Towards a structure-based typology

Hubert Haider

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

Abstract

The following issues are raised and resolved with the conclusion that the predictive power of Greenbergian word order typologies can be enhanced by grounding them on structural properties. *First*, the S-V-O *linearization* in minimal declarative main clauses is an unreliable marker for the proper identification of the "SVO" type. Accordingly, the rate of *false positive* type assignments is too high. *Second*, if the identification of the subject of transitive clauses is not based on (morpho-)syntactic criteria, languages with different alignment systems get typologically misidentified. *Third*, type profiling benefits from taking into account the structural organization of the major lexical phrases of a given language, at least in terms of the positioning of the heads of phrases. In particular, the [S[VO]] languages, with rigid word order and head-initial verb positioning, must not be conflated with languages in which S-V-O is a frequent serialization option due to non-syntactic order preferences for a grammatically variable word order potential.

1. Introduction

Type assignments should amount to empirically valid predictions. The accuracy of such predictions can readily be improved if the essential word order patterns are joined with a minimum of structural information, that is, information about the phrase structures that underlie the respective patterns. Presently, Greenbergian types are weak predictors, for several reasons. First, they are defined in terms of linearization patterns in minimal clauses. Such patterns are structurally ambiguous and therefore cross-linguistically associated with potentially incompatible grammatical correlates. Second, the crucial notion "S", viz. "subject", needs to be defined structurally rather than semantically, in order to correctly assort the corresponding patterns of different alignment types. The lexico-semantic identification of "subject" in transitive clauses leads to syntactically inappropriate type assignments.

Ultimately and importantly, the clause type does not fully determine the phrase-structure type. A more reliable predictor is the phrase-structure type, that is, head-*final*, head-*initial*, and crucially, *variable* positioning of the head within its phrase in combination with clause-structure types. The paper lists and analyses eight syntactic properties that correlate directly with the grammatically determined, canonical positioning of the head within its phrase. They serve as diagnostics for more accurate type assignments, with SOV, SVO, VSO, and {S,V,O}, that is, syntactically unconstrained head-positioning, as major syntactic clause types. The predictive accuracy of a phrase-structure-based taxonomy is demonstrably higher than a linearization-based one.

Word order is a correlate of a primary property, namely the structure of linguistic expressions. Structure constrains word order, but word order does not determine structure. There is no one-to-one relation between word order and structure. The very same word order may be compatible with several incompatible structures. Grammars determine the mapping from structures to word orders, that is, from structure to linear arrays, but the inverse mapping is often a one-to-many

relation. The sequence S-V-O, for instance, is compatible with at least *five* different clause structures, each one correlated with different, typologically relevant properties (sect. 1.1). Some structural properties, such as the positioning of the head of a phrase within its phrase (sect. 1.2) correlate closely with word order (variation) cross-linguistically. This paper will focus on syntactic correlates of 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 properties of clauses as well.

1.1 Verb positioning in clauses

When Greenberg (1963) elected the relative order of S, O, and V as a simple and easily accessible marker for differentiating between clausal word order patterns, it was not clear that one of the three major types, namely SVO, is difficult to *reliably* identify by inspecting the surface order of the three diagnostic items in minimal utterances. Cross-linguistically, the mapping of this order onto structure is a one-to-many mapping. The very order S-V-O is compatible with several different, even incompatible clause structures, each of which giving rise to a linear ordering in a simple finite clause, with the subject preceding the main verb, followed by its objects. It is an undesirable consequence of this intricacy that in word-order-based typological surveys, languages are classified as SVO although *structurally*, they are not SVO. They merely happen to share the same serialization in *minimal* clauses with a *single* verb. As a reviewer points out, "*SVO was always a problematic word order for Greenberg since it made no reliable predictions for correlating word orders in other phrases, unlike SOV and VSO.*" In Hawkins (1983: 114-16), therefore, verb position is abandoned altogether as a typological indicator.

An instructive example from the current research literature is the variety of different and incompatible type assignments even for well-studied languages such as the continental West-Germanic languages, that is, the typing of Afrikaans, Dutch, Frisian, and German. Some contemporary typologists classify them as SVO languages (Gell-Mann & Ruhlen 2011; appendix)[1]; for others, these are languages with "*no dominant word order*" (Dryer 2013a), and for yet others they are languages with an OV order (Hawkins 2014:140), which by the way, is the *structurally* adequate and empirically correct categorization since they are OV languages with an additional property, namely the "verb-second" property of the finite verb in declarative clauses, which overrides the base-serialization of the verb.

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

The WALS word order page (Dryer 2013a, feature 81A), indirectly honours the verb-second property 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*

---

[1]   In particular, Afrikaans and German 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).

auxiliary." This characterization is roughly[2] appropriate with regard to the positioning of the finite verb. However, the clause-initial position preceding the finite verb is definitely not a subject position. It is open for *any single* constituent of a given clause, that is, not only for subjects,[3] but alternatively for objects, nonfinite verbal constituents, or adverbials, which may outnumber subjects, as will be shown below; see examples (1).

A second difficulty is the definition of "dominant": "*This means that it is either the only order possible or the order that is more frequently used.*" (Dryer 2013b). Evidently, this definition joins two grammatically *disjoint* properties, namely a *strict* word-order property (viz. "the only order") with a *variable*-word-order property. In the strict-order type, variation is ungrammatical; in the variable-order-type, non-syntactic factors, such as information structure, guide the choice of a variant from the pool of grammatical variants. The grammars underlying these types are different. Nevertheless, such languages end up grouped together in spite of incompatible grammatical properties.

Ultimately, a frequency-based decision would first of all require sizeable corpora, which are not available for the majority of languages typologists deal with. Frequency counts for large text samples of languages without literacy are hardly workable. Moreover, frequency is not the decisive criterion for basic word order. For German, frequency data are easily available. Fabricius-Hansen & Solfjeld (1994:101-102) counted roughly 50% non-subjects in the first position of German declaratives in a corpus of 1000 sentences. Here (1) is, for example, the result of a DWDS corpus[4] search and a Google search (Feb. 5th, 2021, filtered for books of the 20th century). In each case, a non-subject outnumbers subjects in the clause-initial position.

(1) a. "er hat damit" ...   DWDS: 84;   71.800 Google books, 20th cent.
        he$_{Nom}$ has it-with

    b. "*damit* hat er" ...   DWDS: 179;   72.200 Google books, 20th cent.
        it-with has he$_{Nom}$

    c. "Der Fehler hat" ...   DWDS: 2;   330 Google books, 20th cent.
        the$_{Nom}$ mistake has

    d. "Den Fehler hat" ...   DWDS: 11;   2760 Google books, 20th cent.
        the$_{Acc}$ mistake has ...

The presentation of a small sample of well-formed, simple, minimal clauses with an S-V-O serialization is not sufficient for the appropriate assignment of the label "SVO" to a given language. Such a criterion will inevitably produce a large amount of *false positive* typing, for the following reason. The very word order pattern goes together with grammatically different clause structures. The sample of structures in (2) is not exhaustive.

(2)   a. [S [VO]$_{VP}$]$_{clause}$ SVO proper   (e.g. English)

---

b. [S V$_{fin}$ [--$_S$ O --$_V$]$_{VP}$]$_{clause}$[5]   SOV plus V2   (e.g. Dutch)

c. [S V$_{fin}$ [--$_S$ [--$_V$ O]$_{VP}$]]$_{clause}$ SVO plus V2   (e.g. Swedish)

d. [S [V --$_S$ O]$_{clause}$]$_{clause}$ VSO plus subject fronting   (e.g. Syrian)

e. [{S V O}]$_{VP\&clause}$ VHP (= variable head positioning)   (e.g. Slavic)

A justly filed SVO language, that is, a language with an SVO clause *structure* is a language with an obligatorily head-initial VP plus an obligatory, VP-external subject position (2a). This type will be referred to as [S[VO]] hereafter. A prototypical specimen is English.

In (2b-d), the very same surface order results from different structural conditions. In (2b), the verb-second property turns the word order of a simple V-*final* transitive clause into an S-V-O order. This secondary property masks the primary property, that is, the canonical order of the main verb, be it OV (2b), as in Dutch, or VO as in Swedish (2c). Grouping Dutch and Swedish together, joins a language with an SOV clause structure with an [S[VO]] language, merely on the evidence of the word order of minimal finite clauses, which typologists typically rely on.

(2d) is familiar from VSO languages with the option of fronting the subject. Steele (1978: 601) presents a survey on word order variation according to which for at least 50% of languages that are classified as VSO, SVO is an alternative order; see also Dryer (2007: 71) on Yagua and Tlingit (Auk dialect).

(2e), eventually, refers to a clause structure based on a VP with variable head-positioning. This property will be discussed below in due detail on the example of Slavic languages. In these languages, the position of the verbal head is *grammatically* not restricted to the peripheral position of the verb phrase (viz. phrase-initial or final). It may appear in any linear order relative to its subject and objects in the phrase. Nevertheless, for non-syntactical reasons, ranging from pragmatics[6] to sentence processing,[7] the S-V-O *linearization* is a frequent pattern in these languages. The clause *structure*, however, is not the clause structure of [S[VO]] languages and their type-specific properties are not shared. Frequency is a misleading criterion for type assignment. Even if V positioning may be syntactically unconstrained for verbs of a given language, this does not entail that each variant is equally frequent.[8] If grammar permits variability, this grammatical freedom is occupied by other preference systems. The crucial point is not equal frequency but equal grammaticality. While OV and VO orders are equally acceptable in Slavic languages, given an appropriate context, OV is strictly ruled out in languages like English. This is not a question of frequency but of grammaticality.

Any typology that assigns, for instance, Polish and English to the same word order type, namely SVO, sacrifices much of its predictive power because of the conspicuous syntactic differences between prototypical [S[VO]] languages on the one hand, and languages such as the Slavic

---

5   "--" marks the canonical position of a displaced item. Here, the subject and the verb are in displaced positions. The subject is just one option for the choice of the clause-initial phrase.

6   Information structuring preferences concerning the relative order of given vs. new, focus vs. background, topic vs. comment; see e.g. Siewierska (1993) on Polish.

7   Goldin-Meadow et als. (2008) found out that actor-first is a cognitively preferred order since it is the preferred sequence in non-verbal tasks, too. Second, early V, that is, V before objects, is favourable for the parser since the verb contains the grammatical information necessary for accurately predicting objects following the verbal head. So, [VO] is targeted by diachronic changes (see Romance and North Germanic).

8   Here is a non-linguistic illustration: Austrian citizens are free to travel to any other country. Nevertheless, Austrians don't visit each foreign destination with equal frequency. Only a few attract most visitors.

languages, on the other hand. If the SVO type has to provide room for properties of English clauses on a par with properties of Polish clauses, the resulting set of *joint* properties will either be minimal or inconsistent since much of what is grammatical in Polish (and in other Slavic languages) is ungrammatical in English.

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

It is commonplace that linguistic expressions are *structured*, that is, they are not mere concatenations of linearly arranged lexical units with rules operating on these concatenations. 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.

Phrases are endocentric, that is, they consist of a "head" item that is associated with other phrases by various grammatical means. The syntactic category of the head determines the category of the phrase.[9] However, the items that constitute a given phrase are not always contiguous. In many languages, grammar permits order variation or dislocation.

In many languages, the positioning of heads within phrases follows a uniform pattern. The head position of a phrase is peripheral and it either precedes or follows the dependent phrases within its own phrase. When this property holds for *all* head categories (viz. V, N, A, P), such a language will be called *uniformly* head-initial or *uniformly* head-final. In such languages, the head either precedes (= head-initial) or follows (= head-final) all its valency-dependents. The serialization of modifying phrases – attributes of noun phrases or adverbials of verb or adjective phrases – is not strictly tied to the serialization of valency-dependent elements. Hence, 'uniformly head-initial" should not be read as "absolutely head-initial". English is uniformly head initial. Japanese, on the other hand, is uniformly head final.

Neither cross-categorical uniformity nor peripherality of the head is a universal property. First, languages may differentiate the positioning of the head in the phrase along lexical categories.[10] Afrikaans, Amharic, Dutch, Frisian, German, Kurdish, or Persian, to name but a few examples, are head-final for verbal heads of phrases (SOV), but noun phrases are head-initial.

Second, the peripherality property of heads arguably is not a universal property either, contrary to widely shared assumptions in the grammar-theory literature. In word-order typology, this possibility is explicitly foreseen: *"Languages with highly flexible word order are themselves a linguistic type."* (Dryer 2007:113). Dixon (2010, vol.1: 74) notes: "*Many 'word order' typologists like to classify each and every language as SOV, SVO, VSO, OSV, OVS, or VOS. How to decide which, if there is in fact no fixed order?*"

If the peripherality constraint is not universal, this leaves room for a type of phrase structure that has not yet been acknowledged in phrase structure theories. In addition to the types with a fixed, peripheral head position – either phrase-initial or phrase-final – there is evidence for the existence of a type with *variable head-positioning* ("VHP"). Slavic languages are easily accessible and fairly well-described languages with variable head-positioning, at least for verbs as

---

[9]   The debate on the universality of lexical categorization has been sparked by Salish languages (Jelinek & Demers 1994: 698), which allegedly lack category distinctions. This is contested by Koch & Matthewson (2009) and Davis et. als. (2014:199): "*No one working on Salish holds to category neutrality these days*."

[10] In the literature, a common term is "disharmonic" (cf. Hawkins 2014:106-115).

heads of phrases.[11] They are usually classified as SVO languages and thereby as languages with head-initial verb phrases. Here is an example from Polish (3), illustrating its undisputed word order freedom. Leszkowicz (2015:121) refers to Polański (2003), who explicitly states that in principle any order of the verb in relation to subject and objects is a *grammatical* option in Polish.

(3) a. (że) Marek Ewie kwiaty *dał*. Polish
    (that) Marek$_{Nom}$ Eve$_{Dat}$ flowers$_{Acc}$ gave (Leszkowicz 2015:121)
  b. (że) *dał* Marek Ewie kwiaty.
  c. (że) Marek Ewie *dał* kwiaty.
  d. (że) Marek *dał* Ewie kwiaty.

Zabrocki (2016: 140) explicitly lists all 24 order variants of "*Janek przedstawił Marysi Marka*" and notes that they each "*have the (logical) meaning of John introduced Mark to Mary*". Polish is representative of Slavic languages in this respect.[12] All these variants are grammatical but, of course, they are not equivalent with respect to information structuring since they are associated with particular focus, topic, or givenness properties.

The simplest way of characterizing the variation in (3) in contrast to English (VO) or Japanese (OV) is one in terms of *grammatically* unconstrained verb positioning. The verb may surface in a clause-final position (3a), in a clause-initial position (3b), or in intermediate positions (3bc,d). In addition, and as a grammatical consequence of this VP-structure (see sect. 4.2), the relative order of the nominal arguments in such languages may vary, too. This amounts to twenty-four possible different serializations of the four items in a main or embedded clause (3).

An essential ingredient of such a word order freedom on the clause level is the positioning of the verbal head within the verb phrase. Variable positioning of a verbal head within its phrase is a general property not only of Slavic languages but in fact of a large number of languages worldwide. Nevertheless, current descriptions of Slavic languages – in typological as well as in 'theoretical' schools of linguistics – take (3d) to be the 'dominant' or 'base' order[13]. As a consequence, Polish is labelled SVO and grouped with languages like English.[14]

As will be explicated in more detail in Section 4, many syntactic properties of Slavic languages would appear to be highly exceptional properties if they were [S[VO]] languages. However, this impression is deceptive. What appears to be exceptional [S[VO]] properties are regular

---

[11] Slavic languages are not *uniformly* VHP. Adjectives as heads of APs are VHP and nouns show VHP-properties as well, except for the Slavic languages with 'articles' (South Slavic), but prepositions are head-initial.

[12] "*Apart from the location of clitics there are virtually no syntactic constraints on the ordering of phrases in main declarative clauses. Thus in each of the Slavic languages all twenty-four possible combinations of a subject, direct object, indirect object and verb occur as grammatical declarative orders.*" Siewierska & Uhliřová (2010:109).

[13] Generative grammarians would derive the variation by 'scrambling' that is, by shifting phrases. In any of these accounts, a language like Polish appears to be an exceptional specimen of an alleged [S[VO]]-type language at first sight. Uncontroversial languages of this type, e.g. North-Germanic languages, do not 'scramble' at all.

[14] Curiously, a neighbouring language of Polish, with virtually *identical* word order properties, namely Belorussian, is classified as language with "*no dominant word order*" in WALS (accessed Jan. 16, 2021). Given the uniformity of word order patterns across Slavic languages with respect to the order of subjects and objects, Mayo's (1993) assessment of Belarusian is appropriate also for Russian, Polish and Ukrainian. "*Belarusian and Russian are genetically very close and structurally very similar languages.* [...] *The structural distance between Belarusian and Russian is of the kind prototypically acknowledged for different dialects of one language.*" (Hentschel 2014: 93-94).

properties of a type of its own, 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 *variable-head-positioning* (VHP) type.

Let us turn now to the other phrasal categories. A typologically under-represented phrase is the adjective phrase. The *internal* order of items in adjective phrases is rarely documented in typological surveys. Information on the phrase-internal position of adjectives as heads of complex adjective phrases is usually missing. 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 or Dutch, adjective phrases are head final. In uniformly head-initial languages such as English or Scandinavian languages, APs are head initial. This has an immediate effect on the positioning of APs as adpositions since 'left' adjuncts of 'left-headed' phrases are subject to an adjacency requirement (see sect. 4.1). The head of a modifier phrase that precedes a head-*initial* phrase must be *adjacent* to the phrase it is adjoined to (Haider 2021b). Consequently, adjective phrases with complements that *follow* the adjectival head are ruled out in prenominal positions (4). In English and French, adjective phrases are head initial. In German (4e), they are head final. Even if Dryer's (1992: 96) assessment – "*I conclude that <noun, adjective> is not a correlation pair*" – is right, regularities such as the adjacency requirement for modifiers preceding head-initial noun phrases will be recognizable only when the head-positions of the phrases involved are taken into consideration. Here is an illustration.

(4) a. a [*tired* (*of silly jokes)]AP audience   – ein [(der dummen Witze)] *müdes* Publikum
                                                 a   (of) the silly jokes *tired* audience

    b. a [*smaller* (*than a thumb nail)] portrait – ein [*kleineres* (*als ein Fingernagel)] Porträt
    c. une [[*fière* (*d'elle-même)]AP nation]   – eine [(auf sich) *stolze*] Nation
       a proud (of itself) nation   a   (of itself) *proud* nation
    d. une [nation [*fière* (d'elle-même)]AP]

In languages with head-initial noun phrases and head-initial adjective phrases, and therefore in any uniformly head-initial language, adjective phrases cannot appear in prenominal positions (4c,d) if the adjective phrase contains a complement. The adjacency requirement holds for adjuncts of head-*initial* phrases only.[15]

As for other adnominal items, typological surveys typically focus on the positioning of a noun relative to the serialization of a "genitive" in possessor constructions.[16] However, this construction is no reliable diagnostics for the position of the head in a noun phrase because the structure of head-*initial* NPs in many languages provides *two* positions (cf. Koptjevskaja-Tamm [1993: 71]), one preceding and one following the head noun (5a,b). As a consequence, "*SVO & GenN languages are as common as SVO & NGen languages.*" (Dryer 2013c). A more precise diagnostic property than possessor expressions would be the serialization of NP-internal

---

[15] Apparent counterevidence such as "a [higher *than average/*last year*] percentage", "an [easy *to spell* (*correctly)] word", or "is [*at the same time/*at noon*] changing the conditions" is analysed in sect. 4 of Haider (2021b).

[16] "The genitive noun phrase is often called the possessor (phrase) and the head noun is sometimes called the possessee (noun), and the construction itself is known either as a *genitive construction* or as a *possessive construction*." (Dryer 2013c).

complements of the head-noun (5c). (5d), contrasting with (5b), shows that the German NP is head-initial, preceded by attributes, and by determiners and equivalent items.

(5) a. die neuen Kleider [*des Kaisers*]
    the new dresses [the emperor's]*Gen*

    b. *des Kaisers* neue Kleider
    [the emperor's]*Gen* new dresses

    c. die unbefugte Benützung [der Daten]
    the unauthorized use [the data's]*Gen*

    d.* [der Daten] unbefugte Benützung
    [the data's]*Gen* unauthorized use

In sum, two kinds of essential information should not be missing in typological surveys, namely first, information on the position of the head[17] of each major phrasal category of a given language, not only of verb phrases and PPs. The second piece of information is information on the rigidity vs. variability of order patterns within complex phrases, that is, verb phrases, noun phrases, and adjective phrases. Presently, the information on phrase-internal serialization in typological descriptions typically refers to the verb relative to its subject and an object, in terms of "*dominant order*". It is the taxonomic fuzziness underlying pure word-order classifications that weakens their predictive accuracy and makes it difficult to extract precise cross-linguistic information on the head positioning within noun phrases from typological surveys.

2. Grammatical structure matters more than content[18]

Content is easy to grasp; structure is difficult to assess. Greenberg (1963: 59) has been very clear about this and about his own – preliminary – recourse to content criteria for the identification of the observables, viz. the relative order of V, O, S, and the correlates thereof.

> "*I fully realize that in identifying such phenomena in languages of differing structure, one is basically employing semantic criteria. There are very probably formal similarities which permit us to equate such phenomena in different languages. However, to have concentrated on this task, important in itself, would have, because of its arduousness, prevented me from going forward to those specific hypotheses.*"

Since then, work on this "*task, important in itself*" has continuously made progress. Half a decade after, we dispose of a much better understanding of how grammars determine structures, and how structures constrain the linguistic form of the presentation of content. Today, Greenberg could more comfortably rely on a set of "*formal similarities which permit us to equate such phenomena in different languages.*"

An instructive example of the disadvantage of a primarily content-based strategy is the misidentified sample of "OVS" languages. Greenberg (1963: 76) described OVS as one of the types that "*do not occur at all or, at least are excessively rare*", and this has proven correct, contrary to positions held in the typological literature, based on about a dozen of alleged OVS languages. Most of them are not Object-Verb-Subject languages. What is the source of the misperception? It is the equation of a lexico-semantic stereotype, viz. agenthood,[19] with "syntactic subject". Agent-V-Patient serialization is taken as a marker for the identification of the SVO type and,

---

[17] Is the head position fixed or variable, and, if fixed, is it phrase-final or phrase-initial?

[18] This section is a condensed version of Haider (2021a).

[19] Queixalós and Gildea (2010:8) are explicit in this respect: "*So for now we adopt the theoretically problematic but heuristically useful practice of relying on intuitive-impressionistic identifications of A and P.*"

conversely, Patient-V-Agent as the identifier for an OVS-type language. This strategy fails for Erg-Abs-languages.[20]

This section will demonstrate that nearly all of the hitherto undisputed candidates for the type "OVS language" are Abs-Erg languages. However, in an ergative language, the patient argument is not "O". Under ergative alignment, the syntactical subject is exactly this *non-agentive,* so-called "O" argument, viz. the absolutive noun phrase. Hence, in syntactic reality, a putative "OVS" language with abs-ergative alignment is an ergative SVO language.

The pivotal point is not so much the basic linguistic description but the subsequent interpretation, that is, the step from "agent" or "patient" to "subject" and "object". Dixon (1994), like other field linguists, is cautious in this respect and separates description from interpretation. He reserves "S" for the single argument of a finite intransitive clause. For transitive clauses, he uses "A" and "O". The source of the subsequent confusion is this. "A" is a *content-based* category while "O" is the customarily used *formal* syntactic category "object". Word order typologies take OVA as input information and interpret it as OVS. What they should do, however, is take patient-V-agent as input, check the alignment system, and then interpret it, based on syntactic criteria, either as SVO in an abs-erg language or, as OVS in a nom-acc language. Dixon (1994: 22) characterizes an (absolutive-) ergative alignment system as follows: "*The term 'ergativity' will be used in the standard way, for referring to S and O being* [grammatically]HH *treated in the same way, and differently from A. 'Ergative' is then used in relation to A, the marked member of such an opposition, and 'absolutive' in relation to S and O, the unmarked term.*"

From this characterization of 'ergativity' it follows that in a clause with ergative alignment, "O" denotes the *subject* of the clause: If S and O are "*treated in the same way*" in an abs-ergative system, and S is the subject of a finite clause, then inevitably, the absolutive "O" will qualify as syntactic subject.[21] Hence, there is no syntactic justification for classifying an ergative language with patient-V-object serialization as "OVS". Dixon (1994: 49-50) explicitly notes that for languages "*with syntactic function shown by constituent order*", OVA is likely to be a sign of ergativity.

Let us recapitulate. In ergative languages, the patient-role of a prototypical agentive transitive verb is represented by the grammatical subject of a simple finite clause. The grammatical subject is – in Mel'čuk's (2014: 179) words[22] – the morpho-syntactically privileged argument of the verbal predicate. For instance, it agrees with the finite verb in languages with subject-verb agreement; it occurs in a structurally uniquely position in languages that identify the subject structurally, and it cannot be omitted[23] without signalling this morpho-syntactically (i.e. by

---

[20] Dryer (2007:70) has pointed this out when referring to Päri: "*Characterizing such languages as OVS is somewhat misleading in that the word order really follows an ergative pattern Abs-V-(Erg).*"

[21] It depends on the grammar of a given language whether the argument marked with ergative case behaves like an oblique noun phrase or is treated like a structural case. In the latter case, it will surface as absolutive in the anti-passive construction (which, in fact is the passive construction of an ergative alignment system, since it signals the morpho-syntactic elimination of the original subject argument).

[22] "*What exactly are syntactic privileges in L has to be indicated by a specific list of SyntSubj privileges elaborated for L.*" (Mel'čuk 2014: 179).

[23] "Omission" must not be confused with the pronominal null-subject phenomenon. In the following example (i), objects of 'forget' and 'forgive' are *omitted*. The subject, however, cannot be omitted (ii.):

passive in nom-acc languages, in direct correspondence to anti-passive languages with absolutive-ergative alignment). Anti-passive, as defined below, is the "passive" of Abs-Erg-languages. It is a term for the morpho-syntactic means for syntactically eliminating the argument that would surface as subject otherwise. Dixon (1994: 146), (2010) and Dixon & Aikhenvald (2009: 9) characterize it as follows:

i. The antipassive construction is formally explicitly marked.
ii. Antipassive forms a derived intransitive from a transitive verb.
iii. The otherwise ergative-marked NP becomes S.
iv. The otherwise absolutive-marked NP goes into a peripheral function and can be omitted.

As for (ii), the appropriate term for the derived antipassive form is not 'intransitive' but 'unergative', since the ergative-marked argument of the active construction switches case and surfaces as absolutive, as stated in (iii).[24] This is in direct correspondence to the acc-to-nom switch in the passive of Nom-Acc languages,[25] if the given alignment system is a system with structural cases. In such systems, a dependency relation holds between the assignment of subject case and the direct-object case. The object case is assigned only in the presence of the subject case (Haider 2000). If the primary subject candidate is syntactically unavailable, subject case is passed on and assigned to the object. The consequence is the familiar acc-to-nom (= object-to-subject) switch or an ergative-to-absolutive (= non-subject-to-subject) switch respectively, in passive. Let us turn now to several samples of alleged OVS languages.

Greenberg's (1963) original sample of thirty languages contained only two languages classified as OVS (with VOS as alternative word order), namely Siuslaw and Coos (s. Greenberg's Appendix II). Both languages are ergative, see Mithun (2005).

Dixon (1994: 50-52) itemizes the following languages as instances of SV/OVA languages, all of which are ergative: Kuikúro (Franchetto 2002), Macushi (Abbot 1991),[26] Maxakalí (Popovich 1986), Päri (Andersen 1988), and Nadëb (Martins & Martins 1999). Dixon also refers to a second pattern, namely VS/AVO, and names Huastec and Paumarí.

Huastec is a Mayan language which Edmonson (1988: 116, 570-75) describes as an ergative language, with the basic order SV/AVO. However, her crucial sample consists of exactly *five* sentences with a structure in which *both* arguments of a transitive verb are present as full noun phrases. "*Sixteen clauses have a variant order* (*O TV, TV A, etc.*)" (Edmonson 1988: 568). Since Mayan languages are predominantly V-initial (England 1991), the Huastec data do not serve as compelling evidence for a strict OVS structure, especially since fronted interrogatives are counted on a par with other NPs (Edmonson 1988: 570).

Paumarí has been characterized as split-ergative by Chapman & Derbyshire (1991: 267, 271) with nom-acc alignment for pronominal arguments. This deserves attention, since in an *ergative* setting, "AVO" would structurally be XVS. Chapman & Derbyshire (1991: 164, 250) declare "SVO" as the basic word order. In Paumarí, a noun phrase is case-marked only in the

---

(i) But Beijing never forgets and certainly does not forgive.
(ii) *But never forgets anything and certainly does not forgive anyone anything.
[24] Likewise, in Nom-Acc systems, the passivized verb becomes *'unaccusative'* rather than intransitive.
[25] "*Antipassive is the exact opposite of the passive in terms of case change.*" Primus (1995: 1090).
[26] Dixon (2010, vol.1: 73) criticizes Ethnologue: "*Macushi [...] is given as OVS, despite the excellent grammar of this language specifying that the 'basic orders' are OVA (although AOV also occurs frequently) and SV.*"

immediately preverbal position. The language has a passive construction, but no antipassive. Zwart & Lindenbergh (in press) note that its coding is incomplete, for case as well as for agreement and conclude "*It seems, therefore, that the pattern is basically accusative (agreement only with S^T/S^I), and that on top of that verbal agreement is sensitive to transitivity (in the 3rd person singular)*. It does not qualify as an ergative language.

In a study on word order type and alignment, Siewierska (1996) lists four languages as "OVS" out of a sample of 237 languages, namely Makushi, Hixkaryána, Päri and Southern Barasano. Makushi is ergative according to Abbot (1991). Päri is an ergative language, according to Andersen (1988). As for Southern Barasano, Jones & Jones (1991) presented a syntax monograph that has been reviewed by Dryer (1994). He points out a crucial weakness[27] of their type assignment and concludes: "*It is possible that it is best treated as indeterminately SOV/OVS, a word order type that appears to be quite common in the Amazon basin.* (Dryer 1994: 63). Hixkaryana (Derbyshire 1979, 1985) will be discussed below.

In WALS (Dryer & Haspelmath 2013), eleven languages are listed as "OVS". Four of them are plainly ergative, namely Kuikúro (Franchetto 1990), Macushi (Abbot 1991), Päri (Andersen (1988), and Tuvaluan.[28] Four are caseless (i.e. 'neutral' alignment) but show ergative properties: Asurini (Jensen 1997),[29] Selknam,[30] Tiriyo,[31] Ungarinjin.[32] The three languages to be discussed further are Kxoe, Urarina, and Hixkaryána.

For Kxoe, Fehn's (2015:214) grammar of Ts'ixa (Kalahari Kxoe) is clear cut: "*There are three patterns available for transitive clauses: AOV, AVO and OAV, with the latter occurring less frequently than the other two. Although the dominant word order of the Khoe languages is thought to be AOV […], AVO is just as frequent.*" Kxoe does not qualify as a reliable testimony of OVS.[33]

The essential issue to be settled for Urarina and for Hixkaryana, too, is this: Are these languages head-initial or head-final? If their VP is head-final, [OV] is a constituent. If they are head-initial, [VA] is a constituent preceded by O. The latter case would make them [O[VA]] languages, with

---

[27] "*A count of all examples in the grammar shows both SV and VS order common, with SV slightly more common, though numbers of examples cited in a grammar is a poor source of data. But the frequency of SV examples both in the grammar and in the text examined does suggest that the claim that subjects tend to follow the verb is based on both noun and pronoun subjects rather than just noun subjects. If we interpret the notion of an OVS language as referring to clauses with a noun object and a noun subject (the standard usage in word order typology), it is not clear that Barasano qualifies.*" (Dryer 1994: 63).

[28] Besnier 1986: 245: "*Despite the word-order freedom exhibited by Tuvalan, there is a basic order, and this order is verb initial.*" Besnier (2000: xxiv): "*Case marking follows an ergative-absolutive pattern*".

[29] Primus (1995:1089): "*The Tupi-Guarani languages Asurini and Oiampi have ergative marking in dependent clauses.*

[30] "*Selk'nam seems to be an ergative language as to word order and verbal marking. Nevertheless, case marking is still an issue that remains to be debated, since the data now available is not sufficient to determine the typological nature of the language, which appears to have been an S marking/A-O unmarked language till the beginning of the twentieth century.*" Rojas-Berscia (2014: 23).

[31] Rill (2017: 430): "*In the end, Tiriyó verb agreement is best analysed as ergative in alignment.*"

[32] Rumsey (1982:145) summarizes the "ordering norms": S precedes V, O precedes V, while A follows. This is exactly the order one expects to find if a language is an SVO language with ergative alignment.

[33] Type-assignment in WALS follows an earlier source, viz. Köhler (1981). "*He himself reduced the richness of Khwe cultural and linguistic expressions in his documentation by increasingly limiting field methods.*" (Boden 2018: 142).

"O" being the structurally highest argument in the clause. This would presuppose ergative alignment. What are the relevant facts?

Both, Urarina and Hixkaryna, are post-positional. According to Dryer (2007: 69) "*the fact that the characteristics in other languages pattern with the order of object and verb would lead us to expect both OVS and OSV languages to pattern with SOV languages. In so far as we have evidence, this prediction seems to be true. For example, Hixkaryana is postpositional and GN.*" The same is true for Urarina. In addition, as Kalin (2014: 1096) emphasizes, the adjective phrase is head-final, too. Olawsky (2006: 667-668) provides additional information on the V+Aux order of Urarina, an order that is completely absent in V-initial languages. Finally, Olawsky (2006: 662) notes that in negated sentences, AOV is an unmarked order, that is, A is not focussed. "*In a transitive clause, constituent order can be AOV as the result of negation.*" Taken together, these grammatical features are good indicators of a head-final organization of the verb phrase in both languages.

The conspicuous evidence for a head-final VP has led Kalin (2014) to the conclusion, that Hixkaryana is an [[OV]SX] language, with the VP[34] in a *secondary*, that is, fronted position. This would support Derbyshire's (1981) conjecture that the OVS clause structure is the result of the loss of ergative case marking in the Carib languages. An [[OV] ... S ...] structure is the likely outcome when in an ergative Abs-V-Erg system, case distinctions are lost and the alignment system is restructured as nom-acc, while the relative order of patient and agent is preserved. The result is a nom-acc system, with OVS order, at the price of a complication in clause structure.

What the preceding discussion boils down to is this: Out of a total of 1377 languages in the WALS data base, not more than a tenth of a percent show a word order that qualifies for OVS linearizations, namely Hixkaryana and Urarina. However, if the analysis of Kalin (2014) turns out to be robust enough, no language is known whose *basic* clause structure is [O[VS]], which would be the structure of a genuine OVS language, that is a language, with a head-*initial* VP, on a par with [S[VO]].

In sum, the *structural* identification of grammatical relations seems to be the proper and inevitable basis for cross-linguistic comparisons. 'Semantic' classifications of grammatical relations lead astray. They rest on a hidden but wrong premise, namely, that universally, for verbs with an agent and a patient argument, the agent argument is the 'subject' in a 'plain'[35] clause. This is true for Nom-Acc-languages, but crucially not for languages with ergative alignment. The equation of Agent with Subject works for Nom-Acc languages, but not for Abs-Erg languages. It is not astonishing at all that such misidentified "subjects" do not share relevant *grammatical* properties.

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

---

[34] "*Transitive clauses have a tightly bound OV verb phrase constituent that is usually followed by the subject NP. Des had actually said so in a dense 1961 paper I had not seen (IJAL 27, 125-142), packed with obscure formulae.*" (Geoffrey Pullum, Obituary: Desmond Derbyshire, *Linguist List* 19.1, Jan. 03, 2008).

[35] 'Plain' means non-passivized, non-middle, etc., or, in other words, a non-reordered serialization.

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 to be of any practical use and theoretical significance. The invariant core amounts to a set of predictions for grammatical properties that must be met by any language when it is assigned to a given type.

For instance, if it were legitimate to file Slavic languages as [S[VO]] languages, they ought to share a substantive set of the defining properties of [S[VO]] languages, that is, the properties that follow directly from the particular clausal architecture of an [S[VO]] language. Slavic languages, however, systematically differ from uncontroversial [S[VO]] languages in numerous properties that are core properties of [S[VO]] languages, as will be pointed out in section 4; see Haider & Szucsich (2021), Szucsich & Haider (2015). However, typologically, Slavic languages are by no means exceptional, since "*more of the world's languages are like Russian than are like English.*" (Dixon 2011:183).

In the following section, eight grammatical properties will be reviewed that correlate with the head positioning in phrases, and in particular in verb phrases. It will be argued that such properties provide a more adequate basis for defining the major Greenbergian types – SOV, SVO, VSO – once this set is supplemented and thereby completed with a type that Greenberg's scheme did not explicitly foresee as a *separate* type, namely the type with variable V positioning, that is, the VHP type with 'no dominant order'. The three major Greenbergian types (6a-c) are types with a fixed verb position within the verb phrase. In the type (6d), the positioning of the verbal head in its phrase is not constrained to a single position. Cross-linguistically, the type (6d) seems to be as sizeable as the two major types SOV and [S[VO]], since many languages that belong to (6d) are currently classified as SVO.

 (6)a. SOV: *fixed V-positioning*: phrase-final
    b. VSO: *fixed V-positioning*: phrase-initial
    c. [S[VO]]: *fixed V-positioning*: phrase-initial; obligatory preverbal subject position
    d. {S,V,O}: *variable V-positioning* (VHP): initial, final, or phrase-medial position

4. Head-positioning as a predictor of grammatical properties

This section highlights eight syntactic properties that immediately correlate with the structural position of the head of a phrase. It will be shown that a head-initial structure is subject to constraints that are absent in head-final phrases and also in phrases with variable head-positioning. Slavic languages will serve as examples for the VHP-type. These languages share properties with head-final languages, but not with uniformly head-initial languages, that is, with [S[VO]] languages. What they have in common is the *absence* of certain restrictions. The shared properties are predictably absent in the [S[VO]] type because they are incompatible with this type.

Table 1 itemizes properties correlating with head-positioning in a phrase for three structure types, namely head-initial, head-final, and 'variable' head-positioning. The properties listed in Table 1 are intended to be both *theory based* and *distinctive* for VO. Therefore, the list does not include "SVO"-properties that are shared by numerous SOV languages, such as *Prepositions, N-Gen, clause-initial complementizers, noun-relative order* (cf. Dryer 2007). This overlap (see

Dryer 2019: 65-69) is expected and predictable since many languages differentiate phrase-internal head-positions by lexical category. They are head *final* for verbal heads, but head *initial* for nouns, prepositions, and lexical functional categories such as complementizers. Consequently, they share properties that correlate with head-initial phrases (cf. Fahrsi, Kurdish, all continental West-Germanic languages, and many others).

Table 1	*head-initial*	*head-final*	*variable*	*section*
(1) strictly head-adjacent preverbal adjuncts	☑	no	no	4.1
(2) order variation of nominal arguments	no	☑	☑	4.2
(3) obligatory structural subject position	☑	no	no	4.3
(4) auxiliary-and-verb order variation	no	☑	☑	4.4
(5) compactness of Aux-V orders	no	☑	no	4.5
(6) positional filler-gap restrictions	☑	no	no	4.6
(7) interrogative subject left behind	no	☑	☑	4.7
(8) preverbal interrogatives left behind	no	☑	☑	4.8

4.1 Head-adjacency of 'left side' adjuncts of head-initial phrases

As already mentioned in section 1.2, adjuncts that *precede* head-*initial* phrases are subject to a constraint that is absent for adjuncts of non-head-initial phrases (Haider 2021b). The head of the adjunct must be *adjacent* to the phrase it is adjoined to. In SVO languages, for instance French (7), this constraint applies to preverbal adverbial phrases as well as to prenominal attributes (11) of nouns.

(7) a. Il doit être *très soigneusement* analysée   French
       it must be very carefully analysed
    b.*Il doit être *avec soin* analysée
       it must be with care analysed
    c. It must be *very carefully* analysed[36]
    d.*It must be *with great care* analysed

The difference between (7a,c) and (7b,d), respectively, is one in terms of form. Semantically, these adverbial expressions are roughly equivalent. Phonologically and morphologically, the deviant version is even less complex than the acceptable form. They differ syntactically. In (7a,c), the head of the adverbial is adjacent to the noun phrase; in (7b,d) it is not, since the adverbial phrase is a prepositional phrase and therefore head-initial.

The adjacency constraint is absent for adjuncts of head-*final* phrases, as for instance Dutch (8a) or German (b), or any other OV language that admits postponed phrases:

---

[36] Here are results from corpus search: The sequence "*should more *ly*", with "*" as a joker for a single word, is attested in each consulted corpus: BNC: 17, CocA 53, NOW 254. However, a PP in the pre-VP position, such as "*should with care*", "*should with great care*" is absent. [BNC = British National Corpus (100 million); CocA = Corpus of contemporary American English (520 million); NOW = News on the web (5.2 milliard)]

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

In (8), adverbial phrases precede, but the head of these phrases is not adjacent to the verb phrase. In VO, this is excluded, in OV it is standard. Slavic languages (9), which are testimonies of VHP languages at least with respect to verbal heads,[37] pattern like OV and not like VO, that is, they admit preverbal adjuncts with non-adjacent heads; see Haider & Szucsich (2021), Szucsich & Haider (2015):

(9)   a. V prošlom godu [gorazdo *bol'še* (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]]   B/C/S[38]
       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

The adjacency property is a property that identifies uniformly head-initial languages in general, and in particular, languages with a head-initial verb phrase or noun-phrase, since 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 head-initial. In Russian (10a), Bulgarian (10b), Polish (10c), however, adjacency does not matter. This shows that the noun phrase does not behave as *structurally* head-initial. In German (10d,e), the noun phrase is head-initial, and the attribute must be head-adjacent.

(10) a. [*vernyj* svoej žene] muž   Russian
      faithful his wife$_{DAT}$ man
      'a man faithful to his wife'
    b. [verni-jat (na žena si)] măž   Bulgarian
    c. [wierny (swojej żonie)] mąż   Polish
    d. ein [auf seine Frau *stolzer*] Mann   German
      a   [of his wife *proud*] man
    e.*ein [*stolzer* auf seine Frau] Mann

The adjacency requirement produces collateral effects that have been mentioned already in section 1.2. In uniformly head-initial languages, prenominal attributes cannot contain complements (11a,c) since these phrases would make the head of the attribute non-adjacent, as illustrated in (11). In French, a language with prenominal as well as post-nominal placement of adjectival attributes, complex attributes are post-nominal (11b) or else they cannot be complex (11a) . In English, such phrases (11c) are replaced by post-nominal appositions (11d).

(11) a. un [curieux (*de tout)] homme   French
      a curious$_{AGR}$ (about everything) man
    b. un homme [curieux$_{AGR}$ de tout]
      a man curious about everything

---

[37] They are not uniformly VHP, since PPs are head-initial in Slavic languages.
[38] B/C/S = Bosnian, Croatian, Serbian.

    c. some [well-known (*to anyone)]$_{AP}$ facts

    d. some facts, [well-known to anyone]

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 data are 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 phrases involved. Here is a correlation: In uniformly head-initial languages, *complex* adnominal attributes are post-nominal (or ungrammatical). The reason is the following:

If a language is uniformly 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 any complement of the adjective would follow the adjectival head, which subsequently destroys adjacency (Haider 2021b).

4.2 Variable relative order of the arguments of a predicate

Languages such as German, in which head-positioning is differentiated by the lexical category of the head, provide minimal pairs of head-initial and head-final phrases. Head-final phrases display variable order; head-initial phrases are strictly ordered.[39] These facts clearly indicate that order variation is a property of phrase structure and not a holistic property of languages. German displays variable order 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 (12d) is unacceptable. This contrast between verb phrases and noun phrases is a general contrast between head-final and head-initial phrases.

(12) 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$_{Gen}$

[S[VO]] languages are uniformly 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 from in her field study that "*the inverted order was rejected*", i.e. (13b), and it was rejected by all her informants, without exception.

---

[39] This shows most clearly with nominal arguments since PPs and clausal arguments may be optionally postponed in languages that are not strictly head-final.

(13) 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. (14a-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.

(14) 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

Languages that permit the permutation of objects represented by noun phrases, also permit the permutation of objects and the subject. All in all, 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 uniformly 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 the 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. In sum, if a language admits the permutation of S,V, and O, salva grammaticalitate, the language cannot be an [S[VO]] language.

4.3 Obligatory structural subject position

Obligatory subject expletives are reliable indicators of an obligatory structural subject position. In the absence of a subject candidate, the position reserved for the subject in an SVO-type clause is filled with an item whose only grammatical function is to serve as a dummy for a missing subject. In [S[VO]] 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 be vacant. In [S[VO]], the obligatory subject position precedes the head position of the verb. The objects follow (15a). In the other types, the subject plus the objects uniformly precede (SOV) or follow the verb (VSO). Genuinely subjectless[40] clauses are common in languages with head-final VPs, such as German (15b), and in VSO languages, cf. Syrian Arabic (15c) or Irish (15d). In SVO languages, the subject position must not remain genuinely empty (15e).

(15) 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. nuair a bhí tráite síos uaidh   Irish
       when COMP was ebbed down from-it   (McCloskey 1996: 261)
       'when the tide ebbed down from it'

    e.*Out of this glass was not drunk.

A subject expletive is typically a personal pronoun, such as French ‚il' (16a) or Norwegian ‚det' (16b), or a locative adverbial, such as English 'there' (16c) or Danish ‚der' (16d).

(16) 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

German and Dutch are particularly instructive. On the one hand, they are SOV and consequently there is no obligatory structural subject position. On the other hand, they are V2 languages, with an obligatory clause-initial position in declarative clauses. This position must be filled with an expletive (17a,c) if no other item is placed there. However, these expletive items do not occur as subject expletives in otherwise subjectless clauses (17b,d).

(17) 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

---

[40] "*Genuinely subjectless*" must not be confused with "*without a lexical subject*", since in null-subject languages, pronominal subjects are not lexically present but a subject is (morpho-)syntactically identifiable (see below).

Another instructive combination of two typological traits is the combination of SVO and the null-subject[41] property. The latter property eliminates subject pronouns as candidates for subject expletives, since in null-subject languages, unstressed subject pronouns are obligatorily omitted, and expletives would always be unstressed. Romance languages provide a good example. Most of them 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 (18a,b). French, however, does not share the null-subject property and therefore a pronoun can serve as an expletive (18c). 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 a subject expletive and consequently, intransitive verbs may be passivized in this language (18d).

(18) 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[42]
      there has been spoken about you

In sum, expletive subjects in otherwise subjectless constructions are reliable identifiers of [S[VO]] languages. In [S[VO]] languages, genuinely subjectless sentences do not exist. If a language allows genuinely subjectless sentences, it cannot be an SVO language. Conversely, in an [S[VO]] language, the subject position in a clause without a subject argument is represented by a dummy subject, that is, an obligatory expletive.

Slavic languages, once more, are informative. They allow for subjectless clauses without expletive subjects. This is incompatible with [S[VO]] but expected for languages with flexible head-positioning (VHP). Here are pertinent examples from Russian (19a,b) and Bulgarian (19c,d). The prepositional phrase in (19) is not obligatorily clause-initial.

(19) a. V komnate bylo nakureno   Russian
     in room was smoked

    b. Ob ètom bylo napisano v gazete
     about this was written in newspaper

    c. V stajat e vlizano   Bulgarian (Desclés & Guentchéva 1996:56)
     in room is entered

    d. Po trevata est xodeno
     on grass_{def.} is walked_{IMPF-PPP}

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 (20a),

---

[41] In "null-subject" languages, unstressed pronominal subjects are not pronounced. Such clauses do not 'show' their subject pronoun, but it is recoverable, for instance, by verbal agreement. 'Subjectless' clauses as in the Spanish and Italian examples (i) have a recoverable subject. Passivized intransitives, as in (18a,b), however, are subjectless. (i) "(We) have won": Hemos ganado (Sp.) – Abbiamo vinto (It.).

[42] I am grateful to Cecilia Poletto, as a native of this language AND a syntactician, for having confirmed this fact.

turned into a derived subject. In (20b), the so-called pseudo-passive, the complement of the prepositional object is turned into a subject. Eventually, in (20c) the subject position is filled with a dummy subject in spite of there being available candidates for the role of a syntactic subject. (20d) is unacceptable since the subject position would not be filled. This is true for main clauses as well as embedded ones.

(20) 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.
       that was pasted stamps on letter_DEF.

In sum, the obligatory presence of an unequivocal expletive subject is an [S[VO]] identifier. On the other hand, no language can be of the [S[VO]] 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 [S[VO]] language. Note that this criterion is not affected by the null-subject property. In either case, a subjectless [S[VO]] clause is unacceptable since a null-subject could not serve as an expletive. This is exemplified by Romance languages.

4.4 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 government directionality of verbs in a given language (Haider 2015).

Dryer has been misled by the fact that typologists look at content and therefore tend to classify auxiliaries as "modifiers" and thus they expect them to pattern like modifiers, which they don't. They pattern the way a *governing* verb patterns. 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 what he calls "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.

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 variation of the relative order of 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 (a subset[43] of) OV languages (21) and in VHP languages (22). 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 (21). Alternative positions may be available for the very same auxiliary, as in Dutch and German (21a,b), or positions may vary with the kind of auxiliary, as in Persian (21c,d). While the auxiliary for passive obligatorily follows (21c), the future tense auxiliary (21d) obligatorily precedes (Goldberg 2002, §6.1).

(21) a. dat hij niets gezien *heeft* / *heeft* gezien   Dutch
     that he nothing seen *has* / *has* seen

    b. dass ich es nie beantworten *würde* können / beantworten können *würde*   German
     that I it never answer *would* be-able-to/ answer be-able-to *would*

    c. ānhā gošude *šodænd*   Persian
     they opened$_{Past-Partic.}$ became$_{3P}$ ('They were opened')

    d. ānhā gošude *xāhænd* šod
     they open$_{Past-Partic. FUT-3rd}$ 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. Again, a case in point is the Slavic language family (22).

(22) a. We wtorek *posprzątać musisz* w szafie.   Polish
      on Tuesday *tidy-up must$_{2nd.sg.}$* in wardrobe

    b. We wtorek *musisz posprzątać* w szafie.

    c. Sutra *pospremiti moramo* samo našu sobu.   B/C/S
      tomorrow *tidy-up must$_{1st.pl.}$* only our room

    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.

    g. Ivan ne *trjabva da igrae* futbol.   Bulgarian (Schürcks 2009: 31)
      Ivan not must to- play-$_{PRESENT-3SG}$ football
      'Ivan mustn't/needn't play football'

    h. Ivan futbol *da igrae ne trjabva.*

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.

4.5 Compactness of V-aux orders

In languages with head-initial VPs, adverbs typically may be interspersed between stacked VPs.

---

[43] In languages with fixed verb positions, the order of auxiliaries is fixed, too, as for instance in most East Asian OV languages. The crucial point is this: If there is verb order variation, the *relative* order remains unchanged in [S[VO]] languages but it is variable in SOV languages.

The English example (23a) by Quirk et als. (1986: 495, §8.20) presents an adverb in front of each of the four verbal heads. In an SOV clause structure such as in German (23b), these adverbs precede the whole sequence of verbs. Any interpolation of adverbs into the canonically ordered sequence of verbs in (23b) 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 and in VHP languages, the verbs in a simple clause (i.e. main verb, auxiliaries, quasi-auxiliaries) may be separated by intervening adverbs.

(23) 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' (Haider 2010, ch. 7) 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.

4.6 Positional restrictions on filler-gap relations

Grammatical restrictions constrain the filler-gap relation of phrases fronted to the clause-initial position in languages that employ phrasal displacement in interrogative, comparative or relative clauses, and in some languages also in declarative clauses. Germanic languages, for instance, are V2-languages and so, they front phrases in declarative (24a) as well as in interrogative clauses (24b). In each case, there is a filler-gap relation between the fronted item as the filler and is its canonical position as the gap, indicated by "--" in (24).

(24) 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, but with clear-cut differences between [S[VO]] languages and the other types. A good indicator of an [S[VO]] clause structure is the following constraint. The gap must be within the same directionality domain of a (verbal) head. In [S[VO]], 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.

(25) 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 (25a,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 (25b). In (25c,d), the filler of the gap is

22

the target of comparison[44] and the relation is blocked in (25d), for the same reason as in (25c). In OV languages such as German (Haider 2010, sect. 3.5), this subject-object discrepancy does not exist (26).

(26) 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 classifies Slavic languages as SVO languages, falsely predicts the typical pre- vs-postverbal 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.

(27)   a. Kakuju$_i$ Ivan [--$_i$ mašinu] *kupil* svoej žene?   Russian
         which$_i$ Ivan [--$_i$ car] *bought* his wife
     b. Èta*k* včera [t$_i$ devočka] *pogladila* kota?   Bondarenko & Davis (2021: 1)
         this yesterday girl *stroked* cat
     c. Koju$_i$ Petar [--$_i$ knjigu] *daje* svojoj ženi?   Bosnian/Croatian/Serbian
         which$_i$ Petar [--$_i$ book] *gives* his wife
     d. Jaki$_i$ Paweł [--$_i$ samochód] *kupił* swojej żonie?   Polish (Wiland 2010: 335)
         which$_i$ Paul [--$_i$ car] *bought* his wife

In each example in (27), 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.

4.7 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 (28) is representative.[45] This restriction is absent in SOV and in languages with variable verb positioning, that is in VHP languages.

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

---

[44] In English, it is a null pronoun, like in relative clauses, as in (i); for German, see Haider (2010: 102-104).
   i. the people [~~who~~ I talked [to --]].
[45] The 100-million-words British National Corpus does not contain a *single* hit for "what did who" or "what has who". Google books lists 5150 entries for "*was hat wer*" ('what has who'). [search: Dec. 18, 2020].

In a language with a head-final verb phrase (29) or one with variable head positioning (30), the subject has no priority for the clause initial position in interrogative clauses. Both patterns are attested in corpora.

(29) a. *Was* hat wen schockiert? – Es ist unklar, *was* wen schockiert hat.   German
    what has who shocked   it is unclear what who shocked has
  b. *Wen* hat was schockiert? – Es ist unklar, *wen* was schockiert hat.
    who has what shocked   it is unclear who what shocked has

In SOV (29b) and in VHP languages, such as the Slavic languages (Liakin & Juvénal 2001: 210), interrogative subjects are structurally not privileged over non-subjects (30). An interrogative subject may precede or it may follow another interrogative phrase which is placed in a fronted position. In the unacceptable patterns (28b), 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 which is absent in VHP languages in clauses with S-V-O serialization.

(30) a. *Kdo* ho kde viděl je nejasné?   Czech, Toman (1981: 298)
    who him$_{Clitic}$ where saw is unclear
  b. Kde ho *kdo* viděl je nejasné?
    where him$_{CLITIC}$ who saw is unclear
  c. *Ko* je koga vidio?   B/C/S
    who is whom seen ('Who has seen whom?')
  d. Koga je *ko* vidio?
    whom is who seen
  e. *Kto* kogo ljubit?   Russian
    who whom loves
  f. Kogo *kto* ljubit?

4.8 Preverbal interrogative adverbial phrases left behind

This property is a correlate of the adjacency property of adjuncts discussed in section 2.2, 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 (31a) is unacceptable in this position in a multiple question (31b). In fact, there exists no acceptable alternative serialization of (31b) at all[46] since fronting the adjunct as in (31c) would leave the interrogative subject behind, which would turn the result deviant for the reason discussed in the preceding section.

(31) 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 or 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

---

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

ask for is semantically of a higher type than the semantic type of interrogative pronouns like 'who', 'what', 'when' or 'where'.[47] Higher type interrogatives must precede the canonical position of the verb. Consequently, such an expression cannot follow the verb in SVO or VHP languages. (32a,b) illustrates this property for English.

(32) 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 necessity of the preverbal positioning (31b) 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 [S[VO]] 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 (31c), 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 post-verbal position (32) as well as in the preverbal position (31b).

If Slavic languages were [S[VO]] languages, they ought to pattern like uncontroversial [S[VO]] 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. (33a) shows that the grammatically well-formed position is the preverbal position. Sentences with higher-order interrogative adverbials in the postverbal position (33b) are ill-formed, as expected. In this case, as in (32), the finite verb would not be in the scope domain. Unlike in uniformyl head-initial languages, semantically higher-order interrogative adverbials may occur in the position immediately preceding the verb phrase (33). This separates VHP language from SVO languages.

(33) a. Kto *kak* spit?   Russian
       who how sleeps   Stepanov (1997)
    b. *Kak* kto spit?
       how who sleeps
    c. Mne interesno, kakoj fil'm Boris *kak často* smotrel
       me interests which film Boris *how often* saw
    d.*Mne interesno, kakoj fil'm Boris smotrel *kak často*
       me interests which film Boris saw *how often*

(33d) is instructive for yet another reason. It also indicates that the verb position in (33c,d) 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 interrogatives[48] without affecting the acceptability of the construction.

---

[47] *"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; see Haider (2010:119).

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

5. Summary

The predictive power and accuracy of word-order typologies in the Greenbergian tradition can be enhanced by exploiting syntactically relevant information about the clausal and phrasal organization. For major phrases, the positioning of the head is a relevant information since it does not strictly correlate with the word order patterns of simple transitive clauses which are used as type markers. SVO is the least reliable type for unfailing correlations and accurate predictions, for several related reasons.

First, the S-V-O word order is compatible with two different systems of phrasal organization. In one system, phrase structure entails strict word order. In the other system, phrase structure admits variation in word order. Presently, the set of languages assigned to the SVO type is heterogenous in this crucial respect. One subset consists of languages with a head-initial verb phrase, preceded by the subject. This is the subset with a *strict* S-V-O word order, viz. the [S[VO]] type. The other major subset consists of languages in which S-V-O is a frequent serialization option, with other, equally grammatical serialization options. This is the set of "VHP" languages. The first set patterns to a large extent with VSO, due to the uniformly head-initial serialization of major phrases. The VHP set patterns differently since is does not share the *obligatorily* head-*initial* organization of major phrases of the [S[VO]] type. The aggregate of the two subsets under the single type "SVO" is an aggregate of heterogenous grammatical properties, which destroys the predictive accuracy of the SVO type assignment.[49] The term "SVO" should be reserved for the first subset, i.e. the type of [S[VO]] languages. The second subset constitutes a separate type, namely the type of VHP languages, that is, languages in which at least the verb is a variably positioned head of a phrase.[50] *Uniformly invariant* head-positioning versus *uniformly variable* head-positioning are the opposite regions in a system space of phrase structuring. In between are languages that differentiate head positioning by lexical category. In these languages the property of the verb phrase determines clausal properties, since the verb phrase is constitutive for clause structure.

Second the concept denoted by "S" in the denomination of the Greenbergian types needs to be syntactically grounded. A content-based identification is misleading for languages with an abserg alignment. Presently, languages are classified as "OVS", although, structurally, they are SVO languages with ergative alignment. What matters syntactically is not content, viz. agent or patient roles, but the syntactic function, that is, syntactical subject or object. This is reflected cross-linguistically in the shared subject properties across alignment systems.

Third, attention to head-positioning in major phrases is a means of capturing differences resulting from cross-categorically differentiated head-positions. Many SOV languages are not *uniformly* head-final. If, for instance, the noun phrases are head-initial, these language pattern with VO languages, but not with OV, in all properties that relate to a head-initial noun phrase. Such languages are classified as SOV languages due to the head-final VP as the basis of the clause structure, but they show orderings such as *genitive-noun, article-noun, noun-relative clause,*

---

[49] Here is an example: Dryer (2002) tested Hawkins' (1990) prediction on the correlations of SVO with little or no morphological case-marking vs. SOV with rich marking. He did not find confirmation, because the SVO sample is 'polluted' with too many AHP languages misfiled as SVO. Most of them are languages with rich morphological case and agreement systems, like Slavic languages (except Bulgarian and Macedonian).

[50] Languages in which *every* phrase is VHP have been termed "*non-configurational*" by Hale (1983).

plural word-noun, etc., which is the inverse of the orderings expected for OV languages and characteristic of VO languages. This is fully understandable once one acknowledges that the word order properties, viz. "OV", "VO", VHP, do not reliably predict the word order property of the major phrases, except for uniformly head-initial or uniformly head-final languages. If the category-specific ordering relations are not taken into consideration, the accuracy of predictions based on types suffers. In order to do this, information on the positioning of the heads of major phrases must be made available by grammar writers.

Bibliography

Abbott, Miriam. 1991. Macushi. In Derbyshire, Desmond C. and Pullum, Geoffrey K. (eds.), Handbook of Amazonian Languages 3. 23-160. Berlin: Mouton de Gruyter.

Åfarli, Tor. 1992. *The syntax of Norwegian passive constructions*. Amsterdam: Benjamins .

Andersen, Torben. 1988. Ergativity in Parï, a Nilotic OVS language. *Lingua* 75: 289-324.

Avgustinova, Tania 1997. *Word order and clitics in Bulgarian*. Unpubl. doctoral dissertation. Saarbrücken: Universität des Saarlandes.

Bech, Gunnar 1955. *Studien über das deutsche verbum infinitum*, Copenhagen: Munksgaard. (Second ed. 1983, introduced by Cathrine Fabricius-Hansen, Tübingen: Niemeyer).

Besnier Niko. 1986. Word order in Tuvaluan. In Paul Geraghty and Lois Carrington and Stephen A. Wurm (eds.), FOCAL I: Papers from the Fourth International Conference on Austronesian Linguistics, 245-268. Canberra: Research School of Pacific and Asian Studies, Australian National University.

Besnier, Niko. 2000. Tuvaluan: A Polynesian Language of the Central Pacific. London: Routledge.

Boden, Gertrud. 2028. The Khwe collection in the academic legacy of Oswin Köhler: Formation and potential future. *Senri Ethnological Studies* 99: 129-146.

Bondarenko, Tanya & Colin Davis. 2021. Sub-extraction asymmetries and linearization in Russian. (to appear in) *Proceedings of FASL* 28. lingbuzz/004915

Chapman, Shirley & Desmond C. Derbyshire. 1991. Paumarí. In Desmond C. Derbyshire & Geoffrey K. Pullum (eds.) *Handbook of Amazonian languages*, vol. 3. 161–352. Berlin: De Gruyter.

Davis, Henry, Carrie Gillon & Lisa Matthewson. 2014. How to investigate linguistic diversity: Lessons from the Pacific Northwest. *Language* 90(4), 180-226.

Dehé, Nicole 2004. On the order of objects in Icelandic double object constructions. *UCL Working Papers in Linguistics* 16:85-108.

Derbyshire, Desmond C. 1979. *Hixkaryana*. (Lingua Descriptive Studies 1) Amsterdam: North-Holland.

Derbyshire, Desmond C. 1981. A diachronic explanation for the origin of OVS in some Carib languages. *Linguistics* 17: 209-220.

Derbyshire, Desmond C. 1985. *Hixkaryana and linguistic typology*. Dallas: Summer Institute of Linguistics.

Desclés, Jean-Pierre & Guentcheva, Zlatka. 1996. Passivization in Bulgarian: A formal approach. *Univ. of Trondheim Working Papers in Linguistics* 28: 55-74.

Dixon, Robert M.W. 1994. *Ergativity*. Cambridge: Cambridge University Press.

Dixon, Robert M. W. 2010. *Basic Linguistic Theory*. vol. 1-3. Oxford: Oxford University Press.

Dixon, Robert M.W. 2011. *I am a linguist*. Leiden: Brill.

Dixon, Robert M. W. & Aikhenvald, Alexandra Y. 2000. Introduction. In R.M.W. Dixon & A.Y. Aikhenvald (eds.) *Changing valency: Case studies in transitivity*. 1-29. Cambridge: Cambridge University Press.

Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68(1): 81-138

Dryer, Matthew S. 1994. Review of Wendell Jones and Paul Jones, Barasano Syntax. *Word* 45(1): 62-65. DOI: 10.1080/00437956.1994.12098343

Dryer, Matthew S. 2002. Case distinctions, rich verb agreement, and word order type. *Theoretical Linguistics* 28(2): 151-157. DOI: https://doi.org/10.1515/thli.2002.28.2.151. |

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. 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. 2013a. Order of Subject, Object and Verb. In: Dryer & Haspelmath (eds.) (http://wals.info/chapter/81, Accessed 2018-03-27)

Dryer, Matthew S. 2013b. Determining Dominant Word Order. In: Dryer & Haspelmath (eds.) (Accessed 2018-04-01)

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. 2019. Grammaticalization accounts of word order correlations. In Schmidtke-Bode, Karsten, Natalia Levshina, Susanne Maria Michaelis & Ilja A.Seržant (eds.). Explanation in typology. 63-95. Berlin: Language Science Press.

Dryer, Matthew S. & Haspelmath, Martin. 2013. (eds.) The World Atlas of Language Structures online. (http://wals.info).

Edmonson, Barbara Wedemeyer. 1988. *A Descriptive Grammar of Huastec* (*Potosino Dialect*). Ph.D. dissertation. New Orleans: Tulane University.

England, Nora C. 1991. Changes in basic word order in Mayan languages. *International Journal of American Linguistics* 57(4): 446-486.

Erdmann, Oskar. 1886. *Grundzüge der deutschen Syntax nach ihrer geschichtlichen Entwicklung*. Erste Abt. Stuttgart: Cotta. (Reprint 1985. Hildesheim: Georg Olms Verlag).

Fabricius-Hansen, Cathrine, & Kåre Solfjeld. 1994. Deutsche und norwegische Sachprosa im Vergleich: Ein Arbeitsbericht [Comparing German and Norwegian nonfiction]. *Arbeitsberichte des Germanistischen Instituts 6*. Oslo: Olso University.

Farhat, Haytham 1991. *Empty categories in Syrian Arabic*. Unpubl. PhD-Dissertation. Bangor: Univ. of Wales.

Fehn, Anne-Maria. 2015. *A Grammar of Ts'ixa* (*Kalahari Khoe*). Doctoral dissertation. Cologne: Universität zu Köln.

Franchetto, Bruna. 1990. A ergatividade Kuikúro (Karíbe): algunas propostas de análise. *Cadernos de Estudos Lingüísticos* 18: 57-77.

Gell-Mann, Murray and Merritt Ruhlen. 2011. The origin and evolution of word order. *PNAS* 108 (42): 17290-17295. https://doi.org/10.1073/pnas.1113716108

Goldberg, Adele. 2002. Persian Auxiliaries. Ms. Princeton University. https://www.princeton.edu/~adele/papers/Papers (accessed 2018-04-27)

Goldin-Meadow, Susan, Wing Chee So, Aslı Özyürek & Carolyn Mylander. 2008. The natural order of events: How speakers of different languages represent events nonverbally. *PNAS* 105 (27): 9163-9168. https://doi.org/10.1073/pnas.0710060105

Greenberg, Joseph. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (ed.) *Universals of Language*. 73-113. Cambridge, MA.: MIT Press.

Haider, Hubert. 2010. *The Syntax of German*. Cambridge: Cambridge University Press.

Haider, Hubert. 2013. *Symmetry breaking in Syntax*. Cambridge: Cambridge Univ. Press.

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. 2020. Grammar change – a case of Darwinian cognitive evolution. Target paper (in press), *ELT – Evolutionary linguistic theory*. [lingbuzz/005134]

Haider, Hubert. 2021a. *"OVS" – A misnomer for SVO languages with ergative alignment*. Ms. *lingbuzz*/005680.

Haider, Hubert. 2021b. 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. [*lingbuzz/004970*].

Haider, Hubert and Luka Szucsich (2021). *Slavic languages – "SVO" languages without SVO qualities*? (in press in) *Theoretical linguistics*. [*lingbuzz/004973*].

Hale, Kenneth 1983. Warlpiri and the grammar of non-configurational languages. *Natural Language and Linguistic Theory* 1, 5-47.

Haspelmath, Martin. 2014. Comparative syntax. In Carnie, Andrew, Yosuke Sato, and Dan Siddiqi (eds.) *The Routledge handbook of syntax*. 490-508. London: Routledge.

Hawkins, John A. 1983. *Word order universals*. New York: Academic Press.

Hawkins, John A. 1990. A parsing theory of word order universals. *Linguistic Inquiry* 21(2): 223-261.

Hawkins, John A. 2014. *Cross-linguistic variation and efficiency*. Oxford University Press

Hentschel, Gerd. 2014. Belarusian and Russian in the mixed speech of Belarus. In Besters-Dilger, Juliane, Cynthia Dermarkar, Stefan Pränder & Achim Rabus (eds.), *Conguence in contact-induced language change*, 93-121. Berlin: De Gruyter.

Jelinek, Eloise, and Richard Demers 1994. Predicates and pronominal arguments in Straits Salish. *Language* 70: 697–736.

Jensen, Cheryl. 1997. *Coreferential marking in Tupí-Guaraní languages*. Paper presented at the XIII International Conference of Historical Linguistics. Düsseldorf, August 10-15.

Jones, Wendell & Jones, Paul. 1991. *Barasano Syntax*. Studies in the languages of Colombia 2. Summer Institute of Linguistics & University of Texas at Arlington.

Kalin, Laura. 2014.The syntax of OVS word order in Hixkaryana. *Natural Language & Linguistic Theory* 32: 1089–1104. DOI 10.1007/s11049-014-9244-x

Koch, Karsten & Matthewson, Lisa. 2009. The lexical category debate in Salish and its relevance for Tagalog. *Theoretical Linguistics* (35,1): 125–137.

Köhler, Oswin. 1981. La langue kxoe. In Perrot, Jean (ed.) *Les langues dans le monde ancien et moderne*. 483-555. Paris: Centre National de la Recherche Scientifique.

Koptjevskaja-Tamm, Maria. 1993. *Nominalizations*. London: 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.

Liakin, Denis & Juvénal Ndayiragije. 2001. New look at an old problem: Multiple Wh-Fronting in Slavic. In Gerhild Zybatow, U. Junghanns, G. Mehlhorn & L. (eds.), *Current Issues in Formal Slavic Linguistics*. 206-213. Frankfurt/Main: Lang.

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.

Mayo, Peter. 1993. Belorussian. In Comrie, Bernard and Corbett, Greville G. (eds.), *The Slavonic Languages*, 887-946. London: Routledge.

Mel'čuk, Igor, A. 2014. Syntactic Subject: Syntactic Relations, Once Again. In Plungjan, V. with Danièl, M., Ljutikova, E., Tatevosov, S. Fedorova, O. (eds), (eds.) *Jazyk. Konstanty, Peremennye. Pamjati*. 169–216. Sankt-Petersburg: Aletejja.

Mithun, Marianne. 2005. Ergativity and language contact on the Oregon Coast: Alsea, Siuslaw and Coos. In Andrew K. Simpson (ed.), *Proceedings of the Twenty-Sixth Meeting of the Berkeley Linguistics Society*. 77-95. Berkeley: Berkeley Linguistics Society.

Muysken, Pieter, Harald Hammarström, Olga Krasnoukhova, Neele Müller, Joshua Birchall, Simon van de Kerke, Loretta O'Connor, Swintha Danielsen, Rik van Gijn & George Saad. 2016. *South American Indigenous Language Structures (SAILS) Online*. Jena: Max Planck Institute for the Science of Human History. (https://sails.clld.org).

Olawsky, Knut J. 2006. *A Grammar of Urarina*. (Mouton Grammar Library, 37) Berlin: Mouton de Gruyter.

Olawsky, Knut J. 2007. ObViouS OVS in Urarina syntax. In: Simpson, A. & P. Austin (eds.), *Endangered languages*. Special volume of *Linguistische Berichte*. 45-72. Hamburg: Buske Verlag.

Polański, Kazimierz 2003. *Encyklopedia Językoznawstwa Ogólnego*. Wrocław: Ossolineum.

Primus, Beatrice. 1995. Relational typology. In Jacobs, Joachim & Arnim von Stechow (eds.) *Syntax. Ein internationales Handbuch zeitgenössischer Forschung*. 2nd vol. 1076-1109. Berlin: De Gruyter.

Queixalós, Francesc & Gildea, Spike. 2010. Manifestations of ergativity in Amazonia. In Gildea, Spike & Queixalós, Francesc (eds.) *Ergativity in Amazonia*. 1-25. Amsterdam: Benjamins.

Quirk, Randolph, Sidney Greenbaum, Geoffrey N. Leech and Jan Svartvik (1986[4]) *A comprehensive grammar of the English language*, London, Longman.

Rill, Justin. 2017. *The morphology and syntax of ergativity: a typological approach.* Newark: University of Delaware PhD dissertation.

Rivière, Nicole 1981. *La construction impersonnelle en français contemporain*. Documents de Linguistique Quantitative no. 41, St. Sulpice-de-Favière: Éditions Jean-Favard.

Rojas-Berscia, Luis Miguel. 2014. *A Heritage Reference Grammar of Selk'nam*. Nijmegen: Radboud University. Master thesis.

Rumsey, Alan. 1982. An intra-sentence grammar of Ungarinjin. *Pacific Linguistics*. Series b - no. 86.

Schürcks, Lilia. 2009. *Binding and discourse: where syntax and pragmatics meet.* Frankfurt: Lang.

Siewierska, Anna. 1993. Syntactic weight vs. pragmatic factors and word order variation in Polish. *Journal of Linguistics* 29: 233-265

Siewierska, Anna / Uhliřová, Ludmila 2010. An overview of word order in Slavic languages. In: Siewierska, Anna (ed.) *Constituent order in the languages of Europe*. 105-150. Berlin: Mouton DeGruyter.

Steele, Susan 1978. Word order variation: a typological study. In Joseph H. Greenberg, Charles A. Ferguson and Edith A. Moravcsik (eds.) *Universals of Human Language: IV: Syntax*. 585–623. Stanford: Stanford University Press.

Stepanov, Arthur 1997. On wh-fronting in Russian. In Pius N. Tamanji & Kiyomi Kusumoto (eds.) *Proceedings of NELS* 28: 453-467. Amherst: GLSA, University of Massachusetts.

Szucsich, Luka & Haider, Hubert. 2015. Freie Wortstellung in slavischen Sprachen und die VO/OV-Unterscheidung. In Elena Dieser (ed.) *Linguistische Beiträge zur Slavistik: XX. JungslavistInnen-Treffen* 2011. 94-124. München: Verlag Otto Sagner.

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.

Toman, Jindřich. 1981. Aspects of multiple wh-movement in Polish and Czech, in: Robert May & Jan Koster (eds.) *Levels of Syntactic Representation*. 293-302. Dordrecht: Foris.

Vikner, Sten. 1995. *Verb movement and expletive subjects in the Germanic Languages*. Oxford: Oxford University Press.

Wiland, Bartosz 2020. Overt evidence from left-branch extraction in Polish for punctuated paths. *Linguistic Inquiry* 41(2): 335-347.

Zabrocki, Tadeusz 2016. Syntactic diacrisis in a rigid and a free word order language. *Investigationes Linguisticae* 34.113-150. Poznań: Adam Mickiewicz University.

Zwart, Jan-Wouter & Lindenbergh, Charlotte. (in press). Rethinking alignment typology. In Bárány, András, Theresa Biberauer, Jamie Douglas & Sten Vikner (eds.) *Inside syntax III: Syntactic architecture and its consequences*. Berlin: Language Science Press.