porto* 5: 145-173.
- Sleeman, Petra and Harry Perridon (eds.) 2011. *The noun phrase in Romance and Germanic: Structure, variation, and change*. Amsterdam: Benjamins.
- Taraldsen, Tarald K. 1979. Remarks on some central problems of Norwegian syntax. Review article of T. Fretheim (ed.) Sentrale problemer i norsk syntaks. *Nordic Journal of Linguistics* 2:23-54.
- Zabrocki, Tadeusz 2016. Syntactic diacrisis in a rigid and a free word order language. *Investigationes Linguisticae* 34: 113-150.