Eliciting Data for the OV/VO Alternation

Andreas Hölzl
Nina Adam
Andreas Pregla
University of Potsdam

24th January 2024

Abstract

This descriptive paper presents an ongoing process of data collection for the cross-linguistic documentation and study of syntactic phenomena using a specifically designed questionnaire. The study is part of the project *Consequences of head argument order for syntax* (CHAOS), which is part of the Collaborative Research Centre *Limits of Variability in Language: Cognitive, Computational, and Grammatical Aspects* at the University of Potsdam, funded by the German Research Association (DFG). The overall goal is to collect high-quality data for languages from around the world and to use this data to test hypotheses of cross-linguistic correlations with the relative order of transitive verbs (V) and their objects (O). The study describes the methodological background and the questionnaire. It illustrates the data collection process with the help of data from selected languages, including Chinese, Czech, Marathi, Nepali, Oromo, Upper Sorbian, and Tunen.

Keywords
word order correlations; head-directionality; data collection; elicitation; questionnaire

1 Introduction

Word order typology has found several robust correlations between the order of different types of head and their dependents. Among the most widely discussed features are those that correlate with the serialization of a transitive verb (V) and its object (O). Many of the correlating features are instances of a harmony in the serialization of head and dependent, such as the order of adpositions (head) and noun phrases (dependent). The present study, however, which is part of a project entitled *Consequences of head argument order for syntax* (CHAOS), focuses on features that are known or suspected to correlate with OV/VO but go *beyond the linearization of head and dependent*. An example for this is the position of oblique elements (X) such as adverbial nominal or adpositional phrases in a transitive clause. Concerning this feature, OV order shows more flexibility than VO, with the latter excluding, among other things, VXO as dominant order, which cannot be described in terms of head and dependent ordering (Dryer & Gensler 2013). The aim is to collect high-quality crosslinguistic data that can be used to test whether there are such systematic correlations, analyze the type and strength of these correlations and investigate whether there could be alternative explanations, including the structural position of the subject within the clause (Fanselow 2020). As such, the project is a contribution to both language documentation and to the cross-linguistic study of syntactic phenomena. This paper is mostly descriptive in nature and presents the methodological background, some of the features under investigation (e.g., the position of obliques), the language samples, the questionnaire used for the elicitation of the data, and some preliminary interpretations and the coding of the data.

The paper is structured as follows. Section 2 illustrates different types of cross-linguistic correlations with the order of verb and object. Section 3 presents a few examples of the features investigated in the project, and Section 4 sketches the selection of the language sample. Section 5 describes the data collection procedure with the help of a questionnaire, which is introduced in Section 6. The following Section 7 illustrates the application of the questionnaire with data from Nepali. Section 8 sketches some preliminary ideas for a database that contains all the collected information, and Section 9 presents possible theoretical conclusions that result from the data. Section 10 presents a concluding summary. The Appendix lists the references for each of the different features investigated in the project; under Supplementary Material, a link to the full questionnaire used for the data elicitation can be found.

2 Cross-linguistic Correlations

In this section, we describe the types of correlations with verb-object order that are the focus of our research. The following subsections address the harmony in head-dependent ordering which has already been investigated with large language samples (Section 2.1) as well as features that are attributed to other structural properties of VO and OV languages, respectively, and for which broad cross-linguistic studies are lacking (Section 2.2.).

2.1 Order of Head and Dependent

Previous research concerning linear word order identified two different types of correlations: bidirectional correlations, represented by adpositions (Adp), and unidirectional ones, represented by relative clauses (Rel) (Dryer 2007). These are two well-known cases that can be described in terms of cross-categorial harmony of head and dependent.

First, languages with OV order usually have adpositions that follow the noun phrase, while those with VO have prepositions. In this bidirectional correlation, both features predict each other (Dryer 2007: 89):

(1) if OV then NPAdp and if NPAdp then OV

(2) if VO then AdpPN and if AdpNP then VO

Table 1 shows the relevant numbers in a global sample. There is a strong tendency towards harmony of head-initial (VO & AdpN) and head-final types (OV & NAdp). The few exceptions from the correlations indicate that these should be considered statistical tendencies rather than absolute universals.

Table 1: Matrix showing the order of OV/VO and AdpNP/NPAdp (Dryer 2013b, 2013c)

(other: 139)	OV	VO
AdpNP	14	454
NPAdp	472	41

A certain number of languages is not classifiable according to this simple typology. For instance, many languages with SVO order and prenominal genitives, such as Mandarin, have both pre- and postpositions (Dryer 2019).

Second, as opposed to adpositions, adnominal relative clauses only exhibit a unidirectional correlation with the order of the transitive verb and its object. In this case, only the following predictions are cross-linguistically valid (Dryer 2007: 97):

(3) if VO then NRel

(4) if RelN then OV

In contrast, OV cannot predict the relative order of relative clauses and noun phrases, and, knowing that a language has NRel order, one cannot accurately anticipate the order of objects and transitive verbs, although VO is much more common (Table 2).

Table 2: Matrix showing the order of OV/VO and RelN/NRel (Dryer 2013b, 2013e)

(other: 178)	OV	VO
NRel	113	415
RelN	132	5

Apart from a tendency towards harmony (VO & NRel, OV & RelN), there is a cross-linguistic preference for postnominal relative clauses (NRel) that is unrelated to the order of object and verb (Dryer 2011: 342). Therefore, the 113 languages displaying OV & NRel in Table 2 break the head-dependent harmony but are in line with the preference for relative clauses to appear after the noun.

There are only a few exceptions to these correlations, such as RelN & VO order in Sinitic and some surrounding languages (Comrie 2008; Sposato 2014).

(5) Mandarin (Sinitic, Trans-Himalayan)

a) gēge zuótiān mǎi le yì běn shū
e.brother yesterday buy PFV one CLF book
'The elder brother bought a book yesterday.' (elicited)

b) [gēge zuótiān mǎi de] shū
e.brother yesterday buy ATTR book
'the book the elder brother bought yesterday' (elicited)

Nevertheless, the correlations remain cross-linguistically valid as very strong tendencies. Other correlations, including the tendency for prenominal articles (ArtN) in VO and postnominal ones (NArt) in OV languages, are more difficult to detect (Dryer 2007: 94).

2.2 Features beyond Head-Dependent Order

Correlations that go beyond the order of head and dependent can be illustrated with the dominant position of oblique elements (X) with respect to O and V, where X is defined as a noun phrase or adpositional phrase used as adverbial modifier or adjunct (e.g., *with a key*). While the linear relation of X and V would instantiate just another case of head-dependent ordering, the relative position of X and O in this triplet cannot be broken down to a head-dependent relationship. As such, cross-categorial harmony alone would not suffice to explain potential correlations in this domain. Table 3 contains the numbers of a global sample. In this case, languages with OV and VO exhibit a very different behavior.

Table 3: Order of verb, object, and oblique (X) in a global sample (Dryer & Gensler 2013)

Type	Languages
VO**X**	210
no dominant order	167
XOV	48
OV**X**	45
O**X**V	27
XVO	3
V**X**O	-

As further discussed in Section 5, the data need to be treated with caution since the relative order of O and X is often unknown or inferred from information on OV/VO and XV/VX in grammatical descriptions, which leads to a biased sample with OVX being overrepresented. Another problem is the category of "no dominant order", which does not allow for further conclusions.

The dataset in Table 3 exhibits different patterns (Hawkins 2008). There are general tendencies for oblique elements to follow both V and O (VOX, OVX) and for obliques and objects to occur on the same side of the verb (VOX, XOV, OXV). If V stands between X and O, the order is usually OVX. XVO is rare and mostly found in Sinitic and some surrounding languages, such as Hmong-Mien (Sposato 2014). Most importantly, OV allows X to intervene between O and V (OXV), whereas intervening X is not found in languages with VO order (absence of VXO). When VXO order is attested, as for example in (6b), it is not the only or dominant order; in Czech, VOX as in (6a) is more common and perceived as less marked.

(6) Czech (Slavic, Indo-European)

a) Marie namalovala portrét **vodovými** **barvami**.
Marie paint.PTCP portrait.ACC water.ADJ.INS colours.INS

b) Marie namalovala **vodovými** **barvami** portrét.
Marie paint.PTCP water.ADJ.INS colours.INS portrait.ACC
'Mary painted a portrait with watercolors.' (elicited)

One type of pattern can be observed when focusing on the positioning of the oblique element: OV shows much more variability than VO (Table 4). While XOV, OVX, and OXV are all widely attested, VO languages usually have VOX order. A hyphen indicates the position of O and V relative to X.

Table 4: Matrix showing the order of OV/VO and obliques (X)

(other: 167)	OV	VO
-X	45	210
-X-	27	-
X-	48	3

This allows the establishment of unidirectional correlations similar to the relative clauses above:

(7) if VO then -X (VO**X**)

(8) if -X- or X- then OV (O**X**V, **X**OV)

However, the oblique element following both O and V cannot accurately predict the relative order of O and V because both VOX and OVX are widely attested (but see Section 5). Similarly, OV cannot predict the positioning of X because all three possibilities are equally common (**X**OV, O**X**V, OV**X**).

Another pattern is found when considering the relative order of X and O, exclusively, which compresses the dataset to four types as above (Table 5). This obscures the effects of intervening X, but puts into focus the relative order of two dependent elements, X and O.

Table 5: Matrix showing the order of OV/VO and OX/XO

(other: 167)	OV	VO
OX	72	210
XO	48	3

This illustrates an overlapping but slightly different unidirectional correlation, shown in (9) and (10). In contrast, OX and OV do not allow for clear predictions.

(9) if VO then OX (VO**X**)

(10) if XO then OV (**X**OV)

To summarize, while the relative order of noun and relative clause or of adposition and noun can be described in terms of head and dependent, the position of the oblique element within the clause or the relative order of X and O cannot. This is an instance of a correlation that goes beyond the ordering of head and dependent but arguably involves structural or other factors (Section 9). As in the case of relative clauses, Mandarin is one of the few exceptions to the cross-linguistic tendencies against X- & VO and XO & VO (see Section 3).

3 Features and Testable Predictions

The project collects data for over 30 potential features of this sort that are suspected to correlate with the OV/VO distinction, as well as additional background information. Only a subset can be illustrated here (see the Appendix and the Supplementary material for more details). From a methodological perspective, the features belong to at least three different types.

3.1 Features based on European Languages

The first set of features derives from theoretical considerations mostly based on evidence from Germanic, which shows an internal division into OV and VO (e.g., Haider 2013). This can be exemplified by the availability of adverbial intervention, that is, whether a language allows for adverbials to neutrally intervene between the lexical verb and the direct object. Stowell (1981) already notices that Dutch (OV with V2) allows for adverbials to intervene between O and V resulting in neutral OAdvV order, while in English, intervening adverbials, VAdvO, are only available as a marked option. Subsequent work followed up on the idea that the basic word order is the determinant factor due to an interplay of linearity and the licensing of nominal arguments.[1] The corresponding universal claim is given in (11) with a restriction to adverbs.

(11) "In VO, adverbs either precede the VP or follow the verb plus its nominal arguments, but they do not intervene." (Haider 2020: 349)

As a testable prediction about surface orders, this observation could take the following form:

(12) if -Adv- then OV

Unlike oblique elements, adverbial intervention has not been studied using frequency with the help of large cross-linguistic samples (but see Tomlin 1986: 84–88). Table 6 shows the numbers from two available medium-sized samples using dominant word order and adverbs. The first is an unpublished global sample of 94 languages provided by Matthew Dryer (p.c. 2022). The second is a sample of 70 creole languages in Haspelmath et al. (2013). The latter only includes frequency adverbs.

[1] For references, see Reuland & Kosmeijer (1988), Neeleman (1994), Corver & van Riemsdijk (1997), Saito & Fukui (1998), Haider (2013), Janke & Neeleman (2012), and Belk & Neeleman (2017).

Table 6: Order of verb, object, and adverb in two different samples (Dryer, p.c. 2022; Haspelmath et al. 2013); numbers in parentheses refer to languages that also exhibit other orders

Type	Global	Creoles
VO**Adv**	39	12 (39)
AdvOV	4	0 (7)
OV**Adv**	2	0 (1)
O**Adv**V	36	0 (7)
AdvVO	2	12 (36)
V**Adv**O	11	1 (16)

According to Dryer (p.c. 2022), VAdvO is overall dispreferred but still occurs as the dominant order in 11 languages of his sample, many of which are Austronesian (Zulu, Nkore-Kiga, Modern Greek, Khasi, Tukang Besi, Nuaulu, Tigak, Kara (Papua New Guinea), Lenakel, Woleaian, Boumaa Fijian). A few other languages are said to have either VOAdv or VAdvO, depending on the type of adverb. Haspelmath et al. (2013) mention one additional language with VAdvO order for frequency adverbs (Reunion Creole) while 16 more languages exhibit this as well as other word order types. Thus, VAdvO is less strongly dispreferred than VXO (Dryer, p.c. 2022). The data, although limited, are sufficient to show that the prediction in (12) needs to be qualified but could well remain valid as a statistical tendency.

As in the case of the oblique elements (Section 2), the data can be interpreted in different ways. Tables 7 and 8 focus on the position of the adverb (six types) and the relative order of adverb and noun (four types), respectively.

Table 7: Matrix showing the order of OV/VO and adverbs (Global / Creoles)

	OV	VO
-Adv	2 / 0 (1)	39 / 12 (39)
-Adv-	36 / 0 (7)	11 / 1 (16)
Adv-	4 / 0 (7)	2 / 12 (36)

Table 8: Matrix showing the order of OV/VO and OAdv/AdvO (Global / Creoles)

	OV	VO
OAdv	38 / 0 (8)	39 / 12 (39)
AdvO	4 / 0 (7)	13 / 13 (52)

Given the limited data, few conclusions can be drawn on possible correlations. Since the majority of the creole languages have SVO order and are only found in specific geographical areas, the picture is heavily biased. Nevertheless, the preliminary samples indicate the following testable unidirectional correlations:

(13) if OV then -Adv- / OAdv

(14) if -Adv then VO

(15) if AdvO / Adv- then VO

When investigating adverbial intervention, an important methodological difference comes into play: Dryer determines the position of the adverb with reference to dominant word order, defined as the most frequent one. Thus, in Dryer's terms, Rapa Nui, like the 11 languages in his sample, represents a counterexample to (12) by displaying VAdvO as the dominant order:

(16) Rapa Nui (Malayo-Polynesian, Austronesian)

he	hāŋai	**haka'ou**	i	te	mā'aŋa	rikiriki.
NTR	feed	again	ACC	ART	chick	PL:small

'He raised chicks again.' (Kieviet 2017: 183)

In a generative approach, however, it must be shown that VAdvO is present underlyingly to be a valid counterexample to (11). A comparison to Hawkins (2008) illustrates this point: in Hawkins (2008), VXO order is absent because there is no language where VXO is either the only available order or at least twice as frequent as VOX. For the generativist researchers mentioned in footnote 1 however, a mere surface VAdvO token as in the Norwegian (17a) would not count as a sufficient counterexample. It would have to be shown that VAdvO obtains even when controlling for the possibility of movement. In Norwegian, for example, the V2-property must be controlled for. This can be done by using a non-finite construction, thus revealing the ban on adverbial intervention (17b), such that Standard Norwegian is, again, in line with (11) when controlling for movement (see also Sande et al. 2019).

(17) Norwegian (Germanic, Indo-European)

a)

Dronningen	klapper	**ofte**	katter.	
queen.DEF	strokes	often	cats	

'The queen often strokes cats.'

b)

*Dronningen	skal	klappe	**ofte**	katter.	
queen.DEF	will	stroke.INF	often	cats	

c)

Dronningen	skal	**ofte**	klappe	katter.
queen.DEF	will	often	stroke.INF	cats

d)

Dronningen	skal	klappe	katter	**ofte**.
queen.DEF	will	stroke.INF	cats	often

'The queen will often stroke cats.' (own knowledge)

A diverse cross-linguistic investigation of adverbial intervention of this sort has not been undertaken yet; instead, the discussion has been largely restricted to languages of Europe. Our project will fill this gap and investigate predictions about both underlying and surface orders, as well as potential exceptions, controlling for factors such as different types of adverbs (frequency, place, manner, time).

3.2 Features based on Global Samples

The second type of feature is based on typological research. As in the case of the position of the oblique elements discussed in Section 2, this type of feature has already been investigated with the help of much more extensive and balanced cross-linguistic samples (Dryer & Gensler 2013). Here, the project will provide additional high-quality data and more detailed distinctions, such as differentiating between different types of obliques. For instance, in the Cushitic language Oromo, for which the data was provided by Wakweya Olani Gobena, XOV is the natural order for instruments and locations, while OXV is also possible. For destinations, however, OXV is the only possible order.

(18) Afaan Oromo (Cushitic, Afro-Asiatic)

a)

meerii-n	**bruuʃii-ɗaan**	fakkii	haaraa	boʈʃ'-t-e	
Mary-NOM	brush-INSTR	portrait	new	paint-3SF-PFV	

'Mary painted a new portrait with a brush.'

b) meerii-n **paark-ittʃa** **keessa-tti** fakkii boʧ'-t-e
Mary-NOM park-DEF inside-LOC portrait paint-3SF-PFV
'Mary painted a portrait in the park.'

c) meerii-n fakkii **gara** **gabaa-tti** geessi-t-e
Mary-NOM portrait towards market-LOC take-3SF-PFV
'Mary took the portrait to the market.' (elicited)

Drawing more fine-grained distinctions will allow a higher resolution and more qualified predictions concerning potential cross-linguistic correlations.

A methodological problem pointed out by Sposato (2014: 118) for the rare XVO languages is that many of the prepositions that are part of the oblique element derive from verbs. In example (17a) from Mandarin, *yòng* is analyzed as a preposition meaning 'with' but is formally still identical to the verb *yòng* 'to use'.

(19) Mandarin (Sinitic, Trans-Himalayan)

a) gēge **yòng** xiànjīn mǎi le yì běn shū.
e.brother with cash buy PFV one CLF book
'The elder brother bought a book with cash.'

b) gēge wǎng **shàng** mǎi le yì běn shū.
e.brother internet on buy PFV one CLF book
'The elder brother bought a book on the internet.' (elicited)

Therefore, example (19a) could potentially be interpreted as an instance of verb serialization. However, the postposition in (19b) lacks a verbal origin and demonstrates that Mandarin does have XVO order. To rule out potential other explanations of this sort, the project collects extensive background information for every language under investigation.

3.3 Rare or Areally Restricted Features

The last type of feature is specific constructions that are either present or absent in any given language (Table 9). These usually exhibit an areally biased distribution but nevertheless could correlate with the order of O and V.

Table 9: Matrix showing the possible types of binary categories

	OV	VO
specific construction	+/-	+/-

Examples of this type of category are wh-scope marking (see Section 4.2), correlative relative clauses (see Section 4.3), and so-called mermaid constructions (see below). All three have been claimed to show a connection with OV order (Fanselow no date; Lipták 2009; Tsunoda 2020a, among others). Since all three are cross-linguistically rare, they show very different patterns than the other features mentioned above. A comparison with the category of adjectives is informative. Adjectives are usually prenominal (AdjN) in the Eurasian OV languages, except some in Southeast Asia. However, many languages outside of Eurasia also have a category of adjectives, but their position shows no clear correlation with OV/VO from a global perspective (Dryer 2007). For instance, Oromo exhibits OV but, like most of the languages of Africa, has NAdj order (see 18a, Dryer 2013d).

The three features mentioned above, on the other hand, are only attested in certain areas. For instance, correlative relative clauses, which we describe in Section 4.3, are almost

exclusively found in Western Africa, South Asia, and ancient Indo-European languages (Dryer 2013d; Lipták 2009). Mermaid constructions are often illustrated with Japanese:

(20) Japanese (Japonic)

[Boku=wa	gohan=o	tabe-ta]	**tokoro**=da.
1SG.INF=TOP	food=ACC	eat-PST	place=COP.NPST

'I have just eaten.' (elicited)

According to Tsunoda (2020b), mermaid constructions are mono-clausal despite having an embedded clause and usually have the form [Clause] Noun Copula. They are furthermore characterized, among others, by the following features: The clause can be used as an independent sentence, and its subject is not coreferential with the noun (in this case, *tokoro*). Most known languages with this type of construction are located in a large area that stretches from Northeast Asia over Central China to South Asia and also includes a few Austronesian languages and at least one African language (the Cushitic language Sidaama) (Tsunoda 2020a: xii). Most of these languages exhibit OV order, which suggests the following unidirectional prediction:

(21) if + mermaid construction then OV

However, because mermaid constructions seem to be cross-linguistically rare, there are many OV languages without this type of construction. In this case, OV might be a necessary precondition for the emergence of the feature, which is still incompletely understood. Moreover, there are very few exceptions to the correlation in (21), such as Thai or Tagalog.

The CHAOS project also intends to find potential explanations for this type of correlation. The mermaid construction, for instance, appears to be based on copula clauses. Previous research has demonstrated that languages with OV order usually also have clause-final copulas (Dryer 2007: 91), which might be one reason for the connection to OV order. For this purpose, too, many additional features need to be included in the data collection.

4 Language Selection

There are some 7000 languages in the world. Since no typology has used information from all languages, even those correlations based on large samples of several hundred languages can, in principle, still be empirically tested by using a new set of languages. Ideally, this set should not intersect with the first set. In practice, this methodology is limited by the availability and reliability of the data. To avoid this problem, the project collects fresh data from three different but overlapping language samples.

4.1 Global Sample

The first sample is an ideally balanced global sample of languages with OV and VO word order based on Dryer's (1992: 83) six macro areas of Africa, Eurasia, Southeast Asia & Oceania, Australia & New Guinea (excluding Austronesian), North America (including Mayan and Aztecan), and South America (including Central America). The aim is to investigate languages with OV and VO word order from each of these areas. Cross-linguistically valid correlations should remain stable independent of the specific areas (Dryer 2011). Aiming to avoid an exceeding overlap with previous samples, the project includes several understudied and minority languages, such as Cabécar (Chibchan) or Jula (Mande).

4.2 Language Families with Internal Variation

The second sample, following the example of Haider (2015, 2020), consists of languages with OV and VO order that are all from a single language family. Systematic structural differences between such related languages can lend additional support to the cross-linguistic correlations. Stocks of this type include, among many others, Germanic (Haider 2020), Uralic (Schmidt & Surányi 2019; Vilkuna 2022), Slavic, or Sinitic. For instance, Standard Mandarin, like most other Sinitic languages, typically has SVO order. Additionally, under certain conditions, there is marked SOV order in combination with flagging by a preposition derived from verbs:

(22) Mandarin (Sinitic, Trans-Himalayan)

a) zhāxī dǎ le yùlín
Zhaxi hit PFV Yulin

b) zhāxī **bǎ** yùlín dǎ le
Zhaxi ACC/take Yulin hit PFV
'Zhaxi has beaten Yulin' (elicited)

However, several Sinitic languages, such as Zhoutun Chinese, have undergone language change to basic SOV order, possibly due to language contact. Among other features, these underdescribed languages have developed a case marking system involving suffixes or enclitics, which is highly untypical for Sinitic:

(23) Zhoutun (Sinitic, Trans-Himalayan)
tʂaɕi ylĩ=**xa** ta=lɔ.
Zhaxi Yunlin=ACC beat=PFV
'Zhaxi has beaten Yulin' (Zhou 2022: 24)

Examples of this sort need careful qualitative evaluation to differentiate between structural and areal factors. While languages with SVO order like Mandarin usually have zero argument marking (Sinnemäki 2010) and there is a cross-linguistic tendency for SOV languages to develop case systems (Siewierska & Bakker 2009), the case marking in these languages could also have developed due to contact with other SOV languages, including Mongolic, Tibetic, and Turkic. In extreme cases, such as so-called *metatypy*, language contact can lead to a complete copy of the syntactic structure. The Oceanic language Takia, unlike related languages with SVO, exhibits SOV order and many other differences, such as post- rather than prepositions. In this case, however, these features likely originate in the contact language Waskia (Ross 2001: 139-152). Deriving cross-linguistic correlations from such extreme cases alone could be misleading.

Another relevant language family is the Slavic branch of the Indo-European languages, which is largely uniform in displaying basic VO order. However, under the influence of German, Sorbian (and especially Upper Sorbian) has developed neutral OV order. In contrast to German, this applies to both embedded and main clauses (Stone 1993: 653), whereas German displays V2 in declarative main clauses. Arguably also due to the contact situation, colloquial Upper Sorbian has developed other features untypical for Slavic, such as the use of expletives with weather predicates (Scholze 2007), a feature common in Northwestern Europe (Eriksen et al. 2010: 575):

(24) Colloquial Upper Sorbian (Slavic, Indo-European)
To so děšćika dźo. **To** so hrimoce.
it REFL rain go.PRS.3SG it REFL thunder.PRS.3SG
'It is raining. It is thundering.' (Scholze 2007: 322; our glossing)

(25) Czech (Slavic, Indo-European)
Prší. Hřmí.
rain.PRS.3SG thunder.PRS.3SG
'It is raining. It is thundering.' (own knowledge)

Another example is the availability of "wh-scope marking" in Sorbian. In this type of construction, which exists in German (26) but is absent from most Slavic languages, the embedded clause is a content question, whereas the main clause contains a dummy interrogative (typically *what*) referring to and indicating the actual scope of the embedded interrogative phrase. Despite the two interrogatives, the construction is a single question (Fanselow no date).

(26) German (Germanic, Indo-European)

Was	glaubst	du,	**wer**	dieses	Paper	lesen
what	believe.PRS.2SG	2SG.NOM	who.NOM	this	paper	read.INF
wird?						
will.PRS.3SG						

'Who, do you think, will read this paper?' (own knowledge)

A systematic comparison of a range of features in Sorbian, other West Slavic languages such as Czech and Polish, and German can shed light on the relationship between contact-induced changes and changes that might be tied to a more fundamental shift in underlying structure.

A further issue with respect to Slavic is the degree to which the language group as a whole, or some Slavic languages in particular, adhere to patterns typically ascribed to SVO languages. Over the last decade, Haider & Szucsich (2022) and Szucsich & Haider (2015) have argued that Slavic should not be classified as SVO with a number of untypical properties. Instead, Slavic represents a type of language where the order of head and dependent is underlyingly underspecified. Recently, Šimík & Jasinskaja (2022) have pointed out that the discussion requires a more language-specific focus. They show that with respect to the features investigated by Haider & Szucsich (2022), Czech adheres to more SVO properties than Russian. The here-presented project can contribute to this discussion by comparing a greater number of features in a larger language sample, thereby both helping to identify the properties that pattern with head-dependent order and collecting comparable data for a range of languages from Slavic and beyond.

4.3 Languages with Rare or Unusual Features

The third set includes languages with rare grammatical features, such as XVO in Sinitic languages, the unexpected VAdvO order, or the three rare constructions mentioned in Section 3. A detailed investigation of such features can lead to alternative explanations or a better understanding of exceptions to cross-linguistic correlations. In these cases, specific languages with the features in question need to be chosen, such as Marathi for correlative relative clauses (27). Correlative relative clauses are a cross-linguistically rare phenomenon in which a head-internal relative clause commonly precedes the main clause, which contains a resumptive element and, optionally, a copy of the head noun.

(27) Marathi (Indo-Iranian, Indo-European)

dzyā	mulī	tithe	basalyā	āhet,
REL.PL.F	girl.PL	there	sit.PFV.3PL.F	be.PRS.3PL
tyā	(mulī)	ājārī	āhet.	
that.PL.F	girl.PL	sick	be.PRS.3PL	

'The girls who are sitting there are sick.' (elicited)

Another example is S-Aux-O-V-Adv/X order in certain African languages, a combination of Aux-O-V split predication and OVX order (Güldemann 2008: 159–163). For instance, this is the canonical word order in the Bantu language Tunen as spoken in Cameroon. The following examples provided by Elisabeth Kerr illustrate this word order for adverbs (28) and obliques (29). Not only is this word order cross-linguistically rare, but it is also not very common in the area immediately surrounding Tunen, which usually has VOX order. S-Aux-O-V-Adv/X is also unusual from a structural viewpoint because it shows disharmony of clausal headedness: AuxV but VO (Sande et al. 2019).

(28) Malíá **aka** bɛbíə nyákɛna **naánɛkɔl**
/Malíá a-ka bɛ-bíə nyákɛna naánɛkɔla/
1.Maria SM.1-PST3 8-beer drink yesterday
'Maria drank the beers yesterday.'

(29) **mɛ́ndɔ** baná batɛ́˩tɛ́á báfandɛ́ sinə
/mɛ-^{H}ndɔ ba-ná ba-tɛ́^{L}tɛ́á ba-^{H}fandɛ́ sinə
SM.1SG-PRS 2-child 2-small 2-two see
u **miím**.
ɔ miímə/
PREP house
'I see two small children in the house.'

For all three samples mentioned above, a large amount of background information about the languages is required for a complete analysis (e.g., about competing prenominal relative clauses in Marathi or the clause structure of Tunen).

5 Process of Data Collection

A major concern for any cross-linguistic research is the collection of the necessary data. Often, typological studies are based on available grammar books and descriptions. While unavoidable for large language samples, this method is known to lead to distortions in some of the resulting typologies. For instance, Dryer & Gensler (2013) argue that this may have led to the overrepresentation of OVX in their typology with respect to XOV and OXV:

> The frequency on the map is a methodological artifact, and emerges from the fact that many grammars state the ordering of O and V and of X and V but not of O and X. In such cases, if the language is of the type OV&XV, nothing can be concluded about the ordering of O and X, and so such languages had to be omitted from the map. By contrast, if the language is of the type OV&VX, then the ordering OX follows automatically, and so the language is always included on the map.

To avoid this type of distortion, this project is not primarily based on reference grammars but on data elicited under controlled conditions. While this has the disadvantage of being time-consuming and costly, it offers the clear advantage of providing the necessary data for answering the questions. It also offers the possibility to include properties that are not usually found in reference grammars, such as grammaticality judgements or information about specific types of obliques.

The data are collected in three different ways. First, some of the data are elicited from native speakers by the authors of this paper (e.g., Czech, Mandarin, Udmurt). Second, some of the questionnaires are filled in by native speakers with linguistic training and access to further native speakers for verification (e.g., Kazakh, Oromo, Thai). Third, some data is collected by linguists through fieldwork (e.g., Cabécar, Kurux, Tunen). The project offers a financial

compensation for the latter two types of data collection. On occasion, remaining gaps can be filled with available reference grammars (e.g., Kobayashi & Tirkey 2017 on Kurux).

6 Questionnaire

The questionnaire used for the elicitation of the data was originally designed by Gisbert Fanselow and extended by the authors of this paper. It has two main purposes. First, similar to the *Lingua Descriptive Studies Questionnaire* (Comrie & Smith 1977), it can function as a guideline for describing aspects of the syntax of individual languages. Second, for the CHAOS project, it allows the elicitation of specific information concerning the OV/VO alternation. The questionnaire is available through the link provided under Supplementary Material.

The preface of the questionnaire asks about the researcher, the informants, the notation system, and the use of additional references. The questionnaire itself has the following structure:

(30)
- A. Background Information
- B. Data Collection
  - 0. Data Collection Guideline
  - 1. The Order of Verbs and Arguments
    - 1.1. Verb and object in neutral contexts
    - 1.2. Verb and subject in neutral contexts
    - 1.3. Word order variation in non-neutral contexts
  - 2. Pronominals
    - 2.1. Obligatoriness
    - 2.2. Impersonal constructions
  - 3. Operators and Movement: Content Questions
  - 4. Contrastive Focus
  - 5. Formation of Relative Clauses
  - 6. The Verbal Domain
    - 6.1. The VP
    - 6.2. Restructuring verbs
    - 6.3. Further issues
  - 7. Clausal Arguments
  - 8. Marking and Properties of Nominal Arguments
  - 9. Additional Background Information

Section A on the background information concerns the name of the language, the Glottocode, the notation system used in the description, details concerning the number of speakers, the genetic affiliation, and the dominant word order in intransitive and transitive clauses to allow a preliminary classification. A brief guideline for the collection of the data follows in the first part of Section B.

The remainder of the questionnaire concentrates on syntactic aspects that are addressed with specific questions. The following is a representative example of the questionnaire, focusing on oblique elements (Question 6 in Part 1.1). Every topic starts with a specific question:

(31) Question 6: What is the position of obliques (X) with respect to O and V in transitive clauses in a wide focus context?

This is followed by all logical possibilities, referred to as “alternatives to be checked”. In this case, there are six types (also seen in Section 2):

(32) O-X-V / X-O-V / V-O-X / V-X-O / X-V-O / O-V-X

Many questions address a specific context to control for factors pertaining to information structure. In this case, the informant is asked to provide information in an "out-of-the-blue" or "wide focus" context, i.e., without preceding discourse. This is indicated with a "prompt", such as *What's new?* or *What has happened?* to allow an easier understanding of this concept. In other cases, a specific question or situation is provided to check for the role of information structure on syntactic phenomena (for example, involving different kinds of focus).

This in turn is followed by an illustration of the answer in English or another language in case English does not exhibit the feature or construction in question.[2]

(33) English examples
Mary a portrait* [*with a brush***] *painted.* (O-X-V)
Mary* [*with a brush***] *a portrait painted.* (X-O-V)
Mary painted a portrait [***with a brush***]. (V-O-X)
Mary painted* [*with a brush***] *a portrait.* (V-X-O)
Mary* [*with a brush***] *painted a portrait.* (X-V-O)
Mary a portrait painted* [*with a brush***]. (O-V-X)

In this specific case, there are three different sub-questions concerning different types of oblique elements, viz. instruments, places, and directions, that are known to affect the outcome in some languages:

(34) a) Instruments (*with a brush*, *with a pencil*)
Mary painted a portrait [*with a brush*].

b) Place (*in the park*, *at the market*)
Mary painted a portrait [*in the park*].

c) Direction (*to the market*, *towards the hills*)
Mary carried a portrait [*to the market*].

Since the examples sometimes need to be adjusted depending on the structure of the language, several possibilities are provided in parentheses. The questions are followed by a brief instruction on how to answer the question in case this should be necessary:

> Explanation: Please try to re-use the lexical material of [the preceding question] if possible, so that we get minimal pairs. When selecting the oblique, please choose one of the following: adjuncts, adverbial elements, or noun phrases or adpositional phrases marked with adpositions or peripheral case markers that are neither direct, nor indirect Os. Please exclude contrastive focus or topic constructions as well as pronominal arguments.

At the end of most questions, a brief section explains the relevance for the present investigation. This helps to clarify the intention behind the question:

> Background: Languages with V-O order usually also show V-O-X order. V-X-O order is unattested and X-V-O order is exceedingly rare (e.g., Mandarin). O-V order shows much more variation: O-X-V / X-O-V / O-V-X.

The questionnaire was already successfully used for the elicitation of data for languages as diverse as South Sámi, Upper Sorbian, Afaan Oromo, South Bolivian Quechua, or Nepali.

---

[2] These include Evenki (Nedjalkov 1997), French (own knowledge), German (own knowledge), Hindi (Fanselow, no date), Italian (own knowledge), Japanese (Tsunoda 2020b), Korku (Zide 2008), Manchu (own knowledge), Mandarin (elicited), Persian (Windfuhr & Perry 2009), Russian (elicited), and South African Indian English (Mesthrie & Dunne 1990).

7 Illustration of the Questionnaire

The following illustrates the answer to the question about oblique elements with the help of these Nepali data (Indo-Iranian, Indo-European) provided by Dubi Nanda Dhakal. All examples are glossed according to the Leipzig Glossing Rules.

In all cases, both XOV and OXV order are equally possible. Other word orders are considered ungrammatical in a wide focus context. For simplicity, the glossing is omitted in these ungrammatical examples.

(34) Instrument

a) meri-le **burush-le** nəjã tsitrə bəna-i
Mary-ERG brush-INST new picture make-PST.3SG.F.NH
'Mary painted (made) a new picture with a brush.' (XOV)

b) meri-le nəjã tsitrə **burush-le** bəna-i
Mary-ERG new picture brush-INST make-PST.3SG.F.NH
'Mary painted (made) a new picture with a brush.' (OXV)

c) *meri-le nəjã tsitrə bəna-i **burush-le** (OVX)

d) *meri-le **burush-le** bəna-i nəjã tsitrə (XVO)

e) *meri-le bəna-i **burush-le** nəjã tsitrə (VXO)

f) *meri-le bəna-i nəjã tsitrə **burush-le** (VOX)

(35) Location

a) meri-le **park-ma** nəjã tsitrə bəna-i
Mary-ERG park-LOC new picture make-PST.3SG.F.NH
'Mary painted (made) a new picture in the park.' (XOV)

b) meri-le nəjã tsitrə **park-ma** bəna-i
Mary-ERG new picture park-LOC make-PST.3SG.F.NH
'Mary painted (made) a new picture in the park.' (OXV)

c) *meri-le nəjã tsitrə bəna-i **park-ma** (OVX)

d) *meri-le **park-ma** bəna-i nəjã tsitrə (XVO)

e) *meri-le bəna-i **park-ma** nəjã tsitrə (VXO)

f) *meri-le bəna-i nəjã tsitrə **park-ma** (VOX)

(36) Direction

a) meri-le **bədzar-ma** nəjã tsitrə ləg-i
Mary-ERG market-loc new picture take-PST.3SG.F.NH
'Mary carried a new picture to the market.' (XOV)

b) meri-le nəjã tsitrə **bədzar-ma** ləg-i
Mary-ERG new picture market-loc take-PST.3SG.F.NH
'Mary carried a new picture to the market.' (OXV)

c) *meri-le nəjã tsitrə ləg-i **bədzar-ma** (OVX)

d) *meri-le **bədzar-ma** ləg-i nəjã tsitrə (XVO)

e) *meri-le ləg-i **bədzar-ma** nəjã tsitrə (VXO)

f) *meri-le ləg-i nəjã tsitrə **bədzar-ma** (VOX)

The inclusion of three different types of oblique elements allows a more fine-grained classification of languages than the typology shown in Section 2. While there is no difference in Nepali, other languages exhibit different word orders depending on the type of oblique element (see Section 3 on Oromo).

8 Database and Coding

The result of the data collection is a set of extensive language reports for the individual languages. They will be transferred to a uniform file format adequate for long-term storage (CSV). The collected data will eventually be stored in a publicly available repository (e.g., OSF). All examples will be fully analyzed to allow their future use in other cross-linguistic studies.

To make the data accessible for computational approaches, hypothesis testing, and comparisons with already available databases, the individual datapoints need to be compressed into a useful format. For the sake of this coding procedure, individual questions in the questionnaire will be split into different datapoints corresponding to the number of logical possibilities. In the case of oblique elements, each of the three sub-questions on different types of adjuncts will be split into six different datapoints corresponding to O-X-V, X-O-V, V-O-X, V-X-O, X-V-O, and O-V-X, respectively. In this example, there will be a total of 18 (3x6) different datapoints. For each datapoint, one of the following six values will be assigned (108 possibilities in total):

(37)

0	impossible
1	possible
0.5	grammaticality unclear
ND	no data
NA	not applicable
U	unclear

The values 0 or 1 are assigned if a given linearization or feature is absent or present in the language under investigation, respectively. In the case of Nepali, in each of the three sub-questions the value 1 will be assigned to X-O-V and O-X-V and the value 0 to the remaining four categories. The value 0.5 will be assigned in contrast to 1 if a certain order is potentially available in a language but its grammaticality is uncertain. If a given datapoint is missing, the value ND is used. If a given question cannot be applied to a language (e.g., if it lacks a certain category), the value NA is used. If a datapoint is available but for some reason cannot be assigned to any of the other values, the value U is employed. Details of the coding might vary slightly as new languages are included. Additionally, if more than one option is grammatical in a language (the value 1 occurs more than once) but one of them is more natural or preferred, this will be indicated in a separate commentary:

(38)

P	primary/preferred
S	secondary
ND	no data
NA	not applicable

This allows a direct comparison of the different datapoints within a given sub-question in case multiple constructions are present. It also helps to avoid problematic categories, such as "no dominant order" as seen in Section 2. For instance, Oromo prefers X-O-V order (value P) over O-X-V order (value S) in sub-questions 6a and 6b (Table 10).

Table 10: Preliminary coding of the Nepali and Oromo data for oblique elements

Question	**Question name**	**Subquestion**	**Subquestion name**	**Parameter**	**Nepali value**	**Nepali preference**	**Oromo value**	**Oromo preference**	**…**
…									
6	obliques	a	instrument	O-X-V	1	ND	1	S	
6	obliques	a	instrument	X-O-V	1	ND	1	P	
6	obliques	a	instrument	V-O-X	0	NA	0	NA	
6	obliques	a	instrument	V-X-O	0	NA	0	NA	
6	obliques	a	instrument	X-V-O	0	NA	0	NA	
6	obliques	a	instrument	O-V-X	0	NA	0	NA	
6	obliques	b	place	O-X-V	1	ND	1	S	
6	obliques	b	place	X-O-V	1	ND	1	P	
6	obliques	b	place	V-O-X	0	NA	0	NA	
6	obliques	b	place	V-X-O	0	NA	0	NA	
6	obliques	b	place	X-V-O	0	NA	0	NA	
6	obliques	b	place	O-V-X	0	NA	0	NA	
6	obliques	c	direction	O-X-V	1	ND	1	NA	
6	obliques	c	direction	X-O-V	1	ND	0	NA	
6	obliques	c	direction	V-O-X	0	NA	0	NA	
6	obliques	c	direction	V-X-O	0	NA	0	NA	
6	obliques	c	direction	X-V-O	0	NA	0	NA	
6	obliques	c	direction	O-V-X	0	NA	0	NA	
…									

The resulting database will be based on CSV files for each language. It contains additional information, such as the macro-area where a language is spoken, the Glottocode, the affiliation of the language, and a meta-classification of the questions (referring to linear order, structure, the background information, etc.).

9 Theoretical Interpretations

As mentioned in Section 4, correlations between verb-object ordering and other features could in many cases be restricted to certain language families, or they could be based on language contact and thus apply only in a certain area. In controlling for these factors, the present project aims to filter out features that are cross-linguistically tied to VO/OV. For these, we then aim to clarify the source of these correlations.

One potential explanation is that the correlation is based on structural factors. Consider again the placement of adverbs as illustrated in Section 3: for Germanic, Haider (2020) observes that VO languages do not allow the adverb to intervene between verb and object. His theoretical account involves an asymmetric notion of merger (it is universally leftward) and the need for a dependent phrase to be licensed by the head (Haider 2020: 360). In contrast, Belk & Neeleman (2017) assume symmetric merger, but asymmetric restrictions on movement and case assignment (Belk & Neeleman 2017: 19–21). What the two approaches have in common is that

structural explanations for the differences between VO and OV languages are sought in the asymmetry somewhere in the general syntactic architecture of languages.

An alternative explanation is provided by performance-based approaches such as Hawkins (2008) for the patterns of oblique placement in the WALS sample, as described in Section 2. From these, he derives the following three tendencies, all of which have their source in performance preferences: verb and object should be adjacent; object and oblique should be on the same side of the verb; the object should precede the oblique. Extremely rare or unattested patterns are those that meet fewer than two of these requirements. Hawkins (2008) argues that such originally performance-driven patterns result in grammatical conventions.

If exceptions are found to generalizations like the compactness of the verb-object sequence in VO languages, this does not automatically falsify structural and performance-based accounts. After all, also strong statistical tendencies, even when not absolute, require a linguistic explanation. For example, deviations might result from reorderings which a given language employs for the marking of information structure. To attain a thorough understanding of such processes, part 1.3 (questions 12–20) of the questionnaire investigates word order variation in non-neutral contexts, viz. potential variations that occur when a constituent is the information focus or the contrastive focus or topic.

An issue related to these considerations is whether it is possible to draw a sharp line between VO and OV languages, or whether the VO/OV distinction is better viewed as gradual, with "typical" representatives of each group at the two far ends of a scale. Such a view predicts that, even though certain features might be associated with VO languages and others with OV languages, many languages will not display all these features. For instance, as mentioned in Section 4, Šimík & Jasinskaja (2022) argue that Czech displays more features assumed to be typical for VO languages than its relative Russian. In a scalar view, Russian would then be closer to the center of the VO-OV scale than Czech. As shown in Sections 2 and 4.2, Mandarin is an example of a language with VO order but many OV properties, such as prenominal relative clauses. The size and variety of our language sample will allow us to approach the specifics of such a view, for example the question of the scale's division (is it continuous, and if not, how many divisions are there?), and whether certain properties remain stable throughout one side of the scale, whereas others only apply to its extremes. By investigating features that are already known to correlate with VO/OV order, we are able to place individual languages on such a preliminary scale and identify more and less typical VO and OV languages. This is also an important basis for the testing of further hypotheses, which might only apply to typical representatives of a group, as done by Sande et al. (2019).

Generally, the present study largely focuses on the analysis of features beyond the ordering of head and dependent, as described in Section 2.2. This does not mean, however, that correlations in the order of head and dependent in different domains are trivial to analyze; for an overview of different approaches to this issue, see Biberauer & Sheehan (2013). One recent proposal in this long-standing debate is the Final-over-Final Constraint (Biberauer et al. 2014: 171), which prevents a head-final phrase from dominating a head-initial phrase within the same extended projection. It thereby rules out, among others, VOAux orders. The empirical verification of this constraint requires a clarification of whether the investigated elements are indeed heads. With respect to VOAux orders, this applies particularly to the auxiliary. A close inspection of the data is thus required for each individual language, and our genetically, areally and typologically diverse data will be a valuable contribution to this particular question and the general discussion of harmonic head-dependent ordering.

10 Conclusion

This study briefly described an ongoing data collection that focuses on syntactic phenomena, especially as they relate to the OV/VO alternation. It is a contribution not only to the cross-linguistic study of syntax, but also to the documentation and description of languages from around the world, including some that have not been described in detail before. An example is Linxia Chinese (aka Hezhou Chinese, Lee-Smith 1996), which is among those Chinese varieties in Western China that have changed to OV order (Li 1984, Section 4.2). Previous research on Linxia Chinese is very limited and difficult to access for researchers outside of China (Ma 1984; Zhang 2021). Additionally, some of the languages included in the study only have few remaining native speakers (e.g., the Omotic language Sheko, Hellenthal 2010). Some of the features, such as wh-scope marking, are cross-linguistically rare or have not been addressed in global surveys. The documentation will allow such underdescribed or endangered languages and rare features to be included in future cross-linguistic studies.

Acknowledgement

The presented 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 Association (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. To a large extent, the questionnaire was designed by Prof. Gisbert Fanselow. We would like to thank Prof. Dubi Nanda Dhakal for providing data for Nepali, Elisabeth Kerr for the Tunen examples, and Wakweya Olani Gobena for information about Oromo.

Appendix

This appendix lists all features investigated in the project. Numbers in the second column correspond to the questions in the questionnaire.

Features	**Questions**	**Selected references**
position of objects (O arguments)	1–4, 12, 22	Dryer 2013b, Holmberg 2015
position of adverbs	5	Stowell 1981, Cinque 1999, Haspelmath et al. 2013, Haider 2020
position of oblique elements	6, 15	Hawkins 2008, Dryer & Gensler 2013
position of indirect objects (G arguments)	7, 14	Heine & König 2010, Janke & Neeleman 2012
height of the subject	8–11, 13	Fanselow 1987, 2020, Dryer 1997, 2013g
mirror-image effect	16–17	Neeleman 2017
obligatoriness of the subject	21, 23–25	Haider 2010, 2013, Fanselow 2020
long movement and wh-scope marking	18, 19	Fanselow 2006, 2017, no date
discontinuous NPs	20, 70	Fanselow p.c.
A- and A-bar-movement	26, 27	Mahajan 1990, Haider 2010
position of interrogative phrases	28	Fanselow 2006, 2017, Dryer 2013f
superiority	29	Haider 2010, Häussler et al. no date
subject islands	31-34, 36-38	Huang 1982, Stepanov 2007, Haider 2010, Sheehan 2013, Bošković 2015

weak crossover alleviation	30, 35	Haider 1986, Mahajan 1990
(correlative) relative clauses	39, 40, 81	Lipták 2009, Dryer 2013e
partial VP fronting	41-43	Thiersch 1985, Müller 1998, Haider 2010, Janke & Neeleman 2012
VP-internal subject	44-46	Wurmbrand 2006
VP ellipsis	47	Johnson 2001
free variation in the verb complex	48, 49	Haider 2010, Abels 2016
restructuring	50-53	Haider 2010, Wurmbrand 2017
light verbs	54	Polinsky & Magyar 2020
serial verb construction	55	Haspelmath 2016
adverbial subordination and "con-verbs"	56, 57, 83	Haspelmath 1995, Bickel 1998, Longacre 2007
alternative questions	58	Anonymous to appear
mermaid constructions	59	Tsunoda 2020a, 2020b
finiteness of subordinate clauses	60	Cristofaro 2013a, 2013b
placement of clausal arguments	61, 62	Bayer 2001
marking of arguments and align-ment	63–66	Siewierska & Bakker 2009, Dixon 2010, Heine & König 2010, Sinnemäki 2010, Bickel 2011, Comrie 2013
overt marking of information structure	67	Colantoni & Sánchez 2021, Fanselow p.c.
presence and position of articles	68	Dryer 2007, 2013a
nominal splits	69-70	Fanselow p.c.
head-dependent orders	71–85	Dryer 2007, 2022
order in appositions	86	Hackstein 2003

Supplementary Material

The full questionnaire can be found at https://osf.io/5rymf/.

References

Abels, Klaus. 2016. The fundamental left-right asymmetry in the Germanic verb cluster. *Journal of Comparative Germanic Linguistics* 19: 179–220.

Anonymous. To appear. A typology of alternative questions in Chinese and other East Asian languages. To appear in 2025 in: *Language and Linguistics*.

Bayer, Josef. 2001. Two grammars in one: Sentential complements and complementizers in Bengali and other South Asian languages. In P. Bhaskararao & K. V. Subbarao (eds.), *The yearbook of South Asian languages 2001*, 11–36. New Delhi: Sage Publications.

Belk, Zoë & Ad Neeleman. 2017. AP adjacency as a precedence constraint. *Linguistic Inquiry* 48(1): 1–45.

Biberauer, Theresa, Anders Holmberg & Ian Roberts. 2014. A syntactic universal and its consequences. *Linguistic Inquiry* 45(2): 169–225.

Biberauer, Theresa & Michelle Sheehan. 2013. Theoretical approaches to disharmonic word order. In Biberauer, Theresa & Michelle Sheehan (eds.), *Theoretical approaches to disharmonic word order*, 1–44. Oxford: Oxford University Press.

Bickel, Balthasar. 1998. Converbs in cross-linguistic perspective. *Linguistic Typology* 2: 381–397.

Bickel, Balthasar. 2011. Grammatical relations typology. In Song Jae Jung (ed.), *The Oxford handbook of linguistic typology*, 399–444. Oxford: Oxford University Press.
Bošković, Željko. 2016. On the timing of labeling: Deducing Comp-trace effects, the Subject Condition, the Adjunct Condition, and tucking in from labeling. *The Linguistic Review* 33(1): 17–66.
Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective.* Oxford: Oxford University Press.
Colantoni, Laura & Liliana Sánchez. 2021. The Role of Prosody and Morphology in the Mapping of Information Structure onto Syntax. *Languages* 6: 207.
Comrie, Bernard & Norval Smith. 1977. Lingua Descriptive Studies: Questionnaire. *Lingua* 42(1): 1–72.
Comrie, Bernard. 2008. Prenominal relative clauses in verb-object languages. *Language and Linguistics* 9: 723–733.
Comrie, Bernard. 2013. Alignment of case marking of full noun phrases. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online.* Leipzig: MPI for Evolutionary Anthropology. http://wals.info/chapter/98
Cristofaro, Sonia. 2013a. Purpose Clauses. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online.* Leipzig: MPI for Evolutionary Anthropology. http://wals.info/chapter/125
Cristofaro, Sonia. 2013b. 'When' clauses. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online.* Leipzig: MPI for Evolutionary Anthropology. http://wals.info/chapter/126
Dixon, R. M. W. 2010. *Basic linguistic theory, volume 3*. Oxford: Oxford University Press.
Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68(1): 81–138.
Dryer, Matthew S. 1997. On the six-way word order typology. *Studies in Language* 21: 69–103.
Dryer, Matthew S. 2007. Word order. In Timothy Shopen (ed.), *Language typology and syntactic description, volume 1*, 61–131. 2nd edn. Cambridge: Cambridge University Press.
Dryer, Matthew S. 2011. The evidence for word order correlations. *Linguistic Typology* 15: 335–380.
Dryer, Matthew S. 2013a. Definite Articles. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online.* Leipzig: MPI for Evolutionary Anthropology. http://wals.info/chapter/37
Dryer, Matthew S. 2013b. Order of object and verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online.* Leipzig: MPI for Evolutionary Anthropology. http://wals.info/chapter/83
Dryer, Matthew S. 2013c. Order of adposition and noun phrase. Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online.* Leipzig: MPI for Evolutionary Anthropology. http://wals.info/chapter/85
Dryer, Matthew S. 2013d. Order of Adjective and Noun. Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online.* Leipzig: MPI for Evolutionary Anthropology. http://wals.info/chapter/87
Dryer, Matthew S. 2013e. Order of relative clause and noun. Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online.* Leipzig: MPI for Evolutionary Anthropology. http://wals.info/chapter/90
Dryer, Matthew S. 2013f. Position of interrogative phrases in content questions. Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online.* Leipzig: MPI for Evolutionary Anthropology. http://wals.info/chapter/93
Dryer, Matthew S. 2013g. On the six-way word order typology, again. *Studies in Language* 37: 267–301.

Dryer, Matthew S. 2019. Grammaticalization accounts of word order correlations. In Karsten Schmidtke-Bode, Natalia Levshina, Susanne Maria Michaelis & Ilja A. Seržant (eds.), *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence*, 63–95. Berlin: Language Science Press.

Dryer, Matthew. 2022. On the word order properties of universal quantifiers. 14th Conference of the Association for Linguistic Typology (ALT), 2022.12.15–17, Texas at Austin.

Dryer, Matthew S. & Orin D. Gensler. 2013. Order of object, oblique, and verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online.* Leipzig: MPI for Evolutionary Anthropology. https://wals.info/chapter/84

Eriksen, Pål Kristian, Seppo Kittilä & Leena Kolehmainen. 2010. The linguistics of weather. *Studies in Language* 34: 565–601.

Fanselow, Gisbert. 1987. *Konfigurationalität.* Tübingen: Narr.

Fanselow, Gisbert. 2006. Partial wh-movement. In Martin Everaert & Henk Van Riemsdijk (eds.), *The Blackwell companion to syntax*, 437–492. Malden: Blackwell.

Fanselow, Gisbert. 2017. Partial wh-movement. In Martin Everaert & Henk Van Riemsdijk (eds.), *The Blackwell companion to syntax*, 2nd edn., 2885–2941. Malden: Blackwell.

Fanselow, Gisbert. 2020. Is the OV-VO distinction due to a macroparameter? In Tanaka, Masatoshi, Tomoya Tsutsui & Masashi Hashimoto (Eds), *Linguistic Research as an Interdisciplinary Science*, 1–26. Tokyo: Hitsuji Publishers.

Fanselow, Gisbert. no date. *Remarks on the distribution of Wh-Scope Marking.* (Unpublished manuscript.)

Güldemann, Tom. 2008. The Macro-Sudan belt: Towards identifying a linguistic area in northern sub-Saharan Africa. In *A linguistic geography of Africa*, 151–185. Cambridge: Cambridge University Press.

Hackstein, Olav. 2003. Apposition and word-order typology in Indo-European. In Brigitte L.M. Bauer & Georges-Jean Pinault (eds.), *Language in time and space*, 131–152. Berlin: De Gruyter Mouton.

Haider, Hubert. 1986. *Deutsche Syntax – Generativ.* University of Vienna. (Habilitation thesis.)

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

Haider, Hubert. 2015. Head directionality – in syntax and morphology. In Antonio Fábregas, Jaume Mateu & Michael Putnam (Eds.), *Contemporary linguistic parameters*, 73–97. Bloomsbury Academic.

Haider, Hubert. 2020. VO/OV-base ordering. In Michael T. Putnam & Page, B. Richard (eds.), *Cambridge handbook of Germanic linguistics*, 339–364. Cambridge: Cambridge University Press.

Haider, Hubert & Luka Szucsich. 2022. Slavic languages – "SVO" languages without SVO qualities? *Theoretical Linguistics* 48(1/2): 1–39.

Haspelmath, Martin. 1995. The converb as a cross-linguistic category. In Martin Haspelmath & Ekehard König (eds.), *Converbs in cross-linguistic perspective*, 1–56. Berlin: Mouton de Gruyter.

Haspelmath, Martin. 2016. The serial verb construction: Comparative concept and cross-linguistic generalizations. *Language and Linguistics* 17: 291–319.

Haspelmath, Martin & the APiCS Consortium. 2013. Order of frequency adverb, verb, and object. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *The atlas of pidgin and creole language structures.* Oxford: Oxford University Press.

Häussler, Jana, Gisbert Fanselow, Þórhallur Eyþórsson, Radek Šimík & Luis Vicente. no date. *Crossing movement paths: Multiple wh-questions in seven languages.* (Unpublished manuscript.)

Hawkins, J. A. 2008. An asymmetry between VO and OV languages: The ordering of obliques. In G. G. Corbett & M. Noonan (eds.), *Case and grammatical relations: Studies in honor of Bernard Comrie*, 167–190. Amsterdam: Benjamins.
Heine, Bernd & Krista König. 2010. On the linear order of ditransitive objects. *Language Sciences* 32: 87–131.
Hellenthal, Anneke Christine. 2010. *A grammar of Sheko.* Utrecht: LOT.
Holmberg, Anders. 2015. Verb second. In Kiss, Tibor and Alexiadou, Artemis (eds.), *Syntax: An international handbook of contemporary syntactic research*, 342–383. Berlin: de Gruyter.
Huang, Cheng-Teh James. 1982. *Logical relations in Chinese and the theory of grammar*. MIT. (Doctoral dissertation.)
Janke, Vikki & Ad Neeleman. 2012. Ascending and Descending VPs in English. *Linguistic Inquiry* 43(1): 151–190.
Jing, Yingqi, Damián E. Blasi & Balthasar Bickel. 2022. Dependency-length minimization and its limits: A possible role for a probabilistic version of the final-over-final-condition. *Language* 98(3): 397–418.
Johnson, Kyle. 2001. What VP ellipsis can do, and what it can't, but not why. In Mark Baltin & Chris Collins (eds.), *The handbook of contemporary syntactic theory*, Chapter 14, 439–479. Oxford: Blackwell.
Kieviet, Paulus. 2017. *A grammar of Rapa Nui.* Berlin: Language Science Press.
Kobayashi, Masato & Bablu Tirkey. 2017. *The Kurux language.* Leiden: Brill.
Lee-Smith, Mei W. 1996. The Hezhou language. In Stephen A. Wurm, Peter Mühlhäusler & Darrell T. Tryon (eds.), *Atlas of languages of intercultural communication in the Pacific, Asia, and the Americas*, 865–873. Berlin: Mouton de Gruyter.
Li, Charles N. 1984. From verb-medial analytic language to verb-final synthetic language: A case of typological change. *The Annual Proceedings of the Berkeley Linguistics Society* 10: 307–323.
Lipták, Anikó (ed.) 2009. *Correlatives cross-linguistically.* Amsterdam: Benjamins.
Longacre, Robert E. 2007. Sentences as combinations of clauses. In Timothy Shopen (ed.), *Language typology and syntactic description: Volume 2, Complex constructions*, 2nd edn., 372–420. Cambridge: Cambridge University Press.
Ma, Qiping. 1984. Linxia fangyan yufa chutan [A preliminary study of the grammar of the Linxia dialect]. *Lanzhou xuekan*. 1: 79–84.
Mahajan, Anoop. 1990. *The A/A-bar distinction and movement theory*. MIT. (Doctoral dissertation.)
Mesthrie, Rajend & Timothy Dunne. 1990. Syntactic variation in language shift: The relative clause in South African Indian English. *Language variation and change* 2: 31–56.
Müller, Gereon. 1998. *Incomplete category fronting.* Dordrecht: Kluwer.
Nedjalkov, Igor. 1997. *Evenki.* London: Routledge.
Neeleman, Ad. 2015. *Two asymmetries between pre- and post-head order and their implications for syntactic theory.* (Unpublished manuscript.)
Neeleman, Ad. 2017. PP-over-V meets Universal 20. *The Journal of Comparative Germanic Linguistics* 20(1): 3–47.
Polinsky, Maria & Lilla Magyar. 2020. Headedness and the lexicon: The case of verb-to-noun ratios. *Languages* 5(1): 9.
Ross, Malcolm D. 2001. Contact-induced change in Oceanic languages in North-West Melanesia. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Areal diffusion and genetic inheritance*, 134–166. Oxford: Oxford University Press.
Saito, Mamoru & Naoki Fukui. 1998. Order in Phrase Structure and Movement. *Linguistic Inquiry* 29(3): 439–474.

Sande, Hannah, Nico Baier & Peter Jenks. 2019. The syntactic diversity of SAuxOV in West Africa. In Emily Clem, Peter Jenks & Hannah Sande (eds.), *Theory and description in African Linguistics: Selected papers from the 47th Annual Conference on African Linguistics*, 667–701. Berlin: Language Science Press.
Schmidt, Andreas & Balázs Surányi. 2019. Hungarian as an OV language. A reappraisal. Paper presented at SOUL 2019, Tartu, 18–19 June 2019.
Scholze, Lenka. 2007. *Das grammatische System der obersorbischen Umgangssprache.* University of Konstanz. (Doctoral dissertation.)
Sheehan, Michelle. 2013. Some implications of a copy theory of labeling. *Syntax*, 16(4): 362–396.
Siewierska, Anna & Dik Bakker. 2009. Case and alternative strategies. In Andrej L. Malchukov & Andrew Spencer (eds.), *The Oxford handbook of case*, 290–303. Oxford: Oxford University Press.
Sinnemäki, Kaius. 2010. Word order in zero-marking languages. *Studies in Language* 34: 869–912.
Sposato, Adam. 2014. Word order in Miao-Yao (Hmong-Mien). *Linguistic Typology* 18: 83–140.
Stepanov, Arthur. 2007. The end of CED? Minimalism and extraction domains. *Syntax*, 10(1): 80–126.
Stone, Gerald. 1993. Sorbian (Upper and Lower). In Bernard Comrie & Greville G. Corbett (eds.), *The Slavonic Languages*, 593–685. London: Routledge.
Stowell, Timothy. 1981. *Origins of phrase structure.* Massachusetts. (Doctoral dissertation.)
Szucsich, Luka & Hubert Haider. 2015. Freie Wortstellung in slavischen Sprachen und die VO/OV-Unterscheidung. In Elena Dieser (ed.), *Linguistische Beiträge zur Slavistik XX*, 94–124. München: Otto Sagner.
Šimík, Radek & Jasinskaja. 2022. There is no single Slavic word order type. *Theoretical Linguistics* 48(1/2): 85–97.
Thiersch, Craig. 1985. *VP and scrambling in the German mittelfeld.* Storrs, CT: University of Connecticut. (Unpublished manuscript.).
Tomlin, Russell S. 1986. *Basic word order: functional principles.* London: Croom Helm.
Tsunoda, Tasaku (ed.). 2020a. *Mermaid construction: A compound-predicate construction with biclausal appearance.* Berlin: de Gruyter Mouton.
Tsunoda, Tasaku. 2020b. Mermaid construction: An introduction and summary. In Tsunoda, Tasaku (ed.), *Mermaid construction*, 1–62. Berlin: de Gruyter Mouton.
Vilkuna, Maria. 2022. Word order. Marianne Bakró-Nagy, Johanna Laakso & Elena Skribnik (eds.), *The Oxford guide to the Uralic languages*, 950–960. Oxford Oxford University Press.
Windfuhr, Gernot & John R. Perry. 2009. Persian and Tajik. In Gernot Windfuhr (ed.), *The Iranian languages*, 416–544. London: Routledge.
Wurmbrand, Susi. 2006. Licensing case. *Journal of Germanic Linguistics* 18(3). 175–236.
Wurmbrand, Susi. 2017. Verb clusters, verb raising, and restructuring. In Martin Everaert & Henk van Riemsdijk (eds.), *The Wiley Blackwell companion to syntax*, 2nd edn., 229–343, Oxford: Blackwell.
Zhang, Qiang. 2021. Linxia fangyan ge biaoji “xa43” tanjiu [An investigation of the case marker “xa43” in Linxia dialect]. *Huainan shifan xueyuan xuebao* 126: 64–73.
Zide, Norman H. 2008. Korku. In Gregory D. S. Anderson (ed.), *The Munda languages*, 256–298. London: Routledge.