

# Subjects in Word Order and Alignment Typology

Nina Adam & Andreas Hölzl\*

## Abstract

Following a research agenda proposed by Gisbert Fanselow, this study discusses word order and argument marking properties in a small sample of 16 languages. A focus lies on a possible correlation between the basic order of verb and object and a high obligatory subject position. The study shows that the correlation proposed by Haider is not valid and that his recent redefinition of the term “subject” is not applicable to the languages of the world.

## 1. Introduction

Cross-linguistic correlations with the basic order of object and verb (VO or OV) have been discussed since Greenberg’s (1963) observations regarding the consistent ordering of head and dependent across different categories within a language. Subsequent research has found that also beyond such harmony statements, certain morphosyntactic properties of languages might be dependent on the basic order of object and verb; for an overview, see Dryer (2013a) or Biberauer and Sheehan (2013).

For syntactic theory, the crucial question regarding such statements is whether there are underlying structural factors that cause the correlation between two properties. Whilst this is obvious for the head-dependent harmony mentioned in the beginning, it is far less trivial when discussing properties that go beyond such cross-categorial ordering patterns. A structural analysis is proposed by Haider (2013), who observes that VO languages like English display no scrambling, subject-object asymmetries for extraction, and further properties, because syntactic structures are universally right-branching.

---

\*This study is part of the project *Consequences of Head Argument Order for Syntax* (CHAOS, C08), originally granted to Prof. Gisbert Fanselow (1959–2022) by the German Research Foundation (DFG). It is part of the Collaborative Research Centre 1287 *Limits of Variability in Language: Cognitive, Computational, and Grammatical Aspects* at the University of Potsdam (Project-ID 317633480 – SFB 1287). We would like to thank Gereon Müller, Martin Salzmann, and all of the language consultants and collaborators.

This argument is largely based on a comparison of Germanic languages, but has also been extended to Slavic (Haider and Szucsich 2022).

Many of the properties of VO and OV languages that Haider (2013) and Haider and Szucsich (2022) discuss relate to the subject position. This is due to the fact that Haider assumes a bi-directional implication between VO base order and a high obligatory subject position (aka extended projection principle, EPP). Thus, every VO language has such a position (i.e., it is [+EPP]), whilst every OV language can leave the subject within the VP (i.e., it is [-EPP]). This is largely compatible with the Germanic data. However, Fanselow (2020) shows that a broader cross-linguistic investigation quickly leads to a collapse of the correlation. He therefore argues that the EPP value should be considered an independent macro-parameter, as given in (1).

- (1) “Languages differ as to whether the subject must occupy the specifier position of a functional projection.” (Fanselow 2020: 9)

He thereby follows a similar line of argument as found in the typological literature. The traditional word order typology following Greenberg differentiates six different types, namely SVO, SOV, OSV, OVS, VSO, and VOS. However, Dryer has repeatedly argued for an alternative typology based on two separate variables, the relative position of S and V as well as of O and V (e.g., 2013a). This new approach is based on the observation that cross-linguistically certain features do not correlate with any of the six types above but rather with either OV/VO or SV/VS. A third variable, the relative position of subject and object, turned out to be less relevant for cross-linguistic predictions, see Table (1). Among the predictions is that languages with, for example, SOV, OSV, and OVS order according to the traditional typology will pattern alike in certain cases because they share the OV feature. For instance, the order of adposition and noun phrase correlates with OV/VO, and it does so relatively independently of the position of the subject: languages with OV order usually have postpositions while languages with VO order have prepositions (e.g., Dryer 2011: 337).

Based on a preliminary cross-linguistic convenience sample of 16 languages, given in Table 2, the present study addresses the structural position of the subject and its relation to the OV/VO typology. The data were collected with a questionnaire in the broader context of the ongoing project *Consequences of Head Argument Order for Syntax*, which was devised and designed by Gisbert

	SO	OS	
SV & OV	SOV	OSV	verb-final
SV & VO	SVO	-	(verb-medial)
VS & OV	-	OVS	(verb-medial)
VS & VO	VSO	VOS	verb-initial

Table 1: A simplified comparison of two word order typologies (based on Dryer 2013a: 270)

Fanselow. For the purpose of assessing the general validity of the correlation between VO/OV and the EPP, we control for the basic position of main verb and object (VO/OV), the alignment type (accusative/ergative/neutral/tripartite), and the availability of subject drop in the language. The presence of a high obligatory subject position (EPP) is tested for through the (non-)availability of low subjects and subjectless clauses. We focus on these surface properties in order to permit a first classification of typologically diverse languages, to many of which elaborate syntactic tests cannot be applied.

Family	Language	Family	Language
Afroasiatic (Cushitic)	Oromo	IE (Slavic)	Czech
Afroasiatic (Semitic)	Amharic	IE (Slavic)	U. Sorbian
Atlantic Congo (Gur)	Bwamu	IE (Italic)	Italian
Austroasiatic (Vietic)	Vietnamese	Kra-Dai (Tai)	Thai
IE (Germanic)	English	Sino-Tibetan (Sinitic)	Huarong
IE (Germanic)	German	Turkic (Kipchak)	Kazakh
IE (Indo-Aryan)	Marathi	Turkic (Oghuz)	Gagauz
IE (Indo-Aryan)	Nepali	Turkic (Oghuz)	Turkish

Table 2: Preliminary sample of 16 languages (IE = Indo-European)

We first address the different alignment systems found among the languages of our sample in Section 2 and then, in Section 3, present the findings for languages with dominantly accusative or neutral alignment, where the results can be directly compared to Haider’s (2013) claim. Since our sample also includes languages that show ergative alignment patterns, we describe the patterns observed there and critically discuss the notion of the term “subject” for these languages in Section 4, before concluding with Section 5.

## 2. Argument Marking

The 16 languages of our sample employ a wide range of different argument encoding strategies. Since the identity of the core arguments S and O is of central importance to the present study, this section provides a general overview of the marking of arguments. This can be accomplished by different means, including case marking and verb agreement. Newer typological approaches replace the latter two terms with *flagging* (including case and adpositions) and (*verbal*) *indexing* (or *indexation*), respectively (e.g., Croft 2022). For instance, the following examples from Oromo contain a nominative flag on the subject and a verb index referring to the same argument.

### (2) Oromo

- a. abbaa-**n**=koo                      kaleessa    k'ak'k'ab-**e**.  
 father-NOM=1SG.POSS yesterday arrive-3SG.M.PFV  
 'My father arrived yesterday.'
- b. abbaa-**n**=koo                      [konkolaataa haaraa] bit-**e**.  
 father-NOM=1SG.POSS car                      new              buy-3SG.M.PFV  
 'My father bought a new car.'

A key observation in cross-linguistic comparison is that languages not only differ in the type of marking, but also in which arguments are marked in the same way (e.g., Bickel 2011; Croft 2022). The most prominent marking patterns involve the core arguments of intransitive (called S argument) and transitive clauses (called A and O arguments). S is usually defined as the single argument in an intransitive clause and A and O are the most agent-like and other argument in a transitive clause. In Oromo, the S and A arguments are marked with the same flag and index. Such a pattern is referred to as accusative alignment. Oromo represents a special case because of the marked nominative and the unmarked O argument (A=S;O). In other languages like Turkish, the O argument is overtly marked with an accusative instead (which is only present if the O argument is definite or specific; e.g., *kitab* 'a book', *kitab-ı* 'the book (ACC)'). The S and A arguments remain without flag but can be represented in the verb with the same index.

Bwamu has neutral alignment without indexing or flagging (A=S=O), except for the pronominal system, which exhibits an accusative-aligned flagging system similar to English for the singular and the second person plural (e.g., *ĩ* '1SG', *mi* '1SG.ACC'). Chinese varieties (e.g., Huarong) likewise exhibit

no marking (neutral alignment) but can have a special prepositional marker *pa*<sup>21</sup> (also a verb today meaning ‘to give’) for the O under certain conditions (accusative alignment). In this language, the differential O marking (DOM) entails a shift from VO to OV.

In some languages it is not the A argument but the O argument that patterns with the single argument in intransitive clauses (A;S=O). This phenomenon is known as ergative behavior. However, many languages of this type also exhibit additional splits that can sometimes also lead to what is called a tripartite system in which S, A, and O are all marked differently. Such a mixed system can be found in the Indo-Aryan language Marathi (e.g., Pandharipande 1997; Dhongde and Wali 2009). Marathi has flagging and indexing on the verb, but several features represent a challenge for the alignment typology. Marathi has an unmarked nominative and a marked oblique stem (also marking gender) for overtly marked flags like the instrumental-ergative (SG *-ne*, PL *-nī*) and the dative-accusative (SG *-lā*, PL *-nā*), e.g., male singular *mitra* ‘friend (NOM)’, *mitrā-ne* ‘friend (INST-ERG)’, *mitrā-lā* ‘friend (DAT-ACC)’ (Pandharipande 1997: 273-277). S arguments are always in the unmarked nominative. A arguments are either in the nominative or, in perfective clauses, in the ergative.

### (3) Marathi

- a. [to māṇus] gāḍī wika-t āhe.  
that.M man[M] car[F] sell-IPFV be.PRS.3SG  
‘That man is selling a car.’
- b. [tyā māṅsā-**ne**] gāḍī wik-1-ī.  
that.M.OBL man[M]-ERG car[F] sell-PFV-3SG.F  
‘That man sold the car.’

O arguments can remain unflagged (3) but are more likely to be marked with the dative-accusative with animate and specific nouns (4). The indexing system (involving person, number, and gender) is intricately linked to this pattern and appears to jump from argument to argument. Usually, the unflagged argument is indexed on the verb. This can be both the O argument (3-b) or the A argument (4-a). In case both arguments are marked, flagging resorts to neutral indexing referring to neither argument (4-b).

## (4) Marathi

- a. *sāyalī anuj-lā mār-t-e.*  
 Sayali[F] Anuj[M]-ACC hit-IPFV-3SG.F  
 ‘Sayali hits Anuj.’
- b. *sāyalī-ne anuj-lā mār-l-a.*  
 Sayali[F]-ERG Anuj[M]-ACC hit-PFV-3SG.N  
 ‘Sayali hits Anuj.’

Languages like Bwamu, Huarong, and Marathi illustrate that alignment is not a global feature of an entire language but rather item- or construction-specific (Bickel 2011; Croft 2022). For the mere description of the alignment system, the terms “subject” and “object” are thus unnecessary. But for the purpose of word order and syntactic analysis, the terms continue to be important. Given the fact that the alignment typology operates with three categories (S, A, O) but the other fields with only two (subject, object), there is a disagreement about the mapping of the two sets that will be discussed in the following.

### 3. Correlations within the Sample

We now turn to the investigation of Haider’s prediction that VO correlates with a high obligatory subject position. The dependent variables we investigate are:

1. Subject height – Can subjects appear in the verb phrase in an information structurally neutral context?
2. Subject obligatoriness – Can the subject position remain empty when the subject is displaced or missing?

Regarding subject height, we use psych verbs as the critical test case because their thematic structure makes the availability of low subjects most likely (Temme 2018). Regarding subject height, expletive subjects are a strong indicator of the obligatoriness of the high subject position. A potential confounding factor is the possibility to drop unstressed subject pronouns, therefore we control for this feature as well; see below for details.

Following Haider, we would expect that in VO languages, low subjects are not possible, and the subject position is obligatorily filled. In contrast, OV languages should permit low subjects, and the subject is not obligatory. The results for our sample can be found in Table 3; cells where Haider’s prediction is not borne out are backgrounded in grey. Since the notion of subject is

especially debatable for the two languages with partially ergative alignment, we will discuss these separately in Section 4.

Language	word order	Alignment	Low S	Oblig. S	S-drop
Amharic	OV	acc	–	–	+
Bwamu	VO	neu/acc	–	+	–
Czech	VO	acc	+	–	+
English	VO	neu/acc	–	+	–
Gagauz	VO/OV	acc	+	–	+
German	OV	acc	+	–	–
Huarong	VO	neu/acc	+	?	+
Italian	VO	acc	+	–	+
Kazakh	OV	acc	+	–	+
Marathi	OV	acc/erg/tri	+	–	–
Nepali	OV	acc/erg/tri	+	–	–
Oromo	OV	acc	+	–	+
Thai	VO	neu	–	+	+
Turkish	OV	acc	+	–	+
U. Sorbian	OV	acc	+	–	+/-
Vietnamese	VO	neu	+	–	+

Table 3: Results for our preliminary sample of 16 languages

We find that there are a four “untypical” VO languages in our sample: Czech, Huarong, Italian, and Vietnamese. An example for low subject placement in a VO language can be found in (5).

(5) Czech

- a. Přišel                    **můj**                    **otec.**  
 arrive.PTCP.M my.M.NOM father.NOM  
 ‘My father arrived.’
- b. Mému    otci            se                    líbí                    **auto.**  
 my.M.DAT father.DAT REFL.ACC please.PRS.3SG car.NOM  
 ‘The car pleases my father.’

Within the group of OV languages, only Amharic displays unexpected subject behaviour. Amharic and Huarong additionally show that the two subject-related features height and obligatoriness need not correlate: a language without an

obligatory subject position can nevertheless ban low subjects (Amharic); and a language that permits low subjects can nevertheless have an obligatory subject (possibly Huarong).

The investigation of subject obligatoriness is complicated by so-called “pro-drop languages”, for our purposes: languages which allow for subject drop. Such languages might have an obligatory functional subject position that can be occupied by *pro*, which means that apparently subjectless sentences might have a covert subject after all. However, if Haider (2019) is correct in his assumption that there are no covert expletives, impersonal passives can serve as a test: here, valency reduction leads to the dropping of the subject, as shown for Amharic in (6). Thus, the languages listed as having subject drop in Table 3 only received a “–” for subject obligatoriness if they allow for impersonal passives without an expletive.

(6) Amharic

mata bīzu tə-tʃ'əffər-ə.

last.night much PASS-dance.IPFV-3SM.SUB

‘Last night, there was much dancing.’

Some languages require further comments. For instance, Upper Sorbian, a Slavic minority language spoken in the east of Germany, allows both VO and OV; OV is preferred in embedded clauses. We classify it as OV here because in VP topicalisation, only OV order is possible. Yet, further investigation is necessary. As Table 3 shows, Upper Sorbian generally permits low subjects. Regarding subject-drop, however, standard and colloquial language must be distinguished: the former has subject-drop, like its direct relative Czech, whilst the latter is more similar to German in not having subject-drop and using an expletive-like element in some subjectless constructions. Both varieties allow for subjectless passives.

One language that could not be classified with regard to the VO/OV distinction is Gagauz. According to our consultant, it might be developing from OV towards VO, with older speakers preferring the former, younger speakers the latter order. However, the patterns regarding the subject are just as in Turkish. This might be a further indicator that the EPP is independent of verb-object order – if VO entailed a high obligatory subject position, these two properties would have to change simultaneously.



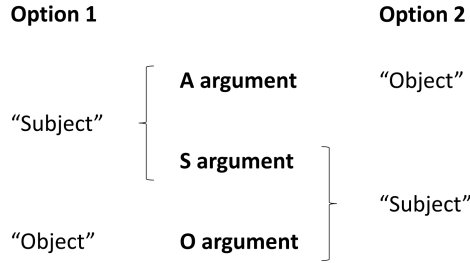


Figure 1: Possible uses of the terms “subject” and “object”

#### 4. Ergative and Tripartite Alignment

In his word order typology of transitive clauses, Dryer (2013*b*) explicitly defines the notions of “subject” and “object” semantically as referring to the agent (i.e., A argument) and patient (i.e., O), respectively. In an unpublished manuscript, Haider (2023) has recently proposed a redefinition of the terms “subject” and “object” in purely morphosyntactic terms and in relation to different alignment systems. According to him, the S argument is the defining category: if an A argument is marked the same then this should be considered the subject for that particular language; but if the O argument marked the same then this would qualify as the subject instead. This means that the term “subject” would not be universal and relative to the alignment type (either A=S or S=O). A consequence for this approach would be that the word order for several languages would have to be reclassified. Instead of SOV word order, languages with ergative alignment would exhibit OSV order, etc. Thus, Haider argues for a more flexible classification depending on the alignment type as shown in Figure 1: Option 1 for languages with accusative alignment and Option 2 for languages with ergative alignment. However, the approach runs into major problems, only some of which will be briefly addressed here.

As shown in Section 2, the global classification as “ergative language” is highly problematic. A language like Marathi can have several different alignment systems simultaneously. In terms of flagging, Marathi has neutral (all unmarked), accusative (only O is marked), ergative (only A is marked), or tripartite alignment (S unmarked, A marked, O marked differently).

However, according to the definition, Marathi ergative-marked A arguments

should clearly qualify as “object” if the O argument is unmarked. Haider argues that the “subject” should be the omissible argument in both passive (A) and antipassive constructions (O). However, independent of the alignment type, it is the A argument and not the O>S argument that can be left out in a passive sentence in Marathi. Passives are periphrastically marked with the verb ‘to go’.

## (7) Marathi

- a. to **gāḍī** tsālawa-t āhe.  
3SG.M car[F] drive-IPFV be.PRS.3SG  
‘He is driving a car.’
- b. **gāḍī** (tyā-tS-yā-kaḍun) tsālaw-l-ī dzā-t  
car[F] 3SG-GEN-OBL-from drive-PFV-3SG.F go-IPFV  
āhe.  
be.PRS.3SG  
‘The car is being driven (by him).’

## (8) Marathi

- a. tyā-ne **gāḍī** tsālaw-l-ī  
3SG.M-ERG car[F] sell-PFV-3SG.F  
‘He drove the car.’
- b. **gāḍī** (tyā-tS-yā-kaḍun) tsālaw-l-ī ge-l-ī.  
car[F] 3SG-GEN-OBL-from drive-PFV-3SG.F go-PFV-3SG.F  
‘The car got driven (by him).’

Furthermore, Haider argues that languages with tripartite alignment (like Marathi) or with an active/stative system (the S being either marked as A or O) do not possess a “subject”, among other things because these are missing “the grammatical management of subject omission” (Haider 2023: 7). Apart from the fact that Marathi certainly has such a system, as a consequence the word order of these languages could no longer be straightforwardly compared with those that do have a “subject”. For these and similar reasons, we follow Dryer in using Option 1 in Figure 1. Thus, we treat Marathi and Nepali in parallel with the other languages and classify them as OV languages that are [-EPP], i.e., where the subject can appear low and is not obligatory.

## 5. Conclusion

Through a cross-linguistic comparison, this paper has shown that there is no direct correlation between the basic order of object and verb and the EPP, i.e., the height and obligatoriness of the subject. In consequence, as proposed by Dryer (2013*a*) and Fanselow (2020), the language sample must be divided into four groups instead of just two, as shown in Table 4. The cells for VO languages without the EPP and OV languages with the EPP are the ones that are predicted to be empty by Haider (2013) and later publications. Whilst in our sample, Amharic is the only OV language that potentially breaks the pattern, Pregla (2023) provides a thorough analysis of Uralic and other languages that also belong to this group.

	VO	OV
[−EPP]	Czech, (Huarong), Italian, Vietnamese	German, Kazakh, Marathi, Nepali Oromo, Turkish, U. Sorbian
[+EPP]	Bwamu, English, Thai	(Amharic)

Table 4: Classification of investigated languages into four groups. Languages in brackets are not uniform in their subject properties.

The findings for Amharic and perhaps Huarong are particularly interesting. They suggest that different subject-related features that are discussed under the umbrella of the EPP do not necessarily correlate: subject height and subject obligatoriness can be independent of one another. However, it should be noted that these are preliminary results that are part of an ongoing investigation. Therefore, as proposed by Fanselow (2020), the project continues to investigate these and other subject-related properties in a growing sample of languages in order to test their relation to verb-object order as well as the general homogeneity of features attributed to the EPP.

## References

- Biberauer, Theresa and Michelle Sheehan (2013): Introduction. In: T. Biberauer and M. Sheehan, eds., *Theoretical approaches to disharmonic word order*. Oxford University Press, Oxford, p. 1–41.
- Bickel, Balthasar (2011): Grammatical relations typology. In: J. J. Song, ed., *The Oxford handbook of linguistic typology*. Oxford University Press, Oxford, p. 399–444.

- Croft, William (2022): *Morphosyntax*. Cambridge University Press, Cambridge.
- Dhongde, Ramesh Vaman and Kashi Wali (2009): *Marathi*. John Benjamins, Amsterdam.
- Dryer, Matthew S. (2011): ‘The evidence for word order correlations’, *Linguistic Typology* **15**(2), 335–380.
- Dryer, Matthew S. (2013a): ‘On the six-way word order typology, again’, *Studies in Language* **37**(2), 267–301.
- Dryer, Matthew S. (2013b): Order of subject, object and verb. In: M. S. Dryer and M. Haspelmath, eds., *WALS Online (v2020.3)*. Max Planck Institute for Evolutionary Anthropology, Leipzig.  
**URL:** <http://wals.info/chapter/81>
- Fanselow, Gisbert (2020): Is the OV-VO distinction due to a macroparameter?. In: M. Tanaka, T. Tsutsui and M. Hashimoto, eds., *Linguistic research as an interdisciplinary science*. 2nd edn, Hitsuji Publishers, Tokyo, p. 1–26.
- Greenberg, Joseph Harold (1963): Some universals of grammar with particular reference to the order of meaningful elements. In: J. H. Greenberg, ed., *Universals of language*. Cambridge MIT Press, p. 73–113.
- Haider, Hubert (2013): *Symmetry breaking in syntax*. Cambridge University Press, Cambridge.
- Haider, Hubert (2019): On absent, expletive, and non-referential subjects. In: P. Herbeck, B. Pöll and A. Wolfsgruber, eds., *Semantic and syntactic aspects of impersonality*. Helmut Buske Verlag, pp. 11–46.
- Haider, Hubert (2023): ‘“OVS” – A misnomer for SVO languages with ergative alignment’. (Unpublished manuscript.).
- Haider, Hubert and Luka Szucsich (2022): ‘Slavic languages – “SVO” languages without SVO qualities?’, *Theoretical Linguistics* **48**(1–2), 1–39.
- Pandharipande, Rajeshwari V. (1997): *Marathi*. Routledge, London.
- Pregla, Andreas (2023): Word order variability in OV languages. A study on scrambling, verb movement, and postverbal elements with a focus on Uralic Languages. PhD thesis, Universität Potsdam.
- Temme, Anne (2018): The peculiar nature of psych verbs and experiencer object structures. PhD thesis, Humboldt Universität zu Berlin.