Case marking and constituent order variation in Nilo-Saharan: Insights from a typological survey and a case study of Datooga (Nilotic)

Inaugural-Dissertation zur Erlangung des Doktorgrades der Philosophischen Fakultät der Universität zu Köln

im Fach Linguistik

vorgelegt von
Mandy Lorenzen

geb. am 20.03.1996
in Kiel

Köln, September 2024

This is a slightly revised version of my doctoral dissertation that was accepted by the Faculty of Arts and Humanities of the University of Cologne in September 2024. The research for this dissertation was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) *SFB 1252 Prominence in Language*, project B02.

First supervisor: Jun.-Prof. Dr. Alice Mitchell
Second supervisor: Prof. Dr. Birgit Hellwig
Third supervisor: Prof. Dr. Alena Witzlack-Makarevich
Chair of the doctoral committee: Prof. Dr. Daniel Bunčić
Date of the defense of the dissertation (Defensio): December 10, 2024

Acknowledgements

A PhD is a long and challenging journey, and I would not have made it without the support of the following people.

First and foremost, I would like to thank my supervisor, Alice Mitchell, for her guidance and support throughout the different stages of this dissertation. I am grateful for her availability and advice whenever I had questions or needed direction.

I also thank my committee members, Birgit Hellwig and Daniel Bunčić, for their valuable feedback and encouragement. Special thanks go to my third supervisor, Alena Witzlack-Makarevich, who has supported me since my bachelor's degree in 2017, continued through my master's, and accompanied me during the final stages of my PhD. Her input over the years has been immensely valuable.

I am deeply grateful to Monika Feinen for creating the beautiful maps featured in this dissertation. Her work has not only been of great practical help but also enhanced the way this study can be presented visually.

In Tanzania, I am deeply grateful to Herman Malleyeck and his family for their generosity and hospitality during our stay. I would also like to thank all participants who took part in our elicitation study.

My heartfelt thanks go to Trudel Schneider-Blum, to whom I owe a great deal. She always had an open ear and took the time to give me feedback on everything - presentations, posters, chapters - never hesitating to point out critical issues directly and constructively. At the same time, she was always receptive to my concerns and personal struggles. I have learned so much from her guidance and support.

I would like to thank Elli Clages, our student assistant, for her open and enthusiastic approach, as well as the consistent interest and dedication she showed throughout the project. Her work was crucial in making the data I worked with usable and accessible. Additionally, her presence during the fieldwork in Tanzania made the experience both more enjoyable and enriching.

I feel fortunate to have had such supportive peers on this journey. Sharing experiences and struggles was incredibly helpful. I especially want to thank Magdalena Repp and Isabel Zimmer for the many productive hours during our weekly "Days of Writing," as well as Janne Lorenzen (the "real" Lorenzen) and Barbara Zeyer, who accompanied me from beginning to end and stood by my side even during the most challenging times. Many thanks also to the best floor in the

best office building - the second floor of the “House of Prominence.” Every single coffee break with Mark T. Ellison, Sophia Meyer, Timo Buchholz, and Katharina Gayler was a delight! My appreciation also goes to former colleagues from the same floor, who were no longer physically present towards the end but continued to support me in other ways: Maria Bardají – the best office mate (or Mitbüronerin, even if the term never caught on beyond our conversations) I could have wished for – and Fafa Same, who still helped me navigate “crazy data” issues, even though I apparently scared her constantly (unintentionally, of course).

Beyond my immediate academic environment, I am also grateful to Sarah Müller, who pursued her PhD in a completely different field at another university. Our exchange, despite being outside of linguistics, was a source of valuable input and encouragement.

Lastly, I would like to thank my dad. Danke, dass du immer an mich geglaubt hast - und das, obwohl du wahrscheinlich nicht wirklich viel davon verstanden hast, woran ich eigentlich geforscht habe.

Contents

List of Figures

List of Tables

Abbreviations

1	1st person
2	2nd person
3	3rd person
ABS	Absolute/Absolutive
ACC	Accusative
ADV	Adverb
AFF	Affirmative
AGT	Agentive
ANAPH.PRO	Anaphoric pronoun
AP	Antipassive
ASSOC	Associative
ATT	Attenuative
BEN	Benefactive
CAUS	Causative
CF	Centrifugal
COMPL	Completive
COND	Conditional
CONJ	Conjunction
CONT	Continuous
COP	Copula
CP	Centripetal
CVB	Converb
DAT	Dative
DECL	Declarative
DEF	Definite
DEM	Demonstrative
DEP	Dependent
DET	Determiner
DIR	Directional
DIST	Distal
DSC	Discourse marker
DS	Directional suffix
ERG	Ergative
EXCL	Exclusive
E	Evidential
fERG	Focal ergative
FOC	Focus
FUT	Future
GP	General preposition
H	Honorific
HAB	Habitual
HIAF	High affectedness
IMPRS	Impersonal
IMP	Imperative
INCH	Inchoative
INCP	Inceptive
INCOMPL	Incompletive
INDEF	Indefinite
INF	Infinitve
INS	Instrumental
INTERJ	Interjection
IPFV	Imperfective

IR	Individual reference	PRS	Present tense
IS	Inflectional suffix	PST	Past tense
K	K-series agreement	PS	Primary suffix
M	Maculine	PTC	Participle
MIN	Minimal number	P	Patient
mN	Modified noun form	REAL	Realis
MR	Multiple reference	REC	Recent past
NARR	Narrative	RED	Reduced agreement
NEG	Negation	REFL	Reflexive
NMLZ	Nominalizer	REL	Relative
NN	Non-nominative (agreement feature)	RN	Relational noun
NOM	Nominative	R	Repeated action
NONFUT	Non-future	SAP	Speech act participants
NT	New topic	SBJV	Subjunctive
N	Nominative (agreement feature)	SBJ	Subject
OBJ	Object	SEQ	Sequential
OBL	Oblique	SGV	Singulative
PAR	Partitive	SG	Singular
PFV	Perfective	SIML	Simultaneous
PLUR	Pluractional	SS	Secondary suffix
PL	Plural	TEL	Telic
POSS	Possessive	TERM	Terminal
PP	Postposition	TR	Transitive
PREP	Preposition	UR	Unit reference
PRET	Preterite	VEN	Venitive
PRF	Perfect tense	VOC	Vocative
PROX	Proximal		

Chapter 1

Introduction

Although case marking is found in every language phyla across Africa, the phyla vary in the number of languages exhibiting case marking. In East Africa, a notable phenomenon occurs where constituent order influences case marking across a range of more or less related language families. In these languages, arguments that serve as subjects or agents typically receive case marking when they occur after the verb. Conversely, when these arguments appear before the verb, they are no longer marked for case. This phenomenon, although not unique to East African languages, is significant in that area and suggests an areal feature that transcends genealogical relations (Casaretto et al. 2020: 118). This is illustrated with the following example of the Western Nilotic language Päri, an ergative language spoken in South Sudan (Andersen 1988).

(1) Päri (Western Nilotic; South Sudan; Andersen 1988: 292–293)[1]

a. ***ùbúr*** *á-túuk`*
Ubur COMPL-play
'Ubur played.'

b. *dháagɔ̀ á-yàaɲ* ***ùbúrr-ì***
woman COMPL-insult Ubur-ERG
'Ubur insulted the woman.'

c. ***ùbúr*** *dháagɔ̀ á-yáaɲ`-ɛ̀*
Ubur woman COMPL-insult-3SG
'Ubur insulted the woman.'

The intransitive clause in Example (1a) has a canonical SV order, i.e. S *Ùbúr* precedes the verb *túuk`* 'play' and is not marked with case. Concerning the transitive clauses, Päri has a basic OVA constituent order, as illustrated in Example (1b). In this example, A *Ùbúrr* is following the verb *yàaɲ* 'insult' and is marked with the ergative suffix *-ì*. By contrast, in an AOV order where A *Ùbúr* precedes the verb, as in Example (1c), A is not marked with the ergative suffix. Thus, when S and A precede the verb, they are not marked for case in Päri.

This feature, as exemplified by Päri above, is described as the 'no case before the verb' rule by König (2008a: 240) and represents a context in which case marking is neutralized.

[1] The author made the following changes in the glosses: 3SG instead of 3S.

König (2008a) illustrates this rule with data from eight languages: Päri, Teso, Shilluk, Dinka, Chai, Baale, Tennet, and Tima. She further explains that this phenomenon likely arose due to historical processes involving cleft structures and topicalization, which became grammaticalized over time.

The importance of understanding this phenomenon is underscored by the role that constituent order variation plays in these languages, as all of these languages are flexible to some extent. Cross-linguistically, constituent order is influenced by factors such as animacy (Tanaka et al. 2011), argument realization (Dryer 2007), negation (Payne 1995), and the weight of arguments (Arnold et al. 2000). Additionally, discourse pragmatics and information structure (Halliday 1967; Chafe 1979; Lambrecht 1994) also contribute significantly to this variation. Understanding the reasons for constituent order variation is crucial for comprehending the case marking patterns of these languages.

Despite its significance, the contexts driving constituent order variation in these languages remain unclear, with limited studies explicitly addressing this subject. One notable exception is Tima, the only known Niger-Congo language exhibiting case marking that interacts with constituent order. In Tima, constituent order is affected by the animacy of the agent and patient, the visibility of the agent, and the givenness (Schneider-Blum 2023). In the data that emerged from a picture elicitation task, constructions where an animate agent was not fully visible but could be identified as being animate, regardless of the animacy of the patient, OVA order came up more frequently, as in Example (2a). By contrast, if the referent of the A argument was visible, AVO order was preferred, as in Example (2b) (Schneider-Blum 2023: 105).

(2) Tima (Niger-Congo; Sudan; Schneider-Blum 2023: 103)

a. *pʌɽʌŋ=lí ɔ́-páwà ŋ̀=kʌhúnèn ɲ=ìdʌwún*
big.bowl=FOC P-take.hold ERG=woman INS=hands
'There is a big bowl (that) a woman took hold of with both her hands.'

b. *kʌhúnén=lì ɔ̀-kɔ̀kwɛ́ɛ́ pʌɽʌŋ ɲ=ìdʌwún*
woman=FOC P-take:TR big.bowl INS=hands
'There is a woman holding (lit. took) a big bowl with (her) hands.'

While Schneider-Blum (2023) has identified some of the factors determining constituent order in Tima, more comprehensive research is needed to understand these dynamics across a broader range of languages.

1.1 Aims and objectives

This thesis addresses the research gap as described in the previous section by focusing on two main objectives: an extended typological survey of case marking and constituent order variation in the Nilo-Saharan phylum, as well as an in-depth study of constituent order variation in Datooga.

The extended typological survey of the Nilo-Saharan phylum is based on the findings of König (2008a), who introduced the 'no case before the verb' rule. My decision to focus on

this phylum was primarily driven by convenience, as it provided a broad and diverse sample of languages distributed across East Africa. While I am aware that the classification of a Nilo-Saharan phylum has been disputed by a range of scholars (e.g. Güldemann 2018, Hammarström 2018), I make use of this term as a way to capture as broad a selection of languages as possible. For this reason, I chose to include all languages that had been classified as Nilo-Saharan at some point, using Dimmendaal's (2020) classification as my basis.

The research questions for the typological survey are as follows:

> RQ 1: What is the distribution of differential case marking in combination with constituent order variation in Nilo-Saharan languages?
>
> RQ 2: Which features are relevant for the variation, and how do they interact with each other?
>
> RQ 3: What is the impact of information structural categories, such as topic and focus, when it comes to constituent order variation?

The typological survey aims to build on the findings of König (2008a) by extending the analysis to language descriptions that have become available after 2008. This survey will also provide a quantitative analysis of all available grammars and other language descriptions of the Nilo-Saharan phylum, focusing on how the conditions related to variation are described and what can be deduced from the study of the provided examples.

As constituent order variation is a key property of languages exhibiting the 'no case before the verb' rule, I focus on this phenomenon in the second part of the book. Chapters 4 and 5 concern the in-depth study of constituent order variation in one of the Nilo-Saharan languages, namely the Southern Nilotic dialect cluster Datooga, spoken in Tanzania. Datooga is generally described as a verb-initial language which allows for some degree of flexibility in constituent order (Kießling 2007: 171 for Gisamjanga-Datooga; Griscom 2019: 31 for Asimjeega-Datooga). According to Kießling (2007: 171), constituent order "can be manipulated for pragmatic purposes". However, the exact reasons for these pragmatic purposes have remained unclear so far. This thesis will therefore provide an in-depth study of constituent order variation in a corpus of the Datooga dialects Barabaiga and Gisamjanga.

For the Datooga study the following research questions were pursued:

> RQ 1: What are the conditions for constituent order alternation in Datooga?
>
> RQ 2: What is the impact of information structural categories, such as givenness, topic, and focus, for the different constituent orders in Datooga?
>
> RQ 3: How does constituent order vary across different text genres, i.e. monologues and multi-party conversations?

The research questions for the typological survey and the in-depth Datooga study have in common that they focus on the factors important to the variance in case marking and constituent order variation for the survey and constituent order variation in the Datooga study, respectively. Both studies attempt to further investigate the role of information structure on constituent order variance.

1.2 Overview on data

To address the research questions of the typological survey, information on the influence of information structure is inferred from grammars and other language descriptions. In total, data were drawn from 115 languages. All languages considered in this survey are listed in the Appendix A. For a more detailed description of the data as well as the methods used, see Section 3.2.

In contrast to the typological survey, the Datooga study draws on analyses of existing corpus data from multiparty conversations and monologic narratives, as well as on stimuli-based data, as summarized in Table 1.1. In total, the dataset comprises 1959 clauses from nearly two hours of recorded speech produced by more than six speakers.

File name	**Content**	**Length**	**Word count**	**# clauses**	**Speakers**
Rituals for a new born	Description of rituals conducted for a newborn baby	00:04:42	427	119	UD
Sheep, beam, and dog	Dilemma tale	00:01:27	127	43	UX
Donkey and hyena	Short folk tale	00:01:08	68	28	HM
Samoota	Folk tale	00:06:20	561	242	UD
Qeambeemeeda	Folk tale	00:06:54	677	222	YG
Honey beer	Procedural text	00:04:40	476	130	YG
Yonas hulanda	Multi-party 'visiting' talk	00:06:20	1104	368	multiple
Maandazi	Domestic interaction	00:55:11	1769	505	multiple
HM videos	Video stimuli based on QUIS	00:06:35	359	96	HM
UN videos	Video stimuli based on QUIS	00:08:09	271	85	UN
MA videos	Video stimuli based on QUIS	00:07:33	456	121	MA
	Totals	01:51:59	6295	1959	

Table 1.1: Annotated texts of the corpus; green highlighted texts: monological narratives, orange highlighted texts: multi-party conversations, red highlighted texts: elicited data.

The two datasets therefore offer different but complementary perspectives: the typological survey provides a broad cross-linguistic overview based on grammatical descriptions, whereas the Datooga study allows for a more detailed qualitative examination of information-structural effects based on recorded speech data.

1.3 Content and structure of this book

This book is structured as follows: Before presenting the two studies, Chapter 2 lays the groundwork by introducing the key concepts, case marking and constituent order variation, as well as terminology related to this topic.

Chapter 3 is concerned with the typological survey of Nilo-Saharan languages conducted in this dissertation. The chapter also includes an overview of case marking and constituent order variation specific to East Africa, including previous findings of König (2008a) on the 'no case before the verb' feature.

The following Chapter 4 provides an overview of the Datooga language and its speakers. This includes information about the speakers, as well as on the typological profile, which is necessary to follow the forthcoming study of constituent order variation. In addition, this chapter provides an introduction to syntax on clause structures that are not in the focus of the subsequent study, including constituent order in subordinate clauses, as well as structures allowing for a preverbal placement of entities which are non-arguments and thus not relevant for the future study, such as dislocation and cleft structures. As will be further discussed in Chapter 3, dislocation and cleft structures are argued to be a likely origin for grammaticalized case loss of S and A arguments in preverbal but not postverbal position. Therefore, it is necessary to introduce these structures and show their forms to distinguish them from those relevant to the subsequent analysis.

The study of constituent order variation in Datooga declarative main clauses is the subject of Chapter 5. In this chapter, the conditions for variation that played a role in the previous chapters are analyzed for their importance in constituent order variation in Datooga. It further includes a qualitative study of the information structural categories topic, focus, and contrast.

Finally, Chapter 6 offers the conclusions of this dissertation by summarizing the main findings and providing an outlook for future research.

Chapter 2

Preliminaries

The current chapter establishes the foundation for the rest of the dissertation, aiming to discuss crucial concepts and define the terminology to be employed in subsequent sections. Two pivotal components, namely case marking and constituent order (variation), hold particular significance in the languages examined in this work.

Section 2.1 offers an overview of case marking, pertinent alignment patterns, and the concept of differential argument marking, all of which are crucial for this dissertation. Section 2.2 delves into constituent order, covering aspects such as basic constituent order, flexibility in constituent order, and the factors influencing variation observed in languages globally. This section places special emphasis on information structural categories, including topics, focus, and givenness, which are frequently associated with changes in constituent order.

2.1 Case marking

This section serves as an introduction to various elements related to case marking essential for comprehending the ongoing dissertation. Firstly, in Section 2.1.1, a definition of grammatical case, as applied in this dissertation, is provided. Following that, Section 2.1.2 offers an overview of alignment patterns observed across the world's languages, with a specific focus on patterns pertinent to this dissertation, namely, accusative, ergative, and marked nominative alignments. Lastly, Section 2.1.3 delves into the discussion of differential argument marking.

2.1.1 Definition of case

Case is a system of indicating the dependency of nouns or noun phrases on the larger structures they belong to. At the clause level, this dependency is between noun phrases and verbs, whereas at the phrase level, case also marks the relationship of a noun to adpositions or other nouns (Blake 2004: 1). Case is commonly seen as an inflectional system that may be expressed through tone, reduction of roots, affixes, or accent shift. König (2008a) further includes adpositions "only insofar as they encode core participants such as S, A, and O" (König 2008a: 5). Since the main

concern of my study is on the case marking of these core participants, I will follow König (2008a) and thus also include adpositions—if they encode core participants—into my definition of case.

In the literature on case marking, there are a few terms that need to be distinguished when speaking of case. Following Blake (2004: 2–3), these terms are *case markers* and *case forms*. According to him, a case marker is the marker that is applied to indicate case, whereas the case form refers to the whole inflected word. This distinction will be necessary in later chapters of this dissertation, as case marking is often indicated by tonal inflection in languages of East Africa, including the language of the later case study in this work, Datooga. In autosegmental phonology, tone patterns are treated as separate entities that get imposed on the noun and thus should be considered case markers. However, with tonal inflection, it may be analytically more straightforward to talk about case forms rather than case markers.

Another distinction that must be made is that between *case systems* and *case patterns*. Following König (2008a: 16), the term 'case system' is restricted to grammaticalized case in the sense of inflection. 'Case patterns', by contrast, can also refer to alignment patterns concerning the syntactic relations S, A, and O in contexts outside of grammaticalized case, e.g. constituent order. English is an example of a language that expresses syntactic function by means of word order and where a word order change results in a change in the assignment of grammatical relations. For example, the sentence *The woman saw the dog* has a different meaning than *The dog saw the woman* in English. As S or A appear before the verb while O follows the verb, English shows an accusative pattern for its word order (Dixon 1994: 10–11). Core case inflection in English is absent, however, except within the pronominal paradigm.

Linguists have traditionally identified three primitive syntactic relations in language: S (intransitive subject), A (transitive subject), and O[1] (transitive object) (Dixon 1994: 6). These syntactic relations are especially relevant when it comes to case patterns or systems (see Section 2.1.2). While S refers to the single core argument of an intransitive verb, it is not always easy to distinguish the two core arguments of a transitive clause, A and O. One way to distinguish these two syntactic relations is on a semantic basis. Following Dixon (1994: 8), there are certain semantic roles that can be associated with either A or O depending on the semantic nature of the verb. Dixon (1994: 124) defines A as the role that "is *most relevant to the success* of the activity (normally human, and this then equates with: *could initiate or control* the activity)". In turn, O is "the *other core* role or—if there are more than one non-A core roles—that role which is *most saliently affected* by the activity" (Dixon 1994: 124).

Using the syntactic relations S, A, and O rather than grammatical functions such as subject and object allows us to make distinctions that are really important for describing languages with ergative case systems. According to Dixon (1994: 111), there are certain semantic properties associated with subject NPs, such as control and agentivity, as well as grammatical criteria, such as their obligatoriness in sentences, their ability to be the pivot in coordinative and subordinative constructions, or that they can be in cross-reference with the verb. If S and A align, these

[1]Other scholars, such as Comrie (1989), use P instead of O. In the following, I will stick to the terminology of Dixon (1994) and use O.

properties are met. However, this notion of 'subject' developed in linguistic theory within the context of more familiar European languages, which predominantly exhibit accusative alignment (see Section 2.1.2 for an overview of alignment patterns).

In languages with ergative case systems where S and O align, these properties are not shared to the same extent, and the notion of subject becomes more problematic. One reason for this are the mentioned semantic properties that are typically associated with subjects in accusative languages, i.e. having the semantic role of an 'agent'.

Following Dowty (1991) 'agents' and 'patients' are the only semantic roles that are actually necessary for argument selection, having the canonical features of agentivity and control for the agent, and affectedness for the patient (Dowty 1991: 571–572). In his work he defines them as *Proto-Agent* and *Proto-Patient* which are illustrated in Table 2.1.

Proto-Agent:	**Proto-Patient:**
• volitional involvement in the event or state	• undergoes change of state
• sentience (and/or perception)	• incremental theme
• causing an event or change of state in another participant	• causally affected by another participant
• movement (relative to the position of another participant)	• stationary relative to movement of another participant
• (exists independently of the event named by the verb)	• (does not exist independently of the event, or not at all)

Table 2.1: Semantic features of Proto-Agents and Proto-Patients, adapted from Dowty (1991: 572). The last feature for each, Proto-Agents and Proto-Patients, is in parentheses because of the uncertainty for the proto-role definition.

The properties of a Proto-Agent and a Proto-Patient are said to be semantically independent from each other, and not all Proto-Agent properties will apply for the arguments of a given verb. Nonetheless, in some cases all the properties are relevant, such as the English verb 'build' (Dowty 1991: 572). When I speak of agents or patients in the following course of the dissertation, it is meant in the sense of Dowty's (1991) Proto-Agent and Proto-Patient definition.

In accusative languages, S and A behave alike in terms of case marking, and both syntactic relations can serve the semantic function of the agent. While the latter is also true for ergative languages, i.e. S and A generally have the semantic function of an agent, S and A act differently in terms of case marking. A typically receives a dedicated case marker, while S and O stay unmarked, thereby, S and O align. This distinction in case marking might make it difficult to apply the usual definition of 'subject' to ergative languages. In the following I will therefore

avoid the term subject while describing ergative languages and either use the syntactic relations S, A, and O, or refer to S and A as agents.

2.1.2 Alignment patterns

The alignment or grouping of the previously introduced core participants S, A, and O is important for the subsequent study, as it deals with different case-marking systems that are distinguished in terms of the alignment of S, A, and O. Linguists have identified five typological alignment patterns across the world's languages. In the first two patterns, S either aligns with A, as in accusative patterns, or with O, as is the case for ergative patterns. The neutral pattern treats S, A, and O the same. Another option is the three-way or tripartite pattern in which S, A, and O are all treated differently from each other (Dixon 1994: 39). The last logical possibility is a system in which A and O align and are distinct from the intransitive subject S. However, this alignment, which is called double-oblique or horizontal alignment (Haude & Witzlack-Makarevich 2016), is very uncommon and can so far only be found in some Iranian languages of the Pamir region (Payne 1980).

Figure 2.1 illustrates these alignment patterns. As indicated in the previous section, alignment patterns are relevant for both case patterns, i.e. in contexts outside of grammaticalized case, such as constituent order, but also for case systems, i.e. with grammaticalized case. In the following, I will focus on case systems rather than case patterns.

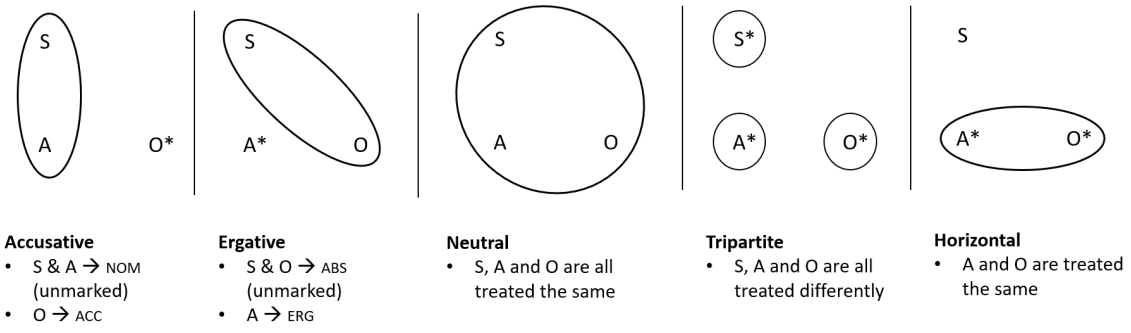

Figure 2.1: Alignment patterns found in the world's languages. The asterisk indicates the case-marked argument(s). Keep in mind that the unmarked case in accusative and ergative languages is usually the nominative or absolutive case, respectively. However, there are certain languages in which both cases in both systems get dedicated case markings.

The case systems most relevant to this thesis are introduced in the following subsections. The first subsection looks at accusative case systems, followed by a subsection on ergative case systems. Marked nominative case systems, which combine elements of accusative and ergative systems, are covered in the last subsection.

2.1.2.1 Accusative alignment

According to Dixon (1994: 40), the accusative alignment pattern is the most frequent pattern among languages that apply case marking. In accusative case systems, S and A are treated in the same way while O behaves differently. S and A receive what tends to be called the nominative case, which is usually the unmarked case. Unmarked here means that the case form is not different from the citation form of a word. By contrast, the transitive object O receives dedicated accusative case marking. An example of a language with an accusative case system is Hungarian, as in Example (3).

(3) Hungarian (Uralic; Hungary; Kenesei et al. 1998: 195, 208)

a. *jö-tt* *egy levél az igazgaló részére*
come-PST.INDEF.3SG DET letter DET director BEN
'A letter came for the director.'

b. *a lány ir-ja* *a level-**et***
DET girl write-DEF.3SG DET letter-ACC
'The girl is writing a letter.'

In the intransitive clause in Example (3a) S *levél* 'letter' is not marked with dedicated case which is also true for the noun phrase *lány* 'girl' functioning as A in the transitive clause in Example (3b).[2] In Hungarian, accusative case is marked with the suffix *-et* which is applied on O *level* 'letter' in the transitive clause.

Case-marking forms that index argument roles can serve additional functions. Accusative marking, for example, can also attach to adverbials in many languages, such as in Finnish in Example (4) where the accusative case marker *-n* applies to adverbials of duration and distance.

(4) Finnish (Uralic; Finland; Kittilä & Malchukov 2008: 553)

juoks-i *tunni-n/kilometri-**n***
run-3SG.PST hour-ACC/kilometre-ACC
'ran for an hour/a kilometre'

As pointed out by Kittilä & Malchukov (2008: 553), in other languages accusative marking can also be found on manner adverbs, such as in Standard Arabic or on nominal predicates as is the case for Ge'ez. In Even accusative marking is further applied to nominalized complement clauses (Malchukov 2005). Other contexts of accusatives include the marking of infinitives (e.g. Quechua; Adelaar & Muysken 2004) or the indication of switch-reference (e.g. Kosati; Kimball 1991).

According to WALS (World Atlas of Language Structures) (Comrie 2013b,a) languages with accusative case systems are distributed all over the world and can also be found in languages of Africa, including some branches of what is called the Nilo-Saharan phylum (e.g. Fur (Jakobi 1990), Runga (Nougayrol 1990), or Kunama (Bender 1996)).

[2] The stress marker (´) in Hungarian, as in *levél* or *lány*, does not indicate case (Kenesei et al. 1998: 428–430), in contrast to tonal inflection, which can be a case marker in languages covered in the work that follows.

2.1.2.2 Ergative alignment

In terms of alignment patterns, ergative case systems behave differently to accusative case systems in that S and O are treated the same instead of S and A. In languages with ergative case systems, the case for A is typically called ergative while S and O receive absolutive case. In these languages it is usually the absolutive case that is formally unmarked and it is the form that is commonly used as the citation form.

An example of a language with an ergative case system is Hunzib. The case marking in Hunzib is illustrated in the two clauses below. The S argument *ože* 'boy' in the intransitive clause in Example (5a) does not receive dedicated case marking. By contrast, in the transitive clause in Example (5b) the A argument *už* 'boy' is morphologically distinguished from the other core argument O *kid* 'girl' with the ergative suffix *-l*. Additionally, A differs from S in the preceding phrase due to the oblique singular suffix *-di*, which results in the loss of the *e* in the stem. In terms of case marking, S *ože* in Example (5a) and O *ože* in Example (5c) are aligned and differ from A *už-di-l* in Example (5b).

(5) Hunzib (Daghestanian; Russia, Georgia; van den Berg 1995: 80, 122)

a. *ože ut'-ur*
boy sleep-PRET
'The boy slept.'

b. *už-di-**l*** *kid hehe-r*
boy-OBL-ERG girl hit-PRET
'The boy hit the girl.'

c. *iyu-l* *ože gəl-ər*
mother-ERG boy put.down-PRET
'Mother put the boy down.'

Just like accusative marking, ergative marking can have functions beyond case marking. According to Palancar (2008: 568), ergative case can additionally be used to mark instruments in certain languages. In languages that mark instruments with the ergative marker, it can further be used to indicate cause in intransitive constructions. This is for instance the case for Sanumá in Example (6). In Example (6a) the agent argument *kamisamakö* '1PL.EXCL' is marked with the ergative case marker *-nö*. However, Example (6b) and Example (6c) show that instruments or causes can also be marked with the same marker *-nö* that is used for agents.

(6) Sanumá (Yanomamic; Brazil, Venezuela; Borgman 1990: 29, 123)

a. *kamisamakö-**nö*** *hama sama* *töpö se kite*
1PL.EXCL-ERG/AGT visitor 1PL.EXCL 3PL hit FUT
'We will hit the visitors.' (Agent)

b. *kusiali te-**nö*** *sa ia pia* *kule*
spoon 3SG-ERG/INS 1SG eat intend PRS
'I am about to eat with a spoon.' (Instrument)

c. *kamakali te* *wasu-**nö*** *ipa ulu a* *noma-so-ma*
high.fever 3SG deadly-ERG/CAUS my son 3SG die-FOC-COMPL
'My son died from a deadly high fever.' (Cause)

Another function of ergative, although less frequent than the marking of instruments, is the expression of possessors (e.g. Ladakhi; Koshal 1979). In some Australian languages ergative can further mark spatial categories. In languages with nominal case, such as in North West Caucasian languages, ergative can further serve different functions which results in it being seen as an 'oblique' marker. In addition to its locative function, ergative can also mark adverbial functions, dative participants and other demoted patients (Palancar 2008: 569).

The geographical distribution in WALS (Comrie 2013b,a) shows that languages with an ergative case system can mostly be found in Australia, Austronesia as well as South-America and South-Asia. In Africa languages with ergative case-marking systems are not frequent, according to Creissels et al. (2007: 90). However, as the typological survey in Chapter 3 will show, there are languages, especially in East Africa, with ergative systems.

2.1.2.3 Marked nominative

Marked nominative systems share features of both ergative and accusative case systems. Like accusative languages, S and A are formally aligned, i.e. marked in the same way. However, unlike accusative languages, it is S and A rather than O that receive dedicated case marking. Languages with a marked nominative system are then similar to ergative languages because O is functionally unmarked in both case systems (König 2006: 657). Figure 2.2 illustrates the similarities of marked nominative systems to both accusative and ergative systems.

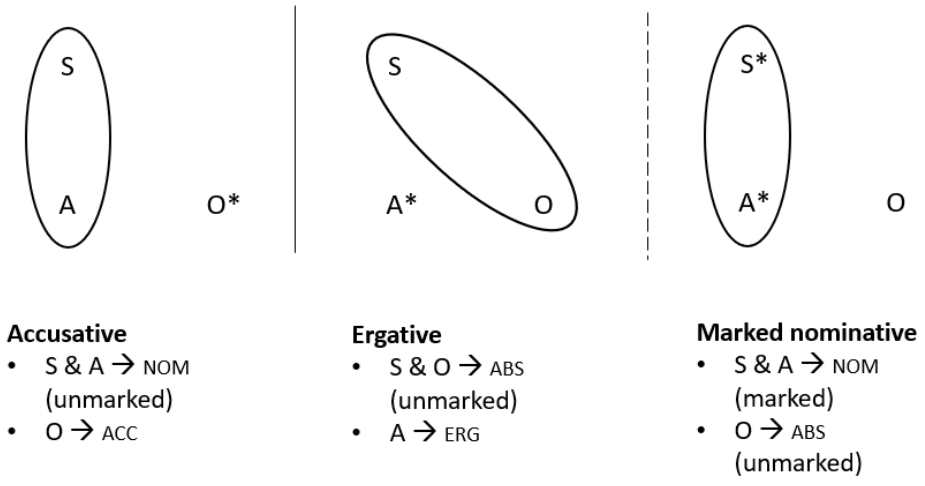

Figure 2.2: The marked nominative pattern in comparison to the accusative and ergative patterns. The asterisk indicates the case-marked argument(s). Keep in mind that the unmarked case in accusative and ergative languages is usually the nominative or absolutive case, respectively. However, there are certain languages in which both cases in both systems get dedicated case markings.

In the literature there is some confusion about the terminology especially with regard to the case forms. First of all, the term for the whole case system *marked nominative* is difficult.

Marked nominative is sometimes referred to as 'extended ergative' (e.g., Andersen 1988) as it could be argued that the overt case marking for A extends to S in intransitive clauses. For Dixon (1994: 64) both terms, marked nominative and extended ergative, are difficult as the case system in question behaves differently than accusative and ergative systems but according to him, the term marked nominative is the less confusing of the two. With regard to the terminology for the cases there is even more disagreement in how to name them. König (2006, 2008a) uses 'accusative' for the unmarked case or the citation form in marked nominative systems. However, other scholars, such as Kießling (2007), prefer the term 'absolute' to refer to the unmarked case. He further argues that this term is less confusing than terms like 'absolutive' or 'accusative' as they are already used for accusative and ergative alignment patterns and therefore should be avoided (Kießling 2007: 152). In the following, I will follow this proposal and use the term 'absolute' to refer to the unmarked case in marked nominative systems.

An example of a language with a marked nominative case system is the Southern Nilotic language Cherang'any, spoken in Kenya. The marked nominative case system in Cherang'any is illustrated in Example (7).

(7) Cherang'any (Southern Nilotic; Kenya; Mietzner 2016: 62, 181)

a. *riré* *lɛ̀kwɛ̀ɛ̀t*
3SG.cry child.DEF.NOM
'The child is crying.'

b. *ìrɔ̀-í* *lɛ̀kwɛ̀ɛ̀t* *tɛ̀ɛ̀tá*
see-IPFV child.DEF.NOM cow.DEF.ABS
'The child sees the cow.'

c. *á-rʌ̀↓í* *lɛ̀kw↓ɛ́ɛ́t*
1SG-see-IPFV child.DEF.ABS
'I see the child.'

In Cherang'any, case is marked through tonal inflection, i.e. each case has its own tone pattern. The S argument, *lɛ̀kwɛ̀ɛ̀t* 'child', is marked with the nominative case in Example (7a). The same tone pattern occurs with A in the transitive clause in Example (7b) while O, *tɛ̀ɛ̀tá* 'cow' is in the absolute case which is also the citation form in Cherang'any (Mietzner 2016: 212). Example (7c) shows the absolute case form of the noun 'child' which differs from the nominative tone pattern.

Marked nominative languages can be found in different language families of Africa and especially north-east Africa. This includes languages of the Cushitic, Omotic, Berber, Surmic families, as well as Nilotic languages (Dixon 1994: 64; Kießling 2007: 151). Outside of Africa marked nominative is also attested in the Yuman family of California (Dixon 1994: 65), as well as in Savosavo spoken on the Solomon Islands (Wegener 2012).

2.1.3 Differential argument marking

This picture of case systems is further complicated by the fact that arguments can behave differently depending on various linguistic contexts, which brings us to the discussion of differential

argument marking. The thesis addresses differential argument marking in the context of the typological survey in Chapter 3, so I will explore this topic in some detail here. Differential argument marking (DAM) refers to the situation in which an argument of the verb is variably marked, i.e. through dedicated case marking or indexing on the verb, depending on various factors. In the following I will use the broad definition of DAM by Witzlack-Makarevich & Seržant (2018):

> Any kind of situation where an argument of a predicate bearing the same generalized semantic argument role may be coded in different ways, depending on factors other than the argument role itself, and which is not licensed by diathesis alternations. (Witzlack-Makarevich & Seržant 2018: 3)

According to Witzlack-Makarevich & Seržant (2018), the concept of DAM covers a range of phenomena. The first mention of differential marking itself occurred in the context of *differential object marking* which concerns different marking of object and which has so far received the most attention in research (Bossong 1982, 1985). Compared to differential object marking, research has only recently began to discuss differential marking on roles other than objects, such as *differential subject marking* (e.g. de Hoop & de Swart 2009). However, the notion of 'subject' is not unproblematic as has been pointed out in Section 2.1.1. Hence, Fauconnier (2011) proposes the concept of *differential agent marking* to describe the differential case marking of A in transitive clauses exclusively. In contrast, *differential subject marking* describes the differential case marking of S in intransitive clauses in her terminology. There is another frequently discussed phenomenon that is related to DAM and is even older than differential object marking, which is *split*, as in *split ergativity*. This concept was introduced by Silverstein (1976) and then made popular by Dixon (1979, 1994). In the context of ditransitive clauses, distinct approaches are employed to differentiate the case marking of the various arguments involved in these constructions. Terms such as *differential recipient marking, differential theme marking*, or *differential goal marking* have been used in the literature to categorize the differential case marking patterns in ditransitive clauses (Haspelmath 2007, Kittilä 2008).

Other phenomena that involve a change in case marking and fall under the broad definition of DAM given above, are *optional* and *alternating case marking*. However, following Chappell & Verstraete (2019: 8), optional and alternating case marking should be differentiated from other phenomena covering changes in case-marking, i.e. reference-based splits and construction-based splits. McGregor (2010: 1610) defines optional case marking as "the situation in which, in specifiable lexical or grammatical environments, case marking morpheme (inflectional affix, clitic, or adposition) my be either present or absent from an NP of a specifiable type without affecting the grammatical role borne by the NP." By contrast, alternating case marking is "the situation where two overt case markers alternate in the same environment, similarly without

affecting grammatical roles" (Chappell & Verstraete 2019: 2).[3] Thus, with optional case marking there is a contrast between the absence and presence of at least one overt case marker, while alternating case marking contrasts between at least two different overt case markers for the coding of the same grammatical role. According to Chappell & Verstraete (2019: 10), both, optional and alternating case marking, are conditioned by information structure. This contrasts with reference- and construction-based splits which are conditioned by lexical or grammatical properties of either the referent or the whole construction Chappell & Verstraete (2019: 3–4), such as the ones discussed in the following.

As described in Witzlack-Makarevich & Seržant (2018: 4), a range of factors can trigger differential marking on arguments. First of all, the arguments themselves can initiate DAM. Features of a single argument or a combination of features of multiple arguments within a clause may influence differential marking. Argument-triggered DAM is distinguished into inherent and non-inherent factors. Inherent properties of arguments include animacy (Silverstein 1976, Bossong 1991, Aissen 2003), person (Dixon 1979, Croft 2003), discreteness (Bossong 1991), uniqueness (Croft 2003), number, as well as argument realization or part-of-speech-distinction (pronoun vs. noun) (Bickel et al. 2015). These properties are often scaled on so-called hierarchies (Witzlack-Makarevich & Seržant 2018: 6). An example of such a hierarchy is the *potentiality of agency scale* by Dixon (1979):

> 1st pronouns > 2nd person pronouns > 3rd person pronouns > proper names > humans > animates > inanimates

This hierarchy is based on Silverstein's (1976) frequently discussed animacy hierarchy (hierarchy of "inherent lexical content of noun phrases" Silverstein 1976: 113) and shows the potential of a pronoun or noun phrase to be in A function. Similar versions to this hierarchy are the empathy hierarchy (DeLancey 1981), the hierarchy of reference (Zwicky 1977), or the personal hierarchy (Siewierska 1993).

An example of a language where case split is conditioned by the animacy hierarchy as discussed above is Dyirbal. Dyirbal is a split-ergative language which shows an ergative alignment for case marking on nouns while first and second person pronouns show an accusative pattern (Dixon 1994: 14), as illustrated in the clauses below.

(8) Dyirbal (Pama-Nyungan; Australia; Dixon 1994: 10, 14)

a. *yabu* *banaga-n^{y}u*
mother.ABS return-NONFUT
'Mother returned.'

b. *nyuma* ***yabu-ŋgu*** *bura-n*
father.ABS mother-ERG see-NONFUT
'Mother saw father.'

[3] In the literature, the distinction between optional and alternating case marking is also found under the terminology of asymmetric vs. symmetric case marking (e.g. de Hoop & Malchukov 2008, McGregor 2010, Iemmolo 2013).

c. *ŋana* *banaga-n^{y}u*
we all.NOM return-NONFUT
'We returned.'

d. *ŋana* *n^{y}urra-na* *bura-n*
we all.NOM you all-ACC see-NONFUT
'We saw you all.'

In the intransitive clause in Example (8a), S ,*yabu* 'mother', is in the absolutive case which is unmarked. S is marked like O, *nyuma* 'father' in the transitive clause in Example (8b) as both syntactic roles are unmarked while A receives ergative case marking with the suffix *-ŋgu*. The Examples (8c) and (8d) show the accusative alignment with first and second person pronouns. In these clauses, S and A ,*ŋana* 'we all', are in their nominative form which is unmarked whereas O, *n^{y}urra* is marked with the accusative suffix *-na*.

Witzlack-Makarevich & Seržant (2018: 6) point out that these inherent characteristics should not be viewed as the only prerequisites for DAM, though. Actually, in most languages, DAM is more dependent on the interactions of several features.

Non-inherent properties, on the other hand, are based on the discourse context and include definiteness and specificity, as well as information structural features (Witzlack-Makarevich & Seržant 2018: 9). Warrwa is an example of a language in which discourse properties are a condition for DAM. There are two ergative markers in Warrwa. The first one *-na* is used when the agent is already established and expected, such as in Example (9a). The other ergative marker *-nma*, which can be seen in Example (9b), occurs whenever the agent is unexpected (McGregor 2006: 402).

(9) Warrwa (Nyulnyulan; Australia; McGregor 2006: 402)

a. *nyinka jurrb ø-ji-na-yina* *kinya wanyji kwiina iri,*
this jump 3min.NOM-say-PST-3min.OBL this later big woman
ka-na-ngka-ndi-ø *ø-ji-na,* *kinya-**na** wuba,*
1min.NOM-TR-FUT-get-3min.ACC 3min.NOM-say-PST this-ERG small
'The little one jumped at her then, at the big woman, and tried to get her.'

b. *kinya kwiina-**nma** iri* *marlu laj* *ø-ji-na-ø*
this big-fERG woman not throw 3min.NOM-say-PST-3min.ACC
kinya wuba, laj, *marlu laj* *ø-ji-na-ø,*
this little throw not throw 3min.NOM-say-PST-3min.ACC
'But no, the big woman threw the little man away.'

It is further possible that not only the properties of one single argument in a clause cause a split but rather that it is triggered by a combination of properties of different arguments, which is sometimes referred to as a scenario (Witzlack-Makarevich & Seržant 2018: 12). An example of this comes from Aguaruna. In the first Example (10a), O *nĩ* '3SG' does not have a dedicated case marker. In the other two examples (10b) and (10c) O is marked with the accusative case marker *-na*.

(10) Aguaruna (Chicham; Peru; Overall 2007: 443, 444)

a. *nĩ ~~ii~~ma-ta*
3SG carry.PFV-IMP
'You(SG) carry him!'

b. *ata u-na yu-a-tata-ha-i*
chicken-ACC eat.HIAF-FUT-1SG-DECL
'I will eat chicken.'

c. *hutii a-ina-u-ti daka-sa-tata-ham -i ami-**na***
1PL COP-PL.IMPFV-REL-SAP wait.for-ATT-FUT-1SG>2SG.OBJ-DECL 2SG-ACC
'We will wait for you.'

In Aguaruna the case marking of O is not only conditioned by its own properties but also by that of A. If A is 1SG or third person, O is always marked with accusative case, such as in Example (10b). By contrast, if A is 1PL or second person, O only receives accusative case when it is higher on the animacy hierarchy than A, such as in Example (10c) where A is 1PL and O is 2SG (Overall 2007: 443).

In addition, event semantics are included in the category of argument-triggered DAM. Features that are important to mention here are for example volitionality, agentivity, as well as affectedness (Næss 2004, McGregor 2006, Fauconnier 2012). Volitionality, agentivity and control are typically triggers in the context of differential subject marking whereas affectedness can be a condition for differential object marking (Witzlack-Makarevich & Seržant 2018: 14). In Tsova-Tush, 1SG pronouns have different forms depending on whether an action was intended or not. In Example (11a) the falling was the fault of the first person agent and therefore takes the ergative form. However, in Example (11b) the falling was not the fault of the first person agent which is why the argument is in the nominative form.

(11) Tsova-Tush (Nakh; Georgia; Holisky 1987: 105)

a. ***as*** *wože*
1SG.ERG fell
'I fell.' (It was my own fault that I fell down.)

b. ***so*** *wože*
1SG.NOM fell
'I fell.' (No implication that it was my fault.)

Further, DAM can be triggered by features of the predicate. Apart from DAM that is triggered by the properties of an argument without a change of the predicate, predicate-triggered DAM occurs when the form of the predicate changes. One of such features is the clause type. There are languages where arguments are marked differently depending on whether they occur in a main or a dependent clause. In Maithili S and A in main clauses are in the nominative case *ham* '1NOM', as illustrated in the intransitive main clause in Example (12a). In converbial clauses, S and A are in the dative case, which is illustrated for S *hamrā* '1DAT' in Example (12b).

(12) Maithili (Indo-Aryan; India, Nepal; Bickel & Yādava 2000: 350, 353)

a. ***ham*** *cail-je-b-ah*
1NOM move-TEL-FUT[1N]-2mhNN
'I will go away (if you don't want me to stay).'

b. *[**hamrā*** *ghar āib-kẽ] pitā-jī khuśī he-t-āh*
1DAT home come-CVB father-hNOM happy be(come)-FUT-3hN
'When I come home, father will be happy'

Other frequent conditions in terms of predicate-triggered differential marking are tense, aspect and mood. In some languages, arguments are for example marked differently depending on the tense. An example of this is Georgian. In the present tense A *Vano* occurs in the nominative form with no dedicated case marker, such as in Example (13a). However, in the aorist tense A *Vano* receives the ergative case marker *-m*, as in Example (13b).

(13) Georgian (Kartvelian; Georgia; Harris 2009: 40)

a. ***vano*** *pikrobs mariḳaze*
Vano.NOM he.thinks.3 Marika.on
'Vano is thinking about Marika.'

b. *vano-**m** ipikra mariḳaze*
Vano-ERG he.thought.3 Marika.on
'Vano thought about Marika.'

Additionally, the polarity of a clause can lead to DAM. This is for example the case in Finnish. As Example (14a) shows, the O argument in an affirmative clause can either receive accusative *-n* or partitive case marking *-a*. However, negative clauses only allow for O to be in partitive case as in Example (14b).

(14) Finnish (Uralic; Finland; Sulkala & Karjalainen 1992: 115)

a. *söln omena-**n** / omena-**a***
eat.IPFV.1SG apple-ACC / apple-PAR
'I ate/was eating an apple.'

b. *en syönyt omena-**a***
NEG.1SG eat.2PTC apple-PAR
'I didn't eat/was not eating an apple.'

When it comes to predicate-triggered DAM, information structure can play a role which is the case in the following examples from Somali. In the first Example (15a) A is not in focus and marked with the subject clitic *=u* and with verbal agreement. Compared to this, the A argument in the second clause Example (15b) is in focus which can be seen by the preverbal focus marker *báa* and reduced verbal agreement. Further, A does not take the subject clitic *=u* (Green 2021: 309).

(15) Somali (Cushitic; Somalia; Green 2021: 310)

a. *adí=g=**u** w=áad rab-t-aa*
2SG=K.DEF=SUBJ DECL=2SG want-2SG-PRES
'You want it.'

b. *adí=ga* *báa* *rab-á*
2SG=K.DEF FOC want-PRS.RED
'YOU want it.'

The languages examined in this study exhibit another factor contributing to case splits, which is variation in constituent order. The specific case split phenomenon observed in these languages is characterized by the marking of case on postverbal S and/or A, while preverbal S and/or A typically do not receive case marking. This feature is not only reported for languages with case split but is also observed in languages with optional ergativity, such as in certain Australian languages (e.g. Meakins & O'Shannessy 2010). This distinctive feature adds to the complexity of case marking patterns in these languages and is explored in further detail in the subsequent chapters of the work.

2.2 Constituent order variation

Constituent order holds significant importance in the context of case marking within the languages covered in this study. It will also be a central focus in the case study of Datooga later in Chapter 5. Therefore, it is crucial to clarify several notions in advance.

Section 2.2.1 initiates the discussion by examining the concept of 'basic constituent order' across languages worldwide. Following that, Section 2.2.2 delves into the flexibility observed in constituent order. Finally, Section 2.2.3 explores the conditions relevant to variation in languages globally, encompassing factors such as information structural categories like topic, focus, or givenness.

2.2.1 Basic constituent order

In studying constituent order variation, linguists often attempt to establish a "basic constituent order" for a given language. Traditionally, this attempt is based on the 6-way distinction coined by Greenberg (1963). Following this approach, there are six possible constituent orders in transitive clauses in the languages of the world: SOV, SVO, VSO, VOS, OVS, and OSV. In these types S is associated with the 'subject', O with the object and V with the verb in transitive clauses. However, as explained in Section 2.1, this notion is not unproblematic when it comes to ergative languages. Thus, I will use A instead of S when referring to transitive clauses in the following, i.e. I will use AVO instead of SVO, as well as SV and VS for intransitive clauses.

As outlined by Greenberg (1963: 43), all six of these orders are theoretically feasible in the world's languages. Nevertheless, he highlights that three of these orders, namely VOA, OAV, and OVA, are exceedingly rare or entirely absent. Notably, they all share the characteristic of having O preceding A. As a result, Greenberg's (1963) universals primarily focus on the remaining three orders: AOV, AVO, and VAO.

There are a few arguments against this traditional approach. First, this distinction focuses on a type of clause that is relatively rare: a clause containing both a lexical subject and a

lexical object (Du Bois 1987, Dryer 1997). Further, there are languages for which it is not possible to determine a single basic consituent order. This is for example the case for some verb-initial languages that are either VAO or VOA order following Greenberg's (1963) typology. Dryer (1997) puts forth the argument that these languages should be categorized as belonging to the same type. Languages with VAO and VOA order share common linguistic properties. For instance, they commonly employ prepositions instead of postpositions, have relative clauses positioned after the noun phrase, and use articles that precede the noun. Furthermore, both VAO and VOA constituent orders are prevalent within the same language families and geographic regions (Dryer 1997: 76–78). This suggests that these languages share significant linguistic characteristics, justifying their classification into a unified type. Finally, the 6-way distinction ignores intransitive clauses although intransitive clauses are more likely to contain a lexical subject than transitive clauses (Dryer 1997: 87).

An alternative to Greenberg's (1963) 6-way distinction is a three-way typology that distinguishes between verb-initial, verb-medial, and verb-final, based on the position of the verb in the clause (e.g. Nichols 1986). However, Dryer (1997: 92) argues that this distinction is also not fully convincing as uncommon constituent order OVA is merged with the other verb-medial order AVO in this typology and they do not share the same features. Therefore, Dryer (1997) proposes an approach based on two distinct two-way typological parameters that identify four types: VS & VO, SV & VO, SV & OV, and VS & OV. In this framework, the VS & VO type roughly aligns with VAO and VOA in Greenberg's (1963) typology. The combination of SV and VO roughly matches AVO. SV & OV can be associated with the two constituent orders AOV and OAV. Lastly, the VS & OV type is linked to the rare order OVA (Dryer 1997: 92).

Nevertheless, the 6-way framework introduced by Greenberg (1963) appears to be the most widely adopted approach in linguistic typology, including among scholars studying the languages under consideration in this study. To prevent any confusion, I will adhere to the six potential orders mentioned earlier, as these are the standard classifications commonly employed to depict constituent orders in the languages discussed. However, it is important to note that this study will not exclusively concentrate on transitive orders; it will also encompass intransitive clauses, which were left out in the traditional approach.

2.2.2 Constituent order flexibility

Languages can be divided into languages with rigid constituent order and languages with flexible constituent order. Rigid constituent order languages either only allow for one grammatical constituent order or other constituent orders are used only infrequently or in specific pragmatic contexts. This work will focus on the other type of languages which are the ones with flexible or 'free' constituent order. These languages can either have one dominant constituent order that could be considered the basic constituent order or they do not have a dominant constituent order at all.

In the case of languages that demonstrate flexibility in constituent order but still possess a basic constituent order, various approaches can be employed to identify this basic order. The first method is to establish the frequency a certain constituent order has over other possible constituent orders. According to Greenberg (2005: 67), the most frequent constituent order in a language is always the basic constituent order. An argument in favor of frequency calculations is that it is relatively easy to conduct and does not require a lot of knowledge of the language and its grammar or pragmatics, though a large corpus coded for word class would be necessary. However, this approach has been questioned with counter examples of languages for which the most frequent constituent order could not be argued to be the basic one. Following Dryer (1997), frequency of linguistic structures in a language is not just a result of arbitrary usage but is influenced by both the grammatical rules that guide their use and the real-world contexts in which those structures are employed. He suggests that frequency alone is not sufficient to fully understand the underlying grammatical principles of a language. When one constituent order is favored over another, it could be due to the higher prevalence of discourse conditions that require its use (Dryer 1997: 72). In Papago, for example, the definiteness of noun phrases determines constituent order. Indefinite noun phrases typically occur before the verb, while definite noun phrases are usually used after the verb (Payne 1987: 794–795). Thus, this finding primarily indicates the higher occurrence of definite noun phrases rather than being a characteristic of the Papago grammar, and the notion of a fundamental constituent order lacks strong substantiation in this particular case (Dryer 1997: 73). Moreover, the frequency of a certain constituent order could be different depending on the genre of the text that has been analyzed. This means that the most frequent constituent order in narratives might not be as frequent as in e.g. conversational texts. Claims about basic constituent orders are often based on the frequency of constituent orders in narratives (Payne 1995: 454).

Another way to find out the basic constituent order of a language is the markedness of certain constituent orders over others. The basic constituent order is thereby the constituent order which is considered to be unmarked. According to Dryer (1995: 106), this could also be understood as the "elsewhere" case of a constituent order which will be known when the contexts for other constituent orders are clear. The basic constituent order then is used in all contexts which cannot further be defined. Moreover, the unmarked constituent order is the constituent order that continues the flow of a text. In contrast, a marked constituent order is seen to provide a certain element of "unexpectedness" to the text (Dryer 1995: 108).

An argument in favor of considering markedness in the determination of constituent order is that the possible genre bias could be avoided when one understands the reasoning and contexts of different constituent orders. However, whether something is "marked" or not requires a lot of knowledge about the language. Moreover, testing whether a constituent order is pragmatically unmarked is usually done through elicitation which is not unproblematic as one has to judge pragmatical markedness by the consultant's reaction for problematic cases. According to Lambrecht (1994), the idea of an 'unmarked' constituent order is further flawed because it wrongly assumes that certain grammatical structures do not have pragmatical implications relative to

others. This claim lacks meaning without a clear explanation of what 'normal presuppositions' with which these 'unmarked' orders are expected to occur (Lambrecht 1994: 16).

In order to determine the pragmatical markedness of an utterance, Lambrecht (1994: 17) therefore proposes the following: "given a pair of allosentences, one member is pragmatically unmarked if it serves two discourse functions while the other member serves only one of them. While the marked member is positively specified for some pragmatic feature, the unmarked member is neutral with respect to this feature." Lambrecht (1994) further explains this with two sentences applying different constituent orders and the property of 'argument focus': In the basic AVO clause *She likes* GERMANS there is no argument focus on *Germans* since it may be construed in both broad or narrow (contrastive) focus. In contrast, the cleft construction with an OVA order, *It is* GERMANS *she likes*, applies argument focus as *Germans* here can only be regarded as having narrow focus (Lambrecht 1994: 17).

There is some variation in terms of the distribution and frequency of the logically possible constituent orders illustrated above, i.e. AVO, AOV, VAO, VOA, OVA, OAV, and languages with no dominant basic constituent order. According to Hammarström (2016: 25), of the 5252 languages with available information on constituent order, the majority apply a basic AOV constituent order (2275), closely followed by languages with a basic AVO order (2117). AOV order is significantly more widely distributed across language families (239) than AVO order, which, while common in terms of overall counts, is only present in 55 language families. Verb-initial languages, i.e. languages with a basic VAO or VOA order, which play a crucial part in my work, are less frequent than the aforementioned orders, with 503 and 174 occurrences, respectively. Concerning the distribution across language families, VAO order is found in 27 language families while VOA is only found in 15. Languages with a basic OVA or OAV order are the rarest among the languages of the world, appearing in only three or one language family, respectively. These two orders are even less common than a non dominant basic constituent order which is found in 124 languages across 26 language families. Overall, this distribution highlights the diversity and complexity of constituent orders in the world's languages, underscoring the unique characteristics and importance of less prevalent orders in linguistic studies.

For the in-depth study of Datooga, I will use frequency measures to determine the basic constituent order. It is important to note that the corpus could ideally be larger to make more robust claims about the basic order. However, this calculation is balanced by the inclusion of data from both multi-party conversations and monologic narratives, which ensures that the analysis is not overly biased despite the corpus size.

2.2.3 Conditions for variation

2.2.3.1 Semantic and syntactic features

There are several semantic and syntactic conditions for constituent order variation attested in the languages of the world. One of these is the heaviness of arguments as a lot of languages show a preference for placing heavier elements towards the end of a clause. When speaking of

heaviness or *weight* one has to distinguish between *syntactic* and *phonological weight*. In English, syntactic weight is noticeable in various structural choices, such as in what Arnold et al. (2000) call 'Heavy NP-shift'.

Compare the following clauses:

Heavy NP-shift (Arnold et al. 2000: 28)

a) The waiter brought **the wine we had ordered** to the table.

b) The waiter brought to the table **the wine we had ordered**.

In these clauses, O *the wine we had ordered* is heavier than the prepositional phrase *to the table*. In the first example O precedes the prepositional phrase, while O is placed at the end of the clause in the second example. In their study, they define *heaviness* "as the difference in length between two constituents, in terms of number of words" Arnold et al. (2000: 29). Drawing from a corpus study, Arnold et al. (2000: 36–37) demonstrate that shifted orders, i.e. those that place O at the end, as in example, become more common when the O constitutent consists of more words than the prepositional phrase.

In terms of phonological weight, there are several conditions that contribute to the heaviness of constituents and their placement at the end of the clause. These conditions include syllable length, vowel length, onset length, as well as the type of onset and coda: elements in later position tend to have more obstruent onsets and more sonorous codas as elements placed in earlier position (Ryan 2016: 724–725). Further, there is a relation between givenness and the heaviness of constituents. Referents that have already been introduced to the discourse can usually be referred to with shorter forms. By contrast, in order for the hearer to form a suitable mental image of the referent, new discourse referents need to be introduced with longer or more complex form (Ward & Birner 2006: 158).

Besides heaviness, negation is also a condition for constituent order variation and it is a common feature for verb-initial languages, according to Payne (1995: 456). One example for such a language is the Mayan language Mam. In Mam, negated clauses or phrases directly follow the negative particle *miyaa?*, which appears sentence-initially. This places the entity under negation at the front, emphasizing it (England 1983: 244).

(16) Mam (Mayan; Guatemala, Mexico; England 1983: 245)

a. *ma chi tzaj t-tzyu-?n Cheep kab´ xiinaq*
REC 3.PL.ABS DIR 3SG.ERG-grab-DS José two man
'José grabbed some[4] men.'

b. *miyaa? Cheep saj x-0-tzaj tzyuu-n ky-e kab´ xiinaq*
NEG José REC DEP-3SG.ABS-DIR grab-AP 3PL-RN two man
'It wasn't José who grabbed some men.'

[4]Although the glosses offer a precise number, the translation is directly drawn from England (1983: 245), which suggests a more generic meaning.

In the affirmative clause in Example 16a, A *Cheep* 'José' follows the verb *t-tzyu-ʔn* 'grab'. When negated, as in Example 16b, *Cheep* 'José' occurs before the verb, right after the negative particle *miyaaʔ*. However, although the constituent order technically changes from affirmative to negative clauses in Mam, there is also a change noticeable in the verbal morphology. In the affirmative clause, A is indexed on the verb with the prefix *t-* for the third person singular ergative which does not occur in the negative clause in the second example. Instead, the anti-passive suffix *-n* is attached to the verb.

Another condition that could be involved in a change of constituent order is animacy. According to Bock & Warren (1985), animacy is identified as one of several factors influencing the ease of retrieving conceptual information and therefore, animacy also has an impact on constituent order preferences. For instance, in a psycholinguistic study with Japanese speakers Tanaka et al. (2011: 323) found out that when the animate entity came before the inanimate object rather than the other way around, participants were more likely to re-utter OAV clauses as AOV clauses. The research carried out by Branigan & Feleki (1999) supplies further compelling backing for the impact of animacy on Greek constituent order. Study participants exhibited a distinct inclination towards remembering sentences such that an animated entity preceded an inanimate one, regardless of grammatical function. This tendency manifested for both the favored AVO order and the less favored OVA order. Notably, when the subject was animate, participants significantly remembered OVA sentences as AVO sentences, positioning the animated entity first (47%), in contrast to cases where the subject was inanimate, and the inanimate entity took the initial position (36%). Similarly, participants notably favored recalling AVO sentences as OVA sentences when the subject was inanimate, leading to the animated entity being placed first (10%), as opposed to situations where the subject was animated, and the inanimate entity came first (2%) (Branigan & Feleki 1999: 99–100).

The effect of animacy on constituent order variation can also be observed in African languages. For instance, in Southern Sotho, a Bantu language, the order of the direct and the indirect object depends on the animacy of the referents (Morolong & Hyman 1977), as illustrated in the following clauses in Example (17).

(17) Southern Sotho (Southern Bantu; Lesotho, South Africa; Morolong & Hyman 1977: 202–203)

a. *ke-phehétsé mokété lijó*
1SG-cooked feast food
'I cooked the food for the feast.'

b. *ke-phehétsé lijó mokété*
1SG-cooked food feast
'I cooked the food for the feast.'

c. *ke-phehétsé ngoaná lijó*
1SG-cooked child food
'I cooked the food for the child.'

d. **ke-phehétsé lijó ngoaná*
1SG-cooked food child
'I cooked the food for the child.'

When both the direct object and the indirect object are nonhuman as in the Examples (17a) and (17b), both orders are possible. However, when one of these referents is human and the other nonhuman, as in Example (17c), the more animate argument has to precede the less animate. Example (17d) is considered ungrammatical in this context, because the direct object *lijó* 'food' is less animate than the indirect object *ngoaná* 'child' (Morolong & Hyman 1977: 203).

Another semantic condition that can affect constituent order cross-linguistically, is definiteness. In Section 2.2.2, this was briefly discussed with Papago, an Uto-Aztecan language spoken in Mexico and the United States, in the context of how to find the basic constituent order cross-linguistically. In Papago, indefinite noun phrases typically precede the verb while definite noun phrases follow it (Payne 1987: 794–795). A similar procedure is observable in certain Mayan languages, as discussed in the work of England (1991). The following examples from K'iche', a language spoken in Guatemala and Mexico, illustrate this.

(18) K'iche' (Mayan; Guatemala, Mexico; England 1991: 469)[5]

a. *xril le achi le wuj*
saw DEF man DEF book
'The man saw the book.'

b. *xril jun wuj le achi*
saw INDEF book DEF man
'The man saw a book.'

c. *?xril le wuj jun achi*
saw DEF book INDEF man
'A man saw the book.'

d. **xril jun achi le wuj*
saw INDEF man DEF book
'A man saw the book.'

[5]There are no glosses provided in England (1991). They were deduced from the examples provided.

The examples provided above vary in the definiteness of both A and O. According to (England 1991: 469), when both arguments are definite, there is a preference for VAO order, as in Example (18a). Yet, VOA order is favored if A is definite and O is indefinite, as in Example (18b). In instances when A is indefinite and O is definite, such as in Example (18c), VOA order is considered slightly acceptable, but VAO order, as in Example (18d), would be considered grammatically incorrect.

The conditions discussed so far do not independently trigger changes in constituent order. Rather, these changes are often influenced by discourse pragmatics and information structure. In other words, certain conditions are typically associated with specific information-structural properties, as observed for example with syntactic weight and givenness where new arguments tend to be generally heavier. The following section will delve deeper into the role of discourse pragmatics and information structure in shaping constituent order.

2.2.3.2 Discourse pragmatics

Constituent order in languages with flexible order is further sensitive to discourse pragmatics. In a well-functioning discourse, new information must be linked to previously established knowledge in the listener's memory, following Harold (1995: 139). As a result, these previously established facts may be related to what the listener already knows or to information that was discussed earlier in the conversation. But as Harold (1995) points out, this does not mean that the speaker always has the speech prepared in advance. In this context, Chafe (1979) proposes that the organization of discourse is closely tied to the flow of thought as it transitions from one focal point of attention to another.

According to Harold (1995: 140), the organization of a discourse is linked to certain pragmatic conditions. The first one involves the accessibility, or availability in Harold's (1995) terminology. Accessibility refers to how easily a concept can be accessed by the listener and is influenced by factors such as prior knowledge, previous mention, and derivability (Harold 1995: 141). Entities can also gain accessibility through inference. This occurs when a concept is not explicitly stated in the discourse but is linked to and made available through concepts that have already been mentioned. For instance, in a conversation centered around the concept of "sailing," the concept of a "boat" might become accessible through implicitly shared expectations, even if it hasn't been explicitly mentioned before (Harold 1995: 141). As the most accessible constituents in a clause typically occur early in an utterance, accessibility thus has an influence on constituent order variation (Harold 1995: 141; Branigan et al. 2008: 174).

Another pragmatic factor involved in the organization of a discourse, according to (Harold 1995: 140), is the importance of an entity. Important elements tend to reappear and those previously introduced are generally deemed more important than novel ones. Within this framework, Harold (1995) distinguishes between *discourse importance* and *external importance*. Discourse importance is contingent on the frequency of reference throughout the discourse: If a segment of the discourse centers on a specific entity, that entity is deemed to possess discourse importance.

In contrast, external importance relates to concepts that were rarely mentioned in the preceding discourse but indicate the speaker's awareness that they are presently of concern to the listener (Harold 1995: 140).

Following (Harold 1995: 142), there is a correlation between the accessibility and the importance of an entity: If an entity is important, it usually returns to the discourse while entities are typically more important when they became accessible earlier in the discourse compared to those that are new. Another factor that affects the accessibility of entities is their predictability. Harold (1995: 142) defines predictability as "the degree to which the listener can predict the presence and role of a particular concept." An entity is very predictable if it was recently mentioned, is in the same role, and no other concept fits that role. Predictability further works against the need to mention concepts repeatedly. Highly predictable concepts are often not mentioned again because it is assumed the listener already knows about them. Unpredictability, on the other hand, helps organize information by highlighting new or unexpected concepts (Harold 1995: 142).

So far, this section has dealt with discourse pragmatics and how it affects communication. Another important aspect of languages, however, is information structure. Following Lambrecht (1994: 3) information structure is "a component of grammar, more specifically of sentence grammar" and further "a determining factor in the formal structuring of sentences." Although the mental state of the interlocutors is important, it is only relevant for the interpretation of information structural categories when it is formally expressed in morphosyntactical or prosodic structures, which include pitch accents, the choice of lexical expressions, or the order of constituents within a clause (Lambrecht 1994: 6). Thus, whereas discourse pragmatics considers the interlocutors' mental states and how certain communication goals are met, information structure concerns the structuring of information within a sentence and how this is conveyed in grammatical structures.

In the following, I will discuss the information structure categories givenness, topic, focus, and contrast. Understanding the terminology and concepts behind information structure will be crucial for the typological survey in the next chapter as these information structural categories play a role in constituent ordering and are often the only explanation for a change in linear order. On the other hand, discourse pragmatics are relevant for the study of constituent order variation in Datooga which is based on corpus data and thus allows for the consideration of the discourse context while interpreting utterances.

Givenness The first concept that is crucial for understanding information structure and thus also plays a role when it comes to constituent order variation, is givenness. Krifka (2008: 262) defines givenness as follows: "A feature X of an expression α is a Givenness feature iff X indicates whether the denotation of α is present in the CG [common ground] or not, and/or indicates the degree to which it is present in the immediate CG".

Following Chafe (1976), information that is given is usually presented in a "weaker" fashion than new information. More precisely, as compared to new information, this manifests in a lower

pitch and weaker stress with given information in English. Given information is further more likely to be expressed with pronouns (Chafe 1976: 31).

Chafe (1976) notes that whether an entity is given or new depends on what the speaker assumes to be in the addressee's mind due to both the extralinguistic and the linguistic context. With the extralinguistic context, the speaker may believe that they share a perspective with the addressee and that they both are aware of their surroundings at the time of speaking. For instance, if the speaker observes the addressee looking at a specific picture on the wall, they might unexpectedly say, "I bought it last week", knowing that they both know the speaker is referring to the picture on the wall. Thus, it is possible to use the pronoun *it* when referring to the picture although it was not previously mentioned. In the linguistic context, whether something is given depends on whether it was previously mentioned (Chafe 1976: 31–32).

According to Ward & Birner (2006) a common pattern in many languages is to organize discourse according to the "old/new", i.e. given and new, principle. As a result, information that is assumed to be unfamiliar or new to the hearer, is usually provided after information that is considered to have been known to them before (Ward & Birner 2006: 154).

Topic One of the notions associated with information structure is topic. Krifka (2008: 265) defines topic as follows: "The topic constituent identifies the entity or set of entities under which the information expressed in the comment constituent should be stored in the CG [common ground] content". In other words, when people are communicating, the topic constituent helps establish what they are talking about, and the comment constituent provides information about that topic. Together, they contribute to building a shared understanding or common ground among the participants in the conversation. The notion of topic is therefore also associated with the concept of "aboutness".

Topics are further often considered to be "old" information in the sense that they are already given to the discourse in one way or another because as Lambrecht (1994: 164) notes: "One cannot "add" information about a referent unless this referent is in some important sense already available in the discourse as a starting point". However, he further states that although there is a correlation of the function of the topic and the representation of "old" information, this is not always the case, as "old" or given referents can also represent foci (Lambrecht 1994: 164–165). On the other hand, topics can also be new, as in the sentence "A good friend of mine married Britney Spears last year", where *a good friend of mine* is the topic and new to the discourse (Krifka 2008: 265).

According to Halliday (1967), topics often appear first in sentences because they provide a setting and direction. Lambrecht (1994: 199) notes that the link between sentence-initial position and the status of being the topic is especially prevalent in free word order languages, such as Russian or Czech. In these languages, any constituent can occur first in a sentence and thus, the argument appearing first is typically associated with the topic of the particular clause (Lambrecht 1994: 200).

However, the claim that topics are the elements that appear first in a sentence has been questioned due to the existence of VAO and VOA languages. In these languages it is typically the verb, i.e. a non-topical element, that occurs first in the default constituent order, which challenges the theory that sentence-initial position indicates the topic. Even if these languages have rules allowing noun phrases to precede verbs, it does not fully align with the general principle that sentence-initial position marks the topic as such constructions are most likely considered to be marked in these languages (Lambrecht 1994: 200). According to Lambrecht (1994: 200), another reason to question the universal validity of the topic-first principle arises from languages such as English or German. In these languages, focus elements can be placed at the beginning of a sentence, and topical constituents that are not subjects may follow the verb in their typical argument position without any syntactic marking, such as in "PETER said it" where the capitalized element *Peter* is in focus while *it* represents the topic of the clause. The marking of information structure in such sentences in English and German is done solely through prosody, not through syntactic construction (Lambrecht 1994: 200). Lambrecht (1994: 201) also states that there is no functional need for the topic to appear at the beginning of the sentence after a topic referent has been pragmatically established. In such cases, the primary role of the topic expression shifts from announcing the topic referent to indicating its status as an argument in a proposition. In conclusion, while the topic-first principle captures a general tendency, its applicability faces challenges in accounting for certain language patterns and may not be universally valid.

Focus Another concept related to information structure is focus. Following Krifka (2008: 248) "a property F of an expression α is a Focus property iff F signals that alternatives of (parts of) the expression α or alternatives of the denotation of (parts of) α are relevant for the interpretation of α". According to this definition, focus indicates the existence of alternatives to an expression.

In the literature, focus is also often associated with representing 'new' information (e.g. Gundel & Fretheim 2006) or the 'highlighting' of information. Neeleman et al. (2009: 16–17) note that a proposition's 'focus' is the information that is highlighted, such as the element that corresponds to a WH-expression. The rest of the phrase then serves as a background for this point of the focus. For example, for a question like "What did Rutger buy?", the answer could be "Rutger bought A GUN", where *A GUN* corresponds to the WH-expression *what* in the question and is in focus (Neeleman et al. 2009: 17). Further elaborating on that, Lambrecht (1994: 210) states that the main role of focus marking is to highlight a correlation between a specific element within a proposition and the proposition as a whole.

Diverging from that, Krifka (2008: 257) argues that features like 'highlighting', 'emphasis', 'importance' and 'newness' are statistical correlates frequently associated with focus, but they should not be used as definitive criteria for defining what focus is. The essence of focus lies in its specific linguistic function, and these associated characteristics are secondary aspects that often accompany but do not define focus. Essentially, focus marking is used to draw attention to a

particular part of the proposition, emphasizing its significance or relevance within the broader context of the information being conveyed.

In terms of position within a clause, there is cross-linguistic variety just like was the case with topics. In English, a focused element could appear in sentence-initial position but it is not obligatory, according to Gundel & Fretheim (2006: 185), although it might mostly be expressed through prosodic mechanisms, following Lambrecht (1994: 224). However, in other languages, such as Italian, a change in constituent order combined with prosodic marking can indicate a focus entity Lambrecht (1994: 224). Thus, as both, the focus and the topic, can appear in sentence-initial position, the position within a clause is not the strongest indicator for an entity representing the topic or the focus of a construction.

Contrast Contrast is another category that is relevant for information structure. Following Chafe (1976), an example of contrast can be found in a sentence like "Rónald made the hamburgers" with the accent on *Ronald* (Chafe 1976: 33). In this sentence, *Ronald* is contrasting with other possible people who could have made the hamburgers and might be in the mind of the listener. However, by placing the accent on *Ronald*, the speaker is indicating to the listener that *Ronald* is the right choice in this context (Chafe 1976: 33).

Chafe (1976) lists three factors that contribute to contrastiveness. Firstly, there is an assumed shared awareness between the speaker and the addressee, such as with the previous example, that the hamburgers were made by someone, which he refers to as "background knowledge". Secondly, there's a set of potential candidates, indicating the speaker's assumption that the listener has a limited number of options in mind. Finally, the proper candidate is selected, which serves as the "focus of contrast". It's worth noting that a sentence can have more than one focus of contrast (Chafe 1976: 32–35).

Although called "focus of contrast" by Chafe (1976), contrast can manifest in focus and topic entities (Lambrecht 1994, Krifka 2008, Neeleman et al. 2009). Following Neeleman et al. (2009: 18), both contrastive topics and foci are marked through distinct intonation patterns in English. More specifically, contrastive foci are marked by a high tone preceding a low tone while in contrastive topics a L+H* pattern precedes a low tone and high boundary tone (Neeleman et al. 2009: 18).

An example of a clause containing a contrastive topic could be "**Maxine** was introduced to the queen on her birthday" where *Maxine* serves as the contrastive topic emphasizing that the speaker is aware that Maxine was introduced to the queen on her birthday, distinguishing her from other relevant individuals. Following Krifka (2008: 267), contrastive topics "consist of an aboutness topic that contains a focus, which is doing what focus always does, namely indicating an alternative." Thus, contrastive topics share properties of both topic and focus. *Maxine* in the example above is therefore not only a topic in the sense of aboutness, i.e. the proposition is about her, but due to the rising accent, also signals that there are possible alternatives in the addresses mind.

"Rutger bought **A GUN**" is an example of a clause containing a contrastive focus on *a gun*, signalling that *Rutger* purchased a gun and not something else (Neeleman et al. 2009: 18). According to Krifka (2008: 259), contrastive focus is commonly employed for correction. However, it can also be used with an additive meaning, as illustrated in the following example where the contrastive focus is indicated by capitalized letters:

- A: John wants coffee.
- B: MAry wants coffee, TOO.

This chapter has provided an overview of the foundational concepts essential for understanding the subsequent chapters of this thesis. It introduced the relevant case-marking systems—accusative, ergative, and marked nominative—necessary for the typological survey in the next chapter, as well as triggers for case splits in various languages. The section on constituent order variation covered the concept of basic word order and methods for determining it. Additionally, it examined cross-linguistic factors influencing constituent order variation, including discourse pragmatics and information structural properties such as givenness, topic, focus, and contrast.

The following Chapter 3 will delve deeper into case marking and constituent order variation, focusing specifically on the languages of East Africa, which are central to this thesis. After reviewing previous research on these topics, a quantified typological survey will be conducted.

Chapter 3

Case and constituent order in East Africa

The previous Chapter 2 provided an introduction to critical concepts associated with case marking. This included an exploration of case marking systems such as ergative and marked nominative, as well as an examination of the motivating factors for constituent order variation. The discussion of constituent order variation encompassed an analysis of information structural properties that contribute to variations in sentence structure. These aspects are foundational for the subsequent chapter.

In this chapter, the first Section 3.1 offers an overview of the distribution of case marking and constituent order variation in East Africa. A specific emphasis is placed on introducing the prevalent 'no case before the verb' rule, a feature notably common in this geographical region. The languages conforming to this rule will be examined in more detail in a typological survey of Nilo-Saharan languages in Section 3.2. This section extends the scope of König's (2008a) research on languages exhibiting interactions between case and constituent order. Finally, Section 3.3 provides a summary of the results of the survey and engages in a discussion of these findings.

3.1 Previous research

The upcoming section explores the existing research on East African languages, focusing on case, constituent order variation, and the specific feature addressed in this study—the 'no case before the verb' rule (König 2008a). This rule delineates a phenomenon predominantly observed in East African languages. For languages following this rule, case marking exclusively occurs on postverbal S and/or A, with no marking on preverbal arguments. Thus, it becomes imperative to provide an overview of the grammatical landscape in East Africa. The overview starts with an exploration of case marking in East African languages in Section 3.1.1, followed by an examination of constituent order variation in Section 3.1.2 which also discusses the 'no case before the verb' rule in more detail, exploring its origins and implications.

3.1.1 Case in Africa

Languages featuring case marking can be identified across all four language phyla that are indigenous to Africa: Afroasiatic, Niger-Congo, Khoisan, and Nilo-Saharan. However, the frequency of languages exhibiting case marking varies within these four phyla.

It is important to note that the classification of the Khoisan languages, as well as of the Nilo-Saharan phylum, both proposed by Greenberg (1963), are seen as controversial today. Following König (2008b), the term 'Khoisan' does not unit of genetically related languages and rather consists of language families with different origins. However, these languages are only discussed briefly as they are spoken outside the geographical area of interest of this book. Thus, I will use the term for the sake of simplicity.

As for Nilo-Saharan, Güldemann (2018: 254) notes that languages of the claimed Nilo-Saharan phylum "are typologically very diverse, and it is hard to find traits that either are synchronically shared across the entire spectrum of the group and/or can be easily argued to be historically related to each other by principles of diachronic typology". As of today, not all branches of Nilo-Saharan are agreed on in the literature. Exceptions to this include branches such as Eastern Sudanic, which is widely acknowledged (Güldemann 2018: 299) and, as subsequent chapters will demonstrate, is particularly significant to the typological survey. For convenience, I will employ Dimmendaal's (2020) classification as basis in the following course of this book.

The classification of the Afroasiatic phylum is illustrated in Figure 3.1. Afroasiatic languages are distributed all over northern and central Africa and the phylum has a lot of branches with languages that distinguish case. The branches with languages spoken in East Africa, and thus falling in the scope of the current study, are Omotic, Semitic, and Cushitic.

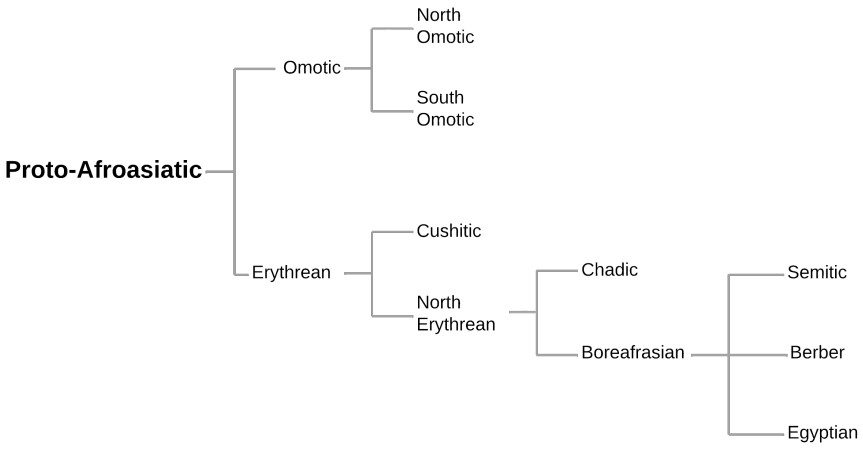

Figure 3.1: Classification of the Afroasiatic phylum following Ehret (1995).

In terms of case-marking systems, all three branches spoken in East Africa, i.e. Omotic, Semitic, and Cushitic, include languages with accusative and marked nominative systems. Accusative case systems, i.e. with an unmarked nominative case form and a marked accusative

form, can be found in languages such as Dime, Sheko and Yemsa within the Omotic branch (Amha 2017: 823). For the Semitic branch, Frajzyngier (2018: 9) notes that the languages with accusative case systems typically have distinct overt case markers for both the nominative and the accusative. However, there are also a few modern Semitic languages with unmarked nominative and marked accusative case, such as Amharic, spoken in Ethiopia, which further applies differential object marking determined by definiteness (Frajzyngier 2018: 9). Marked nominative case systems are evident in the majority of Omotic languages (Amha 2017: 823). Within the Cushitic branch, marked nominative is also present in languages like Saho and Afar (Frajzyngier 2018: 23).

Case marking within the biggest African phylum, Niger-Congo, is rare. However, according to Creissels et al. (2007: 88) there are a few Western Bantu languages (Benue-Congo branch in Figure 3.2) that form a compact area from Gabon to Angola and which have tonal cases, such as Ngangela (Zavoni 2003). In accordance with Casaretto et al. (2020: 119), it is noteworthy that case marking is also present in certain languages spoken in the Nuba mountains, where examples include the accusative case in the Heiban group, which belongs to the Kordofanian branch in Figure 3.2. Additionally, within the Katloid group (Kordofanian branch in Figure 3.2), there is a notable instance of a Niger-Congo language, Tima, which features a split-ergative case system. So far, no Niger-Congo language has been described as having an ergative case system.

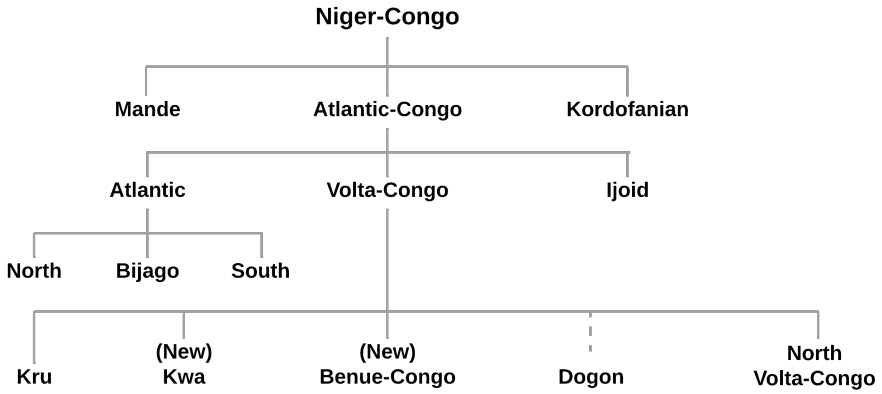

Figure 3.2: Classification of the Niger-Congo phylum following Williamson (1989).

Khoisan languages generally lack case marking to distinguish grammatical relations. Creissels et al. (2007: 88) highlight the uncertainty in some language descriptions regarding whether certain morphemes mark discourse roles like topic or focus or actually distinguish grammatical relations. For example, the topic marker *má* in the W2 dialect of the Khoisan language !Xun is considered a potential nominative case marker, as it is obligatory for S and A in declarative clauses (König 2008a: 273–275). Additionally, in the Khoe languages Khwe and Khoekhoe, the clitic *ǂa* is a candidate for being a case marker encoding objects (König 2008a: 276). According to McGregor (2018: 243) this clitic evolved from functioning as a copula into an optional accusative marker in the majority of Khoe languages. For example, for Khwe, König (2008a)

states that the clitic *'a* serves as a copula, highlighting new or focal information. However, it is also commonly used to introduce objects, exhibiting characteristics of an object marker even when the object does not convey focal information. Thus, in specific contexts, *'a* functions as an object marker while preserving its copula and focal functions in other situations. In Khoekhoe, *'a* is used as both a copula and to indicate objects. However, the focus function is no longer productive in Khoekhoe compared to Khwe (König 2008a: 276–277).

In the Nilo-Saharan phylum, languages exhibiting case marking are distributed across various branches, including Eastern Sudanic, Central Sudanic, Maban, Fur, Kunama, Saharan, Kuliak, and Gumuz. Within this phylum, accusative case is present in certain groups of the Eastern Sudanic branch (e.g., Nubian), languages of the Saharan branch (e.g., Kanembu), Central Sudanic, Maban, Kunama, Kuliak, and Fur. Marked nominative systems are exclusive to the Nilotic and Surmic groups and Berta within Eastern Sudanic, as well as the Gumuz branch. Ergative case is found in Nilotic, Surmic, and Jebel within Eastern Sudanic, as well as in the Koman language Uduk. A more detailed exploration of case marking and constituent order variations within Nilo-Saharan will be provided in Section 3.2.

In addition to the case systems mentioned earlier, some languages also exhibit case marking splits. Common conditions for split in the Afroasiatic phylum include definiteness, gender, and tense (Amha 2017: 823; Frajzyngier 2018: 23). However, in the languages of East Africa, another condition that appears to be especially common is constituent order variation. For example, the previously mentioned language Tima is a Niger-Congo language that features this condition (Casaretto et al. 2020: 112), though it is the only Niger-Congo language exhibiting this feature. This condition is prevalent in various languages of this region across different language families and is notably common within the Nilo-Saharan phylum, which is the primary focus of this study. In languages where case marking is conditioned by the constituent order, only postverbal agents/subjects receive dedicated case marking, while these arguments are typically unmarked in preverbal positions. Section 3.1.2 will provide an overview of this feature, including its presumed origins. Before delving into this topic, the section will discuss constituent order variation in Africa.

3.1.2 Constituent order variation in Africa

As mentioned in the previous Section 3.1.1, constituent order plays a role when it comes to case split in the languages of East Africa. This section will give an overview of basic constituent order, flexibility of constituent order, as well as conditions for variation, such as information structure. As this 'no case before the verb' (König 2008a) condition is predominantly found in East Africa, the following two sections will concentrate on languages within this area. This includes the Omotic and Semitic branches of the Afroasiatic phylum, certain branches of Niger-Congo, such as Kordofanian, and Bantu languages of East Africa, as well as the Nilo-Saharan phylum.

The different phyla deviate in terms of their basic constituent order as well as the flexibility of these orders. Within the Afroasiatic phylum, the basic constituent order varies across the different branches in that certain branches typically have languages with verb-final orders, such as the Omotic or Semitic branches, while there are also languages with dominant verb-medial orders in other branches, such as Semitic (Wellens 2003, Amha 2017, Frajzyngier 2018, Frajzyngier & Shay 2020). In contrast, the basic constituent order in Niger-Congo is AVO for the majority of languages with only a few exceptions (Good 2017: 476). In terms of flexibility, Creissels et al. (2007: 127) observes a prevalent occurrence of languages in Africa characterized by a strict constituent order, meaning they exhibit little to no variation. Exceptionally flexible constituent orders, as found in languages like Sandawe in Tanzania, are rare across the African linguistic landscape (Creissels et al. 2007: 127).

As mentioned before, there is a case split feature which is special for African languages of northeast Africa and that is the 'no case before the verb' rule (König 2008a: 240). In languages where this phenomenon occurs there is usually no case distinction before the verb, i.e. the core arguments S, A, and O are always in the morphologically most unmarked case preverbally but distinguish case when they occur after the verb. So far, this phenomenon is attested with ergative and marked nominative languages. In her study, she explores this phenomenon through an examination of seven example languages and briefly references another language, Tima.

For convenience, I will reuse the example previously discussed in Chapter 1 from the Western Nilotic language Päri to further illustrate this feature. Intransitive clauses in Päri always have a SV constituent order as in example (19a). In this example, S *Ùbúr* does not receive dedicated case marking. Transitive clauses, on the other hand, are more flexible when it comes to the order of arguments. Example (19b) shows the dominant constituent order for transitive clauses, OVA. Notice that in this instance A *Ùbúr* is marked by the ergative suffix *-ì*. However, Päri also allows for an alternative order, AOV, as in example (19c). In this clause, preverbal A *Ùbúr* does not receive dedicated case marking (Andersen 1988: 292–293).

(19) Päri (Western Nilotic; South Sudan; Andersen 1988: 292–293)[1]

a. *ùbúr á-túuk`*
Ubur COMPL-play
'Ubur played.'

b. *dháagɔ̀ á-yàaɲ ùbúrr-ì*
woman COMPL-insult Ubur-ERG
'Ubur insulted the woman.'

c. *ùbúr dháagɔ̀ á-yáaɲ`-ɛ̀*
Ubur woman COMPL-insult-3SG
'Ubur insulted the woman.'

This split is predominantly identified in languages spoken in northeast Africa, with the majority of these languages belonging to the Nilo-Saharan phylum, of which, as mentioned earlier, not all branches are equally accepted in the literature regarding their genetic relation.

[1]The author made the following changes in the glosses: 3SG instead of 3S.

One of the branches that is generally accepted as belonging to Nilo-Saharan is Eastern Sudanic. The split is especially common within the Southern Eastern Sudanic branch (see Figure 3.3 for a classification of the Eastern Sudanic branch). Following Dimmendaal (2017: 459), the southern branch of Eastern Sudanic deviates from the northern branch in that it developed new morphosyntactic structures especially regarding the constituent orders of these languages, but also their case marking systems which developed due to the change in constituent order. While the the common constituent order in the northern branch is verb-final, also typical for other languages of Northeastern Nilo-Saharan, verb-initial and verb-second orders are typical in certain families of the southern branch, such as Surmic or Nilotic (Dimmendaal 2017: 459). According to Casaretto et al. (2020: 118), this is more of an areal than a genealogical feature because of the diversity of the Nilo-Saharan phylum and the concentration of languages exhibiting this feature in this particular area, while languages outside of it do not show the feature.

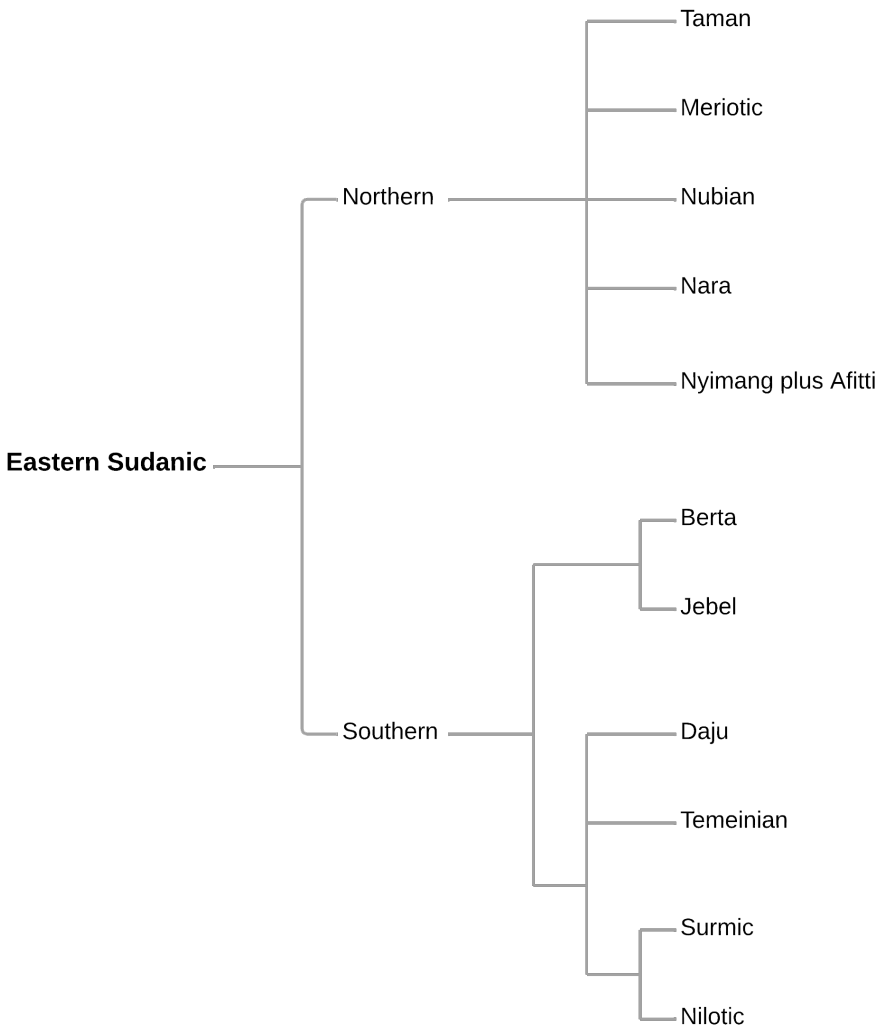

Figure 3.3: Classification of the Eastern Sudanic language family following Jakobi & Dimmendaal (2022: 65).

As for the distribution of ergative and marked nominative systems in the languages of that region, Casaretto et al. (2020) note that it is not likely that it happened through contact. The pattern of marking postverbal A and S arguments but not preverbal ones might be shared across the languages, but the forms of the case marking differ. This suggests that the (proto) languages must have employed internal linguistic resources to develop the observed patterns. The source of case marking for postverbal subjects (marked nominative) and postverbal agents (ergative) cannot be definitively traced to particular subareas or languages. Nevertheless, it is probable that this feature has ancient origins, and the emergence of split ergativity is thought to have taken place, at a minimum, at the Proto-Southern Eastern Sudanic level, considering its manifestation in the majority of its subgroups (Casaretto et al. 2020: 118).

Casaretto et al. (2020) further explain this with the example of Tima. Casaretto et al. (2020: 129) argue that the ergative proclitic *N=* has the same origin as the instrumental clitic as both have the same form and thus suggest an internal origin of the ergative suffix for Tima rather than the borrowing of forms. In addition, features, such as the marking of postverbal agents, are not found in languages where a contact would be likely or that are closely related to Tima, such as Katla or Julut (Casaretto et al. 2020: 129).

In the study by König (2008a), the evolution of 'no case before the verb' is delineated in three stages, using Bari as an illustrative example—an Eastern Nilotic language lacking case marking. In the initial stage, Proto-East Nilotic exhibited case marking and a marked nominative system, evident in all Eastern Nilotic languages through tonal inflection, excluding Bari. The dominant constituent order in these languages likely involved a verb-initial structure. Additionally, an alternative constituent order emerged in stage 1, seen in topic or focus-cleft constructions where participants preceded the verb. Many languages exhibited focus constructions that resembled cleft-constructions, in the form of *'It is X who does Y'*. In marked nominative languages, the unmarked accusative case was assigned to predicates in copula clauses. Therefore, in the focus construction mentioned, X would receive the accusative case. Moving to stage two, the verb-medial order persisted as pragmatically marked, as observed in contemporary languages such as Teso or Turkana. In stage three, a reinterpretation occurred, establishing the former pragmatically marked order as the dominant constituent order. Consequently, case loss transpired due to the absence of context for the nominative. A parallel strategy was identified in topicalization (König 2008a: 257–258).

The languages that follow the 'no case before the verb' rule allow for constituent order variation, i.e. there are contexts in which a preverbal unmarked S or A argument occurs and others where they are postverbal and case-marked. While research on the specific reasons for this variation is limited, there have been some studies on the topic (e.g. Dimmendaal 1983, 1985 for Turkana; Creider & Creider 1989 for Nandi; Jerono 2011 for Tugen). Notably, the Niger-Congo language Tima has been examined in more detail.

The basic constituent order in Tima is SV/AVO, which is the most frequent order and also the most unmarked order in terms of morphology. As mentioned earlier, Tima is a split-ergative language. In a picture-elicitation task, Schneider-Blum (2023) found out that animacy

of the agent and patient, as well as the identifiability of the agent are decisive for the order of constituents in Tima. AVO order was preferred, when the human agents presented and the stimuli were completely visible. When the human agent was not completely visible but could be categorized as human, e.g. through a visible hand, there was a preference for OVA order. This condition could further lead to a possible decomposition of the events, i.e. into two separate clauses. However, when the agent was clearly visible and inanimate and the patient human, a decomposition was favored (Schneider-Blum 2023: 105).

OVA constituent order with an ergative-case-marked A argument is further favored in contexts where there is a shift of agent. In these contexts, a participant which is already given in the discourse, i.e. a previous agent, occurs as the O argument in preverbal position with the new agent being introduced in postverbal position with case-marking. In the following discourse, the new agent then becomes the topic (Casaretto et al. 2020: 122). Furthermore, it is worth noting that in Tima, there is the possibility of a preverbal A argument being marked with the ergative case. This occurs when O is the topic, while the A is focus-marked, leading to an OAV constituent order (Casaretto et al. 2020: 124).

Although this 'no case before the verb' split is common to northeast African languages, it is important to note that there is also constituent order variation which does not lead to a case split. This will further be addressed in the following section, consisting of the typological survey of the Nilo-Saharan phylum.

3.2 Typological survey

The following section deals with the previously introduced 'no case before the verb' feature in more detail. Specifically, it will concentrate on the Nilo-Saharan phylum[2], which has languages that distinguish case marking in almost every branch, as Section 3.1.1 showed, and the languages that follow the 'no case before the verb' rule are frequent here.

The structure of the section is as follows: Section 3.2.1 introduces the research questions behind this survey. Section 3.2.2 describes the data used in this survey, while Section 3.2.3 explains the methods behind it. The results will be described in Section 3.2.4 and will further be discussed in Section 3.3.

3.2.1 Aims and objectives

This survey builds on König's (2008a) study of case marking in Africa. The following study will expand König's (2008a) findings with language descriptions that became available after 2008. Further, this survey provides a quantified analysis of case marking and constituent order variation in Nilo-Saharan languages and includes a discussion on the conditions that lead to case split and/or constituent order variation in these languages.

[2]As has been discussed before in Section 3.1.1, the status of a Nilo-Saharan phylum is challenged. However, as I am not interested in the 'no case before the verb' phenomenon as a genealogical feature, the epistemological status of Nilo-Saharan is less important and I am using it rather as a sampling criterion.

The study will additionally involve a quantified typological survey encompassing the languages within Nilo-Saharan phylum adhering to the 'no case before the verb rule'. To achieve this, it is crucial to identify languages that overtly mark case, thereby presenting the potential for case-split. Subsequently, an examination will be undertaken to ascertain whether these case-marking languages also exhibit splits prompted by changes in constituent order. Regarding constituent order, the study exclusively considers declarative independent clauses. In other words, variations in constituent order within subordinate clauses, distinct from the order in main clauses, are not taken into account when discussing changes in constituent order. The decision to exclude constituent orders in subordinate clauses that differ from those in main clauses is based on the potential syntactic restrictions of these orders. In some languages, subordinate clauses may exhibit a constituent order that would be ungrammatical in main clauses. For instance, in German, the basic constituent order in transitive main clauses is AVO, with the possibility of shifting to OVA for pragmatic reasons. However, in transitive subordinate clauses, the typical order is AOV, which would be ungrammatical in main clauses. Because AVO order in main clauses could never change to AOV without resulting in ungrammaticality, these orders are not comparable as they operate on different syntactic levels. As a result, the AOV order is specifically restricted to subordinate clauses and cannot be directly compared to constituent orders in main clauses.

As previously mentioned in Section 3.1.2, the absence of case marking in preverbal positions may be historically attributed to focus constructions expressed through cleft structures. While these cleft structures persist in certain languages when arguments are moved to preverbal positions, this study will specifically focus on unmarked variations in constituent order—excluding clauses involving cleft or dislocated constructions. This sets the stage for the initial research question of the survey:

> RQ 1: What is the distribution of differential case marking in combination with constituent order variation in Nilo-Saharan languages?

The second aim of this survey concerns the conditions for case-marking split and constituent order variation in the languages that show an interaction of case marking and constituent order variation. Although König (2008a) deals with these conditions, it is only done briefly and there is no generalization across the languages. As case-marking and constituent order are closely linked in these languages, there will be a quantitative analysis on both case-split and constituent order variation. This leads to the second research question:

> RQ 2: Which features are relevant for the variation and how do they interact with each other?

One aspect frequently explored in discussions of constituent order variation across languages is information structure. This encompasses categories such as topicalization, focus, or emphasis (see Section 2.2.3.2 for an overview). Consequently, the final research question for this survey is as follows:

RQ 3: What is the impact of information structural categories, such as topic and focus, when it comes to constituent order variation?

3.2.2 Data

The languages analyzed in this survey all belong to the Nilo-Saharan phylum according to the classification proposed by Dimmendaal (2020), as illustrated in Figure 3.4. The inclusion of Koman and Gumuz in the Nilo-Saharan language family is currently subject to ongoing investigation. However, for the purposes of the following study, they have been included.

Due to the vastness of the Nilo-Saharan phylum and the focus of this study being exclusively on languages spoken in East Africa, languages spoken outside of this region have been omitted from the survey. This is for example the case for the Songhay branch which is distributed across Northwestern Africa. Songhay also exhibits a weak and debatable connection to the Nilo-Saharan phylum (see e.g. Dimmendaal et al. 2019, Dimmendaal 2020).

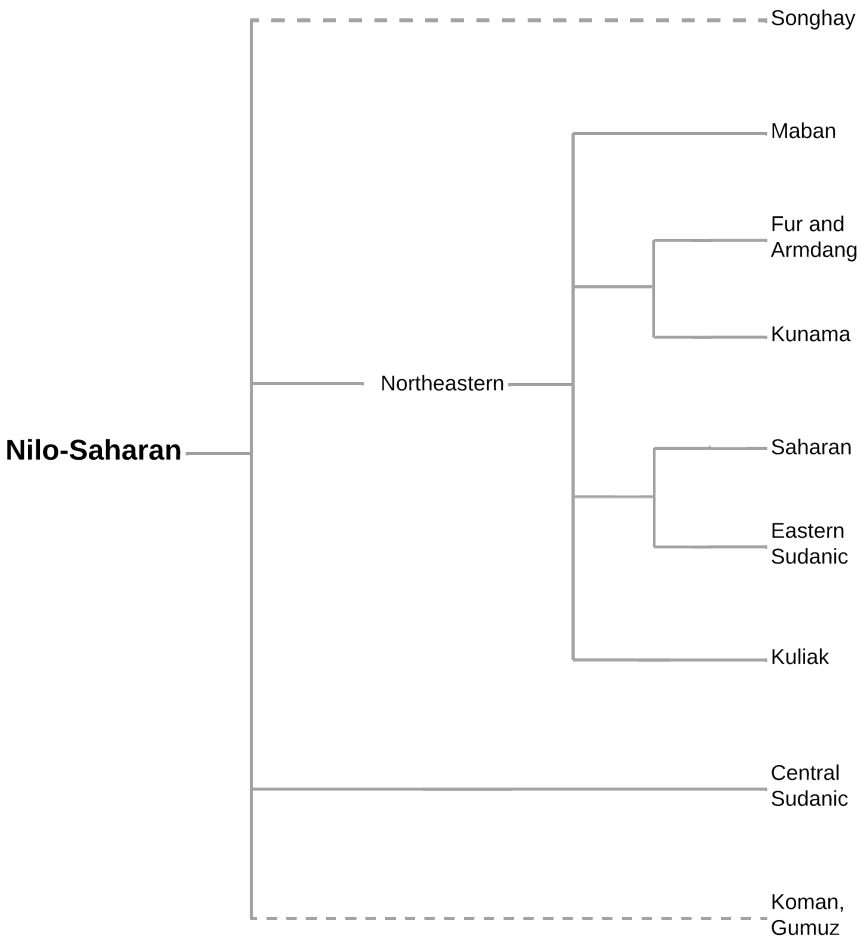

Figure 3.4: Classification of the Nilo-Saharan phylum following Dimmendaal (2020).

The overall data for the analysis stems from available language descriptions, i.e. grammars, grammar sketches, articles about specific features in these languages, as well as information found in overviews of language families. In total, 54 grammars, 20 articles, six book chapters, 45 grammar sketches, seven dissertations on specific features, as well as four overviews on language families were considered. There were a few issues I had to face while gathering the information for the survey. First of all, not every language of this area is sufficiently described and a lot of languages even lack a description at all. Second, even if a language was described, these descriptions were not always available at the time of the study. The languages are listed in Table A.6. Lastly, the amount of information that could be deduced from the available descriptions varied and some features particularly relevant for this study, i.e. case marking and constituent order, were not always discussed in detail. One of these features was tone, and especially the tonal marking of case, which plays a great role in some languages, such as in the Eastern Sudanic branch of Nilo-Saharan.

The remaining languages in the sample are distributed across Northeast Africa, spanning from Tanzania in the South to Egypt in the North and from Chad in the central part of Africa to Eritrea in the East. This distribution is illustrated in Figure 3.5. A list of languages and their descriptions investigated in this study can be found in Table A.1 of the appendix.

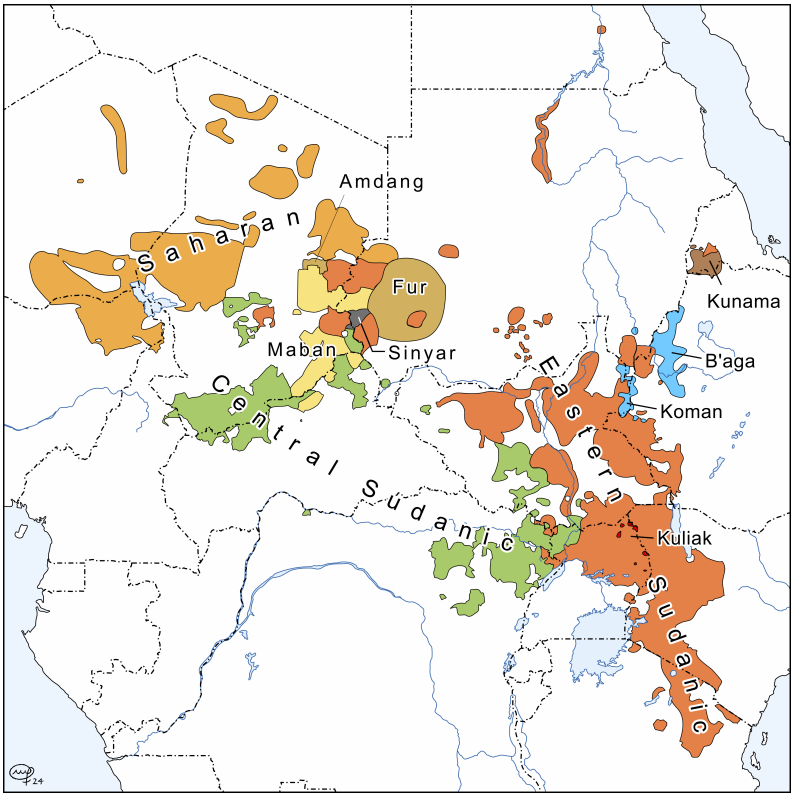

Figure 3.5: Map of the languages within the Nilo-Saharan phylum spoken in East Africa; Map by Monika Feinen.

3.2.3 Methods

The data extracted from the language descriptions was compiled into a spreadsheet. The languages were marked for the variables detailed in Table 3.1. The initial variable #1 annotated was whether a language permits case marking, categorized as 'yes', 'no', or 'unclear' when certainty was lacking. The subsequent variable #2 pertains to the case-marking system of languages distinguishing case, with possible values being 'ergative', 'marked nominative', 'accusative', or 'unclear' again.

Following that, the study delved into the interaction of case marking and constituent order with variable #3. For a 'yes' value, case marking had to affect constituent order, causing S and/or A arguments to lose their preverbal case marking and appear in their citation form. However, some languages exhibited a change in case marking without S and/or A appearing in their citation form in preverbal position, labeled as 'yes*'. Languages without any interaction were labeled 'no', and unclear instances received the 'unclear' designation.

The subsequent table variable #4 addressed the conditions for case split in languages, encompassing information from authors or deduced from provided examples in language descriptions. The case marking strategy (variable #5), describing how a language encodes case marking, had possible values like 'tonal inflection', 'morphological marking' (segmental morphology), 'combined strategy' (a combination of different strategies), and 'syntactic'.

Variables #6–#8 focused on the forms of case markers for S, A, and O, respectively, with possible values such as 'affix', 'tone', 'adposition', 'combined form' (a combination of different forms), and 'unmarked' (i.e. the citation form).

Variables #9 and #10 indicated dominant constituent orders for transitive and intransitive constructions, while #11 reflected the flexibility of constituent order in declarative main clauses. 'Yes' indicated allowance for more than one additional constituent order, 'partly' allowed only one additional order, and 'no' permitted only the dominant order. Instances with unclear possibilities were tagged as 'unclear'.

Variable #12 listed constituent order options mentioned in language descriptions, and finally, variable #13 detailed conditions for variations in constituent order.

The quantitative analysis was then carried out in R (R Core Team 2022).

#	Variable	Values
1	Presence of case marking	Yes, no, unclear
2	Case marking system	Ergative, marked nominative, accusative, unnclear
3	Interaction of case marking and constituent order	Yes, yes*, no, unclear
4	Conditions for case split	Constituent order, verb semantics, animacy, clause type, definiteness, agentivity, topicality, argument realization
5	Case marking strategy	Tonal inflection, morphological marking, combined strategy, syntactical
6–8	Form of case marker for S, A, and O	Affix, tone, adposition, combined form, unmarked
9	Dominant transitive constituent order	VAO, AVO, OVA, AOV, OVA AOV, AVO OVA, VAO VOA, unclear
10	Dominant intransitive constituent order	SV, VS
11	Flexible constituent order	Yes, partly, no, unclear
12	Constituent order options	Any constituent order other than the dominant order (#9 and #10) mentioned in the respective description(s) of a language
13	Conditions for constituent order variation	Animacy, definiteness, argument realization, verb semantics, negation, clause type, information structure

Table 3.1: Survey variables and values.

Figure 3.6 shows for how many languages in the different language families within the Nilo-Saharan phylum linguistic descriptions are currently available. Languages that were annotated with 'yes' had descriptions in the form of grammars, grammar sketches, and/or articles, regardless of their availability, which makes a total of 123 languages. The figure shows that most of the described languages belong to the Eastern Sudanic family (80; 65.04%) which also happens to be the biggest family within the Nilo-Saharan phylum. The following analysis only considered languages that were annotated with 'yes'. However, for eight of these languages, the descriptions were not available at the time of writing, five of them belonging to Eastern Sudanic, two

to Central Sudanic languages, and one to a Maban language. The total number of languages with available descriptions that are included in the further analysis is 115.

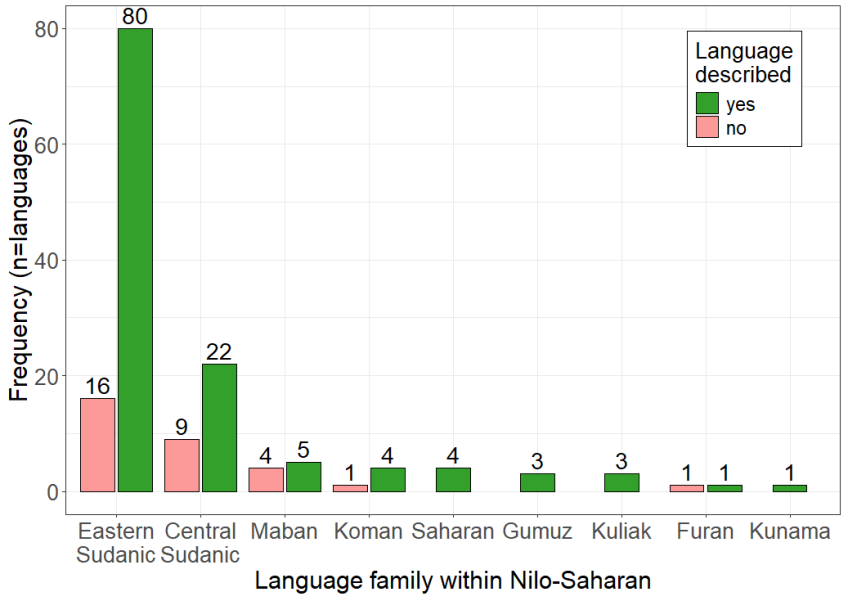

Figure 3.6: The state of language description within Nilo-Saharan.

3.2.4 Results

Figure 3.7 shows how case marking is distributed over the language families of these 115 languages. In total, 64 of these 115 languages distinguish case (55.64%) and they can be found in every language family within Nilo-Saharan. However, the majority belong to Eastern Sudanic (49; 76.56%). The figure also shows that there are languages for which it was not clear whether they have case marking or not (25; 21.74%). In the case of some of these languages, the lack of sufficient discussion on case marking might serve as an argument against the presence of case marking in those languages. However, this possibility was sometimes not explicitly stated by the authors, leading to the annotation of 'unclear'. Another reason to annotate the case marking status for a language as 'unclear', especially for languages within the Eastern Sudanic family, was an insufficient description of tone in these languages. In some Eastern Sudanic languages, as can also be seen in Figure 3.13 later in this section, tonal inflection is sometimes the only marker of case. The languages with unclear case marking are listed in Table A.5.

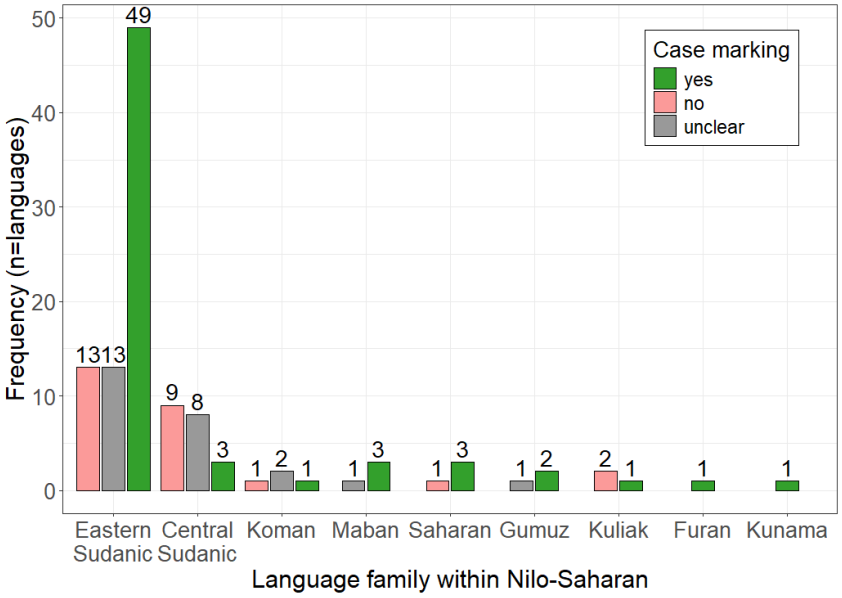

Figure 3.7: Presence of case marking within Nilo-Saharan, deduced from all the available descriptions.

The different case-marking systems within the 64 languages that distinguish case are illustrated in Figure 3.8. The case-marking systems found in Nilo-Saharan are marked nominative, ergative, as well as accusative. Marked nominative systems occur in the majority of languages (27; 42.19%), followed by languages with an accusative system (23; 35.94%), while only eight languages show an ergative system (12.5%). However, while marked nominative and ergative languages are mostly found within the Eastern Sudanic family, languages with accusative systems are found in every language family, except for Koman and Gumuz.

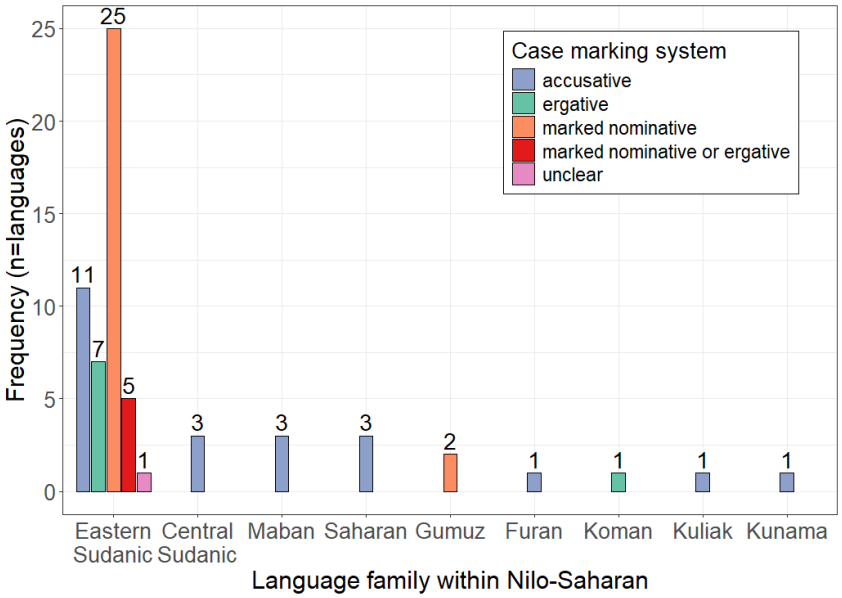

Figure 3.8: Case marking systems in Nilo-Saharan languages, deduced from all the available descriptions.

Similar to the last figure, there are also languages for which the type of case-marking system was not entirely clear. First of all, there are five languages which are annotated exhibiting either a marked nominative or ergative case system. These languages all belong to the Nilotic branch of Eastern Sudanic where only marked nominative or ergative systems are found among the languages distinguishing case. However, an investigation of the language descriptions did not give a sufficient answer on whether these languages should be classified as marked nominative or ergative. This will be explained in more detail later in this chapter when it comes to the interaction of case marking with constituent order. Another language belonging to a different branch of Eastern Sudanic is further classified as 'unclear', distinguishing it from the five languages above. The reason for this is that in other branches of Eastern Sudanic, accusative marking is also found and it could not be ruled out in this instance. These six languages are found in Table A.4.

Figure 3.9 shows the dominant order in transitive clauses in the 64 languages distinguishing case and the 26 languages with no case distinctions, as well as in the 25 languages where case marking is unclear. In the majority of languages with case marking, VAO order is the dominant transitive order with 20 instances (31.25%), followed by AOV with 14 occurrences (21.88%), AVO (13; 20.31%), and OVA (4; 6.25%). For three languages (4.69%) it was not clear whether they have a basic AVO or OVA order, hence the 'AVO/OVA' annotation. Two more languages (3.13%) were classified as having a basic verb-initial order which is why they are annotated as 'VAO/ VOA'. There are a further eight languages (12.5%) for which the dominant constituent order was unclear. In the languages with no case distinction, AVO order is the most common dominant transitive order (19; 73.08%). Compared to that, a dominant AOV order is only found in four languages (15.38%), and VAO only in two (7.69%). One language (3.85%) further applies a dominant order where O directly precedes the verb which could either lead to an OVA or an AOV order. For the majority of languages with unclear case marking, the dominant order was equally unclear (17; 68%). Overall, there is a greater variation of dominant constituent orders in case marking languages. In these languages, orders where both arguments either follow or precede the verb, i.e. verb-initial or verb-final orders, are the most frequent. Although these orders do occur in certain languages without case marking, they are rare compared to the frequency of languages with dominant AVO order where grammatical roles are likely distinguished by their position in relation to the verb. That is, case marking allows for greater flexibility of constituent order.

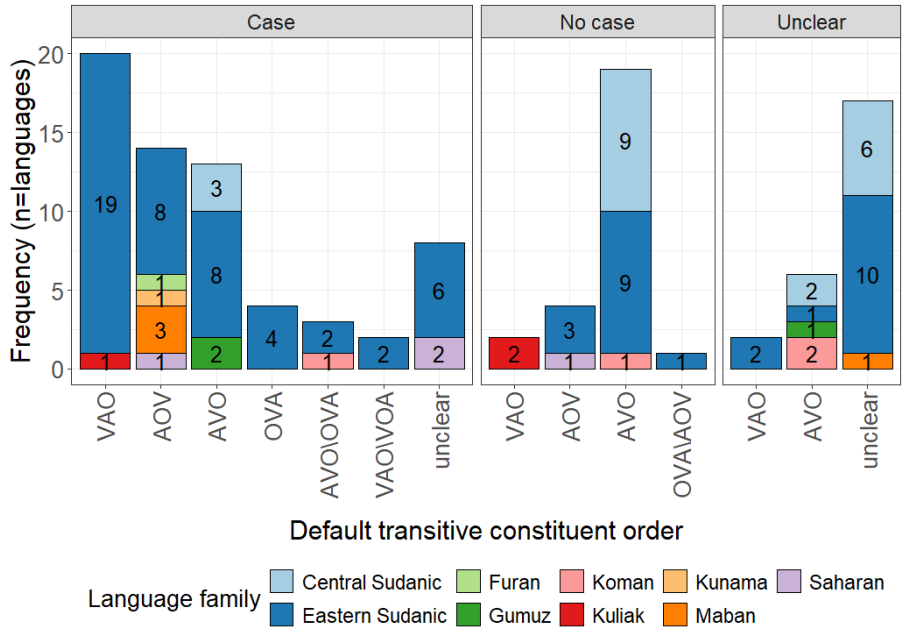

Figure 3.9: Dominant transitive constituent orders in Nilo-Saharan languages with case, no case, and unclear case marking, deduced from all the available descriptions.

The dominant order in intransitive clauses, as illustrated in Figure 3.10, shows a similar pattern to the transitive clauses. In contrast to the transitive order within case languages, the verb-initial order in intransitive clauses, i.e. VS, is not the most common dominant order with its 17 occurrences (26.56%). Instead, the majority have a dominant SV order (26; 40.63%), followed by 21 unclear instances (32.81%). Nonetheless, VS order in case languages is still a lot more frequent than in the languages without a case distinction where this order only occurs in two of the 26 languages (7.69%). With 19 occurrences (73.08%), SV order dominates in this category. There are further less unclear instances regarding the dominant intransitive order in languages without a case distinction.

Figure 3.11 shows the flexibility of transitive constituent orders in the 115 languages with available descriptions. Among the 64 languages that exhibit case marking, 26 languages (40.63%) have at least two other constituent order options in addition to the dominant order, labeled as 'yes', with the majority of 23 languages belonging to the Eastern Sudanic language family. In 12 of the languages with case marking, only one other constituent order next to the dominant order is possible (31.25%), annotated as 'partly'. Compared to this, the 26 languages that do not distinguish case are less flexible. Only two of these languages allow for at least two additional constituent orders besides the dominant order (7.69%), while 9 have only one additional constituent order (34.62%).

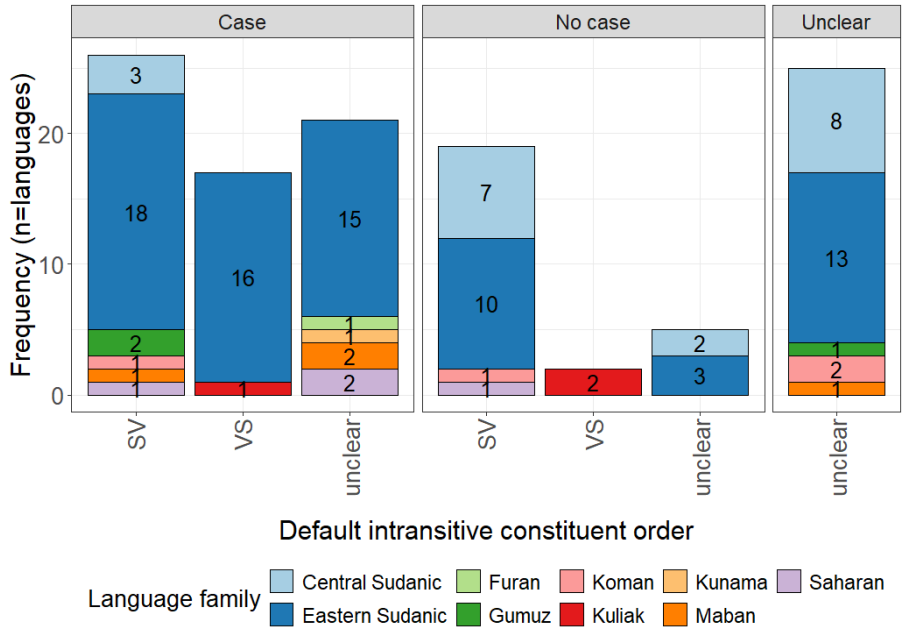

Figure 3.10: Dominant intransitive constituent orders in Nilo-Saharan languages with case, no case, and unclear case marking, deduced from all the available descriptions.

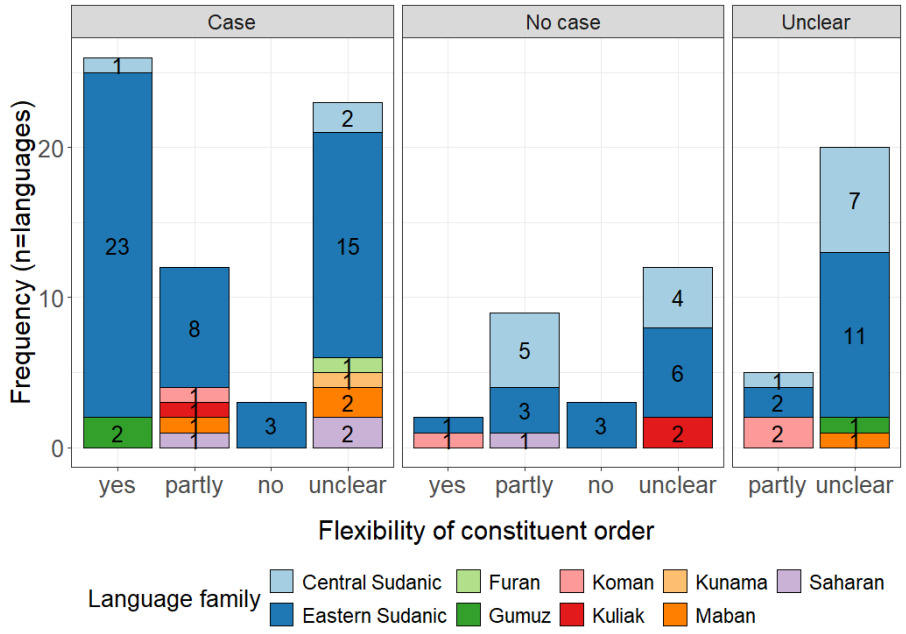

Figure 3.11: Flexibility of constituent order within the Nilo-Saharan language families with case, no case, and unclear case marking, deduced from all the available descriptions.

For both the languages with case marking and those without case marking, there is also a number of languages for which it was not clear whether they allow for additional constituent orders or not. The languages which have an unclear case marking status are also predominantly unclear in terms of the constituent order flexibility. All in all, the figure shows that the case

marking languages of the sample tend to be more flexible in terms of additional constituent orders than languages without case marking.

The language descriptions were also analyzed in terms of whether the languages described show an interaction of case marking and constituent order. Thus, the following results in Figure 3.12 show the 64 languages with case distinctions, as well as their respective case-marking systems. With a total of 30 languages, almost half of the languages with case marking show an interaction of case marking and constituent order variation, annotated with 'yes'. In these languages, arguments are only marked for case in postverbal but not in preverbal position. Although the languages with an accusative case system were the second most frequent in the survey (see Figure 3.8), the results show that interaction of case marking and constituent order variation (with unmarked preverbal arguments) is exclusive to marked nominative and ergative systems. In terms of the language family, the 30 languages with this interaction are only distributed over Eastern Sudanic (27; 90%), in two languages within the Gumuz family (6.67%), and in the only language within Koman that has case marking, Uduk (3.33%).

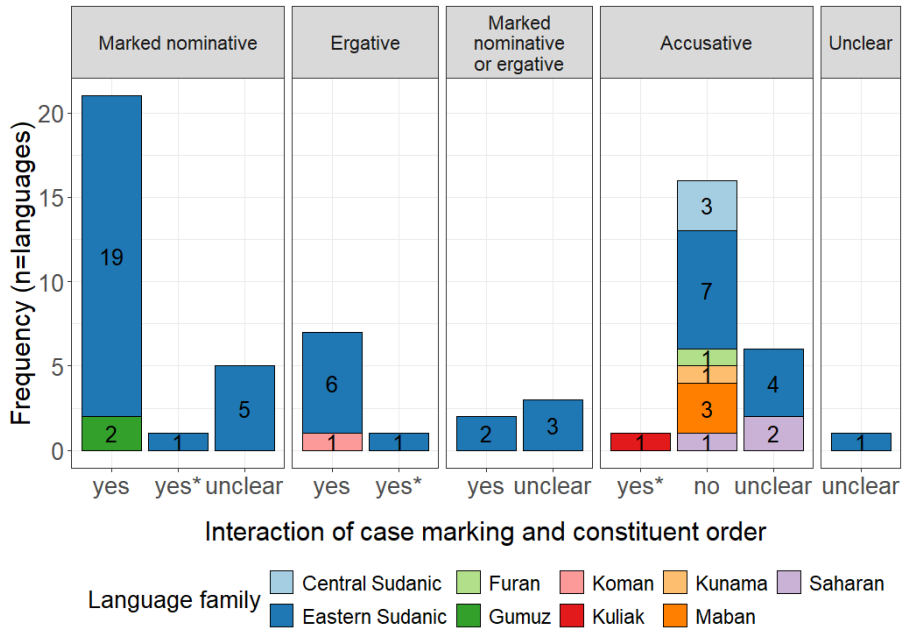

Figure 3.12: Languages with interaction of case marking and constituent order variation and their case-marking systems.

Just like in the previously discussed figures, there are also unclear instances when it comes to the interaction of case and constituent order. In these instances it was not clear whether a change in constituent order also plays a role in case marking as it was not explicitly stated and/or could not be deduced from the examples given in the descriptions. In the three languages marked with 'yes*', there was an interaction of constituent order and case marking but these languages behave differently from the other languages annotated with 'yes', as explained in Section 3.2.3.

One of these languages is Tennet, which was also discussed by König (2008a). In Tennet, a marked nominative language, case marking is not always absent in preverbal S and A arguments. The basic constituent order in Tennet is VAO, as illustrated in Example (20a), where A *Lokori* is marked by the nominative suffix *-i* (Randal 1998: 222). Additionally, Tennet also allows for preverbal S and A arguments under certain conditions, as can be seen in Example (20b). This is an example of a focus construction in Tennet which requires no special marking other than a change in constituent order, which, according to Randal (1998: 261) is not a cleft construction. However, as can be seen, the S argument *wála* 'crow' retains its nominative case marking in this construction. Preverbal S and A arguments are also marked with case in negative clauses where the constituent order is AVO, as illustrated in Example (20c) (Randal 1998: 248). Thus, there are constructions in Tennet in independent declarative main clauses in which the nominative marking is retained on A and S arguments. This distinguishes Tennet from the languages annotated with 'yes' in the data.

(20) Tennet (Surmic; South Sudan; Randal 1998: 222, 248, 261)

a. *á-káti Lokórí-i meger*
PFV-spear Lokori-NOM oribi
'Lokori speared an oribi.'

b. *ijja zin wála-i j-kiya*
and then crow-NOM IPFV-come
'And then Crow came.'

c. *iróng anná k-á-cín-i Lokúli iyókó nékô*
not 1SG:NOM 1-IPFV-see-1SG Lokuli now DEM
'I don't see Lokuli now.'

In the Kuliak language Ik arguments in preverbal position always have a dedicated case marker and Ik therefore differs from the other languages where case marking in preverbal position is absent. However, Ik is still marked with 'yes*', because the case marking shows an interaction with constituent order. In indicative declarative clauses, which have a basic VAO constituent order, A and S are always marked with nominative case, as shown with A *ŋók* 'dog' in Example (21a). Example (21b) shows that nominative case marking of preverbal A and S is retained in conditional or hypothetical subordinate clauses. However, in all other types of subordinate clauses, A and S receive accusative case marking. This is illustrated in Example (21c) where A *oŋori* 'elephant' is marked with the accusative suffix *-a* which is also found with the O argument of the clause, *kɔ́ná* 'scent', and also with O in the previous examples. In addition to the differential subject marking illustrated in the examples, Ik also has differential object marking. The direct object is always marked with accusative case with third-person subjects or in every transitive subordinate clauses, as in the examples below. When A is a first or second person referent, however, O is marked with nominative case, such as *ŋʊr* 'cane rat' in Example (21d) (Schrock 2014: 233-234).

(21) Ik (Kuliak; Uganda, Kenya; Schrock 2014: 233-235)[3]

a. *áts'-á ŋók-á ɔká-k^{a}*
gnaw.3SG-REAL dog-NOM bone-ACC
'The dog is gnawing a bone.'

b. *na=fet-a ŋƙ-ɛt-ɔ-ɔ bi-a*
CONJ=sun-NOM eat-INCH-3SG-SEQ 2SG-ACC
'If you get thirsty, [...]' (lit. 'If the sun starts eating you, [...]'

c. *náa oŋori-a wét-ét-i-e kɔínáá-a*
CONJ elephant-ACC drink-VEN-3SG-SIML scent-ACC
'When the elephant catches the scent, [...]'

d. *áts'-ʊ́kɔt-ím-á=naa ŋʊr-$^{a}_{o}$*
eat-COMPL-1PL.EXC-REAL=REC cane.rat-NOM
'We ate the cane rat up.'

As the Examples (20) and (21) show, there is an interaction of case marking and constituent order variation in these languages. However, the languages behave differently to the 'no case before the verb' rule in that the arguments either do not always appear in the unmarked form when they occur preverbally, as in negative clauses in Tennet, or they receive another dedicated case marker that differs from the one of arguments in postverbal position, which is happening in Ik. To differentiate these languages from the rest, they were annotated with 'yes*' and were not included in the further analysis that follows.

Figure 3.13 shows the case-marking systems and the case marking strategy in the 30 languages where case interferes with constituent order. As the previous figure showed, languages with an interaction are only found within the Eastern Sudanic, Gumuz, and Koman language families. Therefore, in the following, I will concentrate on the branches of these families. As explained above, there are only two case-marking systems relevant when it comes to the interaction with constituent order, namely marked nominative and ergative. The majority of these languages has a marked nominative case system (21; 70%), which can be found in the Nilotic (14; 66.67%) and Surmic (4; 19.05%) branches, as well as Berta (4.76%), and in the Gumuz branch (2; 9.52%). Ergative systems are found within Eastern Sudanic family in the Nilotic (4; 57.14%), Surmic (1; 14.29%) and Jebel (1; 14.29%), as well as in Central Koman (1; 14.29%). The most common case marking strategy in the observed languages is tonal inflection (17; 56.67%) followed by segmental marking (9; 30%) and syntactic[4] marking (2; 6.67%).

However, there are some exceptions to these generalizations. There are once again unclear cases in the data which concern the case-marking system in this instance, as was already briefly addressed earlier in this section. These two unclear instances are two Dinka varieties which have an interaction of case and constituent order but it is not quite clear whether they are marked nominative or ergative languages. In the Agar variety of Dinka, postverbal A is marked with a different tone pattern than postverbal O and preverbal S. The tone pattern of postverbal O

[3]The author made the following changes in the glosses: .3SG instead of [3SG]; 2SG instead of you.SG; REC instead of PST1.

[4]In the two languages classified as having a syntactic marking, case is indicated through adpositions (Burun; Andersen 2015) or particles (Uduk; Killian 2015).

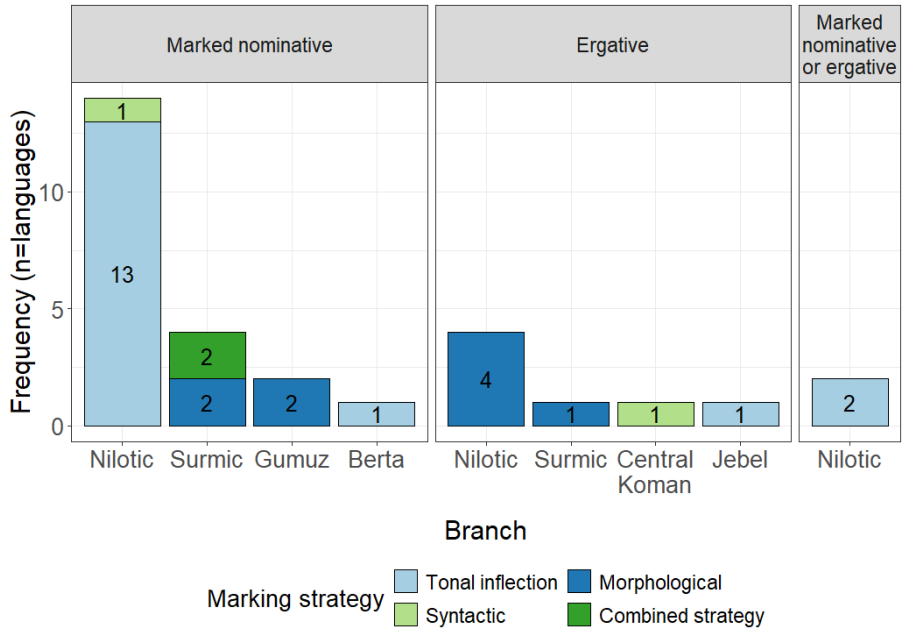

Figure 3.13: Case-marking systems and strategies in language families with interaction.

and preverbal S is further used for preverbal A, as well as for nouns in isolation, i.e. it is the citation form of a noun (Andersen 1991: 273). This aligns with the case marking behavior of other Western Nilotic languages, such as Päri (Andersen 1988), where postverbal S is not possible in declarative main clauses which is why they have an ergative case system rather than a marked nominative system. That is, a syntactic restriction on S means that it never occurs in a position where it could be case-marked and thus pattern with A, instead only patterning with O. According to Andersen (1991: 293), Agar-Dinka is not an ergative language. However, there are no examples of postverbal S provided and it is also not stated whether a VS order would be possible. A similar situation is found in the other Dinka variety, Southeastern Dinka (van Urk 2015). Another reason to consider the Dinka varieties as unclear instances concerning their case-marking systems is their case marking strategy. If they are indeed ergative languages, i.e. if postverbal S is impossible, they would be an exception within Western Nilotic since they mark case through tonal inflection and not through morphological marking like the other Western Nilotic languages. In order to classify these languages as marked nominative languages instead, where tonal inflection within Nilotic is the default, it would be necessary to know whether VS order in declarative clauses would be possible and, if so, how S behaves in this position.

Another exception to the generalizations from Figure 3.13 are the two languages that mark case with a combination of (segmental) morphological strategies and tonal inflection which are the Surmic (Eastern Sudanic) languages Tirma and Baale. In Tirma, not to be confused with the Niger-Congo language Tima, the dominant constituent order in transitive clauses is AVO where A is not marked for case, as in Example (22a). Like the other 29 languages marked with 'yes' for the interaction, case is only marked on postverbal arguments. The most common case marking strategy in Tirma is by means of the suffix *-o*, which is marking *ŋatuŋ* 'lion' for

nominative case in Example (22b). However, when A or S are proper nouns or common nouns that end with *o*, case is marked by tonal inflection. In Example (22c) the preverbal A argument *zugo* 'people' is in its citation form, i.e. it is unmarked for case. In postverbal position, such as in Example (22d), there is a tonal change to a final high tone on *zugo* that marks the nominative case (Bryant 1999: 45-46).

(22) Tirma (Surmic; Ethiopia; Bryant 1999: 45-46)[5]

a. *ŋatuŋ ám cíggín*
lion.SG eat.IPFV.3SG.SBJ hartebeest.SG
'The lion eats the hartebeest.'

b. *cíggín ám ŋatuŋ-o*
hartebeest.SG eat.IPFV.3SG.SBJ lion.SG-NOM
'The lion eats the hartebeest.'

c. *zugo ɗák-ɛ kono*
people hit.IPFV-3SG.SBJ snake
'The people hit the snake.'

d. *na ɗák-áŋ-ɛ zugó ɓɛ ciŋi*
and hit.IPFV-1SG.OBJ-3PL.SBJ people.NOM place little
'Then the people hit me a little bit.'

Baale, the other language annotated with 'combined strategy' behaves similarly in that some nouns take case suffixes while others indicate case through tonal inflection (Yigezu & Dimmendaal 1998: 298).

The dominant constituent orders of all languages with available descriptions were already discussed earlier in this section. Figure 3.14 focuses on the dominant transitive constituent orders in the relevant branches with an interaction of case marking and constituent order. In addition it shows the distribution of constituent orders across the different case-marking systems. Out of the 21 marked nominative languages in the data, 11 (52.38%) have a dominant VAO order and two are either VAO or VOA (9.52%). In addition, there are seven languages with a dominant AVO order (33.33%) and one language which is either AVO or OVA (4.76%). The languages marked as AVO/OVA and VAO/VOA do not have a single dominant order as both orders occur frequently. Following this, there is a tendency for dominant transitive verb-initial orders in marked nominative languages. In comparison to that, languages with an ergative case-marking system do not have dominant verb-initial constituent orders. Instead, all of the ergative languages in the data are verb-medial with OVA being the most frequent (4; 57.14%), followed by AVO (2; 28.57%) and one language which is either AVO or OVA (14.29%).

In the two Dinka varieties that are either marked nominative or ergative, the transitive dominant orders are both verb-medial. This could be another argument against classifying them as marked nominative languages as these are typically verb-initial within the Nilotic branch. There are only two other Nilotic languages with a marked nominative system that have a dominant

[5]The author made the following changes in the glosses: 3SG.SBJ instead of 3SiSu; 1SG.OBJ instead of 1SiO; 3PL.SBJ instead of 3PlSu; NOM instead of PVS

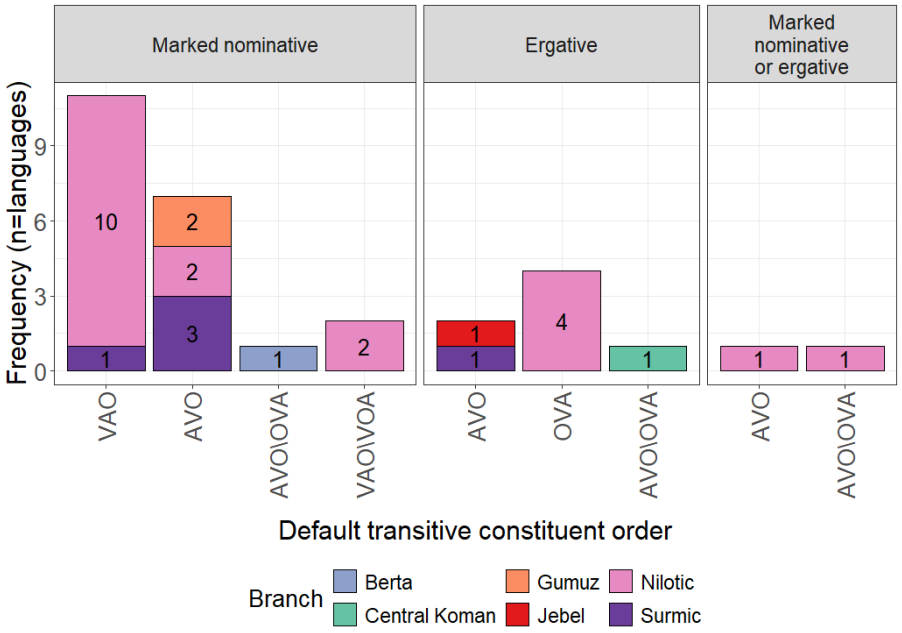

Figure 3.14: Dominant transitive constituent order in language families with interaction of case marking and constituent order variation.

AVO constituent order. The first one is the Datooga dialect Asimjeega (Griscom 2019) and the second one is the Western Nilotic language Burun (Kurmuk) (Andersen 2015). Burun is special in its own right as it is the only Nilotic language that indicates case with adpositions for postverbal S and A arguments. It is therefore also the only clear instance of a marked nominative language belonging to the Western Nilotic branch to which the Dinka varieties also belong. Thus, marked nominative occurs with AVO order in Nilotic and also in the Western Nilotic branch. Since languages with AVO order and marked nominative are attested, we cannot straightforwardly argue that verb-medial Dinka must in fact be ergative, as I suggested above. However, as mentioned above, it is necessary to know whether postverbal S arguments are possible in declarative main clauses.

In addition to dominant transitive orders, Figure 3.15 shows the dominant intransitive constituent orders in the languages that show an interaction of case marking and constituent order allocated to the different case-marking systems. Just as for the transitive clauses in the previous Figure 3.14, the majority of the 21 marked nominative languages are verb-initial in their dominant intransitive orders (12; 57.14%). However, there are also six languages that have a dominant SV order (28.57%), as well as three language for which it is not clear whether the basic intransitive order is VS or SV (14.29%). Compared to this, the ergative languages, as well as the two languages for which the case system is either marked nominative or ergative, all have a dominant intransitive SV order. Thus, while for marked nominative languages, both orders are possible as the basic intransitive order, all of the ergative languages in the data have a basic SV order. Although the majority of Nilotic languages with a marked nominative system have

a basic VS order with an exception of one language, a marked nominative system for the two Dinka varieties can still not be ruled out.

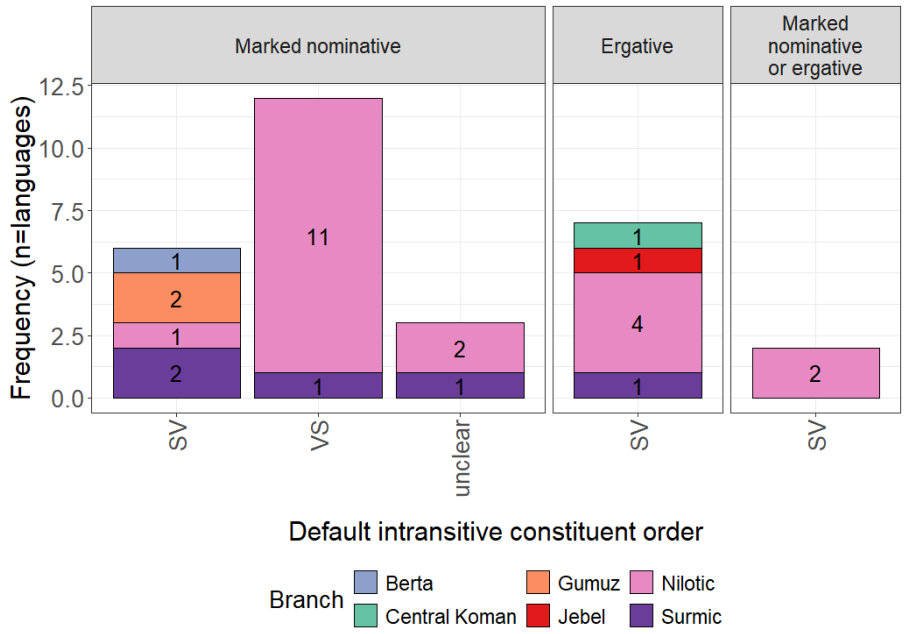

Figure 3.15: Dominant intransitive constituent order in language families with interaction of case marking and constituent order variation.

This section has offered a quantified overview of Nilo-Saharan languages, focusing on their constituent orders and flexibility, case-marking systems, and the interaction between constituent order and case marking. In the following sections, I will discuss the conditions that lead to both case splits and constituent order variations in the 30 languages with an interaction of case marking and constituent order variation.

3.2.4.1 Conditions for case split

In 30 languages of the data, a change in constituent order results in a case split, whereby A and/or S arguments lose their case marking if they appear before the verb. However, some of these languages show additional factors contributing to case split. Table 3.2 presents these conditions for split case marking, as documented in the descriptions of the languages within the dataset. Other conditions for case splits, aside from constituent order variation, are discussed in only a few languages. This does not necessarily mean that no other conditions exist in the languages not listed in the table, but rather that these possible conditions have not yet been analyzed in those languages. The conditions in the languages where additional conditions are discussed include animacy, definiteness, and verb semantics, which will be further examined in the following paragraphs.

The table also lists the condition of clause type. This condition is exclusive to the Western Nilotic languages Päri, Anywa, Jur-Luwo, as well as (at least) one of the Dinka varieties (Agar). In these languages, a change in the clause type, such as non-declarative or dependent clauses,

leads to a case split (Andersen 1988, Reh 1996, Andersen 2002, Storch 2014). These clause types usually also have a constituent order that differs from the declarative order which is why this condition is not completely independent from constituent order but constituent order is not the reason for the case split. However, there will be no further discussion of these languages as the focus of the survey lies on case split in independent declarative main clauses.

Condition	**Language(s)**
Animacy	Jur-Luwo, Chai, Tugen
Definiteness	Anywa
Verb semantics	Turkana, Teso
(Clause type)	Päri, Anywa, Jur-Luwo, Dinka

Table 3.2: Additional conditions for case split apart from constituent order variation. The condition 'clause type' is in brackets because it only applies in subordinate or non-declarative clauses which are not in the focus of the quantified analysis.

In Chai, animacy plays a significant role in case marking, as exemplified in the following instances. Example (23a) demonstrates an indicative clause with a postverbal agent, *Bume*, which is marked with the ergative case suffix *-o*. In contrast, the preverbal agent in Example (23b) does not carry ergative case marking. However, in the last Example (23c), the postverbal agent, *màŋgîsî* 'government', is also devoid of ergative case marking. According to Last & Lucassen (1998: 408), in clauses where O is inanimate and A is animate, only OVA constituent order is permissible, rendering case marking unnecessary. Note that this example also demonstrates that a relationship on the animacy hierarchy triggers a change in constituent order: OVA order is obligatory when the object is inanimate and the agent is animate. However, it is important to clarify that constituent order variation is not the reason for the case split, as the agent in this order typically receives ergative case marking, as evidenced in Example (23a). Thus, there are two splits operating here: Split case marking of postverbal A depending on the animacy of O, and split case marking of A depending on its position relative to the verb. It is further noteworthy that Example (23c) represents a ditransitive clause, whereas the preceding examples are transitive.

(23) Chai (Surmic; Ethiopia; Last & Lucassen 1998: 407-408)[6]

a. *ŋakogine* *haŋae* *bume-o*
Ngakogine.ABS chase.PFV.3PL.3SG.OBJ Bume-ERG
'The Bume chased Ngakogine.'

b. *bume* *haŋae* *ŋakogine*
Bume.ABS chase.PFV.3PL.3SG.OBJ Ngakogine.ABS
'The Bume chased Ngakogine.'

[6]The author made the following changes in the glosses: 3SG.OBJ instead of o3SG; 1SG.OBJ instead of o1SG.

c. *mòlkóɣî ɗónè ájàɲ màŋgîsî*
Moskovitch one give.PFV.3SG.1SG.OBJ government
'The government had given one Moskovitch (=type of Russian gun) to me.'

Another condition for case split observed in the languages within the dataset is definiteness. According to Reh (1996: 137), Anywa does not differentiate core arguments based on case-marking. Instead, the suffix that resembles the ergative case suffix in related languages, such as Päri (Andersen 1988), is primarily used to indicate definiteness. However, according to König (2008a: 114), an interpretation of the marker as a definiteness marker would leave some open questions. This includes the restrictiveness of the marker to S and A and further to the postverbal position of these arguments. König (2008a: 114) argues that these open questions would be resolved when interpreting the marker as a case marker that is restricted to definite S and A arguments instead of a definiteness marker. Hence, in the following discussion, the suffix is regarded as a case marker with definiteness serving as a condition for case split. This is illustrated by the clauses in Example (24).

In Anywa, the S argument *ɲīlàal* 'child' in Example (24a) appears in a preverbal position and is not marked for ergative case. Furthermore, in this position, it can exhibit either definite or indefinite status. In contrast, in Example (24b), where 'child' is functioning as A, postverbally, and is definite, it is marked by the ergative suffix *-lɪ̀*. Example (24c) illustrates that even proper nouns, such as *Acook*, can be marked in postverbal position. However, if the A argument in postverbal position is indefinite, it loses its case marking, as seen in Example (24d). Interestingly, proper nouns acquire an indefinite interpretation without the marker, as demonstrated in Example (24e).

(24) Anywa (Western Nilotic; Ethiopia, South Sudan; Reh 1996: 137, 138)[7]

a. *ɲīlàal ā-àá pàaB*
child PST-leave.for forest
'A/the child went to a/the forest.'

b. *rìŋō ā-cám ɲìlàal-lɪ̀*
meat PST-eat child-ERG
'The child ate the meat (up).'

c. *lúumé ā-wíɲ ācɔ́kí*
word:PL:MN:3SG PST-hear Acook:ERG
'Acook heard her words.'

d. *rìŋō ā-cám ɲìlàal*
meat PST-eat child
'A child ate the meat (up).'

e. *lúumé ā-wíɲ ācɔ́k*
word:PL:MN:3SG PST-hear Acook
'A certain Acook heard her words.'

[7]The author made the following changes in the glosses: PST instead of PA; ERG instead of DEF; PL instead of P; 3SG instead of 3S

Another case split condition found in the languages of the data concerns the semantics of a verb. The Eastern Nilotic language Teso is a marked nominative language, as can be deduced from the following examples. In the intransitive clause in Example (25a) S *àbɛ̀rʊ̀* 'woman' is in the nominative case which is marked through tonal inflection in Teso. This is the same tone pattern used for A, *àbɛ̀rʊ̀* 'woman', in Example (25b) while O, *èkíljòkít* 'man' is in its citation form in the absolute case (Barasa 2017: 206). However, Teso applies case split with specific verbs, typically non-volitional, in intransitive clauses. In these clauses, the only argument of the verb, S, is marked with the absolute rather than the nominative, such as in Example (25c) with *àbɛ́rʊ̀* 'woman' (Barasa 2017: 209–210).

(25) Teso (Eastern Nilotic; South Sudan; Barasa 2017: 206, 209–210)

a. *é-móŋ-í* *àbɛ̀rʊ̀*
3SG-cry-IPFV woman.NOM
'The woman was crying.'

b. *é-míná* *àbɛ̀rʊ̀* *èkíljòkít*
3-love woman.NOM man.SG.ABS
'The woman loves the man.'

c. *kí-cék-í* *àbɛ́rʊ̀*
3SG-tremble-IPFV woman.ABS
'The woman is trembling.'

3.2.4.2 Conditions for constituent order variation

As demonstrated in the preceding section, only a limited number of languages exhibit case split conditions distinct from constituent order variation. The present section explores the factors conditioning changes in constituent order in the languages within the dataset. These conditions are either independent of case marking, as they exclusively pertain to variations in the order of postverbal arguments—typically case-marked—or they involve changes in constituent order leading to a simultaneous case split. These conditions are outlined in Table 3.3.

As was the case for the case split conditions in the previous section, animacy also plays a role when it comes to constituent order variation. This is the case for the Eastern Nilotic languages Turkana, Teso, Lopit, and Toposa. However, in these languages, the variation only concerns postverbal arguments, which is exemplified in the following example from Lopit.

(26) Lopit (Eastern Nilotic; South Sudan; Moodie & Billington 2020: 255, 257)

a. *ɛ-ìlːá* *xótːɔ́ɲí* *ìŋɛ̂*
3-wash.IPFV mother.NOM baby.ABS
'The mother washes the baby.'

b. *eí-ŋór-ò* *nàŋ* *lɔ́xɔ́tɔ́ɾɔ́*
3>1-stingIPFV 1SG.ABS bees.NOM
'The bees are stinging me.'

The first Example (26a) shows the dominant order in Lopit, VAO. However, if the object "is higher in prominence than the subject" (Moodie & Billington 2020: 256), VOA order is employed,

Condition	Language(s)
Animacy	Turkana, Teso, Lopit, Toposa, Tennet
Argument realization	Toposa, Anywa, Gisamjanga-Datooga, Teso, Turkana
Negation	Jur-Luwo, Teso, Lopit
Verb semantics	Shilluk
Information structure	Maasai, Turkana, Cherang'any, Asimjeega-Datooga, Päri, Anywa, Shilluk, Jur-Luwo, Murle, Toposa, Teso, Tirma, Chai, Mursi, Gaahmg, Uduk, Berta, Lopit, Samburu, Burun, Nandi, Tugen, Akie, South Central Dinka (Agar)

Table 3.3: Conditions for constituent order variation described in the 30 Nilo-Saharan languages where case marking interferes with constituent order variation.

as in Example (26b). The constituent order of the postverbal arguments is arranged by the following hierarchy:

> 1st pronouns & 2nd person pronouns > 3rd person pronouns > proper names > humans > animates > inanimates (Moodie & Billington 2020: 257)

The arrangement of postverbal arguments according to this hierarchy is also observed in Turkana, Teso, and Toposa.

Another condition for constituent order variation is the realization of arguments, as illustrated in the Toposa examples below. In Toposa, the order of postverbal arguments is conditioned by whether they are realized as a pronoun or a noun. When both A and O are pronouns, VAO order is applied, as in Example (27a). However, if O is a pronoun in first or second person and A is a noun, O precedes A, as in example (27b) (Schröder 2008: 61-62). This differs from animacy discussed above, as O in Example (27a) occurs after A although it is technically higher on the proposed animacy hierarchy by Moodie & Billington (2020) than A.

(27) Toposa (Eastern Nilotic; South Sudan; Schröder 2008: 61-62)[8]

a. *k-ì-lìm-ókín-î* *íyòŋ* *áyôŋ*
OBJ-2SG.SBJ-tell-BEN-IMP 2SG.NOM 1SG.ABS
'You will tell me.'

b. *k-à-lìm-ókín-î* *áyôŋ* *ló-káàtó* *káŋ*
OBJ-1SG.OBJ-tell-BEN-IMP 1SG.ABS M.SG-brother.NOM 1SG.POSS
'My brother will tell me.'

Argument realization also plays a role in Datooga and Anywa. According to König (2008a: 10, 282), argument realization (or person in König's 2008a terms) serves as another condition

[8]The author made the following changes in the glosses: SBJ instead of SUB; 2SG instead of you; 1SG instead of me; ABS instead of ACC; 1SG.POSS instead of my; . instead of /.

for case split in Datooga, where split case marking only occurs with third person pronouns and never with first or second person pronouns according to Kießling (2007: 160-161). However, first and second person pronouns in Datooga rarely appear independently and these person categories are typically just indexed on the verb. When they do occur independently, Kießling (2007) only attested them in preverbal positions where no case distinction is present. Independent third person pronouns, on the other hand, can ocur postverbally and receive case marking when used in a switch-reference context (Kießling 2007: 160-161). A similar pattern is also observed in the Western Nilotic language Anywa (Reh 1996). Since first and second person pronouns in these languages cannot occur in a position allowing for case marking (i.e., postverbal position), and they do not contrast with third person pronouns in this context, argument realization is not considered a case split condition in these instances. By contrast, argument realization signals a preference for a certain constituent order over another, i.e. preverbal position in contrast to postverbal position, with first and second person pronouns. Therefore, it is considered a condition for constituent order variation rather than case split.

In some languages the polarity of a clause can trigger a different constituent order, such as is the case for Teso. In affirmative clauses, Teso has a basic VAO constituent order, as in Example (28a). In negative clauses, the constituent order changes to AVO, as in Example (28b), where A is in the unmarked absolute case (Barasa 2017: 208). However, it is important to note that the verb *ɲàm* appears in its infinitive form, as indicated by the prefix *à-*. Therefore, the distinction between affirmative and negative clauses is not solely based on the presence of the negation marker *mám'ʊ* and a change in constituent order as the verb form also changes.

(28) Teso (Eastern Nilotic; South Sudan; Barasa 2017: 208)[9]

a. *É-ɲàm àbèrʊ̀ épʊ̀ʊ̀sì*
3SG-eat:PST woman.NOM cat.ABS
'The woman ate the cat.'

b. *mámʊ́ àbérʊ̀ à-ɲàm épʊ̀ʊ̀s(ì)*
NEG woman.ABS INF-eat cat.ABS
'The woman did not eat the cat.'

A similar pattern can be observed in Jur-Luwo and Lopit.

Another condition for constituent order variation observed in the data is the semantic of the verb. In Shilluk, a Western Nilotic language, the dominant constituent order in transitive clauses is OVA, as in Example (29a). In this clause, O *byél* 'grain' precedes the verb and is not marked for case while A, *ɲān dájɔ̀* 'woman (lit. 'person female)' follows the verb and is marked with the ergative marker *yī*. However, constituent order in Shilluk can change for semantic reasons, such as in Example (29b) where the AVO order shows that "the agent chose to accomplish an action with respect to a particular goal" (Miller & Gilley 2001: 36).

(29) Shilluk (Western Nilotic; Sudan, South Sudan; Miller & Gilley 2001: 36)

[9]The author made the following changes in the glosses: PST instead of PAST; . instead of /.

a. *byɛ́l* *á-ˈrākk`* *yī* *ɲān* *dájɔ̀*
grain:PL PST:E-grind:TR:R ERG person female
'The woman ground the durra.'

b. *ɲān* *dájɔ̀* *á-ˈrākk`* *byɛ́l*
person female PST:E-grind:TR:R grain:PL
'The woman chose to grind the durra.'

Up to this point, it is evident that there are relatively few conditions for constituent order variation explicitly mentioned in the languages within the dataset. However, as depicted in Table 3.3, there is a prevalent condition named as a reason for constituent order change in almost every language, and that is information structure. Indeed, information structure emerges as a prominent and widespread factor influencing constituent order variation in almost all of the languages in the dataset. The following section will provide insights into how information structure is described to impact constituent order variation across these languages.

3.2.5 Information structure

In the preceding Section 3.2.4.2, it was elucidated that the primary condition for constituent order variation in the languages under investigation is linked to information structure. Notions that occur in this context include topic (or topicalization), focus, as well as contrast and emphasis, which were introduced in Chapter 2.1.

However, in the context of writing descriptive grammars of underdocumented languages, it is often not possible to conduct in-depth investigations of information structure. Accordingly, grammars often do not give precise definitions of notions such as 'topic' or 'focus'. In the following, I will therefore aim to interpret the findings of these language descriptions in terms of the definitions provided in Section 2.2.3.2 whenever possible.

Concerning the concepts of topic and topicalization, 15 languages in the sample are described as exhibiting a correlation between a change in constituent order and these notions, as detailed in Table 3.4. In these languages, topicalization typically involves the fronting of an argument to the initial position of the clause. However, it is not always the case that the preverbal position is occupied by the topic.

Some scholars interpret the topic as referring to the 'aboutness' of a predication, as observed in the Eastern Nilotic language Turkana (Dimmendaal 1983: 306)[10], and the Western Nilotic language Anywa (Reh 1996: 191).

The concept of topics is further frequently associated with old or given information, as seen in Turkana (Dimmendaal 1983: 64) and Chai (Last & Lucassen 1998: 413). However, as was already discussed in Section 2.2.3.2, topics, i.e. in the sense of 'aboutness', do not necessarily represent old or given information although this is a strong tendency. The introduction of new topics can also lead to a change in constituent order under certain circumstances, as is the case in Chai.

[10]It is noteworthy that Dimmendaal (1983) does not explicitly state this as the definition of a topic, but it can be inferred from the subtext.

Language family	Branch	Language	Case marking system	Dominant order (transitive)	Dominant order (intransitive)
Eastern Sudanic	Nilotic	Turkana	marked nominative	VAO\VOA	VS
		Cherang'any	marked nominative	VAO	VS
		Datooga	marked nominative	VAO	VS
		Nandi	marked nominative	VAO	VS
		Samburu	marked nominative	VAO	VS
		Päri	ergative	OVA	SV
		Anywa	ergative	OVA	SV
		Shilluk	ergative	OVA	SV
		Jur-Luwo	ergative	OVA	SV
		Dinka	marked nominative or ergative	AVO	SV
		Dinka	marked nominative or ergative	AVO	SV
	Surmic	Tirma	marked nominative	AVO	SV
		Mursi	marked nominative	AVO	SV
		Chai	ergative	AVO	SV
		Berta	marked nominative	AVO\OVA	SV

Table 3.4: Languages for which constituent order variation has been associated with the notion of topic or topicalization.

In Chai, an ergative language with a dominant AVO order, OVA order is typically applied whenever a topic is new or resumed. In the first Example (30a), OVA order is used due to a break in the topic chain. Prior to this clause, a different referent was the topic over several clauses. Here, *ɛrro* becomes the new topic, i.e. the clause is about them, thus breaking the chain. The postverbal position further puts *ɛrro* in focus as the order immediately following the verb is associated with focus in Chai (Last & Lucassen 1998: 413). (O)VA order can also be used for the resumption of a topic, as in Example (30b). In this example, *Bume* is a topic that is picked up again and thus occurs in postverbal position. Just like *ɛrro* in Example (30a), there is also focus on *Bume* in this position.

(30) Chai (Surmic; Ethiopia, South Sudan; Last & Lucassen 1998: 413)

a. *na ɓókòn īlgè ɛrro wurlɔ*
NAR half.of.it return.IPFV.3PL kids.NOM.FOC afterwards
'And *my children* took half of it afterwards.'

b. *hangai buméo hangai ngabayo*
chase.PFV.3SG.1PL.OBJ Bume.NOM.NT.FOC chase DEM.place.PP.FOC
'*The Bume* chased us, he chased us to *this place*.'

Thus, it seems like topic and focus interact with each other in Chai in that OVA order is used whenever a topic, in the sense of 'aboutness', i.e. what the clause is about, is resumed or new. Both O referents in the first two clauses are given information while A provides new or 'inactive' information. This shows that constituent order in Chai is also sensitive to the 'old-new-principle' (Ward & Birner 2006), i.e. old information precedes new information.

A similar pattern can be observed in Dinka (Andersen 1991: 279), Burun (Andersen 2015: 509), Berta (Andersen 1995: 44), and Anywa (Reh 1996: 191), but the concept of "topic" appears to be used differently. Following Andersen (1991: 279), the preverbal and clause-initial position has obligatory to be filled with an NP in Dinka. As this position can both be filled with A and O arguments, Andersen (1991: 278) applies the notion of "topic" to refer to the NP in preverbal position. Although the notion of "topic" in these descriptions seems to differ from the information structural category, as defined in Section 2.2.3.2, it still seems to fulfill certain discourse functions. As for Burun, Andersen (2015: 534) notes that the preverbal position is typically assigned to referents that represent given information while the postverbal position is reserved for new referents. This pattern is especially common for the NP in O function.

In two languages, topics are also considered to be more definite and/or animate. This characteristic is observed in Jur-Luwo (Storch 2014: 201) and Turkana (Dimmendaal 1983: 407–408) and is illustrated in the following example. In OVA clauses in Jur-Luwo, such as the one in Example (31a), postverbal A *dhècwɔ̀w-é* 'man' can be interpreted as either definite or indefinite while preverbal O *rɛ́yɔ́* 'fish' is considered definite and topical. With a change to AVO order, as in Example (31b) A *dhècwɔ̀w* 'man' becomes topical and definite (Storch 2014: 201).

(31) Jur-Luwo (Western Nilotic; South Sudan; Storch 2014: 201)[11]

a. *rɛ́yɔ́ à-cám dhècwɔ̀w-é*
fish:SGV PFV-eat:TR man-ERG
'The/a man ate the[12] fish.'

b. *dhècwɔ̀w ù-cʌ́mɔ̀ ké rɛ́yɔ́*
man IPFV-eat:AP PREP fish:SGV
'The man eats (some kind of) fish.'

Another important aspect discussed in language descriptions regarding constituent order variation is focus. Among the languages in the dataset, eleven languages exhibit a connection between a change in constituent order and focus, as detailed in Table 3.5.

In six out of the eleven languages in which a change in constituent order is associated with focus, the argument immediately following the verb is regarded as the focused element. These languages include Jur-Luwo (Storch 2014: 198); Tirma (Bryant 1999: 42); Chai (Last & Lucassen 1998: 412); Mursi (Worku 2020: 476); Päri (Andersen 1988: 308); and Anywa (Reh 1996: 392).

[11] The author made the following changes in the glosses: ERG instead of CASE.

[12] There is no definite article in the translation given by Storch (2014: 201). However, in another context she translates the same form, i.e. the singulative, with 'a single fish' (Storch 2014: 91). Given the explanation above, which considers a definite interpretation of preverbal NPs, I chose to add the definite article in this translation.

Language family	Branch	Language	Case marking system	Dominant order (transitive)	Dominant order (intransitive)
Eastern Sudanic	Nilotic	Maasai (Arusa)	marked nominative	VAO	unclear
		Toposa	marked nominative	VAO	SV
		Lopit	marked nominative	VAO	VS
		Tugen	marked nominative	VAO\VOA	VS
		Shilluk	ergative	OVA	SV
		Jur-Luwo	ergative	OVA	SV
		Dinka (Southeastern)	marked nominative or ergative	AVO	SV
	Surmic	Mursi	marked nominative	AVO	SV
		Tirma	marked nominative	AVO	SV
		Chai	ergative	AVO	SV
	Jebel	Gaahmg	ergative	AVO	SV

Table 3.5: Languages for which a change in constituent order has been associated with focus.

Additionally, the Arusa dialect of Maasai employs three focus orders, with VOA focusing on the immediate post-verbal object argument (Levergood 1987: 43).

In contrast, the other two constituent orders associated with focus in Arusa-Maasai, namely AVO and OVA, place the preverbal argument in focus (Levergood 1987: 43). In other languages this preverbal position is reserved for constituents in contrastive focus. These languages include Toposa (Schröder 2008: 56), Gaahmg (Stirtz 2014: 244–247), and Lopit (Moodie & Billington 2020: 271). Gaahmg even distinguishes between assertive and contrastive object focus, using OVA order for assertive focus and OAV order for contrastive focus on O. Following (Stirtz 2014: 247), a constituent in assertive focus differs to one in contrastive focus "in that the hearers have no values in mind for taking the focus slot". Thus, assertive focus is non-contrastive. The distinction is further illustrated in the following examples. Example (32a) shows an assertive focus construction with O *mīī* 'goat' preceding the verb *mán* 'beat', leading to an OVA order. According to (Stirtz 2014: 248) the ergative marker on the verb, *=ɛ́*, "indicates that an agent follows the verb". The other two examples, (32b) and (32c), have constituents in contrastive focus, namely *mīín* 'the chicken' and *mīīn* 'the goat', which directly contrast with each other. Both clauses have an OAV order.

(32) Gaahmg (Jebel; Sudan; Stirtz 2014: 247–248)

a. *mīī mán nāmán=ɛ́ ɛ́ ūlg=ì mâng wá*
goat certain beat.CONT=ERG GP thirst.ERG=ERG well NEG
'There was once a very thirsty goat. (lit. A certain goat, thirst beat badly.)'

b. *mīín á gàn tú,*
chicken.DEF 1SG.NOM throw.INCOMPL away
'The chicken I am throwing away,'

c. *mīīn á gàfàn ɟɔg fangán ā nə́mdíìgə̀*
goat.DEF 1SG.NOM give.CONT people old.DAT SBJV eat.SBJV
'but the goat I am giving to the old men to eat.'

In his article, Stirtz (2014) only discusses objects in focus. It is not clear how the focus construction would look like if A or S were the arguments in focus given the dominant order of AVO/SV in Gaahmg.

Concerning the position, one could conclude that whenever contrastive focus occurs in these languages, the contrasting referent is moved to a preverbal, clause-initial position. However, there are two languages in the sample where the shift to the immediate post-verbal position is associated with contrastive focus. This is found in Datooga (Kießling 2007: 158) and Shilluk (Miller & Gilley 2001: 36).

The notions of topic and focus are further accompanied by the notion of 'emphasis' in the language descriptions which is typically left without further clarification. For example, in the Eastern Nilotic language Maasai, a constituent order featuring a preverbal object is said to imply "a certain emphasis" (Tucker & Mpaayei 1955: 11), yet the nature of this emphasis is not elaborated upon. Similarly, in the Southern Nilotic language Cherang'any, the notion of 'emphasis' is presented in combination with topicalization: "3rd person is marked in clauses where the VSO order changes to SVO order due to emphasis or topicalization" (Mietzner 2016: 108). Thus, it seems like it is the preverbal position of arguments that emphasizes them in these verb-initial languages.

In the Western Nilotic language Jur-Luwo, AVO order is employed "whenever the O-participant is focused and emphasized" (Storch 2014: 198). In this language, it appears that 'emphasis' occurs with focused elements that immediately follow the verb rather than with topicalized elements which were discussed earlier in this section.

3.3 Discussion

The distribution of differential case marking in Nilo-Saharan languages, in conjunction with constituent order variation, presents a fascinating linguistic landscape. Concerning the first research question, it becomes apparent that case marking is prevalent across all nine relevant language families outlined in Figure 3.4 with the majority of case languages belonging to Eastern Sudanic. Within these languages, marked nominative case systems are the most frequent, followed by accusative systems, while there are only a few languages with ergative systems. However, while accusative case systems are the second most frequent case marking system within the data, the interaction with constituent order variation appears to be exclusive to languages with an ergative or marked nominative case system. Moreover, the 'no case before the verb' feature is found in only three branches: Eastern Sudanic, Gumuz, and Koman.

The second research question concerned the conditions relevant for case split independent of constituent order variation, as well as the conditions for constituent order variation itself. The survey shows that factors impacting case marking mentioned in the language descriptions are animacy, definiteness, as well as verb semantics. There are only a few authors who discuss these conditions in their language descriptions but this does not mean that the other languages do not allow for additional splits to the 'no case before the verb' split.

Animacy and verb semantics also play a role in constituent order variation in certain languages within Nilo-Saharan. Additionally, the realization of arguments, as well as negation can condition a change in constituent order in these languages. However, the most important condition for constituent order variation appears to be information structure which is mentioned in almost every language description when it comes to a change in the order of arguments.

The role of information structure in shaping constituent order was the third research question of the typological survey. Scholars often link changes in constituent order to information structural categories such as topic, topicalization, and focus. As for topic, in the language descriptions it is often linked to properties like aboutness, definiteness, as well as old or given information, and typically occupies the preverbal position. However, in the case of topic shift, the position of the topic can be different, such as in Chai which uses the postverbal position for the change of topics after a topic chain, as well as the resumption of old topics.

In the realm of focus, most scholars attribute postverbal positions to the focused element, though in Arusa-Maasai the preverbal position in AVO and OVA order is associated with focus. This contrasts with languages like Toposa, Gaahmg, and Lopit, where the preverbal position puts the argument in contrastive focus. A postverbal position of contrastive focus can be observed in Gisamjanga-Datooga, as well as Shilluk. Thus, in the majority of languages for which focus is discussed, non-contrastive focus is typically linked to arguments in postverbal position while contrastive focus is more common in preverbal position.

Both topic and focus co-occur with the notion of emphasis in certain languages and there seems to be a difference between languages with a dominant verb-initial order and those with verb-medial order: While for verb-initial languages Arusa-Maasai and Cherang'any, the notion of "emphasis" is used for the topicalized argument in preverbal position, it is the focused element in postverbal position that is emphasized in the verb-medial language Jur-Luwo. However, although this might show a pattern, it only appears to be a weak pattern, if at all, as the notion of "emphasis" only occurs in a few languages and it is moreover a vague concept. The analysis of information structure within languages is a time-consuming task, and as such, the data available does not always provide insights on these categories to the same extent. Therefore, these possible patterns should be viewed as preliminary, which leave room for future research to explore them in greater depth.

Chapter 4

About Datooga

The preceding chapters delved into the topics of case marking and constituent order, initially in a broader context and subsequently focusing on East African languages. It was highlighted that a notable feature common to many East African languages, across various phyla, is the case marking of postverbal agents/subjects while lacking such marking for preverbal ones. Thus, constituent order is important when it comes to case marking in these languages and it underscores the significance of understanding the role of constituent order variation.

One language exemplifying these characteristics is the Southern Nilotic dialect cluster Datooga, which will be the subject of the forthcoming case study of constituent order variation. The following chapter provides an introduction to the language and the properties relevant for the study of constituent order variation. Initial sections will furnish an overview of Datooga and its speakers (Section 4.1), alongside a typological profile (Section 4.2). Subsequently, I will examine how constituent order operates within subordinate clauses in Section 4.3.1. Given Datooga's verb-initial nature, this chapter will also investigate the types of non-argument elements permissible in preverbal positions in Section 4.3.2, as well as structures that lead to a preverbal position of arguments which are not further discussed in the study, i.e. dislocation and clefts, in Section 4.3.3. In Chapter 3 it was argued that the 'no case before the verb' feature has developed from dislocation and cleft structures. In this chapter, they will be discussed to distinguish them from basic clauses.

4.1 The language and its speakers

Datooga is a dialect cluster belonging to the Southern Nilotic branch of the Nilotic language family within the Nilo-Saharan phylum. Datooga is an umbrella term comprising several Southern Nilotic ethnic groups and their languages/dialects (Rottland 1982: 26). According to Rottland (1982) the distinct dialects are Asimjeega, Bajuuta, Barabaiga, Gisamjanga, Bianjiida, Rotigeenga and Bureadiga, whereby Barabaiga and Gisamjanga have received the most attention in research and are further in focus of this study. Both dialects are linguistically fairly similar, and due to their extensive contact in the past, it is difficult for the outsider to dis-

tinguish whether speakers belong to the Barabaiga or Gisamjanga subgroup (Mitchell & Rácz 2021: 391). In the following, I will thus use 'Datooga' to refer to both the Barabaiga and Gisamjanga dialects. As of 2008, Muzale & Rugemalira (2008: 79) estimated the number of Datooga speakers at around 160,000, which are listed under the names Datooga and Taturu. However, it is assumed that the actual number is higher than that.

When Rottland (1982) conducted his research, Gisamjanga and Barabaiga were spoken in the same area. Gisamjanga was concentrated at the border to the Barabaiga territory, namely around Dongobesh to the south of Mbulu and Haydom at the west border of the Barabaiga territory. The Barabaiga group was found in the surroundings of the Hanang' mountain, namely in and around Basoru and Basodesh to the west of the mountain and in the so-called Mangati plains to the south of the mountain (Rottland 1982: 27). However, this data is over 50 years old and although the majority of Datooga speakers still live in northern Tanzania, a rising number have moved south in pursuit of more lucrative grazing areas for cattle (Mitchell 2021: 421). The distribution of the Datooga dialects in northwest Tanzania is illustrated in Figure 4.1.

In terms of multilingualism, Mitchell (2015a: 9) notes that Iraqw and Swahili are the most common second languages of the Datooga, though the most contact happened with Iraqw people (Mitchell & Rácz 2021: 392). Proficiency in Swahili strongly depends on whether people attended school, but also on the place where they live, as Datooga living in urban areas tend to have a better knowledge of Swahili while in rural areas children that did not go to school tend to be monolingual in their early years (Mitchell 2015a: 9, Mitchell & Rácz 2021: 392). Mitchell (2022) notes that men are more likely than women to spend time in urban areas, such as in town or at the market, hence they tend to know more Swahili. According to Mitchell (2015a: 9–10), it is likely that the number of Datooga speaking Iraqw has increased with more Datooga attending school.

According to Blystad & Rekdal (2004), the Datooga people are pastoralists and place a significant cultural value on raising cattle, but this does not necessarily mean that they all have large herds. In fact, most Datooga today only have a few cattle, and some do not have any at all due to decreased herd sizes. As a result, farming has become more important for them, with a lot of Datooga households in Hanang' growing crops like maize, beans, and sorghum. Despite this shift, there are still many myths within Datooga culture that speak negatively about farming and farmers (Blystad & Rekdal 2004: 630).

Similar to other pastoralist communities in East Africa, the Datooga are often characterized as practicing polygyny, where men may have multiple wives. They also follow a patrilineal system, meaning descent and inheritance are traced through the male line. Additionally, there is a notable "warrior" tradition within Datooga culture, and historically, social and political life has been predominantly dominated by males (Blystad & Rekdal 2004: 630).

As noted by Sieff (1999: 3), the Datooga people live in homesteads known as *qeeda* in their language. Within each *qeeda*, an individual household lives independently, typically comprising a married man as head of the household, along with his wife or wives and their children. This homestead represents the primary social and economic structure for the Datooga community.

Figure 4.1: Distribution of the Datooga groups across north and central Tanzania: (1) Rotigenga (Rotigeenga) and Isimijega (Asimjeega), (2) Bajuta (Bajuuta) and Isimijega (Asimjeega), (3) Gisamjanga and Barabaiga, (4) Bianjida (Bianjiida), and (5) Buradiga (Bureadiga); Map by Monika Feinen (based on Rottland 1982).

For more detailed information on the ethnography of the Datooga, see, for example, Mulder (1992), Sieff (1999), and Blystad & Rekdal (2004).

4.2 Typological profile

This section presents an overview of the typological profile of Datooga. The information is based partly on previous research and partly on an analysis of our data and annotations. Unless stated otherwise, the information provided here is derived from our data and annotations, most of which were carried out by Alice Mitchell in the previous years.

4.2.1 Phonology

Table 4.1 shows the consonant inventory in Datooga, both with the IPA symbols but also with their orthographic representation if it deviates from the IPA signs. In the table, the glottal fricative [h] is enclosed in brackets because its phonemic status remains uncertain, as noted by Rottland (1982: 153). It can appear alongside vowels in the initial position of a word or between two vowels. However, in most cases, these contexts are realized without the presence of [h]. This suggests that [h] might be optional in vocalic onsets or sequences of vowels. Nevertheless, there are also instances of vocalic onsets or vowel sequences that do not have an alternative pronunciation with [h] (Rottland 1982: 154).

	Labial	Alveolar	Postalveolar	Palatal	Velar	Uvular	Glottal
Plosive	p b	t d		c <ch> ɟ <j>	k g	q	
Nasal	m	n		ɲ <ny>	ŋ <ng'>		
Trill		r					
Fricative	f	s	ʃ <sh>				(h)
Approximant				j <y>	w		
Lateral approximant			l				

Table 4.1: Consonant inventory of Datooga (Rottland 1982). In cells with two consonants, the left one represents the voiceless counterpart while the right one represents the voiced counterpart. The orthographic representation of the consonants is given in pointed brackets next to the IPA sign if it deviates from it. The glottal fricative [h] is enclosed in brackets because of its uncertain phonemic status.

There are several allophones associated with these phonemes, as outlined by Rottland (1982: 153). The phoneme /q/ can manifest with the following realizations: [q G χ ʁ]. This means that its realizations can vary between voiced and voiceless, as well as between plosive and fricative. Notably, the voiced uvular fricative [ʁ] appears to be the most prevalent realization of this phoneme (Rottland 1982: 153). The phoneme /q/ can also undergo alternation with /g/ in different allomorphs, such as in the affirmative prefix discussed later in this chapter, and this alternation is conditioned by the presence of specific vowels (Rottland 1982: 156). Concerning the other plosives, the phonemes /b d ɟ g/ consistently exhibit voiced realizations when they follow a nasal consonant. When directly following or preceding other consonants, they are predominantly voiceless. In different contexts, these phonemes alternate between voiced and (unaspirated) voiceless forms. In contrast, their counterparts /p t c k/ are invariably voiceless, occasionally with aspiration, and at times geminated. Therefore, only the voiced plosives have alternates, while the voiceless plosives consistently maintain their voiceless quality (Rottland 1982: 153). Creider & Rottland (1997: 81) note that /p t c k/ and /b d ɟ g/ show a distinction between fortis and lenis in that the fortis consonants, /p t c k/, can be pronounced

with aspiration in contrast to the lenis consonants /b d ɟ g/. The lenis consonants are further often realized voiceless, although they occur voiced when in initial position, between vowels or next to sonorant consonants.

There are a few morphological and phonological processes that condition alternative realizations of phonemes in Datooga. Due to certain morphological conditions, there are the following changes in the final position of stems: j → g, s → g, ʃ → l (Rottland 1982: 156). In the initial and final position of stems /b d j g/ can switch to their respective voiceless equivalents under certain circumstances. However, Rottland (1982: 156) notes that this switch cannot be explained phonetically. Additionally, according to Rottland (1982: 156), final stem consonants /d/ and /n/ are omitted when they precede [ʃ].

Datooga features a total of seven short vowels, each of which has a corresponding long counterpart, as described by Rottland (1982: 155). Vowels in Datooga are further distinguished by the feature [±Advanced Tongue Root] (ATR) following the terminology of Hall et al. (1974). According to Creider & Rottland (1997: 72), out of the seven vowels, /i/ and /u/ do not alternate in this category. In addition, /a/ only distinguishes between [+ATR] and [-ATR] when it is long. The [+ATR] alternate of /aː/ is /ɛː/ which also functions as a [-ATR] alternate of /eː/. To avoid confusion, Creider & Rottland (1997: 72) chose the symbol æː for the [+ATR] alternate of /aː/. Following Creider & Rottland (1997: 92) [ɛː] and [æː] have phonetically merged and therefore [æː] does not denote a distinct phoneme. The vowels and their [±ATR] alternates are listed in Table 4.2.

-ATR	**+ATR**
i(ː)	
u(ː)	
a	
ɛ(ː)	e(ː)
ɔ(ː)	o(ː)
aː	æː

Table 4.2: Vowel inventory of Datooga including length and ATR, adapted from Creider & Rottland (1997).

Datooga is a tone language, distinguishing between high tones (ˊ), falling tones (ˆ), and low tones (ˋ) (Kießling 2007: 153). According to Rottland (1982: 155), tone mainly has a grammatical function than a lexical one. However, Mitchell (2015a) found one example in her data where the tonal pattern changes the meaning of the word. According to Mitchell (2015a: 44) the tone pattern is the only difference between *mèandíida* 'young male donkey' and *méandìida* 'look-out hut'.

(33) Datooga (Southern Nilotic; Tanzania; Mitchell 2015a: 44–45)

a. mèandíida
'young male donkey (also 'uncooperative person')'

b. méandìida
'look-out hut'

The major function of tone, however, is grammatical, as it is used to encode case. In Datooga, the absolute case does not involve overt case marking. Rather, it represents the lexically specified underlying tone pattern of the noun and it has several functions, including marking the direct object and the citation form of a noun. In contrast, the nominative case which is used for post-verbal agents shows a tonal alternation whose realization depends on the syllable length of the noun (Kießling 2007).

In disyllabic and trisyllabic nouns, such as *déedá* 'cow' or *ɲáwúudá* 'cat', nominative case is indicated by a high tone on each syllable. With four or more syllables, such as in *qáarèemáŋgá* 'youths', the initial and final two syllables receive high tones, while the syllables between are marked with low tones (Kießling 2007: 153). Case is further discussed in Section 4.2.2.5.

In the examples provided from the corpus, tone is not glossed consistently. Due to time constraints, a phonemic analysis of the language could not be conducted and tone has not been annotated in all of the texts. Thus, not all of the examples distinguish tone. However, as discussed in this chapter, tone plays a significant role in Datooga, particularly in marking case. Therefore, it should be considered in future studies.

In terms of syllables, Datooga allows for CV and CVC structures. When the coda is not empty, it is usually occupied by sonorant consonants, such as [l] or [n], but obstruents like [t] are also possible in this position (Mitchell 2015a: 139). According to Dimmendaal (2012), the presence of obstruents in the coda is less common in Nilotic languages.

4.2.2 Morphology

4.2.2.1 Verbs

Table 4.3 shows the structure of a verb in Datooga.

1	2	3	4	5	6	7	8
Mood	Polarity	Tense	Subject	**Root**	Derivation	Tense	Object
Sequential		Aspect	Impersonal			Aspect	

Table 4.3: Structure of a verb in Datooga. The number indicates the position of the verbal affixes or clitics and the root (Rottland 1982: 173).

The first kind of prefix in Datooga verbs are mood prefixes. Datooga distinguishes five different grammatical moods: conditional, hortative, imperative, indicative, and subjunctive, as deduced from our data. The different mood formatives are exemplified in Table 4.4. As the

table shows, the indicative does not have a dedicated prefix in Datooga. The conditional is marked with the clitic *ii=*. Another grammatical mood is the hortative which is indicated by the prefix *ad-*. The subjunctive prefixes of the 1st person singular is *dá(a)-* ∼ *dɛ́ɛ-* and 2nd person singular arguments receive the prefix *á-*. The plural forms for 1st and 2nd person are *ɛ́ɛ-* ∼ *èe-* and *ɔ́ɔ-*∼*óo-*, respectively. For the 3rd person, there are two different prefixes, *q(w)a-* ∼ *g(w)a-* and *qɔɔ-* ∼ *goo-*, depending on the verb class. In comparison to the other mood prefixes, the subjunctive forms do not occur in the first slot but in the slot for the subject prefix. The structure of a verb in subjunctive mood is less complex and typically attaches directly to the stem but could be preceded by conditional and future markers.

	Form(s)
Indicative	∅
Conditional	ii=
Hortative	ad-
Subjunctive	dá(a)- ∼ dɛ́ɛ- (1SG), á- (2SG), ɛ́ɛ- ∼ èe- (1PL), ɔ́ɔ- ∼ óo- (2PL), q(w)a- ∼ g(w)a-, qɔɔ- ∼ goo- (3rd person)
Imperative	∅ a- (SG/PL) aa- (2PL)

Table 4.4: Grammatical moods and their forms in Datooga. The forms from the subjunctive are adapted from Kießling (2022).

The imperative is indicated by the prefixes *a-* for singular and plural entities and *aa-* for 2nd person plural, which only occur on verbs of class two. Verbs which do not belong to class two are marked with a zero prefix. The structure of an imperative is also less complex and could be illustrated with the following structure:

1	**2**	**3**	**4**
Imperative prefix	**Root**	Inflection (a)	Object

Table 4.5: Structure of an imperative verb in Datooga. The number indicates the position of the verbal affixes and the root.

However, not all of these slots are always filled in the construction of an imperative form. Thus, imperative verbs apply a different structure where the imperative prefix is the only prefix if it is attached.

The first slot in the (basic) verbal structure can also be filled with the sequential clitic *ag=*. If the sequential clitic occurs, it is directly preceding the polarity prefix and it is used to link clauses together.

The next position on verbs is occupied by the polarity prefixes. Datooga distinguishes between *q-* ∼ *g-* for the affirmative, and *m-* for the negative. The prefixes are listed in Table 4.6.

	Form(s)
Affirmative	q- ∼ g-
Negative	m-

Table 4.6: Polarity prefixes in Datooga.

Tense and aspect prefixes occupy the position after the polarity prefix. Table 4.7 shows the tense and aspect prefixes in Datooga. Concerning the tense, the only tenses affixed on the verb are perfect tense and future tense. The perfect prefixes are *n-*, *si-* for first person plural, *nu-* for third person, and *si-* for the impersonal perfect. Future tense is indicated with the prefix *ay-*. Past tense is indicated by a particle in the clause while non-past tense does not have a dedicated marking. However, note that there is no in-depth analysis yet of TAM in Datooga.

Tense		**Aspect**	
	Form(s)		**Form(s)**
Perfect	n-, si- (1PL.PRF), nu- (3PRF), si- (IMPRS.PRF)	**Continuative**	údú-
Future	ay-		

Table 4.7: Tense and aspect prefixes in Datooga.

Datooga distinguishes between the continuative aspect that is prefixed to the root and aspects that are suffixed to the root. The continuative aspect is indicated by the prefix *údú-* (Mitchell 2021).

The last prefix position is filled with the subject prefixes. Table 4.8 shows the subject prefixes in Datooga. In addition to the subject prefixes, this slot can also be filled with the impersonal prefixes. The impersonal forms are *ɛ́ɛ-*, as well as the previously mentioned impersonal perfective form *si-*.

Derivational suffixes in Datooga serve multiple functions, as illustrated in Table 4.9. However, not all of these suffixes can be combined equally as will be discussed later.

	Singular	**Plural**
1	ɛ́ɛ-/a-	ee-
2	ii-/i-	oo-/a-
3	(w)a-/oo-	

Table 4.8: Subject prefixes in Datooga.

The first function of derivational suffixes is to change the valency of the verb, such as with the causative suffix *-j* and with the antipassive suffix *-sh*. Another function of the derivational suffixes includes spatial orientation which is indicated with the centripetal suffixes *-u* and *-n*, as well as with the centrifugal suffix *-d*. Another type of derivation includes the suffix *-aa* for associated locomotion (Kießling & Bruckhaus 2017, Bruckhaus 2021) which is defined as "a verbal grammatical category, separate from tense, aspect, mood and direction, whose function is to associate, in different ways, different kinds of translational motion (spatial displacement/ change of location) to a (generally non-motion) verb event" (Guillaume & Koch 2021: 3). Additional suffixes concerning the lexical aspect, or Aktionsart, in Datooga include the inceptive suffix *-iid*, as well as the pluractional suffixes *-ay, -eay, -ac* and *-ii* (Kießling 1998). Other derivational suffixes include the (multipurpose) oblique *-an*. This term indicates that the suffix serves various grammatical functions in different contexts, such as instrumental, locative, and ablative, rather than having a single specific function. Since there is not a specific term for it, Kießling (2001: 136) suggests calling it '(multipurpose) oblique' to encompass its many functions. Lastly, the category of derivational suffixes further encompasses the purposive suffix *-eew* and the terminal suffix -s.

<table>
<tr><th colspan="2">Valency</th><th colspan="2">Voice</th><th colspan="2">Spatial orientation</th><th colspan="2">Locomotion</th><th colspan="2">Aktionsart</th><th colspan="2">Other</th></tr>
<tr><td rowspan="3">Causative</td><td rowspan="3">-j</td><td rowspan="3">Antipassive</td><td rowspan="3">-sh</td><td>Centripetal</td><td>-u, -n</td><td rowspan="3">Associated locomotion</td><td rowspan="3">-aa</td><td>Inceptive</td><td>-iid</td><td>Multi-purpose oblique</td><td>-an</td></tr>
<tr><td rowspan="2">Centrifugal</td><td rowspan="2">-d</td><td rowspan="2">Pluractional</td><td rowspan="2">-ay, -eay, -ac, -ii</td><td>Purposive</td><td>-eew</td></tr>
<tr><td>Terminal</td><td>-s</td></tr>
</table>

Table 4.9: Derivational suffixes in Datooga, adapted from Kießling (2001: 124). Note that not all of these suffixes can be combined equally.

Following Kießling (2001: 124), not all of the derivational suffixes can be combined with each other. This is for example the case for the antipassive that does not occur with the terminal, centripetal, or centrifugal suffixes.

Another category of suffixes that can be attached to the verbal root are interrogative suffixes. The forms for these are *-aan* and *-n*. However, these have not been discussed in the literature so far but they come up in the data.

The last category of verbal suffixes are the object markers. Table 4.10 shows the object suffixes in Datooga. Note here that object suffixes only occur for 1st and 2nd person objects (Rottland 1982: 190).

	Singular	**Plural**
1	-ean	-easa
2	-éey	-eagwa
3	∅	∅

Table 4.10: Object suffixes in Datooga.

4.2.2.2 Nouns

In Datooga, nominal stems can consist of one, two, or three syllables (Rottland 1982: 162). There are certain affixes that can be attached to nouns, as illustrated in Table 4.11.

1	**2**	**3**	**4**	**5**	**6**
Gender	**Root**	Primary suffix (PS)	Proximal DEM	Secondary Suffix (SS)	Distal DEM
Nominalization			Indefinite		Possessive
Interrogative			Numeral 'One'		Referential

Table 4.11: Structure of a noun in Datooga based on Kießling (2000a: 349) and Creider & Rottland (1997). The number indicates the position of the nominal affixes and the root.

In terms of prefixes, Rottland (1982: 160) notes that nominal stems can be affixed with four types of prefixes: gender, nominalization, interrogative prefixes, and vocative. However, the vocative prefix is only confirmed in the Bianjiida dialect (Rottland 1982: 162), which is why it will not be discussed further.

Concerning gender, Datooga differentiates between feminine and masculine forms, which are mainly used for the formation of names and are attached to nouns or full verb forms (Rottland 1982: 160; Creider & Rottland 1997: 77). The feminine gender prefix appears in two forms: *u-* and *udu-*, as in *úsûːta* 'born during brewing of beer (fem.)' (Creider & Rottland 1997: 77). Semantically, these prefix forms are linked to the singular noun *ûdaː*, denoting 'girl' (Creider & Rottland 1997: 92). The prefixes *gi-* and *gidi-* are used for masculine gender (Rottland 1982: 160; Creider & Rottland 1997: 77), as in *gísûːta* 'born during brewing of beer (masc.)' (Creider & Rottland 1997: 77). Proper nouns lacking a gender prefix are categorically considered masculine, while the feminine gender must be explicitly marked on the noun (Rottland 1982: 161). Gender prefixes are also used with animal denotations. Here, an initial *u-* can be observed with wild

animals, which is identical to the gender prefix (Creider & Rottland 1997: 77). However, in this context, it appears to denote the species rather than the individual animal (Rottland 1982: 161). Mdoe (2024) notes that the choice of gender appears to be based on semantic parameters, such as the size of the referent, rather than biological sex. For example, feminine gender is associated with small referents, while masculine gender is associated with large ones.

Regarding the nominalization prefix in Datooga, class 2 verbs can be nominalized using the *gii-* prefix Rottland 1982: 161; Creider & Rottland 1997: 77; Kießling 2000b), as in *gílíːʃda* 'permission', derived from the verb *liːb* 'to permit' (Creider & Rottland 1997: 77, 92). For verb stems from class 1, there are no corresponding prefixes for nominalization (Rottland 1982: 162).

Finally, the interrogative prefix *ana-*, meaning 'which' can be attached to noun stems. This is illustrated in the following Example (34) from the corpus:

(34)

siisu	*nii*	*andoashi*	*andoashi*
siisu	nii	**ana**-doash	**ana**-doash
person.DEM.PROX	DEM.PROX.SG	which-clan	which-clan

'What clan was this person?'

[2014-09-13_YonasHulanda1_114]

Concerning the suffixes, Datooga, just like the other Kalenjin languages, exhibits a distinction between primary and secondary forms of nouns, which are formed each with a dedicated set of suffixes. A primary form thereby consists either of the bare nominal root or a nominal root combined with primary suffixes. The secondary form is built by attaching the secondary suffix to the primary form Creider & Rottland 1997: 74; Kießling 2007: 154). Following this, the structure of a noun in Datooga can be represented as "root-primary suffix (ps)- secondary suffix (ss)" (Kießling 2000a: 349). Kießling (2000a: 350) notes that secondary forms occur more frequently than primary forms. The secondary form is further considered the citation form of nouns (Kießling 2007: 155). From a semantic perspective, primary and secondary forms differ in terms of specificity. While the secondary form refers to a specific item, the primary form is unspecific and highlights the nominal stem in its most general form (Rottland 1982: 169). In narrative texts, nouns are often employed in their primary form to establish certain protagonists at the beginning. Occasionally, they also serve as terms of address, as in *gwáːrgwèː* 'old man!'. In addition, idiomatic expressions like *diːŋd qàːŋ* 'look at' (lit. 'let the eye climb') use the primary form *qàːŋ* rather than the secondary form *qàːŋ-da* (Kießling 2000a: 350–351).

In Datooga, both the primary and the secondary suffixes provide information about number. Primary suffixes include a broad set of suffixes of which the majority has a numerical meaning as they can be associated with singular or plural meaning. However, some of these suffixes are polysemous, appearing on both individual entities and collectives, such as the suffix *-éː* which can be used both with individuals or nouns with a collective meaning. In contrast to the primary suffixes, the secondary suffixes only comprise two distinctive morphemes, i.e. *-da* used with both singulatives and collectives, and *-ga* for plurals, such as *njòːr-àjéː-ga* 'pumpkin (sp.)'. As the secondary suffix *-da* is used for both individual entities, as in *gìsíl-jáːn-da* 'sheep', and collectives, such as *búnèː-da* 'people', the notion 'singular' does not fit in in this context and it is better to

talk about it as what Hayward (1984: 160f.) terms the "unit reference form", as explained by Kießling (2000a: 349–350). This is also the reason why the language is best described as having not a binary but a ternary number distinction, encompassing singulatives (individual reference form), collectives (unit reference form), and plurals (multiple reference form) (Kießling 2000a: 350). These reference forms, i.e. individual reference, unit reference, and multiple reference, each consist of a combination of primary and secondary suffixes. Thus, while individual reference and unit reference each apply the same secondary suffix *-da*, they are usually distinct in their primary suffix. That is, the noun stem *gìsíl* 'sheep' would have the individual reference form *gìsíl-jáːn-da* while its unit reference form would be *gìsíl-∅-da*. Table 4.12 illustrates the primary and secondary suffixes for singulatives (individual reference), collectives (unit reference), and plurals (multiple reference).

	Primary suffixes	**Secondary suffixes**
Individual reference (IR)	-jaːn ∼ -aːn, -dòː, -èː, -òː, -ùː	-da
Unit reference (UR)	-ìː, -íː, -éː, -èː, -óː	-da
Multiple reference (MR)	(L_0)L-óː, (L_0)H-èː, (L_0)H-ì, (L_0)H-ìn, (L_0)L-àn, (L_0)L-àjéː, (L_0)L-óːjì, (L_0)H-gwàː!, (L_0)H-gáː!, (L_0)L- gwéː, (L_0)H-jeː!, L-jéːn, (L_0)H-áː!, (L_0)H-à!, (H0)H-yèː, (L_0)H-úː!, (L_0)H-úːd ∼ (-áːd, -óːd),(L_0)L-ú(ː)nàd, (L_0)H-úŋgwàː!, tonal conversion into +H	-ga

Table 4.12: Primary and secondary suffixes in Datooga, adapted from Kießling (2000a: 351)

Kießling (2000a) notes that tone also plays a role when it comes to number morphology of nouns. First of all, he distinguishes two lexical tone classes of nouns. Class 1 consists of nouns with a final high tone on the nominal root, e.g. *béːʃ-da* 'elephant', *ŋùʃáb-da* 'tongue'. Nouns of class 2 are indicated by a root-final low tone, e.g. *qàːŋ-da* 'eye', *ùhùː-da* 'head'. The lexical tone class of the noun further has an impact on the choice of the plural suffix form. There are three groups of plural suffixes: the first group of plural suffixes attaches to noun stems of class 1, e.g. *-à!, -gáː!, -gwàː!*, the second group is found on noun stems of class 2 nouns, e.g. *-gwéː!*, and the third group consists of suffixes that cannot be linked to the tone classes, e.g. *-éː!* (Kießling 2000a: 351–352).

Another set of suffixes in Datooga are demonstratives, which show a two-fold distinction between proximal and distal demonstratives. The proximal demonstrative suffix is directly attached to the primary form of the noun, while the distal demonstrative suffix is attached

to the secondary form (Rottland 1982: 169). Table 4.13 shows the demonstrative suffixes in Datooga.

	Singular	**Plural**
Proximal	-ni/-eani	-su
Distal	-ida/-ideani	-iga

Table 4.13: Demonstrative suffixes in Datooga. The suffix for the distal demonstrative plural was not found in the data of this study and was deduced from Kießling (2007: 185).

(35) a. *qaamadooni* *gwasin* *naah* *aba* *hiji*
qaamad-oo-**ni** g-wa-sin naa aba hiji
woman-PS-DEM.PROX.SG AFF-3SG-make what PREP here
'What is this woman doing here?'

[2017-04-19_Maandazi_455]

b. *asi* *gwasunu* *ha* *ng'asheesu* *ha* *gwasunu*
asi g-a-sunu ha ng'ash-ee-**su** ha g-a-sunu
DSC AFF-3SG-be.like.this DSC matter-PS-DEM.PROX.PL DSC AFF-3SG-be.like.this
'These matters are like this, they're like this'

[2016-05-10_Udahadaabu_story_32]

c. *qasa* *heaydida,* *heayda* *madeeda* *duga*
qas-a heay-da-**ida** heay-da madeeda du-ga
hang-IS bull-UR-DEM.DIST.SG bull-UR in.front.of cow-PL
'Catch that bull by the leg, the biggest bull of the cattle.'

[Samoota_073]

d. *qwáŋìɪt* *éenìgìgà*
q-a-ŋɪɪt een-ig-**iga**
AFF-3SG-be.full river.NOM-PL-DEM.DIST.PL
'Those rivers are full.'

(Kießling 2007: 185)

In Datooga, to indicate the indefiniteness of nouns, the suffix *-i* is added to the primary form of the nouns. In the singular, this suffixation results in the strengthening of the final consonant from lenis to fortis (Creider & Rottland 1997: 81).

The possessive suffixes in Datooga are another affix category that is attached to the secondary form of the noun. These possessive suffixes serve to distinguish both the number of the possessee and the number of the possessor, resulting in twelve possible suffixes (Rottland 1982: 170). The possessive suffixes are listed in Table 4.14.

Example (36) shows a possessive construction. As can be seen, the form of the third person plural possessive *-schi* is attached to *iidiga*, the secondary form of *iid* 'ear'.

(36) *qwàlàláktá* *dígéedá* *íídígàaschi*
g-a-lalag-d-a dig-ee-da iid-i-ga-schi
AFF-3SG-flap-CF-IS donkey.ABS-PS-UR ear-PS-PL-3SG.POSS
'So donkey flapped his ear about.'

	Singular	Plural
1SG	-nyu	-su
2SG	-ng'u	-gu
3SG	-nyi	-schi
1PL	-nya	-sa
2PL	-ng'wa	-gwa
3PL	-nyaawa	-nyaawa

Table 4.14: Possessive suffixes in Datooga.

[2013-06-06_HM_digeeda_12]

4.2.2.3 Pronouns

According to Kießling (2007: 160), core arguments of 1st, 2nd or 3rd person are usually only indexed on the verb with pronominal affixes. However, in certain contexts, independent pronouns also occur. The following Table 4.15 shows the personal pronouns in Datooga. As can be deduced from this table, pronouns occur in short and long forms, though their contexts remain unclear.

	Singular	Plural
1	ánìiní/ání	éesèesá/èesà
2	áŋìiŋí/áŋíi	óogòogá/òogà
3	nìŋ	sàawà

Table 4.15: Personal pronouns in Datooga adapted from (Kießling 2007: 160).

Following Kießling (2007: 160), independent pronouns of 1st and 2nd person only occur preverbally due to topicalization whereas 3rd person pronouns can also be placed after the verb. However, as the later chapters will show this is not always true for the data observed in this book. Although independent 1st person and 2nd person pronouns are rare and mostly occur in preverbal position, as in the Examples (37a), (37b), (37d), and (37e), there are also examples of postverbal independent 1st and 2nd person pronouns in subject function which will be discussed in a dedicated section on animacy in the study of the next chapter.

The following examples show clauses that employ independent personal pronouns in the data. The 1st person pronoun singular *ánìini* in Example (37a) and the third person pronoun *ninyi* in Example (37c) are in S role. In Example (37b) and Example (37d) the 2nd person singluar pronoun *áng'ìing'ì*, as well as the 1st person plural pronoun *ásêasà* function as O. A pronoun in A role is illustrated in Example (37e) with *ágòogà*. The pronoun *sàawa* in Example

(37f) is the only occurrence of a third person plural independent pronoun in the data, though it does not appear in a full clause.

(37) a. *gwayeesha* *há* *báabà* ***ániini*** *nɛ́ɛhidu*
g-a-yee-sh-a ha baba **aniini** n-aa-hid-u
AFF-3-say-AP-IS DSC father 1SG PRF-1SG-come-CP
'She said, father, I've come.'

[2014-11-14_YG_story1_097]

b. ***áng'ìing'ì*** *gajeabarnyi*
ang'iing'i g-ay-ee-bar-n-eey
2SG AFF-FUT-IMPRS-kill-CP-2SG.OBJ
'You will be killed.'

[2014-11-14_YG_story1_026]

c. *nea* *gafwi* *geaw* ***ninyi*** *aba* *jeeda*
nea g-a-fuw-i geaw **ninyi** aba jeeda
CONJ AFF-3-hide-IS REFL.SG 3SG PREP PREP
'And he was hiding inside.'

[2014-09-13_YonasHulanda1_339]

d. *diida* ***ásêasà*** *geeyaadaas*
diida **aseesa** g-ee-aad-aas
DEM.DIST.SG 1PL AFF-IMPRS-take-1PL.OBJ
ásêasà *nea* *geewooschi* *haad*
aseesa nea g-ee-wos-ch-i haad
1PL CONJ AFF-1PL-be-AP-IS how.many
'That day they sent us, we were this many'

[2014-09-13_YonasHulanda1_203]

e. ***ágòogà*** *gajea* *ayeateana* *ági*
agooga g-ajea a-aad-aan-a agi
2PL AFF-FUT 2PL-bring.CF-AM-IS one
'You will all get one first'

[2017-04-19_Maandazi_233]

f. *neada* ***sàawa***
neada saawa
epist 3PL
'Didn't they...'

[2014-09-13_YonasHulanda1_035]

4.2.2.4 Demonstrative pronouns

As already mentioned in the context of demonstrative suffixes in the previous subsection on nouns, Datooga makes a two-way distinction between demonstratives, i.e. it differentiates between proximal and distal demonstratives. The following Table 4.16 shows the demonstrative pronouns in Datooga.

	Singular	**Plural**
Proximal	níi	súu
Distal	díitta	giika

Table 4.16: Demonstrative pronouns in Datooga.

The demonstrative pronouns in Datooga are illustrated in the following examples in (38). However, there is no example for the plural distal demonstrative *giika* in the data.

(38) a. *aha qwaameaschi qaqanyi há daduulu* ***níi***
aha q-waa-meas-ji q-a-qany-i ha da-duul-u **nii**
DSC AFF-2PL-wait-IS AFF-3-bite-IS DSC 1SG.SBJV-finish-CP DEM.PROX.SG
'Just wait until I've finished this one.' [Samoota_095]

b. ***súu*** *mii* *gujakka*
suu mii gujaw-ga
DEM.PROX.PL NEG.COP horn-PL
'These aren't horns.' [2013-06-06_HM_digeeda_07]

c. *geebiikta* *eeqas* *há* *heayda* *deena* ***díitta***
g-ee-biig-d-a ee-qas ha heay-da deen-a **diida**
AFF-IMPRS-return-CF-IS IMPRS-hang DSC bull-UR be.equal.to-IS DEM.DIST.PL
'They tied a rope around another bull the same size as that one.' [Samoota_080]

As the examples indicate, independent demonstrative pronouns can occur after the verb but also precede the verb which will be further discussed in Chapter 5. Demonstrative pronouns can also precede copulas as the plural proximal demonstrative *súu* in Example (38b) shows.

4.2.2.5 Case marking

Datooga utilizes a split marked nominative case system, wherein both intransitive (S) and transitive subjects (A) are marked with the same specific case marker, known as the nominative, while the object (O) remains unmarked, retaining its citation form, referred to as the absolute case by Kießling (2007: 152). Marked nominative languages are prevalent within the Nilotic language group, particularly among the Southern Nilotic languages, as detailed in Chapter 3. Similar to numerous other Southern Nilotic languages featuring marked nominative systems, Datooga employs tonal inflection to signify case. Nevertheless, akin to other languages in East Africa discussed in the preceding chapter, subject case marking is only observable with postverbal subjects, whereas preverbal subjects occur in their citation form, i.e. the absolute case.

The following Table 4.17 shows the absolute and nominative tone patterns in Datooga nouns. As can be deduced from the table, the nominative is indicated by high tones throughout in nouns with two or three syllables, such as in *déedá* 'cow'. With four or more syllables, the high tones are placed on the initial and two final syllables with low tones between them, as in *hálàqwèenéedá* 'thorn' (Kießling 2007: 153).

Translation	**Absolute**	**Nominative**
'cow'	déedà (HL)	déedá (HH)
'dog'	gùdéedà (LHL)	gúdéedá (HHH)
'youths'	qàarèemáŋgà (LLHL)	qáarèemáŋgá (HLHH)
'thorn'	hálàqwéenèedà (HLHLL)	hálàqwèenéedá (HLLHH)

Table 4.17: Absolute and nominative tone patterns in disyllabic, trisyllabic, quadrisyllabic, and quinquesyllabic nouns, adapted from Kießling (2007: 152–154).

In contrast to the nominative, the absolute case in Datooga lacks predictability. Instead, Datooga nouns are categorized into distinct lexical tone classes, as outlined by Kießling (2007: 153–155). The primary form, i.e. the nominal stem without secondary suffixes, serves as the domain for lexical tone assignment. One-syllable nouns fall into two distinct tone classes. The first group comprises nouns characterized by a high tone in the stem, exemplified by the previously mentioned *dée-dà* 'cow'. The second group includes nouns with a low tone, such as *gèed-dà* tree'. For disyllabic nouns, four tone classes are identified: HH, LH, HL, and LL (Kießling 2007: 155).

In addition to nouns, there are other word classes in Datooga that exhibit distinctions between nominative and absolute case. These include personal pronouns, personal names, adjectives, numerals and quantifiers, interrogative pronouns, and anaphoric bases (Kießling 2007: 160). In the following, I will provide an overview of the tone patterns of these word classes in their independent use. For a detailed description of their tone behavior when they function as modifiers in complex noun phrases, see Kießling (2007: 175).

Table 4.18 illustrates the tone patterns for personal pronouns in both absolute and nominative cases. As evident from the table, only pronouns in the 3rd person singular and plural exhibit a distinction between absolute and nominative case. This discrepancy arises from their placement within the clause when they appear independently. As noted by Kießling (2007: 160), pronouns in the 1st and 2nd person can be independently used for "empathy or pragmatic prominence, e.g., for topicalisation". Kießling (2007) attested these pronouns only in preverbal positions, which are limited to the absolute case. However, as will be further discussed in Chapter 5 , 1st and 2nd person pronouns can also occur postverbally, though they are rare in this position. In contrast, 3rd person pronouns are permitted in postverbal positions where they take the nominative case.

In terms of adjectives, two types are observed: indefinite (other/s) and definite (the other/s), as illustrated in Table 4.19. As can be deduced from the table, adjectives distinguish between

	Absolute	Nominative
1SG	áníìíní, ání (H(L)H)	
2SG	áŋìiŋí, áŋíı (H(L)H)	
3SG	nìɲ (L)	níɲ (H)
1PL	éesèesá, èesà (HLH/LL)	
2PL	óogòogá, òogà (HLH/LL)	
3PL	sàawà (LL)	sáawá (HH)

Table 4.18: Absolute and nominative tone patterns in personal pronouns, adapted from Kießling (2007: 160).

nominative and a prepausal nominative, which occurs in utterance-final position. However, not all forms are attested and thus marked with '?' (Kießling 2007: 161–162).

Translation	Absolute	Nominative	Prepausal nominative
other (INDEF.SG)	nùut, nʊ̀t (L)	núut, nʊ́ʊt (H)	?
others (INDEF.PL)	sùuk, sʊ̀ʊki (L)	súuk, sʊ́ʊk (H)	?
the other (DEF.SG)	níirjá (HH)	nìirjá (LH)	nìirjà (LL)
the others (DEF.PL)	súurjá (HH)	sùurjá (LH)	sùurjà (LL)

Table 4.19: Absolute and nominative tone patterns in pronominal adjectives, adapted from Kießling (2007: 162). The prepausal nominative is a variant of the nominative that occurs in utterance final position (Kießling 2007: 153).

Table 4.20 displays the tone patterns for the numerals *àgì* 'one' and *dùuwáawì* 'both, two', as well as for the quantifier *sèen* 'all', in both absolute and nominative cases. As was the case with the adjectives, not all forms are attested with numerals and quantifiers (Kießling 2007: 163).

Table 4.21 shows the interrogative pronouns *ŋáa[há]* 'who' and *náahá, -nàa, náahà, -náa* 'what'. Both interrogative pronouns are characterized by an HH pattern in the prepausal nominative, i.e. in utterance final position, while the nominative pattern for the non-final position is not attested (Kießling 2007: 165).

The anaphoric pronouns *daa* (singular) and *gaa* (plural) function as replacements for a noun in sentences with additional modifiers such as possessive pronouns or demonstratives. The specific details of their case marking are depicted in Table 4.22, following Kießling (2007: 165).

While the nominative case is only utilized for marking postverbal subjects (S and A), the absolute case encompasses various functions. First of all, it functions as the citation form of nouns. Moreover, it extends to encompass direct and indirect objects (O), along with subjects positioned preverbally. Additionally, it acts as an indicator for topics and complements in nom-

Translation	Absolute	Nominative	Prepausal nominative
'one'	àgì (LL)	ágí (HH)	?
'both, two'	dùuwáawì (LHL)	dúuwàawí (HLH)	dúuwàawì (HLL)
'all'	sèen (L)	(sêen) (F)	sêen (F)

Table 4.20: Absolute and nominative tone patterns in numerals and quantifiers, adapted from Kießling (2007: 163). The prepausal nominative is a variant of the nominative that occurs in utterance final position (Kießling 2007: 153). The non-final nominative form for the quantifier *sèen* 'all' is just assumed by Kießling (2007) and therefore enclosed in brackets.

Translation	Absolute	Nominative	Prepausal nominative
who	ŋáa[há]	?	ŋàahà
what	náahá, -nàa, náahà, -náa	?	nàahà

Table 4.21: Absolute and nominative tone patterns on interrogative pronouns, adapted from Kießling (2007: 165). The prepausal nominative is a variant of the nominative that occurs in utterance final position (Kießling 2007: 153).

	Absolute	Prepausal absolute	Nominative
SG	dáa	dàa	dàa´
PL	gáa	gàa	gàa´

Table 4.22: Absolute and nominative tone patterns on anaphoric pronouns, adapted from Kießling (2007: 166).

inal predication, as well as prepositional complements. Furthermore, the absolute case identifies possessors within noun phrases and identifies non-initial items within a series of conjoined nouns in S, A, and O arguments (Kießling 2007: 167).

4.2.3 Research on other dialects

Studies on Datooga have mostly been carried out on the Gisamjanga and Barabaiga dialects, such as the research carried out by Kießling (e.g. Kießling 2000a; Kießling 2000b; Kießling 2001; Kießling 2007) and Bruckhaus (Bruckhaus 2015; Kießling & Bruckhaus 2017), and Mitchell (e.g. Mitchell 2015a; Mitchell 2015b; Mitchell 2017; Mitchell 2021; Mitchell & Rácz 2021) on Barabaiga and Gisamjanga.

There is only little linguistic research in Bajuuta, Asimjeega, and Bureadiga. Hieda (2000) seems to be the only linguistic study of the Bajuuta dialect. In this study, Hieda sheds light on the consonantal system in the Bajuuta variety of Datooga. As for the Asimjeega variety, Griscom (2019) is the only linguistic study so far. In this study, Griscom (2019) also discusses constituent order in Asimjeega. In his study on Datooga verbal inflection, Kießling (2022) also analyzes data from the Bureadiga dialect together with data from the Gisamjanga variety.

In addition to linguistic studies, there are further anthropological studies which do not further specify the Datooga group under study, such as Mulder (1992), Sieff (1995), Sellen (2003), and Pike & Patil (2006). However, the focus in these studies is on Datooga groups around Lake Eyasi which is where the Asimjeega and Bajuuta are located according to Figure 4.1.

As of today, there seems to be no research on the Rotigeenga, and Bianjiida.

4.3 Introduction to syntax

The following section serves as an introduction to syntax in Datooga. It consists of a section on three types of subordinate clauses, i.e. complement clauses, adverbial clauses, and relative clauses, (Section 4.3.1), followed by a discussion of elements that can occur preverbally (Section 4.3.2), and finally dislocation and cleft structures are illustrated (Section 4.3.3).

4.3.1 Constituent order in subordinate clauses

Although the main focus of the study of Datooga constituent order concentrates on declarative main clauses, subordinate clauses also deserve a discussion on their own. The following section serves as an overview of constituent order in three types of subordinate clauses, namely in complement clauses, adverbial clauses, as well as relative clauses. Each subsection discusses a dedicated clause type in terms of its form, as well as on possible constituent order variation. Table 4.23 provides an overview of the frequencies of declarative main and subordinate clauses in the analyzed data. Impersonal constructions were excluded from the analysis as they do not have a referent in S or A function.

	Number	**%**
Declarative main clauses	748	77.19
Complement clauses	66	6.81
Adverbial clauses	86	8.88
Relative clauses	69	7.12
Total	**969**	**100**

Table 4.23: Frequency of declarative main and subordinate clauses, excluding impersonal constructions.

Subordinate clauses are distinguished from main clauses in that they are either introduced with specific conjunctions, which is especially the case for adverbial clauses, they follow specific verbs, or they show specific verbal marking, such as clitics or the subjunctive prefix. However, in the majority of instances, subordinate clauses do not show reduced verb forms that contrast with verbs in main clauses. Thus, the distinction is mainly based on the features named above which will be discussed in more detail in the following.

4.3.1.1 Complement clauses

In Datooga, complement clauses can be introduced with the complementizer *gideaba* when they follow verbs of thinking and speaking, such as in Example (39).

(39) *googuul* *há* *Ginaanaan*
g-a-guul ha Ginaanaan
AFF-3SG-scream.with.pain DSC Ginaanaan
gídêabá niisha ***buuneedeanyu***
gideaba n-i-iish-a **buun-ee-da-nyu**
CONJ PRF-3SG-finish-IS folk-PS-UR-1SG.POSS
'Ginaanaana cried out that "my people are finished"'

[Samoota_039–040]

Complement clauses are further used in the context of the verbs *nung'* 'cause, allow, let', as in Example 40a, and *gasay* 'want', as in Example 40b. In contrast to the previous construction, *gideaba* is not applied in these contexts.

(40) a. *qwanung'* *gwahitu* *muusta*
g-a-nung' g-a-hid-u muus-da
AFF-3-let AFF-3-come-CP spirit-UR
'He made the spirits come.'

[2014-11-14_YG_story1_067]

b. *geagasay* *qwayii* *gil*
g-ɛ́ɛ-gasay q-wa-ii gil
AFF-1SG-want AFF-3-hear thing
'I want this one to hear.'

[2017-04-19_Maandazi_124]

Table 4.24 shows the distribution of complement clauses between narratives and conversations and their different constituent orders, as well as their total numbers. For an overview of the size of the analyzed texts in this study, see Section 5.1.2. Similar to main clauses, which will be discussed in Section 5.2, all three possible constituent orders occur in intransitive clauses although they differ in their frequency. For transitive clauses, all orders attested for main clauses are also found in complement clauses with the exception of VA. Concerning the frequency, V

in intransitive and transitive clauses, as well as VO orders are the most common orders with complement clauses just like they are in the main clauses.

	Total		**Narratives**		**Conversations**	
	N	**%**	**N**	**%**	**N**	**%**
Intransitive						
SV	1	2.50	1	8.33	0	0
V	31	79.49	7	58.33	24	89.89
VS	7	17.50	4	33.33	3	10.71
Total	**39**	**≈ 100**	**12**	**≈ 100**	**27**	**≈ 100**
Transitive						
AVO	1	3.70	0	0	1	5.26
OV	1	3.70	0	0	1	5.26
V	6	22.22	1	12.5	5	26.32
VAO	1	3.70	0	0	1	5.26
VO	17	62.96	7	87.5	10	52.63
VOA	1	3.70	0	0	1	5.26
Total	**27**	**≈ 100**	**8**	**≈ 100**	**19**	**≈ 100**

Table 4.24: Distribution of constituent orders within complement clauses

These numbers show that complement clauses are not restricted to a certain constituent order and rather allow for some flexibility. Therefore, they will be included in the analysis in the next chapter.

4.3.1.2 Adverbial clauses

There are different types of adverbial clauses in Datooga. Adverbial clauses of reason are introduced with *(aba) gídêabá* 'because', as the following Examples (41a) and (41b) show. As mentioned in the previous section, *gídêabá* also occurs in the context of complement clauses. However, in contrast to the use of the conjunction in complement clauses, *gídêabá* in adverbial clauses does not follow verbs of thinking or speaking.

(41) a. *gídêabá gwajii* *geamadu*
gideaba g-wa-jii geamadu
CONJ AFF-3-give.birth mother.POSS.2SG
'[...] because he begot your mother.'

[Samoota_009]

b. *aba gídêabá niyii iita eemeedanyaawa*
aba gideaba ni-ii iid-da eem-ee-da-nyaawa
PREP CONJ 3.PRF-hear news-UR ethnic.group-PS-UR-3PL.POSS.SG
'[...] because he listened to the stories of their people.'

[2014-11-14_YG_story1_169]

As for the temporal adverbial clauses, there are different conjunctions to introduce these clauses. First of all, *ama (hit)* 'afterwards, then' occurs in this context, as in Example (42). Although this construction is a clause in itself, it serves as an adverbial modifier with the meaning of 'afterwards, then'.

(42) *gawa ama hita gwayeesha oori*
g-a-wa ama hid-d-a g-wa-yee-sh-a oori
AFF-3-go DSC come-CF-is AFF-3-say-AP-IS boy.VOC
'He went and when he got there he said, "boy"'

[Samoota_024]

Another conjunction that occurs in temporal adverbial clauses is *ea* which is found introducing an adverbial clause with the meaning of 'until', as in Example (43).

(43) *ea qwahita qaamata*
ea q-wa-hit-a qaamad-da
CONJ AFF-3-come-IS mother-UR
'[...] until his mother arrives.'

[2014-09-13_YonasHulanda1_188]

The conjunction *nea* either occurs with the meaning of 'when', as in Example (44a), or 'while', as in Example (44b).

(44) a. *muchu nea qwahidu gwayeesha gatmoadaanyi woo*
much-u nea q-wa-hid-u g-wa-yee-sh-a gadeem-oa-daa-nyi woo
dawn-CP CONJ AFF-3-come-CP AFF-3-say-AP-IS wife-PS-UR-3SG.POSS.SG INTERJ
'One day when he arrived, his wife said "woo"'

[2014-11-14_YG_story1_123]

b. *nea gweadu daraneeda*
nea g-w-aad-u daran-ee-da
CONJ AFF-3-carry-CP beam-PS-UR
'While he was carrying the beam.'

[2016-05-10_Udahadaabu_story_05]

Other conjunctions with the meaning of 'when' are *aba*, as in Example (45a), and the clitic *ii=*, such as in Example (45b). The latter is typically associated with conditional clauses which are discussed below.

(45) a. *aba* *weeshiidi*
aba eesh-iid-i
PREP white-INCP-IS
'When it got light.'

[2014-11-14_YG_story1_071]

b. *nea* *gweargweeda* *iihita* *qwa* *gwarigiischi*
nea gweargw-ee-da ii=hit-a qwa g-wa-rigiis-ji
CONJ elder.male-PS-UR COND=come-IS house AFF-3-brew.beer-IS
'And the elder, when he got home, began brewing honey beer.'

[Samoota_108]

Conditional clauses are marked with the conditional clitic *ii=* on the verb, as in the following clauses in Example (46).

(46) a. *gatmooda* *iijii*
gadeem-oo-da ii=jii
woman-PS-UR COND=give.birth
'If a wife gave birth'

[2014-09-13_YonasHulanda1_169]

b. *asa* *iiyeadu* *siiyeeda* *huda* *há*
asa ii=ead-u siiy-ee-da hu-da ha
? COND=bring-CP husband-PS-UR daughter-UR DSC
'If it was brought by the daughter's husband.'

[2014-11-14_YG_honeybeer_016]

Purposive clauses are indicated by a subjunctive marking on a verb that follows a main verb, as in Example (47). However, dedicated subjunctive marking only occurs with first and second person. In third person, the subjunctive takes the same form as the indicative.

(47) a. *gwayeesha* *habiiyeeda* *héabì* *lalakta* *dadaa*
g-wa-yee-sh-a habiiy-ee-da heabi lalag-d-a da-daa
AFF-3-say-AP-IS hyena-PS-UR DSC FLAP-CF-IS 1SG.SBJV-see
'Hyena said, well, flap them so that I can see.'

[2013-06-06_HM_digeeda_11]

b. *gooneani* *soda* *Dootoosha* *dashuup*
goon-ean-i soda Dootoosha da-shuub
give-1SG.OBJ-IS soda Dootoosha 1SG.SBJV-take.sip
'Give me the soda Dootoosha so I can have a sip.'

[2017-04-19_Maandazi_261]

Manner clauses are introduced with *gawurji*, which is illustrated in Example (48).

(48) *qayi garay diiyanga gabalooli gawurji buuneeda*
qayi garay diiy-an-ga g-a-balool-i g-a-urj-i buun-ee-da
PST PST animal-PS-PL AFF-3-talk-IS AFF-3-be.like-IS folk-PS-UR
'Once upon a time, animals talked like people.'

[2013-06-06_HM_digeeda_01]

Table 4.25 shows the distribution of constituent orders within adverbial clauses in narratives and conversations. In comparison to the complement clauses, adverbial clauses seem to be more evenly distributed across the two genres in the data for both intransitive and transitive constructions. However, just like is the case for complement clauses and main clauses, as discussed in Section 5.2, V (intransitive) and VO are the most common orders. For transitive clauses, VA order is as frequent as V (transitive) order though the numbers are relatively low. Interestingly, there are only a few occurrences of preverbal arguments within adverbial clauses in the data, namely only with SV (6.78% in total). In contrast to complement clauses and main clauses, preverbal A arguments do not occur with adverbial clauses in the data. However, although infrequent, preverbal order in adverbial clauses is possible. Therefore, they will be included in the in-depth study of constituent order variation in the next chapter.

	Total		**Narratives**		**Conversations**	
	N	**%**	**N**	**%**	**N**	**%**
Intransitive						
SV	4	6.78	2	6.45	2	7.14
V	43	74.14	23	76.67	20	71.43
VS	11	18.64	5	16.13	6	21.43
Total	**58**	**≈ 100**	**30**	**≈ 100**	**28**	**≈ 100**
Transitive						
V	5	17.86	3	18.75	2	16.67
VA	5	17.86	3	18.75	2	16.67
VAO	2	7.14	0	0	2	16.67
VO	16	57.14	10	62.5	6	50
Total	**28**	**≈ 100**	**16**	**≈ 100**	**12**	**≈ 100**

Table 4.25: Distribution of constituent orders within adverbial clauses

4.3.1.3 Relative clauses

Relative clauses in Datooga are indicated by a subjunctive marking on the verb immediately following the nominal they modify. For first and second person, as in Example (49a), there is a

dedicated subjunctive marking for all verb classes. For the third person, however, verbs of class one do not have a dedicated subjunctive prefix which is illustrated in Example (49b) while verbs of class two take the prefix *a-* in relative clauses.

(49) a. *goonean* *deeda* *jea* *daqaw*
goon-ean dee-da jea da-qaw
give-1SG.OBJ cow-UR REL.FUT 1SG.SBJV-milk
'Give me a milking cow.'

[Samoota_016]

b. *qeamata* *jiilnean*
qeamad-da jiil-n-ean
mother-UR give.birth-CP-1SG.OBJ
'The mother who gave birth to me.'

[2014-09-13_YonasHulanda1_326]

The distribution of constituent orders in relative clauses is shown in Table 4.26. However, the clause structure of relative clauses is different to those of complement clauses and adverbial clauses discussed above. Relative clauses in Datooga never have all arguments fully realized as they are used to modify a nominal in the main clause which would technically be the argument within the relative clauses. As these are not repeated in the relative clauses, this means that S in intransitive clauses never occurs, leaving V to be the only possible option. For transitive clauses the options are limited to VA when the O argument in the main clause is modified, or VO when the relative clause modifies A in the main clause. Therefore, this table rather shows the distribution of relative clauses modifying A compared to those modifying O, as well as the frequency of intransitive relative clauses. The only conclusion in terms of constituent order that can be drawn from this table is that relative clauses are always verb-initial in the analyzed data. As there is always one argument of the verb that obligatorily cannot be expressed in relative clauses, they will be excluded from the further analysis.

4.3.2 Preverbal position of non-arguments

As already has been mentioned before, Datooga is a verb-initial language. However, as I will discuss in the following sections of this book, it is also a language that allows for constituent order variation and for other elements than verbs to occupy the clause-initial position, i.e. the position before the verb. The following section of this book will delve into this in terms of core arguments of the verb in both intransitive and transitive clauses, namely S, A and O. Before I discuss this kind of variation, I will briefly discuss which elements apart from these core arguments can occur before the verb.

First of all, there are elements apart from noun phrases which can occur preverbally. These include PST- and REC.PST-markers which occurred in the subordinate clauses discussed earlier, such as in Example (48). The preverbal position can further be occupied by adverbials, discourse particles, as well as vocatives.

	Total		Narratives		Conversations	
	N	**%**	**N**	**%**	**N**	**%**
Intransitive						
V	45	100	18	100	27	100
Total	**45**	**100**	**18**	**100**	**27**	**100**
Transitive						
VA	6	25	2	14.29	4	40
VO	18	75	12	85.71	6	60
Total	**24**	**≈ 100**	**14**	**≈ 100**	**10**	**≈ 100**

Table 4.26: Distribution of constituent orders within relative clauses

Noun phrases in other functions than core-argument function, such as locatives, goals or instrumentals, as well as prepositional phrases, usually follow the verb. The clauses in (50) illustrate this. In the first Example (50a) the noun phrase *qeeda giischeawda* 'neighbour's house' has the semantic function of a goal and it is following the verb *gawa* 'go'. The other Example (50b) illustrates this with the noun phrase *ng'uta* 'spear' which serves as an instrument.

(50) a. *gawa* *siida* *akkawa* *qeeda* *giischeawda*
g-a-wa sii-da ag=g-a-wa qee-da giischeaw-da
AFF-3-go person-UR SEQ-AFF-3-go house-UR neighbour-UR
'A man left [his house] and went to his neighbour's house.'
[2016-05-10_Udahadaabu_story_01]

b. *aki* *gaydang'uda* *ng'eanyi* *ng'uta*
ag= g-ay-daa-ng'ud-a ng'eany-i ng'ud-da
SEQ AFF-FUT-1SG.SBJV-stab-IS ground-PS spear-UR
'And I'll stab the ground with the spear.'
[2014-11-14_YG_story1_036]

However, there are a few instances where noun phrases in functions other than core-argument functions, as well as prepositional phrases precede the verb, as in the following clauses in Example (51). Example (51a) has the prepositional phrase *aba dooshta hooweanji* 'over there by the gate' preceding the verb *gajeebiiw* 'feed child', while in Example (51b) it is the locative noun phrase *qeeda Luqmadi* 'house of the Luqmajeega' that precedes the verb.

(51) a. *aba* *dooshta* *hooweanji* *gajeebiiw* *gabeachi*
aba doosh-da hooweanji g-ay-ee-biiw gabeachi
PREP gate-UR ADV AFF-FUT-IMPRS-feed.child ADV
'Over there by the gate they'll just feed the child some milk.'
[2016-05-10_Udahadaabu_story_01]

b. *hóowà bea qeeda Luqmadi madeepa gwanda*
hoowa bea qee-da Luqmadi m-a-deeb-a g-wa-nda
ADV ASSOC house-UR Luqmadi NEG-3-be.possible-IS AFF-3-be
'Over there at the house of the luqmajeega perhaps there's [...].'
[2014-09-13_YonasHulanda1_007]

Both of these examples are different from dislocation structures which will be discussed in the following section. In clauses with left dislocation, there is typically an audible pause between the dislocated phrase and the clause. There are, however, also dislocation structures without a pause but those have an anaphoric pronoun in the main clause. As both of these structures are not found in the examples above, they are not considered dislocation structures.

These recent instances suggest that it is also possible for non-core arguments to appear before the verb, although such instances are infrequent. In the examined data, preverbal locative noun phrases or prepositional phrases occurred only three times. Nonetheless, what prompts the fronting of these phrases is intriguing. For instance, in Example (51b), it could be argued that it is a false start as the speaker does not complete the sentence. However, this does not seem to be the case for the other sentence in Example (51a).

These preverbal orders likely pertain to information structural categories, which will be explored concerning the preverbal position of core arguments in declarative main clauses. A comprehensive discussion of the preverbal position of non-core arguments is beyond the scope of the current book but opens avenues for future research.

4.3.3 Dislocation and cleft structures

When it comes to the constituent order variation of core arguments, other constructions can display a change of the dominant order, such as dislocations and clefts. As discussed in Section 3.1.2, it is assumed that the origin for the lack of case marking of core arguments in preverbal position in certain languages of East Africa lies in dislocation and cleft constructions. In Datooga, these constructions can still be distinguished from other deviations from constituent order due to their form. Therefore, these constructions will briefly be discussed in this section but they will be excluded from the further analysis.

In terms of dislocation, Lambrecht (2001b: 1050) notes that "a dislocation construction (also called detachment construction) is a sentence structure in which a referential constituent which could function as an argument or adjunct within a predicate-argument structure occurs instead outside the boundaries of the clause containing the predicate, either to its left (left-dislocation, henceforth LD) or to its right (right-dislocation, henceforth RD)". Datooga distinguishes both types of dislocation.

Figure 4.2 and Example (52) illustrate a right dislocation in Datooga. As can be seen, the NP *qeamata qee siida* 'woman (lit. 'woman of person's house')' follows the clause *goong'oonyi ninyi* 'she held it' after a pause and although a 3rd person reference is already indexed on the verb with the subject prefix *oo-*. Further, there is an overt co-referential pronoun *ninyi*.

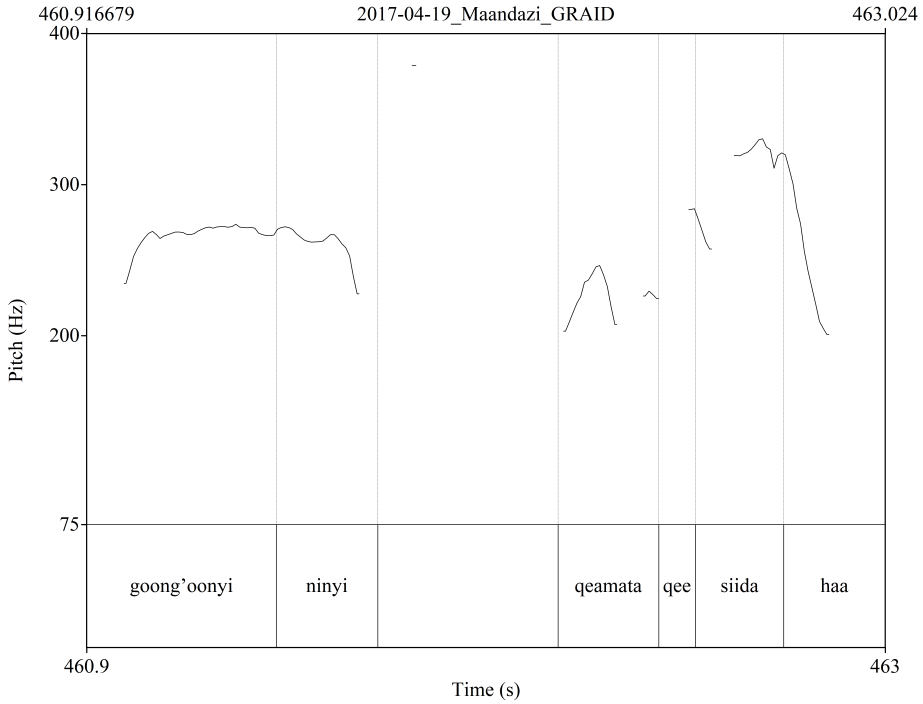

Figure 4.2: Example of a right dislocation.

(52) *goong'oonyi ninyi (.) qeamata qee siida há*
g-oo-ng'oony-i ninyi (.) qeamad-da qee sii-da haa
AFF-3-hold-IS 3SG mother-UR house person-UR DSC
'She held it, that woman.'

[2017-04-19_Maandazi_067]

Concerning left dislocation, the following clauses demonstrate two types of left dislocations in Datooga. First of all, left dislocation can behave similar to the previous example of right dislocation in that it includes a perceptible pause and a constituent that is co-referential with another element in the following clause, as in Figure 4.3 and Example (53). As the figure shows, the pause after the dislocated constituent *Qeambeameeda* is clearly visible. It can also be seen that, in contrast to the example of right dislocation earlier, there is no nominal co-referential constituent in the clause as Datooga also allows for zero pronouns.

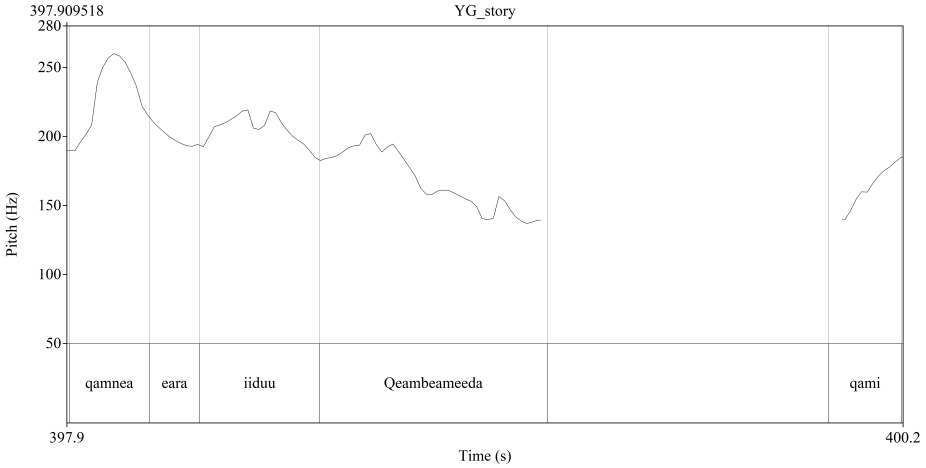

Figure 4.3: Left dislocation with a pause and no co-referential element.

(53) *qamneaw eara iiduu Qeambeameeda (.) qami*
qamneaw eara iiduu Qeambeameeda (.) q-a-mi
now even ADV Qeambeameeda AFF-3-die
'Even now Qeambeemeeda, he died [...].'

[2014-11-14_YG_story1_167–168]

In Figure 4.4 and Example (54) the NP *muhooga* 'calves' is dislocated from the following clause. However, compared to the previous examples, there is no audible pause between the dislocated NP and the clause. Instead, there is the anaphorical pronoun *gea* referencing the calves, which illustrates the second method of forming left dislocations in Datooga.

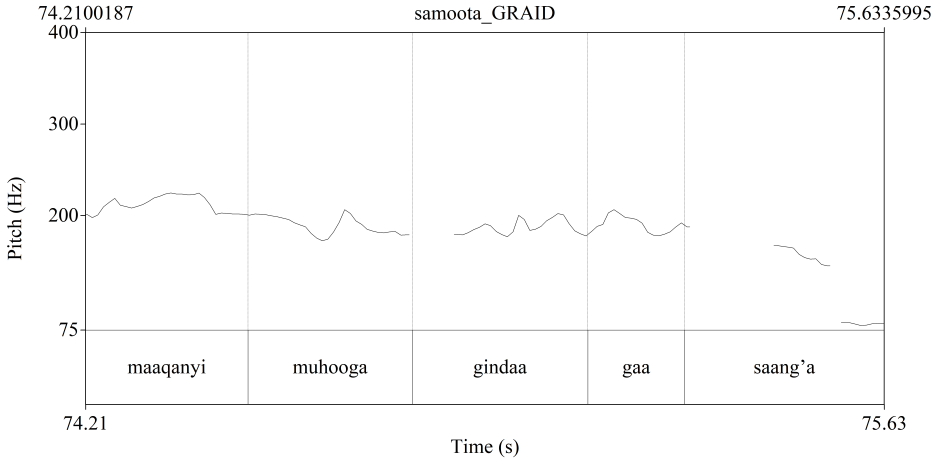

Figure 4.4: Left dislocation with no pause but co-referential element.

(54) *maaqanyi muhooga gindaa gea sàang'a*
maaqanyi meay-oo-ga g-ii-ndeawa gea saang'a
? calf-PS-PL AFF-2SG-have ANAPH.PRO.PL many
'You have a lot of calves.'

[Samoota_20]

Next to dislocation, there is another construction with preverbal constituents and that is the cleft construction. According to Lambrecht (2001a: 467), "a cleft construction (CC) is a complex sentence structure consisting of a matrix clause headed by a copula and a relative or relative-like clause whose relativized argument is coindexed with the predicative argument of the copula. Taken together, the matrix and the relative express a logically simple proposition, which can also be expressed in the form of a single clause without a change in truth conditions."

The structure of clefts in Datooga is still not clear. However, there is one structure which comes close to the proposed definition by Lambrecht (2001a) and this is the structure of "*nea* + NP + *ea* + verb". The following figures and clauses of Example 55 illustrate this. In all of these utterances the pitch contour rises with *nea* which suggests that there is some prominence on this part. However, the status of *nea* and whether it functions as a copula in these instances is not clear yet.

(55) a. *nea lapiya ea mɛ́ɛaadi*
nea lapiya ea m-ɛɛ-aad-i
CONJ money CONJ NEG-1SG-take-IS
'I didn't take the money.'

[2017-04-19_Maandazi_397]

b. *nea duga ea nisàang'ayiit há*
nea dee-ga ea nu-saang'a-ay-iid ha
CONJ COW-PL CONJ 3.PRF-be.many-PLUR-INCP DSC
'[...] but the cattle were many.'

[2014-11-14_YG_story1_168]

c. *asaha nea giing'eataanyi (.) ea gidêabá gwanda maleeda*
asaha nea gii-ng'ead-da-nyi (.) ea gideaba g-wa-nda mal-ee-da
DSC CONJ NMLZ-start-UR-3SG CONJ CONJ AFF-3-be honey-PS-UR
'At its start there's honey.'

[2014_11-14_YG_honeybeer_010–011]

In Example (55a) it is the NP *lapiya* 'money' that follows *nea* and precedes the following *ea* and the verb *aad* 'take'. The clause in the second Example (55b)[1] shows a similar pattern to the previous example as it also starts with *nea* followed by an NP, *duga* 'cows/cattle', followed by *ea* and a verb. Example (55c) also follows this structure although it starts with the discourse particle *asaha* and has a third conjunction *gidêabá* following *ea*.

[1]Note that this is the continuation of Example (53) which was used to illustrate left dislocation.

The following Figures 4.5, 4.6, and 4.7 show the pitch contours of the examples in (55). In the pitch contour of Example (55a) in Figure 4.5 there is a slight increase of pitch at the start of the utterance on *nea*. However, the highest pitch in this example is on *ea* which is then again followed by a decrease in pitch.

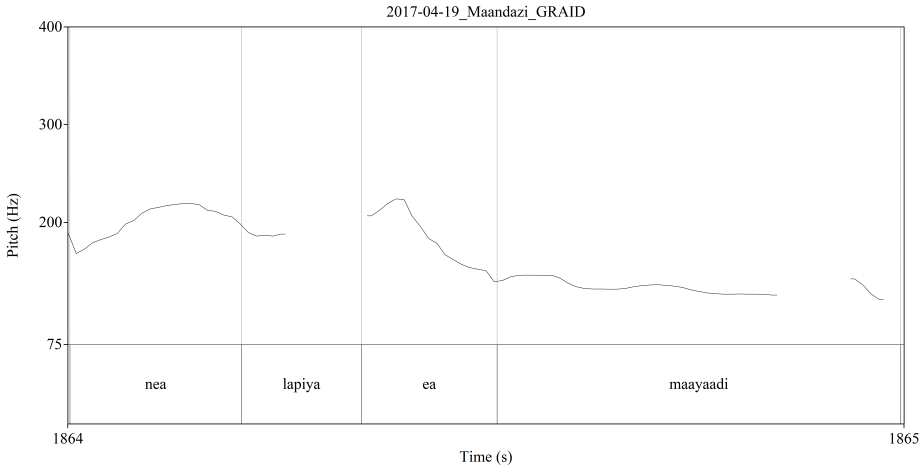

Figure 4.5: Pitch contour of Example (55a)

The pitch contour of Example (55b) illustrated in Figure 4.6 behaves a little differently. Although it also increases on *nea*, the highest pitch in this contour is on the NP *duga* followed by a strong decrease on *ea*. The pitch rises again after *ea* but it is still lower than on the NP *duga*.

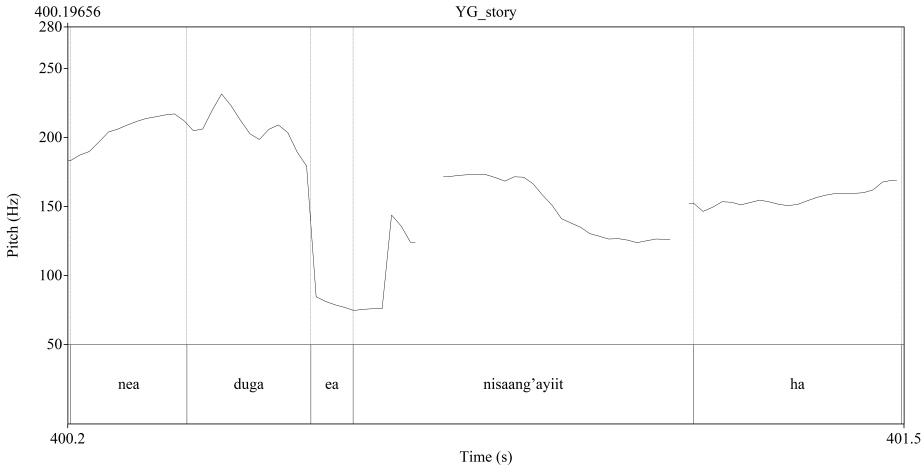

Figure 4.6: Pitch contour of Example (55b)

In the last Example (55c) the pitch contour in Figure 4.7 is similar to the previous ones in that it increases on *nea* where it also is at its peak in this utterance. In contrast to the previous examples, there is also a pause between the NP *giing'eataanyi* 'start' and *ea* which is similar to the dislocated constructions above but deviates from them with the occurrence of *nea* and *ea*.

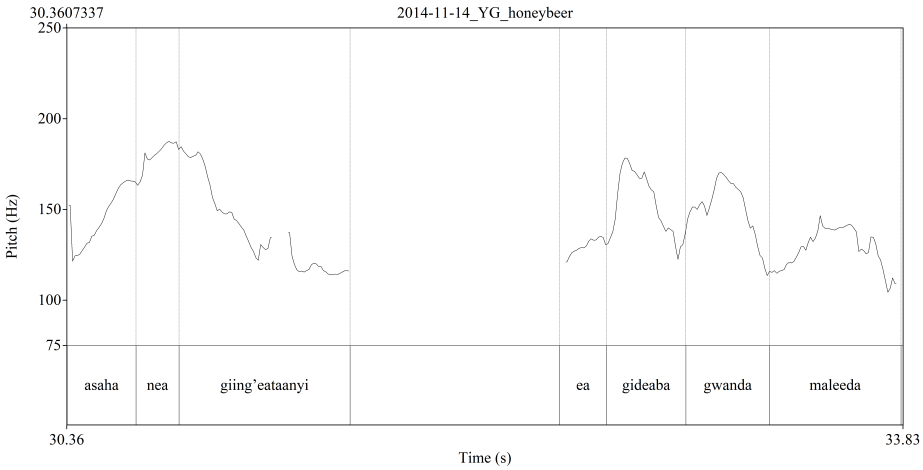

Figure 4.7: Pitch contour of Example (55c)

Usually, *nea* is used as a conjunction in adverbial clauses with the meaning of 'when' or 'while', as discussed in Section 4.3.1.2 with the examples in (44), or as a conjunction that coordinates main clauses, as in Example (45b).

In the presumed cleft examples in (55) *nea* seems not to introduce adverbial clauses, though. If *nea* is used to introduce temporal adverbial clauses, this would suggest that we are dealing with a simple clause with a preverbal argument, e.g. Example (55a) would exemplify the structure OV. However, as was discussed in Section 4.3.1.2 and was supported with Table 4.25, adverbial clauses only infrequently allow for preverbal arguments and that only in intransitive clauses. This could arguably be the case in Example (55b). Nevertheless, none of the examples seem to allow for a temporal adverbial reading.

Following the definition by Lambrecht (2001a), what follows the matrix clause headed by the copula, is a relative or "relative-like" clause. However, *ea* usually does not introduce relative clauses but it could introduce adverbial clauses, as well as main clauses. As discussed in Section 4.3.1.3, the characteristic of relative clauses is the subjunctive marking on the verb. As only first and second person have a dedicated subjunctive prefix, we would only expect this in Example (55a) on the verb *aad* 'take' which is not marked with a subjunctive prefix. This construction thus deviates from the relative clauses discussed above. What speaks against an interpretation of these clauses as adverbial clauses and also of main clauses coordinated by *nea*, on the other hand, is the occurrence of the second conjunction *ea*.

In summary, the construction of "*nea* + NP + *ea* + verb" and whether it constitutes a cleft in Datooga needs further research which lies outside the scope of this book. However, this structure still deviates from the structures of interest of this book through its use of the two conjunctions. Therefore, both the dislocated structures discussed above and this presumably cleft-construction are not included in the further study of this book.

This chapter has offered a comprehensive overview of the Datooga languages, beginning with an introduction to the Datooga-speaking community and followed by a typological profile. It

also provided an introduction to syntax and briefly covered structures that lie outside the main focus of this study, such as dislocation and clefts. Despite being outside the primary scope, discussing these structures was essential to clearly differentiate them from the ones central to the following analysis, namely declarative main clauses, which will be examined in the next chapter.

Chapter 5

Constituent order variation in main clauses

The previous chapter has introduced the Datooga dialect cluster, including its typological profile and an introduction to syntax. As has been mentioned in Chapter 2, the basic constituent order of Datooga in this study is determined by frequency measures. This makes Datooga a verb-initial language and supports the findings in the literature (e.g. Kießling 2007). However, as I will further discuss in the following chapter of this book, it is also a language that allows for constituent order variation and for other elements than verbs to occupy the clause-initial position, i.e. the position before the verb. Having discussed the preverbal placement of non-core arguments, dislocation and clefts in the previous chapter, the following chapter of this book will delve into this in terms of core arguments of the verb in both intransitive and transitive declarative main clauses, namely S, A, and O.

The initial section, Section 5.1 outlines the study, detailing the data and methods utilized in the subsequent analysis. Section 5.2, the first part of the results, provides an overview of the frequencies of constituent orders identified in the dataset, also distinguishing between conversational and narrative data. Following this, Section 5.3 explores the importance of semantic and syntactic features, such as animacy, argument realization, and syntactic weight, in relation to constituent order variation in Datooga. Section 5.4 presents the evaluation of the video stimuli, which were recorded to test the effect of discourse properties, such as givenness or animacy, on constituent order variation. In Section 5.5, the impact of givenness on constituent order is examined. This section also includes a qualitative study on information structure of clauses with preverbal arguments, as well as their discourse context. Finally, Section 5.6 presents a discussion of the findings derived from the analysis.

5.1 Outline of the study

The following sections outline the study on constituent order variation in Datooga. In the first Subsection 5.1.1, I will describe my underlying research questions for the subsequent analysis.

Subsection 5.1.2 presents the corpus utilized in this study and offers an overview of the recorded stimuli that will be incorporated into the qualitative analysis later in the chapter. Finally, Subsection 5.1.3 covers the methodology employed in the study, including a thorough explanation of the employed annotation system, which serves as the basis of the quantitative analysis.

5.1.1 Aims and objectives

Datooga is frequently characterized as a language that possesses a relatively flexible constituent order Kießling 2007: 171 for Gisamjanga-Datooga; Griscom 2019: 31 for Asimjeega-Datooga). Nevertheless, the precise degree of this flexibility has not yet been fully analyzed. Consequently, a primary aim of this part of the dissertation is to examine this flexibility within the Datooga dialects of Barabaiga and Gisamjanga. In various studies of languages, the variation in constituent order is frequently analyzed only within narrative texts, i.e. monologic texts. Nevertheless, as explained in Chapter 2.2, constituent order may diverge based on the genre. This brings us to the first research question:

> RQ 1: How does constituent order vary across different text genres, i.e. monologues and multi-party conversations?

The typological survey on Nilo-Saharan languages in Section 3.2, as well as other briefly mentioned case studies of languages around the world in Section 2.2.3, showed that there are various conditions influencing the choice of constituent order in these languages, including animacy, argument realization, or weight of arguments. Therefore, another research question of this study is the following:

> RQ 2: What are the conditions for constituent order alternation in Datooga?

Both research questions one and two require a quantitative examination of the annotated corpus, excluding data from stimuli-based elicitation. The basis for answering these questions is the GRAID ('Grammatical relations and animacy in discourse') annotation scheme (Haig & Schnell 2014), which will be explained in detail in the following section, but it essentially involves annotating arguments in terms of grammatical relations, syntactic, and semantic features. This scheme was chosen because it allows for the annotation of animacy and realization of the arguments, as well as for the negation of sentences, which were, along with weight, the conditions examined in the study.

Finally, one factor that is often associated with constituent order variation and often also the only explanation given for a variable constituent order, is information structure. The typological survey showed that for many languages especially pre-verbal position is often explained with informational structural processes, such as topicalization, or categories like topic, focus and givenness. Thus, it is necessary to also check for this feature when it comes to constituent order variation in Datooga, which leads to the final research question of this study:

> RQ 3: What is the impact of information structural categories, such as givenness, topic, and focus, for the different constituent orders in Datooga?

To answer this research question, it is necessary to investigate the discourse that surrounds the target utterances which, in this case, are phrases containing preverbal arguments in particular. Therefore, a qualitative analysis will be performed on a sample of texts, including the stimuli.

5.1.2 Data

The majority of data used for the analysis of constituent order variation in Datooga stem from a corpus assembled by Alice Mitchell over a span of almost ten years. To ensure a comprehensive and balanced sample of distinct genres, both narrative monologic texts and multi-party conversations were drawn from this corpus for the analysis with both applying a similar number of clauses in total. In addition, there is also a set of elicited texts in the analysis that was collected during fieldwork in February 2023. We collected these elicited texts using self-produced video-stimuli that were loosely based on the Visibility task in the QUIS ('Questionnaire for Information Structure') manual (Skopeteas et al. 2006). The goal of the stimuli-based elicitation was to specifically highlight certain information structural categories, such as givenness, animacy, or identifiability of the referents. We aimed to investigate whether these categories influence constituent order and we created the stimuli with the expectation that they would reveal such effects.

Before producing our own stimuli, we tried doing the original tasks from the QUIS manual which involved presenting participants with images serving varying information structural roles. For instance, the first picture showed an animate patient and in the next one there was an animate agent who did something to the patient. The problem here was that the participants did not immediately understand that the picture series showed an on-going event and that the people shown in these pictures were the same. We needed to explain this explicitly which undermined the purpose of the whole task.

Therefore, we decided to make our own stimuli embedded within the participants' familiar environment, featuring activities and objects from their daily lives. Furthermore, we chose video format over static pictures to ensure participants' comprehension of the ongoing nature of the events. During the video presentations, playback was paused within the initial moments, so that the participants could describe what they see, i.e. either the patient or agent of an action. We conducted this task with six participants in total of which three recordings have been annotated so far. The stimuli and their descriptions are listed in Table 5.1.

As this table shows, we aimed for a variety of participant properties and discourse properties. For instance, in some of the stimuli both agent and patient are human, as in stimuli one or 17. Other stimuli, such as seven or eleven, have a human agent and an inanimate patient. Some stimuli show the patient first and then agent, as in six or 16, while in other stimuli only the patient is fully visible, as in seven or 15. An example for the beginning and end of a stimulus is provided in Figure 5.1.

Video number	Description	Video number	Description
1	One woman pushes another	**10**	Camera focuses on woman; woman picks up cooking pot
2	Calabash on fence, woman arrives and takes it away and leaves	**11**	Woman empties calabash
3	Calabash inside house, woman arrives, takes lid off, pours in milk, puts it in rope and begins to churn	**12**	Woman standing outside, goes to pick up calabash on fence
4	Bag on windowsill, falls and hits woman sitting below window	**13**	Women arguing, one pushes the other
5	Water pours out of calabash	**14**	Women cooking, knocks over calabash as she walks
6	Cat sitting by house; woman comes and kicks it	**15**	Calabash on ground, woman [not fully visible] kicks it over (on purpose)
7	Calabash on ground, woman [not fully visible] kicks it over	**16**	Cooking pot on ground; woman comes and picks it up
8	Woman and calabash in house; woman hangs calabash	**17**	Woman walking past cat, kicks it
9	Woman and calabash in shot; woman moves and kicks it over (by accident)	**18**	Woman sitting down and bag falls on head

Table 5.1: List of stimuli descriptions.

However, these stimuli presented some challenges during data collection: since all videos were shown to all speakers and we only had two actresses (and a cat) in the videos, they were only new to the speakers in the initial videos. As a result, the effect of givenness was significantly limited. Therefore, the results should be seen as preliminary, providing a basis for future research.

Overall, the sample of texts analyzed in this study includes monologic narratives, multi-party conversations, as well as the recordings of the stimuli-based elicitation. These texts are summarized in Table 5.2.

5.1.3 Methods

The foundation for all the annotations in this study is the GRAID ('Grammatical relations and animacy in discourse') annotation scheme, which is "a system of symbols and conventions for glossing the grammatical relations and overt forms (noun phrases, pronouns etc.) of major clause constituents in texts" (Haig & Schnell 2014: 2). As the name suggests, the GRAID annotations not only capture the annotation of the grammatical relations and their respective morphological forms but also the annotation of the animacy status of these grammatical relations. Conse-

(a) Bag on windowsill

(b) Bag hits woman sitting below window

Figure 5.1: Example of stimuli 4.

File name	Content	Length	Word count	# clauses	Speakers
Rituals for a new born	Description of rituals conducted for a newborn baby	00:04:42	427	119	UD
Sheep, beam, and dog	Dilemma tale	00:01:27	127	43	UX
Donkey and hyena	Short folk tale	00:01:08	68	28	HM
Samoota	Folk tale	00:06:20	561	242	UD
Qeambeemeeda	Folk tale	00:06:54	677	222	YG
Honey beer	Procedural text	00:04:40	476	130	YG
Yonas hulanda	Multi-party 'visiting' talk	00:06:20	1104	368	multiple
Maandazi	Domestic interaction	00:55:11	1769	505	multiple
HM videos	Video stimuli based on QUIS	00:06:35	359	96	HM
UN videos	Video stimuli based on QUIS	00:08:09	271	85	UN
MA videos	Video stimuli based on QUIS	00:07:33	456	121	MA
	Totals	01:51:59	6295	1959	

Table 5.2: Annotated texts of the corpus; green highlighted texts: monological narratives, orange highlighted texts: multi-party conversations, red highlighted texts: elicited data (repeated for convenience).

quently, what emerges from a GRAID annotation is a sequence of symbols that can then be used for quantitative investigations (Haig & Schnell 2014).

The GRAID annotation is done on a separate tier[1] in an ELAN ('EUDICO Linguistic Annotator') (ELAN 2023) file. Figure 5.2 shows the GRAID annotation in the ELAN interface.

[1]Initially, our GRAID annotation was conducted on a tier with the word tier as its parent, a method consistent with the original manual. However, this approach posed challenges when dealing with possessive constructions, particularly as pronominal possessors are typically affixed to the possessee noun phrase. To address this issue, we previously annotated these occurrences with a hyphen to signify separate entities. However, this method proved unsatisfactory, particularly for subsequent statistical analysis, as distinct referents and annotations should be segmented separately. Consequently, we opted to change the parent tier of our GRAID annotation to the morpheme tier. This adjustment allows for clearer delineation of individual elements and facilitates more accurate statistical analysis in the future.

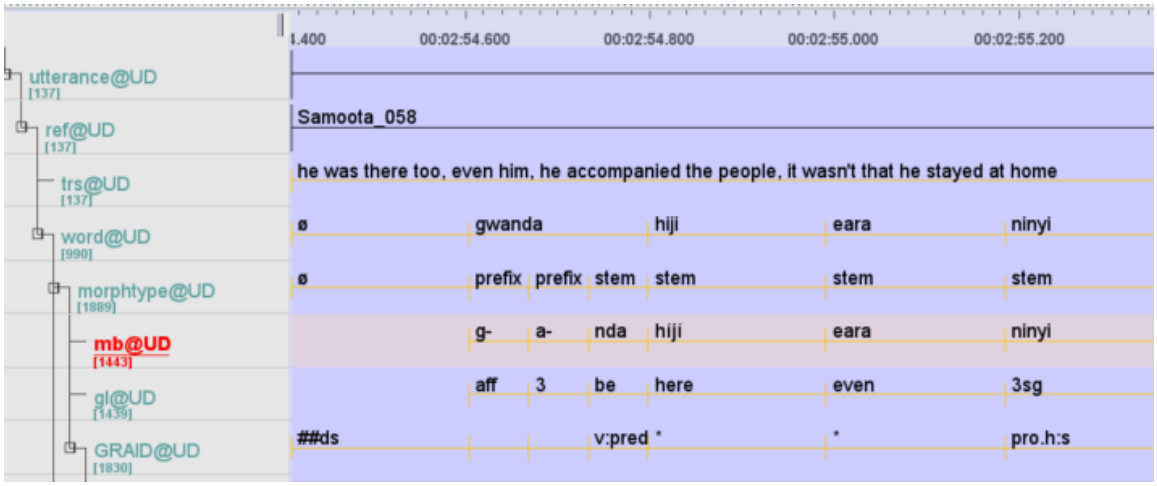

Figure 5.2: ELAN interface with GRAID annotation.

As can be deduced from this figure, there are certain symbols used in the annotation scheme, e.g. there are dedicated symbols used for clause boundaries or arguments. In the following, I will provide an overview about the symbols within the GRAID annotation scheme and whether and how we in our annotation process diverged from the original manual.

First of all, every clause is introduced by a clause boundary which is put in a separate segment before the clause and in the case of embedded dependent clauses, there further is a clause boundary at the end of the embedded clause indicating that it is over and the main clause continues. The GRAID annotation scheme differentiates between main clauses and subordinate clauses and has dedicated annotations for the semantic type of dependent clauses. The annotation scheme further distinguishes affirmative from negative clauses whereby the latter receive a special annotation. Clauses that consist of direct or indirect speech also receive a dedicated annotation tag which directly follows the clause boundaries and negation annotation and, in case of subordinate clauses, the annotation for subordinate clauses. An example of a (complex) clause boundary annotation is [`#ds_ac.neg`], which denotes a negated adverbial clause expressing (in)direct speech. In the original GRAID annotation scheme, the function of subordinate clauses is also annotated (see Haig & Schnell 2014: 24), but we chose to not follow this practice. Table 5.3 shows the different annotations used for clause boundaries.

The annotation for predicates in the GRAID annotation scheme consists of (usually) two elements which are the type of the verb, e.g. a full verb [`v`] or an imperative verb [`vother`], and the function of the verb, which is usually that of a predicate [`pred`] but can also be [`predex`] for marking existential predication. Regarding the predicate type, we also included an annotation for impersonal verbs, which prevail in our data, using [`vimp`]. We decided to distinguish personal from impersonal verbs as the latter behave differently in that they do not allow for referents in S or A function. This also helped excluding them from the further analysis. Following the GRAID annotation scheme, the type and the function are separated with a colon, as in [`v:pred`]. As our annotation is done on the morpheme level, only the stem of the verb receives the annotation

GRAID	Meaning
Boundaries	
##	Beginning of main clause
#	Beginning of dependent clause
%	End of dependent clause
Types of subordinate clauses	
#ac	Adverbial clause
#rc	Relative clause
#cc	Complement clause
Other annotations	
neg	Negative clause
ds	(In)direct speech
nc	Not considered

Table 5.3: Symbols used for clause boundaries in GRAID annotation scheme, adapted from Haig & Schnell (2014).

while the segments for the other morphemes are left empty. The annotations used for type and function of the predicates are listed in Table 5.4.

The next category annotated within the GRAID annotation scheme are the arguments of predicates. The elements annotated for this category are the form of the argument, such as noun phrase or pronoun, the animacy status, such as human or animate, as well as the function or syntactic role of the arguments, e.g. S, A, or P[2]. The annotation of the arguments begins with the form, followed by the animacy status, which is separated from the form with a period. An advantage of the GRAID scheme is its inclusion of zero realized arguments. This is achieved by adding a new annotation element, which is then marked with [0]. It is also possible that an argument does not have one of the mentioned animacy tags, i.e. it is inanimate. In that case there is no period involved and the function is annotated directly after the form, e.g. [0:p]. In the GRAID manual by Haig & Schnell (2014), there is no specific annotation for animate entities apart from the annotation for human referents. To enable a more detailed analysis of animacy status, we decided to annotate animate referents that are not human with [an]. Additionally, we used [*] instead of [other] to annotate forms which are not arguments of the verb, such as particles or interrogative pronouns. The function of the argument is the final part of the annotation and, just as for the predicates, separated from the animacy status with a colon. An example annotation could look like [np.h:s] which means that the annotated argument is a

[2]In the previous chapters I used O to refer to the object of transitive clauses. However, in the GRAID annotation scheme P is used instead.

GRAID	Meaning
Predicate type	
v	Full verb
vimp	Impersonal verb
vother	Deficient verbs, e.g. imperatives
aux	Auxiliary
cop	Copula
lv	Subconstituent of verb complex occurring to the left of verbal head
rv	Subconstituent of verb complex occurring to the right of verbal head
Predicate function	
pred	Predicate
predex	Predicate in existential construction
other	Other

Table 5.4: Symbols used for predicates in GRAID annotation scheme, adapted from Haig & Schnell (2014).

human noun phrase in S-function. The following Table 5.5 shows the different symbols used for the annotation of arguments.

Lastly, there are symbols used for the annotation of non-core arguments, such as obliques, predicates of nonverbal predication, possessors, as well as dislocated topics. All of these categories are annotated as a function, i.e. their annotation follows after a colon, e.g. [`np.h:dt`]. Table 5.6 gives an overview of these categories.

After the annotation of the files in ELAN, the data was imported to R (R Core Team 2022), where the GRAID annotations were extracted from the ELAN files. One of the key steps within the R script is the numbering of the clauses on which every further analysis was based on. For this step, the clause boundaries within the GRAID annotation scheme were important. For every new main or dependent clause, which started with either [`##`] or [`#`], a counter was raised by one. Every [`%`], which indicated the end of an embedded dependent clause, on the other hand, reduces the counter by one so that the parts of the main clause surrounding the embedded clause receive the same number. With the clauses being numbered, it was possible to detect all elements belonging to the same clause as they shared the clause number. For example, this made it possible to label the transitivity of clauses, i.e. if clauses contained a [`v:pred`] or [`vimp:pred`] as well as a [`:p`] annotation in the GRAID column that shared the clause number in the clause number column, they were identified as transitive clauses. By contrast, if there were

GRAID	Meaning
Argument form	
np	Noun phrase
pro	Pronoun
0	Zero argument
dem_pro	Demonstrative pronoun
anaph_pro	Anaphoric pronoun
ln	NP-internal subconstituent occurring to the left of NP head
rn	NP-internal subconstituent occurring to the right of NP head
*	Other forms that are not the previous forms, e.g. interrogative pronouns
Animacy status	
1	First person
2	Second person
h	Human
an	Animate
d	Anthropomorphized referent
Argument function	
s	Single argument of intransitive clause
a	Most agent-like argument of transitive clause
p	Most patient-like argument of transitive clause
p2	Secondary object

Table 5.5: Symbols used for arguments in GRAID annotation scheme, adapted from Haig & Schnell (2014).

[`v:pred`] or [`vimp:pred`] and no [`:p`] annotations in the GRAID column with the same clause number, the clause was labeled as intransitive. Another variable that was annotated with the R-script was the order of core arguments and the verb for both transitive and intransitive clauses. This word order column also included zero arguments, i.e. it includes values such as V with only the verb being realized. Other variables important for the analysis concern the properties of the arguments in the transitive and intransitive clauses, i.e. S, A and O (P in the GRAID annotation scheme). First of all, these arguments were annotated for their animacy statuses

GRAID	**Meaning**
Non-core argument functions	
g	Goal argument of a goal-oriented verb of motion, but also: recipient of verb of transfer, and addressee of verb of speech
l	Locative argument of verbs of location
obl	Oblique argument, excluding goals and locatives
dt	Dislocated topic
poss	Possessor
pred	Predicate (in nonverbal predication)
appos	Apposition
voc	Vocative
other	Other functions that do not fit into the other categories
Other non-core annotations	
refl	Reflexive pronouns (no function added)
adp	Adposition

Table 5.6: Symbols used for non-core functions in GRAID annotation scheme, adapted from Haig & Schnell (2014).

which was also extracted from the GRAID annotation. Within the script, each argument role received their own column, i.e. there is a separate animacy column for S, A and O. Possible values for the animacy status are first person, second person, human, animate, anthropomorphized, and inanimate. Second, each of the core roles received a column for their form or realization which was also extracted from the GRAID annotation column. Arguments can either be realized as noun phrase, pronoun, or as a zero argument.

The quantitative analysis was conducted in several stages. In the first stage, the frequencies of constituent orders were analyzed alongside the effects of the semantic and syntactic features of animacy, argument realization, and negation. This analysis was performed on the entire dataset, as illustrated in Table 5.2, excluding the stimuli-based data, as well as impersonal constructions and relative clauses. The stimuli-based data was excluded because it was elicited data and thus different from the spontaneous data. As impersonal constructions and relative clauses both restrict the number of possible arguments in their own way, the behave differently from the declarative main clauses with personal verbs and are therefore also excluded.

The second stage focused on a subset of the data, specifically all clauses with fully realized arguments, i.e. VS and SV for intransitive clauses, and VAO, VOA, AVO, and OVA for transitive clauses, while also excluding stimuli-based data, impersonal constructions, and relative clauses as in the first stage. This subset was used to annotate the syntactical weight of each S, A,

and O. The comparison unit for syntactical weight was the length of an argument in syllables, which required manual annotation on a case-by-case basis. This was done for each argument, i.e., for S in intransitive clauses and A and O in transitive clauses. After annotation, the data was reimported into R.

Finally, the third stage involved the INCEpTION platform (Klie et al. 2018). INCEpTION is an "annotation platform for tasks including interactive and semantic annotation (e.g., concept linking, fact linking, knowledge base population, semantic frame annotation)" (Klie et al. 2018: 5). The annotation was conducted on a "coreference" layer, as shown in Figure 5.3. Each discourse referent is manually marked, indicated by colorful boxes labeled with (Cor.). If a referent is coreferent with another discourse referent, they can be linked by dragging one box to the other. Each referent is automatically assigned a unique number, followed by a second number indicating its sequence of appearance in the text. For example, a referent might be labeled 4-2, indicating the second occurrence of the fourth discourse referent in the text. With this annotation, the givenness status of references can be tracked, which is relevant for Section 5.5.1.

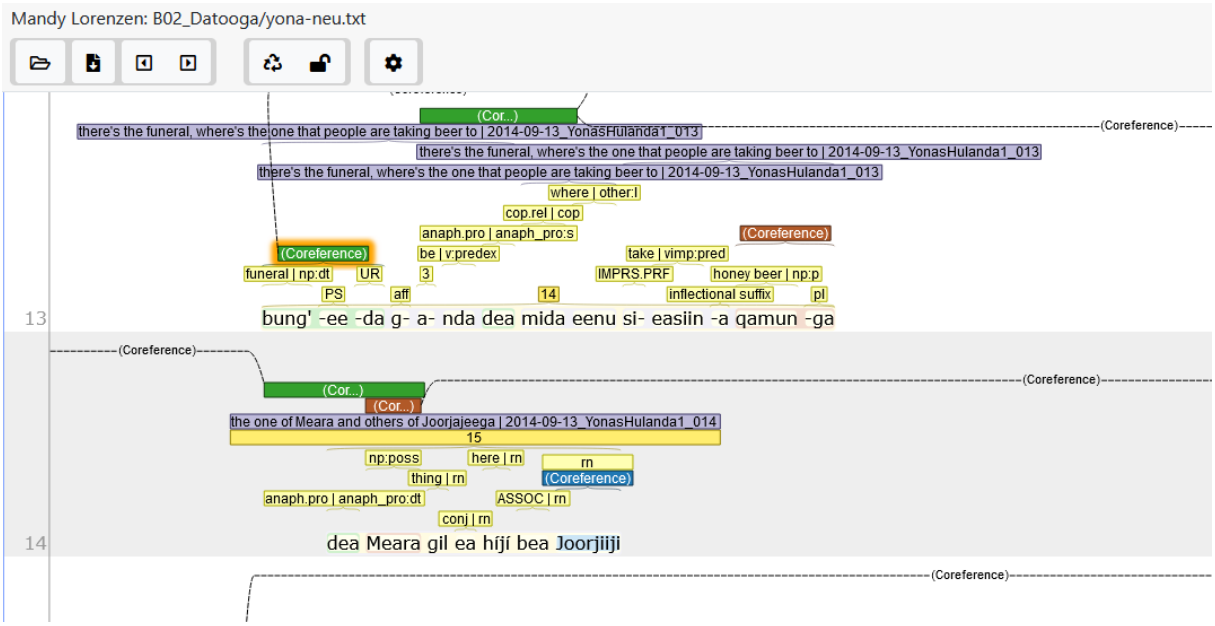

Figure 5.3: INCEpTION interface.

The INCEpTION annotation is implemented on a subset of the corpus, as shown in Table 5.7. This subset mainly consists of the three stimuli-based texts of the corpus but also includes one of the two conversations, as well as two narratives. However, so far only fully realized arguments were considered in the INCEpTION annotation. As the analysis of constituent order in Datooga will show, there are a lot of zero realization of arguments in the language. Thus, it will be necessary to include them in future research.

The qualitative analysis of the data was more fine-grained and also concentrated on those constituent orders with fully realized arguments due to feasibility constraints. For this analysis,

File name	Content	Length	Word count	# Clauses	Speakers
Donkey and hyena	Short folk tale	00:01:08	68	28	HM
Honeybeer	Procedural text	00:04:40	476	130	YG
Yona's hulanda	Multi-party 'visiting talk'	00:06:20	1104	368	multiple
HM videos	Video stimuli based on QUIS	00:06:35	359	96	HM
UN videos	Video stimuli based on QUIS	00:08:09	271	85	UN
MA videos	Video stimuli based on QUIS	00:07:33	456	121	MA
	Totals	00:34:25	2734	828	

Table 5.7: List of analyzed texts in INCEpTION.

the spreadsheet used to annotate the syllable length was used, i.e. a subset of the whole data, with a focus on clauses with preverbal arguments. It further included the clauses with preverbal arguments of the stimuli-based texts. In this step, each argument, i.e. S, A and O, was annotated for its informational category, i.e. focus, topic, or contrast.. As the previous chapters showed, information structure is always connected to the context of an utterance. Therefore, the surrounding discourse of a clause was taken into account as well. The decision for annotating an argument as being the focus, topic or contrast of a clause were based on the definitions in the previous Chapter 2.

It is important to note that for both the quantitative and the qualitative analysis, certain constituent orders only have very few occurrences, such as clauses with both arguments fully realized, which limits the validity and practical significance. Consequently, these values may not provide strong conclusions on their own and should be treated with caution. However, they are included in the analysis as they are of particular relevance to the research question.

5.2 Constituent order frequency

Table 5.8 summarizes the frequency of constituent orders in the data. This table includes data from narratives and conversations and displays constituent ordering in all clause types, without distinguishing between main and subordinate clauses. This computation excludes impersonal constructions and relative clauses[3]. This adds up to 900 clauses in the analysis.

The table shows that the majority of clauses do not contain an overt S or A argument (673; 74.78%) of which the majority are intransitive clauses without S (435; 48.33% of the total count), followed by transitive VO order (148; 16.44% of the total count). As for the preverbal orders, SV has the most occurrences within the data (44; 4.89% of the total count). In terms of preverbal A, AVO is uncommon relative to the overall count, i.e. 24 occurrences (2.67%), but is more frequent than AV without an overt O argument (4; 0.44% of the total count). AV order is even less common than the rarest order within the category of no overt S/A arguments, OV

[3]As already mentioned in Section 4.3.1.3, relative clauses behave differently than other subordinate clauses as they never have both arguments fully realized in Datooga. They were therefore excluded.

(8; 0.89%). The data contains no instances of OVA order. VS has the largest frequency (111; 12.33% of the total count) in relation to the postverbal position of S and A, followed by VAO (18; 2% of the total count), though VS also has more occurrences compared to VAO in general. Within the same group, VOA order is the least prevalent (5; 0.56%), even less so than AVO order.

Constituent order	**N**	**%**
Preverbal S/A		
SV	44	4.89
AV	4	0.44
AVO	24	2.67
Total	**72**	**8**
Postverbal S/A		
VS	111	12.33
VA	21	2.33
VAO	18	2
VOA	5	0.56
Total	**155**	**17.22**
No overt S/A		
V (intr.)	435	48.33
V (tr.)	82	9.11
VO	148	16.44
OV	8	0.89
Total	**673**	**74.78**
Total all groups	**900**	**100**

Table 5.8: Overview of total constituent order frequencies in the data, excluding the stimuli, impersonal constructions and relative clauses.

These results support the "Preferred Argument Structure" principle as proposed by Du Bois (1987). Following this principle, the first constraint says to "avoid more than one lexical argument per clauses" (Du Bois 1987: 819), which, concerning the transitive clauses, would result in a preference for clauses with only one argument realized, i.e. VA, VO, AV, and OV. Taken together, these make up the majority of transitive clauses (181), compared to those where both arguments are zero realized (82), and those with both arguments fully realized, i.e. AVO, VAO, VOA (47). In addition, lexical mentions typically occur with arguments in S and O function rather than A which constitutes the second constraint: "Avoid lexical A's" (Du Bois 1987: 823).

This can also be observed in the data as fully realized A arguments appear less (72; 23.23% of 310 transitive clauses) compared to those in O function (203; 65.48% of 310 transitive clauses). However, although fully realized arguments in S function, i.e. 155; 26.27% of 590 intransitive clauses, are more common than fully realized A arguments when comparing the relative numbers, the difference between these two is not as big as it is between A and O.

Table 5.9 depicts the frequency distribution of constituent order in the two genres, conversations and narratives. A comparison of the three categories, i.e. preverbal S/A, postverbal S/A, and no overt S/A, in the two genres indicates that the general distribution is similar. The category with no overt S/A has the greatest occurrences, followed by postverbal S/A, followed by preverbal S/A. Still, there are a few distinctions between the two categories. The category of preverbal S/A in conversations (11.09%) has a larger proportional frequency of all clauses in conversations than in narratives (4.36%). The deviation between preverbal S/A (11.09%) and postverbal S/A (13.35%) is smaller than in narratives (4.36% vs. 21.79%), indicating a more balanced distribution between these two positions in conversations compared to narratives. AVO order in conversations is also nearly as prevalent as SV order (5.95%). Overall, while the majority of clauses in both conversations and narratives have no overt S/A arguments, conversations have a higher frequency of preverbal S/A and a more balanced distribution of preverbal and postverbal S/A than narratives, which have a stronger tendency for postverbal S/A when comparing the two categories with overt arguments.

5.3 Semantic and syntactical features

The following sections present the results with regard to the semantic and syntactical features that were relevant conditions for variation in the languages of the typological survey, discussed in Chapter 3, and that could be analyzed with the help of the GRAID annotation scheme that was applied in this study. These features are animacy (Section 5.3.1), argument realization (Section 5.3.2), as well as negation (Section 5.3.3). In addition, syntactic weight was analyzed which did not come up as a conditioning factor within the data of the survey earlier but represents a condition for constituent order variation in other languages of the world, as discussed in Chapter 2.

Of interest for the results is the position of the arguments, i.e. S, A, and O, in relation to the verb. Therefore, constituent orders where all the arguments are zero realized, i.e. V in intransitive and transitive clauses, are not further discussed.

5.3.1 Animacy

In the typological survey in Section 3, animacy was found to be one of the conditions associated with a deviation in constituent order in Nilo-Saharan languages. Animacy altered the order of constituents in Turkana, Teso, Lopit, and Toposa. In Lopit, the order of postverbal arguments

Conversations			**Narratives**		
Constituent order	**N**	**%**	**Constituent order**	**N**	**%**
Preverbal S/A					
SV	29	5.95	SV	15	3.63
AV	4	0.82	AV	0	0
AVO	21	4.31	AVO	3	0.73
Total	**54**	**11.09**	**Total**	**18**	**4.36**
Postverbal S/A					
VS	43	8.83	VS	68	16.46
VA	10	2.05	VA	11	2.66
VAO	9	1.85	VAO	9	2.18
VOA	3	0.62	VOA	2	0.48
Total	**65**	**13.35**	**Total**	**90**	**21.79**
No overt S/A					
V (intr.)	243	49.90	V (intr.)	192	46.49
V (tr.)	44	9.03	V (tr.)	38	9.20
VO	77	15.81	VO	71	17.19
OV	4	0.82	OV	4	0.97
Total	**368**	**75.56**	**Total**	**305**	**73.85**
Total of all groups	**487**	**≈ 100**	**Total of all groups**	**413**	**≈ 100**

Table 5.9: Overview of constituent orders in data between narratives and conversations, excluding the stimuli, impersonal constructions and relative clauses. Percentages may not total 100 due to rounding.

is conditioned by animacy in that the argument higher on the animacy hierarchy precedes the one lower on the hierarchy.

Table 5.10 shows the animacy status of fully realized and zero realized S, A, and O arguments in the data. For both, S and A, the majority of fully realized arguments have a 3rd person referent, i.e. 50.32% and 61.11%, respectively. However, this is also the category that has the most zero realizations for both arguments. Compared to that, the majority of fully realized O arguments refer to an inanimate entity, i.e. 74.77%.

	S				**A**				**O**			
	Overt		**Zero**		**Overt**		**Zero**		**Overt**		**Zero**	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
1st person	8	5.16%	90	20.69%	11	15.28%	63	26.47%	0	0	1	0.93%
2nd person	6	3.87%	43	9.89%	2	2.78%	37	15.55%	0	0	1	0.93%
3rd person	78	50.32%	239	54.94%	44	61.11%	118	49.58%	34	16.75%	18;	16.82%
Anthropo-morphized	7	4.52%	0	0	3	4.17%	3	1.26%	0	0	4	3.74%
Animate	6	3.87%	6	1.38%	2	2.78%	0	0	29	14.29%	5	4.67%
Inanimate	50	32.26%	52	11.95%	10	13.89%	17	7.14%	140	68.97%	80	74.77%
Total	**155**	**100%**	**435**	**100%**	**72**	**100%**	**238**	**100%**	**203**	**100%**	**107**	**100%**

Table 5.10: Overview of animacy status in overt and zero S, A, and O arguments. Percentages may not total 100 due to rounding.

Figure 5.4 presents the animacy status of S not distinguished by the genre of the text. It can be observed that 1st and 2nd person referents only occur preverbally. Human, 3rd person referents make up a larger proportion of all postverbal S with 60.4% compared to the 25% S in preverbal position. There are further slightly more inanimates preverbally (36.4%) than postverbally (30.6%). However, in total numbers, S occurs less often in preverbal position compared to when it follows the verb, i.e. 44 occurrences vs. 111.

In comparison to the total perspective, Figure 5.5 indicates the animacy of the S argument distinguished by the genre. In preverbal position (SV) in conversations, the majority of S arguments encode to a human referent, i.e. they are either 1st person, 2nd person, or human (3rd person), whereas 31% refer to an inanimate entity. In narratives, however, humans appear less frequent with a total of 33.3%, i.e. 6.7% 1st person, 13.3% 2nd person, and 13.3% human 3rd person referents, compared to 46.7% inanimate entities and 20% animate, non-human entities. Animate referents further only occur preverbally in narratives. In all genres, first and second person pronouns appear solely preverbally in intransitive sentences. This is consistent with Kießling's (2007) observation that first and second person pronouns are exclusively present

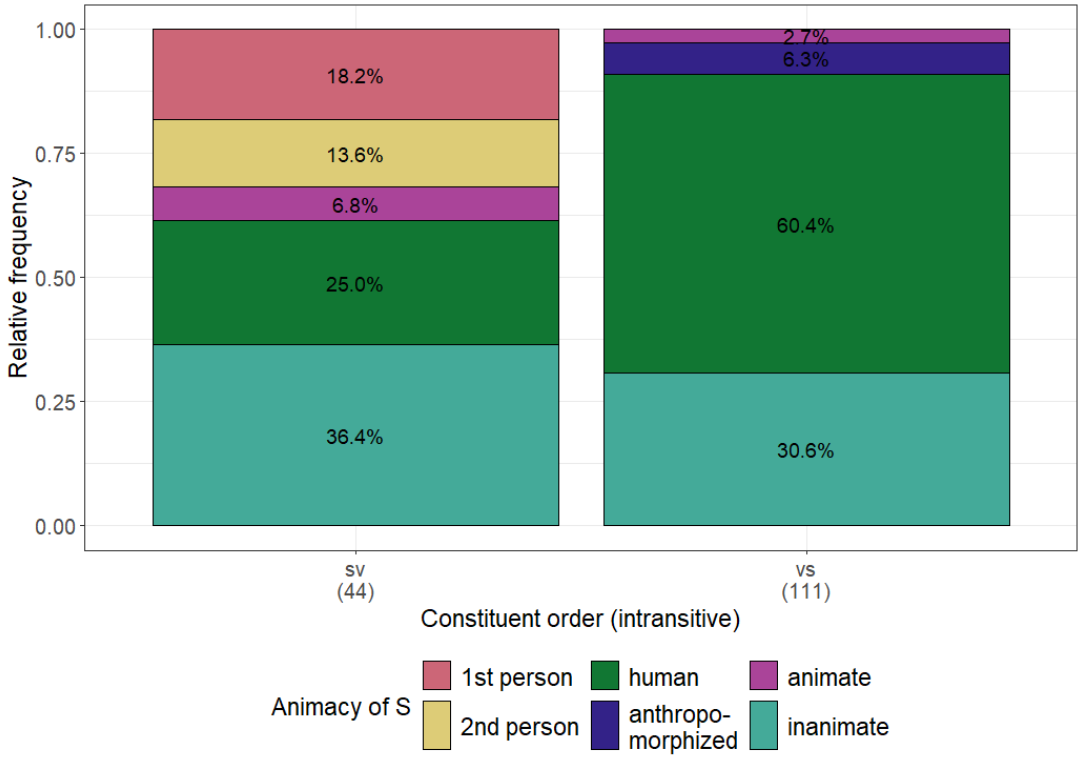

Figure 5.4: Animacy status of S.

in preverbal position where there is no case distinction. In postverbal position, animate and human entities predominate in both genres, but inanimate S arguments are also present. Overall, preverbal position appears to be the more flexible position for the animacy of S, since all values except anthropomorphized referents in both genres, as well as animates in conversations, may be detected here. Inanimate S arguments appear more frequently in preverbal position in narratives, accounting for almost half of all observations, whereas they appear similarly often in preverbal and postverbal positions in conversations. As a result, it appears that although both genres show a restriction for S in 1st or 2nd person in preverbal position, the preference for inanimate S arguments in preverbal position is only found within the narratives.

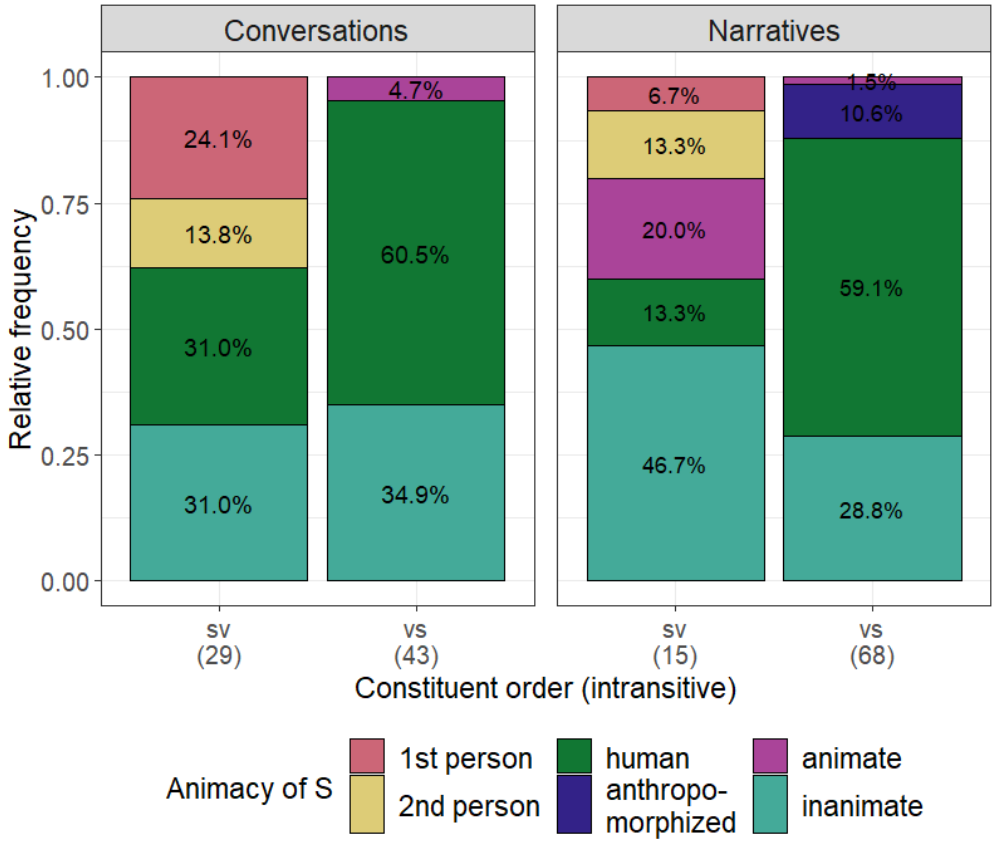

Figure 5.5: Animacy status of S and constituent order between the two different genres.

Figure 5.6 shows the animacy status of A in transitive constituent orders with overt A arguments first without distinguishing the genres.

In addition to the previous figure, Figure 5.7 illustrates the animacy status of the A argument in conversations and narratives. In both conversations and narratives, preverbal position is almost exclusively restricted to animate entities and entities higher on the animacy scale, i.e. human entities, 2nd person, and 1st person. Inanimate preverbal A arguments are only found in conversations and in AVO order; in all the other instances in both genres, they are only found postverbally. 1st and 2nd person pronouns are exclusively encountered in conversations, which is not surprising considering the nature of the genres, and they are usually used in preverbal position, as shown with S arguments. In general, the figure depicts the relative frequency of animate or human A arguments over inanimate arguments, the latter of which often occur postverbally. These results suggest that if A is inanimate, it is more likely to appear postverbally rather than preverbally.

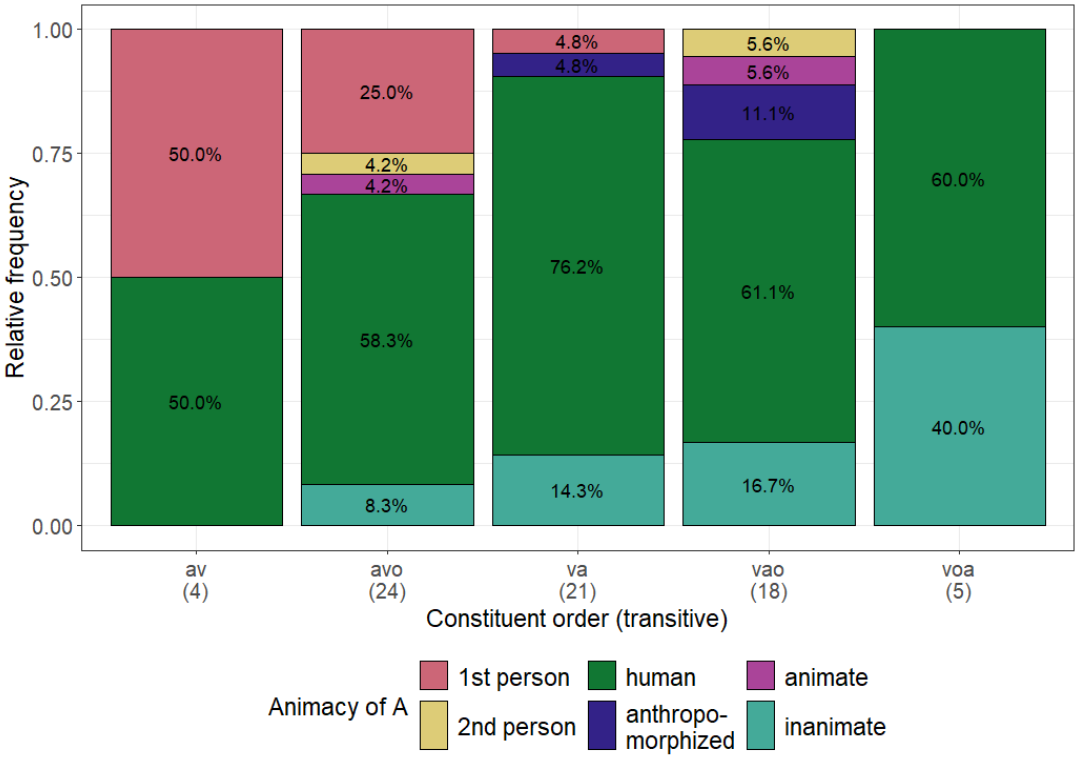

Figure 5.6: Animacy status of A.

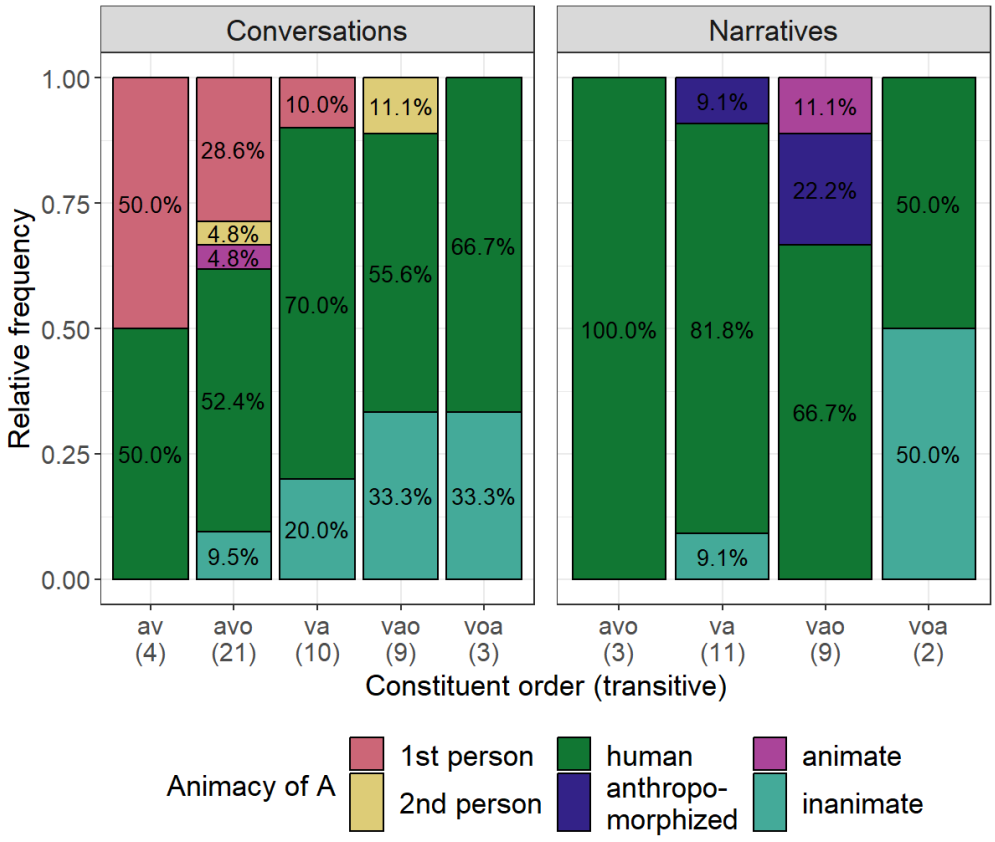

Figure 5.7: Animacy status of A and constituent order between the two different genres.

However, there is also a clause in the conversational data, where the inanimate A argument appears preverbally in AVO order. This is shown in Example (56).

(56) ***fuqareedeanyu*** *qwafkaachi*
fuqar-ee-da-nyu q-a-fkaa-sh-i
mind-PS-UR-1SG.POSS AFF-3-remember-AP-IS
gida *sihindaneas*
gida si-hind-an-easa
thing IMPRS.PRF-do.like-OBL-1PL.OBJ
'My mind remembers what was done to us.'

[2014-09-13_YonasHulanda1_223]

In this example, the inanimate A argument *fuqareedeanyu* 'my mind' precedes the verb *fkaa* 'remember'. Although the NP *fuqar* 'mind' is strictly inanimate, it might be argued that it may become more animate through the first person possessor suffix attached to the NP. Thus, the whole NP *fuqareedeanyu* 'my mind' stands in for the whole person and therefore it can metaphorically be understood as animate. However, this is an unusual example as constructions like these, where body parts are agents, are rare.

There is only one more example in the data with a similar construction but, in this instance, A follows the verb, as shown in Example (57).

(57) *akiisa* *nea* *lapiya* *ea* *neaheakti* *aniini*
ag=g-ii-s-a nea lapiya ea n-ɛɛ-heak-t-i aniini
SEQ-AFF-hear-TERM-IS CONJ money CONJ PRF-1SG-give-CF-IS 1SG
ea ***qam*** ***uhuudeanyu***
ea q-a-mi uhuu-da-nyu
CONJ AFF-3-die head-UR-1SG.POSS
'And I gave him money until my head was confused.'

[2017-04-19_Maandazi_415]

Further, there are two instances of postverbal 1st and 2nd person A arguments which, according to Kießling (2007: 160), never occur postverbally. In the first Example (58a) the first person pronoun *áni* follows the verb *mɛɛsh* with a zero realized object. Example (58b) shows the only occurrence of a postverbal second person pronoun *áng'ìing'ì*.

(58) a. *gaydamɛɛsh* ***áni***
g-ay-daa-mɛɛsh ani
AFF-FUT-1SG.SBJV-lick 1SG
'I'm going to lick.'

[2017-04-19_Maandazi_114]

b. *geeguura* *Meeta* *midaahi* ***áng'ìing'ì***
g-ee-guur-a Meeta m-ii-daa-i ang'iing'i
AFF-IMPRS-call-IS Meeta NEG-2SG-see-IS 2SG

bead-oo-da bak ea qwaraat laini
bead-oo-da bak ea q-wa-ra-a-d laini
?-PS-UR ? CONJ AFF-3-heed-IS-CF line
'He called Meeta, don't you see, until you lost the line.'
[2017-04-19_Maandazi_408]

Although these two sentences demonstrate that first and second person pronouns can appear postverbally, it is important to note that these are the only two instances of these pronouns in postverbal position in the data, and they are both from the same conversation text. Postverbal first and second person pronouns are consequently relatively uncommon.

The animacy status of O in the different constituent orders with overt arguments without a distinction of the genres is illustrated in Figure 5.8.

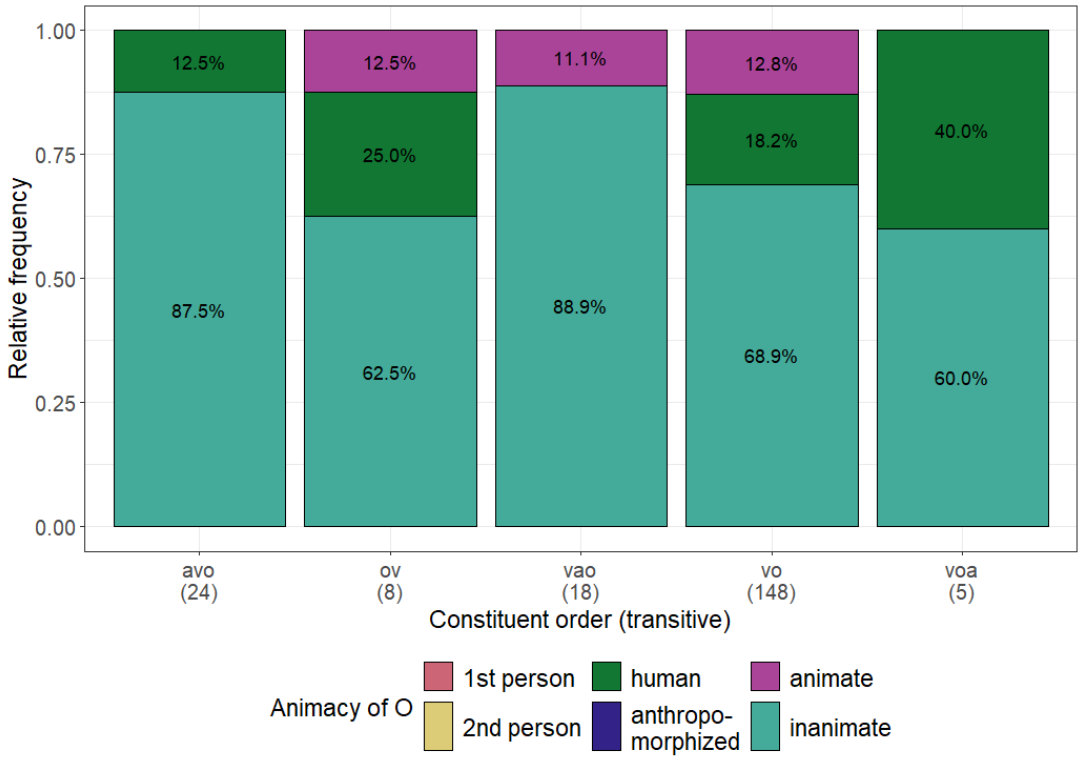

Figure 5.8: Animacy status of O.

Figure 5.9 shows the animacy status of the O argument for the constituent orders with overt arguments. The majority of O arguments in both genres are inanimate, which is unsurprising given their typology. Human and animate, non-human, O arguments can also be found in the data in conversations and narratives. Interestingly, these mainly occur when O is next to the verb, i.e. directly preceding or immediately following it, as in AVO, OV, VO, or VOA in both genres. In VAO order, every instance of O in conversations is inanimate, whereas narratives include 22.2% animate arguments but no human referents. These findings suggest that the position of O may be modified in some way by its animacy status: The higher the referent is

placed on the animacy hierarchy, the closer it seems to appear the verb. However, the results need to be viewed with care because the majority of constituent orders in which O is closest to the verb, i.e. AVO, OV, OV, and VOA, are rather rare.

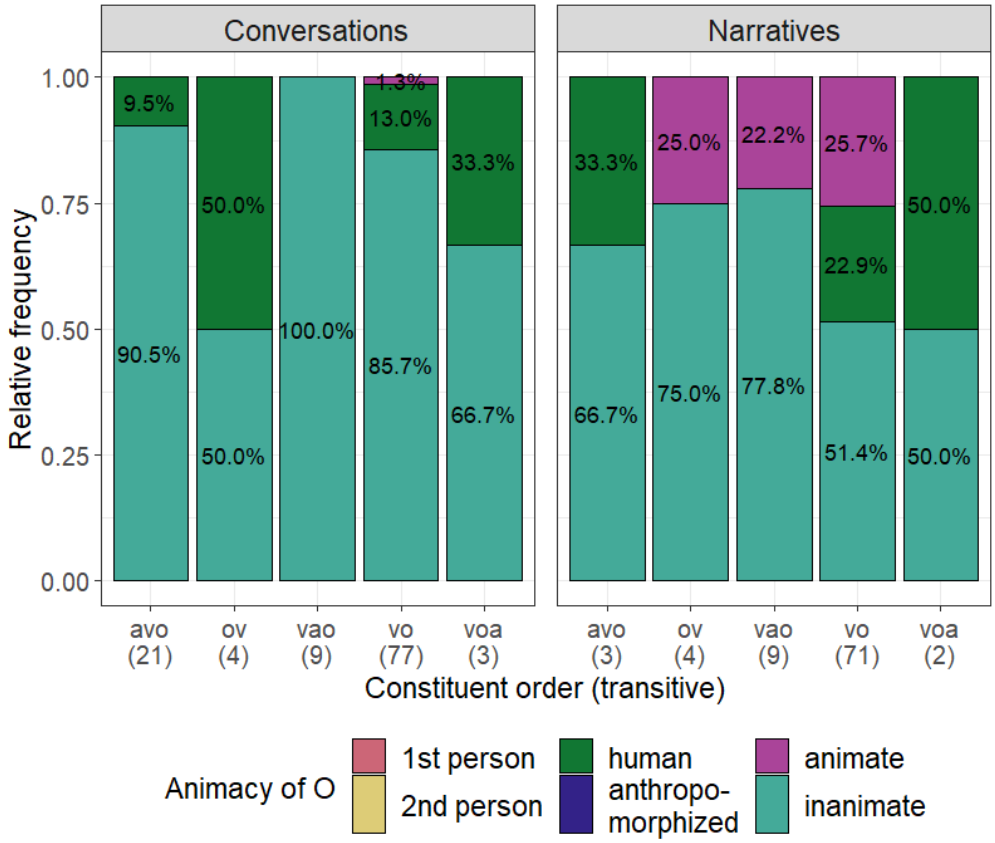

Figure 5.9: Animacy status of O and constituent order between the two different genres.

The overall interaction of animacy between A and O in clauses where both arguments are overt, is illustrated in Figure 5.10. Figure 5.11 only shows the impact of the interaction of the animacy states of A and O in fully realized clauses, i.e. AVO, VAO, and VOA. In both conversations and narratives, A ranks higher on the animacy hierarchy than O in AVO and VAO order. In these orders, there are also some clauses in which A and O share an animacy status and are therefore ranked equally. In VOA order in conversations, the equal rank of A and O is even in the majority while this order has the only instance of O being ranked higher than A in narratives.

These findings align with the results discussed above in that A arguments are predominantly human or animate while most of the O arguments are inanimate and as a consequence A is more often higher ranked than O. The only time O is higher on the animacy hierarchy than A is in VOA order in narratives, when O precedes A, which I presume is connected to its place on the animacy hierarchy, while A is postverbal, i.e. most likely inanimate. Thus, when O has a higher

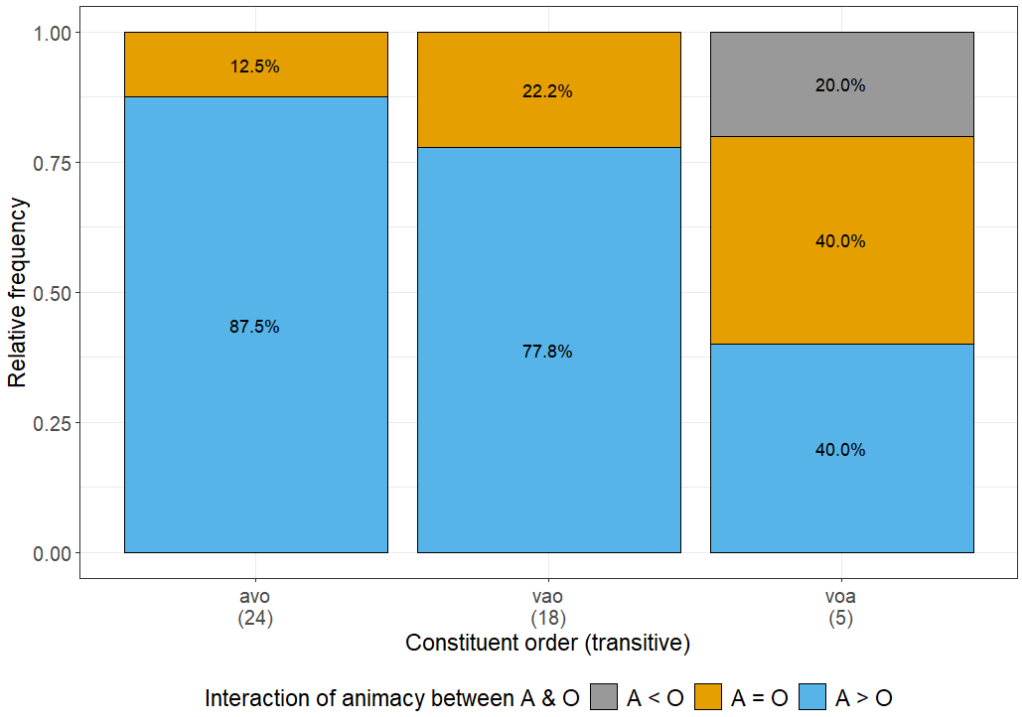

Figure 5.10: Interaction of animacy of A and O with constituent order in clauses with only fully realized arguments.

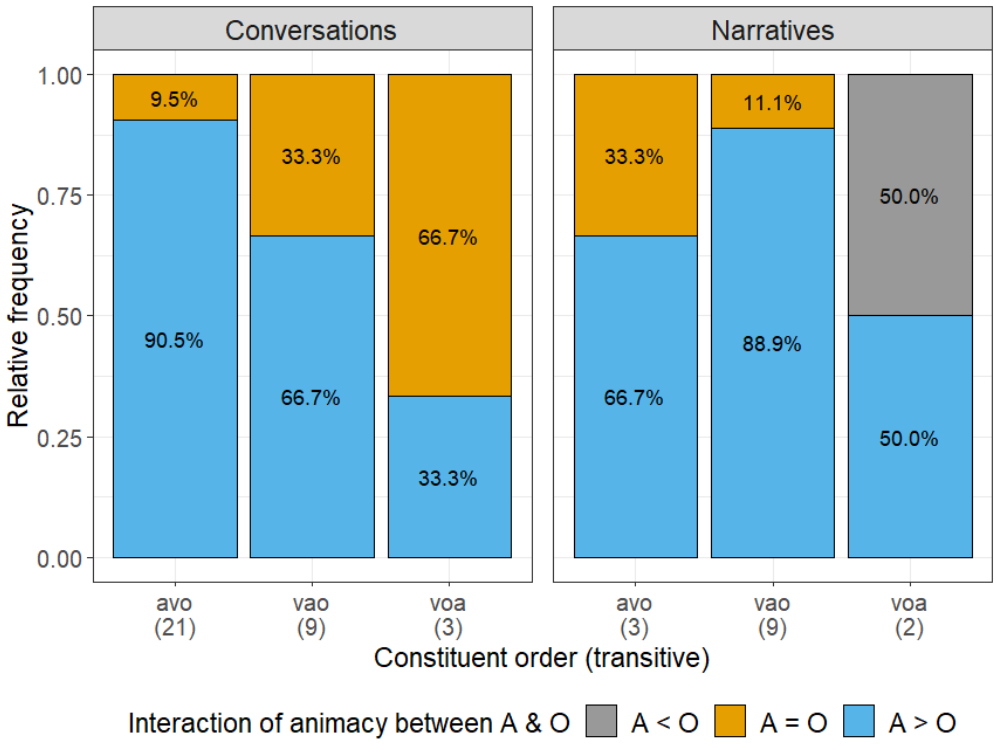

Figure 5.11: Interaction of animacy of A and O with constituent order in narratives and conversations with only fully realized arguments.

or equal animacy status, it seems to appear before A, though this is not always the case. These findings should further be seen with caution as most of these orders have very few occurrences.

Overall, the results suggest an effect of animacy on constituent order. When S and A occur preverbally, they tend to refer to animate, i.e. 1st and 2nd person, human, as well as animate referents. This aligns with previous findings in the literature on the impact of animacy for constituent order. As discussed in Section 2.2.3, animate referents tend to be placed first as they are processed more easily in discourse (Bock & Warren 1985; Branigan & Feleki 1999; Branigan et al. 2008; Tanaka et al. 2011). Inanimate S and A arguments in the Datooga data tend to appear postverbally, although the effect is stronger with A and for S only found within the narratives. For O, the findings suggest that the higher the referent is placed on the animacy hierarchy, the more likely it is for O to precede A. However, the overall numbers are low and these findings should therefore be treated with caution.

5.3.2 Argument realization

The typological survey in Section 3 showed that argument realization also plays a role in constituent order variation, as was illustrated for the languages Turkana, Toposa, and Teso. For instance, the argument realization of postverbal arguments conditions their order in Toposa in that pronominal O precedes A if the latter is a noun. In Datooga, the influence of argument realization is described below. However, because Datooga allows for zero realization of arguments, it is necessary to first determine the number of clauses with fully realized arguments.

Table 5.11 displays the frequency of fully realized and zero-realized, i.e. not overtly expressed, arguments in narratives and conversations. In both genres, O is more frequently fully realized than S and A. S is more fully realized in narratives (30.18%) than in narratives (22.86%), whereas A and O are more frequently non-zero in conversations than narratives. In narratives, the frequency of fully realized A arguments is notably low when compared to the other two argument roles. These results support the typological finding that zero-realized arguments occur more frequently than fully realized arguments (Du Bois 1987; Haig & Schnell 2016). However, the number of zero-realized arguments in Datooga appears to be notably large, particularly for A in narratives.

The following analysis covers constituent orders where one argument is zero realized if it is not the argument in discussion. For example, AV and VA will be included in the debate on the position of A arguments in relation to the verb but not in the discussion of O since this would be the zero realized argument. As previously stated, only constituent orders where all arguments are zero realized, i.e. V for intransitive and transitive clauses, are omitted.

	Total		Narratives		Conversations	
	N	%	N	%	N	%
S						
Non-zero	155	26.27	83	30.18	72	22.86
Zero	435	73.73	192	69.82	243	77.14
A						
Non-zero	72	23.23	25	18.12	47	27.33
Zero	238	76.77	113	81.88	125	72.67
O						
Non-zero	203	65.48	89	64.49	114	66.28
Zero	107	34.52	49	35.51	58	33.72

Table 5.11: Overview of fully vs. zero realized arguments.

Figure 5.12 shows the realization of non-zero S without distinguishing between conversations and narratives.

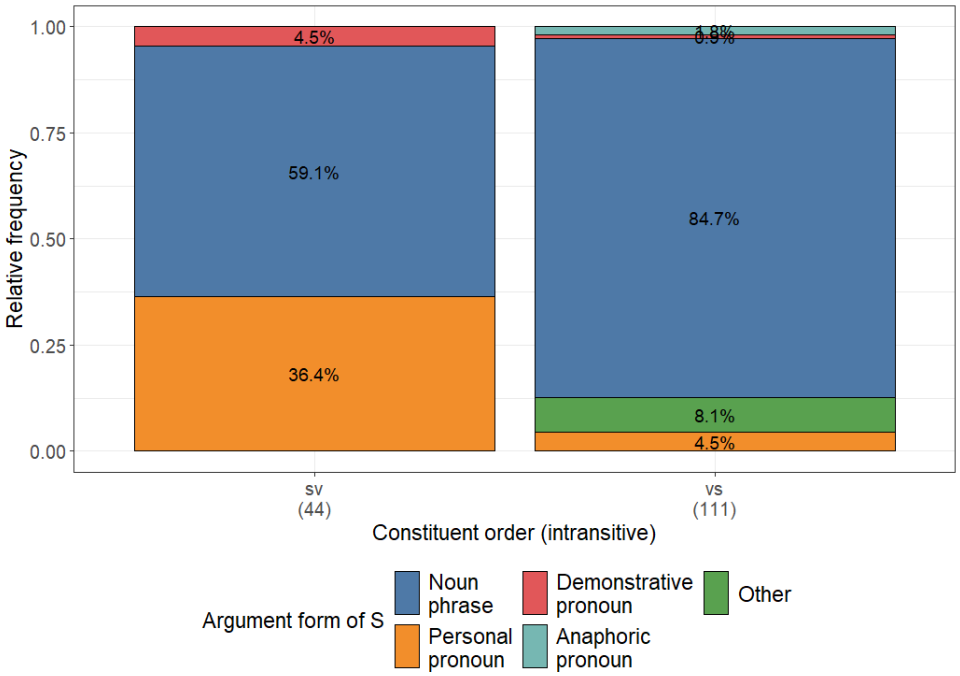

Figure 5.12: Argument form of S.

In addition to the previous figure, Figure 5.13 shows how this is distributed between the two genres. In conversations, S exhibits greater variance in terms of argument form in the data since it may also take the form of a demonstrative pronoun, anaphoric pronoun, and other realizations, such as numerals. Personal pronouns, including first and second person pronouns, as well as third

person pronouns, are more frequently used in preverbal than postverbal positions in both genres. Preverbal pronouns are especially common in conversations, with 44.8% compared to 4.7% in postverbal position. Demonstrative pronouns are uncommon for S (6.9%) and exclusively used preverbally in discussions.

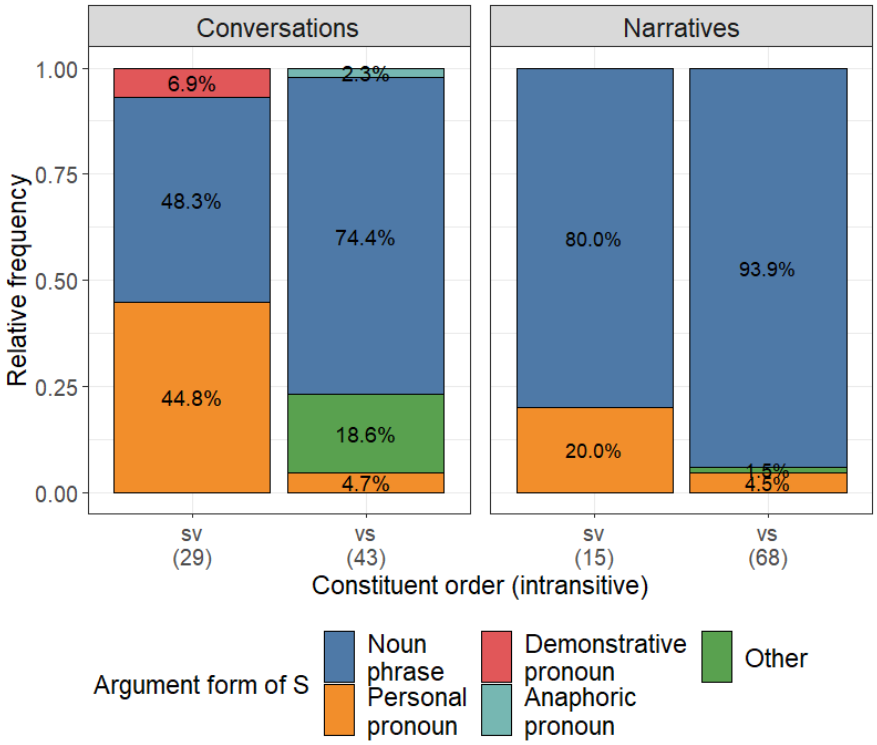

Figure 5.13: Argument form of S in conversations and narratives.

Given the nature of these two genres, it is not unexpected that conversations use more developed personal pronouns than narratives. Conversations consist of multiple people interacting with each other, and personal pronouns are required for directly addressing others, referring to oneself, and sustaining the flow of conversation. Narratives, on the other hand, frequently rely on describing events, people, and locations rather than direct contact between the narrator and the listener. Third person pronouns are more common in this context, although 3rd person referents are frequently zero realized in the data. However, overall personal pronouns show a preference for the preverbal position in the data observed which is prevalent in both genres.

The overall distribution of argument realizations for A in the constituent orders with overt arguments is given in Figure 5.14.

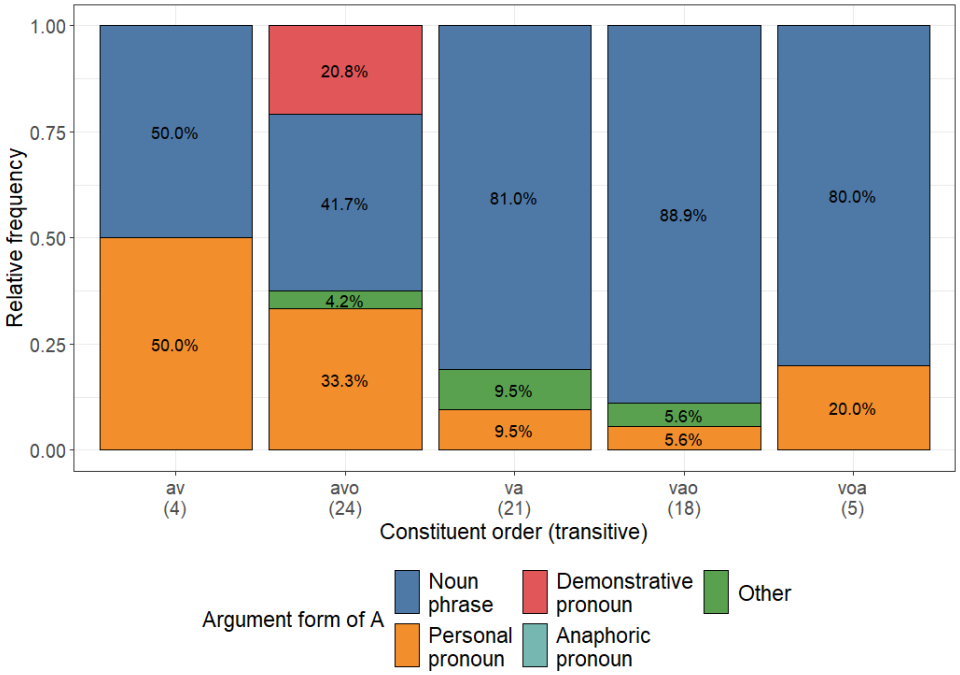

Figure 5.14: Argument form of A.

Additionally, Figure 5.15 presents the argument realization with a distinction of the genres. The variation of the argument form resembles the preceding figure of argument forms for S (see Figure 5.13), with demonstrative pronouns and other realizations only occurring in conversations. Personal pronouns function similarly to S arguments in that they are more prevalent in preverbal than postverbal positions. In narratives, personal pronouns are even reserved for the AVO. Demonstrative pronouns only appear in preverbal AVO order (23.8%). In postverbal position, the vast majority of A arguments are NPs. All in all, A and S behave similarly in terms of argument form variation, with demonstrative pronouns occurring only preverbally and a tendency for preverbal personal pronouns.

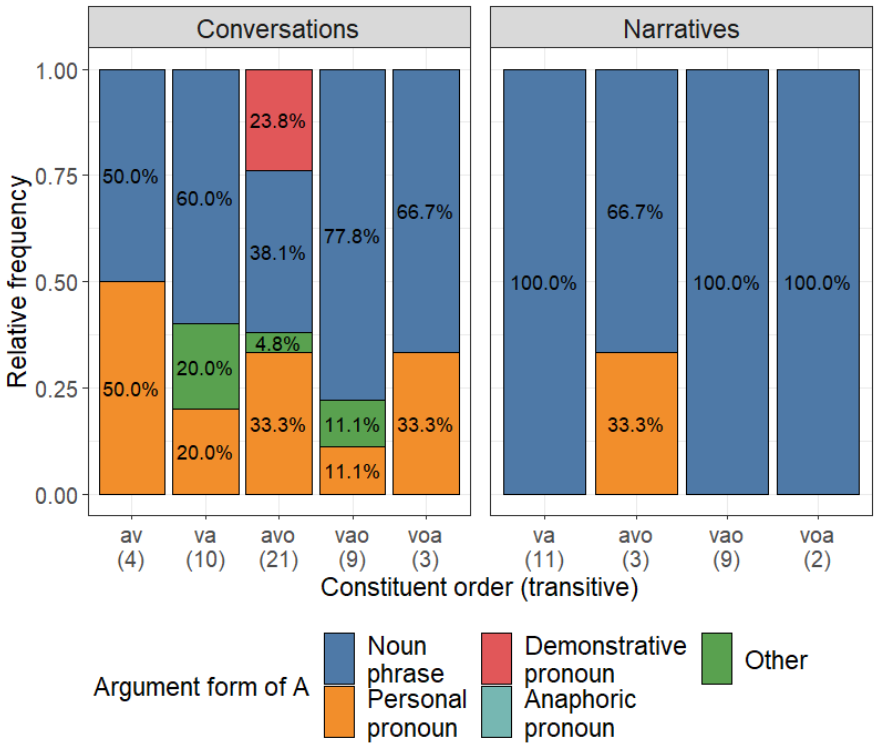

Figure 5.15: Argument form of A in conversations and narratives.

The argument form of O in clauses with overt arguments is shown in Figure 5.16 which does not differentiate between the genres.

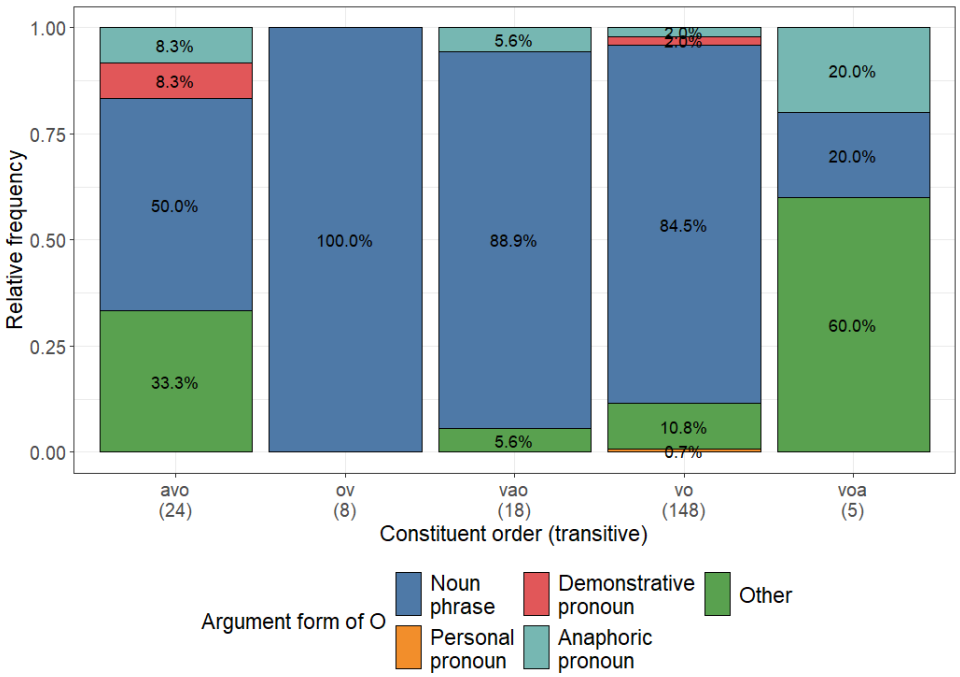

Figure 5.16: Argument form of O.

Figure 5.17 shows the argument realization of O arguments in transitive clauses. Compared to S and A, O is only rarely expressed by a personal pronoun. In fact, O in the form of a personal pronoun only appears in VO order in narratives with 1.4%. In terms of demonstrative pronouns, O is only seen postverbally in that form in VO order in both genres and AVO order in conversations. O is also found with other realizations apart from NPs and pronouns, such as numerals. In the only order with a preverbal O, OV, O is only found to be realized as an NP.

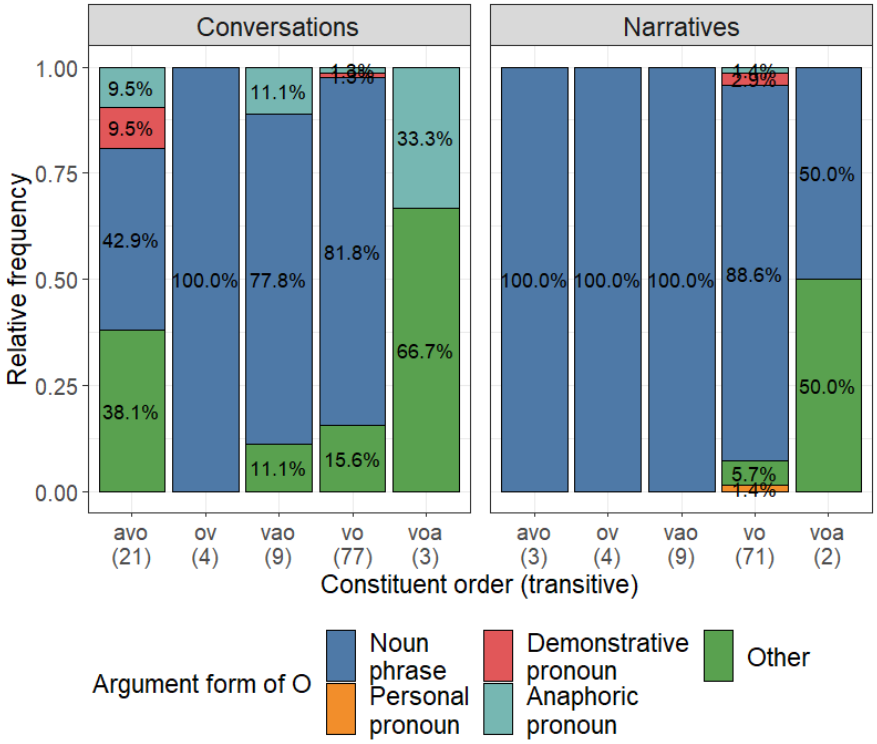

Figure 5.17: Argument form of O in conversations and narratives.

From a typological perspective, it is reasonable that O is mostly realized as an NP than as a pronoun. Concerning the "preferred argument structure" as proposed by Du Bois (1987), two constraints go together here. The first constraint, the "One Lexical Argument Constraint" represent the tendency of languages to avoid more than one lexical argument per clause. Following the second constraint, the "Non-lexical A constraint" it is not the A argument in transitive clauses that follows the first constraint. It is rather more common for O to be lexically realized with a full NP (Du Bois 1987: 822–826). These findings suggest that personal pronouns are much more common in subject than in object function.

To summarize, the findings on the condition of argument realization revealed that the form has an influence on S and A arguments, as both, when realized as personal pronouns, tend to be placed preverbally, but S and A arguments with the form of NPs are more prevalent postverbally. This trend is not seen with O arguments. Whereas demonstrative pronouns only appear preverbally for S and A arguments, demonstrative pronouns that serve as O appear

postverbally. However, while the results demonstrated that the realization of an argument may reveal preferences for some orders over others, the argument realization does not obligatorily lead to a change.

5.3.3 Negation

Another factor influencing constituent order in other Nilo-Saharan languages is the polarity of a clause. This phenomenon is observed in the Nilotic languages Teso and Jur-Luwo, as discussed in Section 3.2.4.2. In Teso, for example, the dominant constituent order VAO changes to AVO in negative clauses.

Table 5.12 illustrates the distribution of affirmative and negative clauses across different genres in the data. Overall, negative clauses are less frequent than affirmative ones. In all three groups—'preverbal S/A', 'postverbal S/A', and 'no overt S/A'—negative clauses appear more frequently in conversations. Specifically, there are seven negative clauses compared to 417 affirmative clauses in conversations, and two negative clauses compared to 16 affirmative clauses in narratives.

	Conversations				**Narratives**			
Constituent order	**Aff**		**Neg**		**Aff**		**Neg**	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Preverbal S/A								
SV	25	53.19%	4	57.14%	13	81.25%	2	100%
AV	2	4.26%	2	28.57%	0	0%	0	0%
AVO	20	42.55%	1	14.29%	3	18.75%	0	0%
Total	**47**	**100%**	**7**	**100%**	**16**	**100%**	**2**	**100%**
Postverbal S/A								
VS	41	69.49%	2	33.33%	65	76.47%	3	60%
VA	8	13.56%	2	33.33%	9	10.59%	2	40%
VAO	7	11.86%	2	33.33%	9	10.59%	0	0%
VOA	3	5.08%	0	0%	2	2.35%	0	0%
Total	**59**	**100%**	**6**	**100%**	**85**	**100%**	**5**	**100%**
No overt S/A								
V (intr.)	205	65.92%	38	66.67%	189	64.07%	3	30%
V (tr.)	36	11.58%	8	14.04%	36	12.20%	2	20%
VO	66	21.22%	11	19.30%	68	23.05%	3	30%
OV	4	1.29%	0	0%	2	0.68%	2	20%
Total	**311**	**100%**	**57**	**100%**	**295**	**100%**	**10**	**100%**
Total of all groups	**417**	**100%**	**70**	**100%**	**396**	**100%**	**17**	**100%**

Table 5.12: Distribution of affirmative and negative clauses between narratives and conversations.

These numbers are further illustrated in Figures 5.18 and 5.19, which show the distribution across the three categories, excluding orders without any overt arguments, i.e. V. Figure 5.18 shows the distribution of negation in intransitive clauses, revealing that both SV and VS orders occur in negated clauses, though it is slightly more common in the SV order in both genres, i.e. 13.8% vs. 4.7% in conversations and 13.3% vs. 4.4% in narratives, respectively. However, it is important to note that the preverbal order overall is less frequent than the postverbal order.

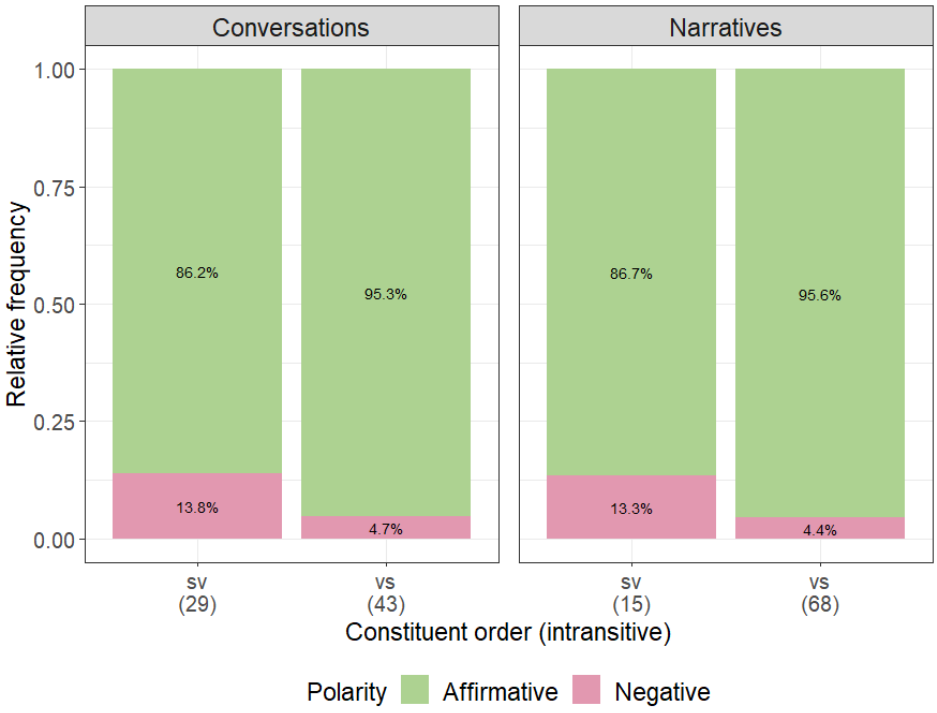

Figure 5.18: Impact of polarity on constituent order in intransitive clauses.

In terms of transitive clauses, as illustrated in Figure 5.19, negation appears in all orders except for VOA in both genres. Negative clauses with OV order are found in narratives but not in conversations, whereas negated clauses with AVO order occur in conversations but not in narratives. In conversations, the AV order is most frequent for negative clauses, accounting for 50%, while in narratives, the OV order is most frequent, also with 50%. However, both of these orders just have four occurrences in total. Thus, the results should be seen with caution.

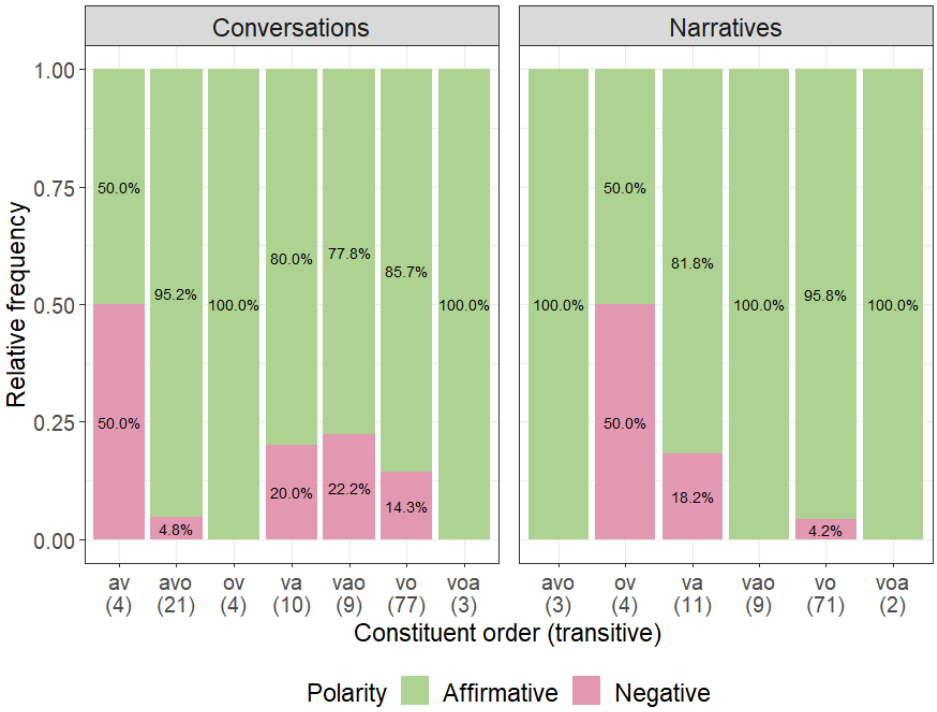

Figure 5.19: Impact of polarity on constituent order in transitive clauses.

The results from this figure support the findings for the intransitive clauses. Negation is found with preverbal A, postverbal A, as well as in clauses without an overt A argument. This suggests that there is no clear constituent order preference within negative transitive clauses. Overall, the frequency of negative clauses is low compared to affirmative clauses. To better observe whether there is a preferred constituent order for negation, a study of a corpus with a higher frequency of negated clauses would be necessary. Nevertheless, as shown by the figures and table, it can still be concluded that negation is not restricted to one order, unlike in other Nilotic languages with this condition.

5.3.4 Syntactic weight

Section 2.2.3 showed that the weight of an argument can affect its position concerning the verb and other arguments, with heavier arguments appearing near the end of the phrase. Figure 5.20 shows the syllable length of S in SV and VS order. It can be observed that S mostly occurs in postverbal position, i.e. in VS order, which was also evident in Table 5.8. The medians of the number of syllables of S in preverbal and in postverbal position are at the same level. However, S in postverbal position shows a greater distribution in terms of syllable length as it ranges from one to eight syllables whereas preverbal S ranges between one and four syllables. Thus, the data suggests that when S has more than five syllables, it tends to occur after the verb, though there are only a few occurrences of S with more than five syllables.

Figure 5.20: Syllable lengths of S in different constituent orders.

Figure 5.21 shows the syllable length of A in the fully realized clauses of the data, namely AVO, VAO and VOA. The fewest data points can be observed for VOA with only five occurrences within the data. In comparison to the previous figure for the fully realized intransitive clauses, a slight tendency in terms of the syllable length of A and the constituent order can be observed: With the preverbal A (AVO) the median of the syllable length is at 2.5, at three for VAO order when A immediately follows the verb and at four in VOA order. This suggests that longer A arguments tend to be placed last in the clause. However, these results should still be interpreted with caution given the uneven distribution of data points within the three groups.

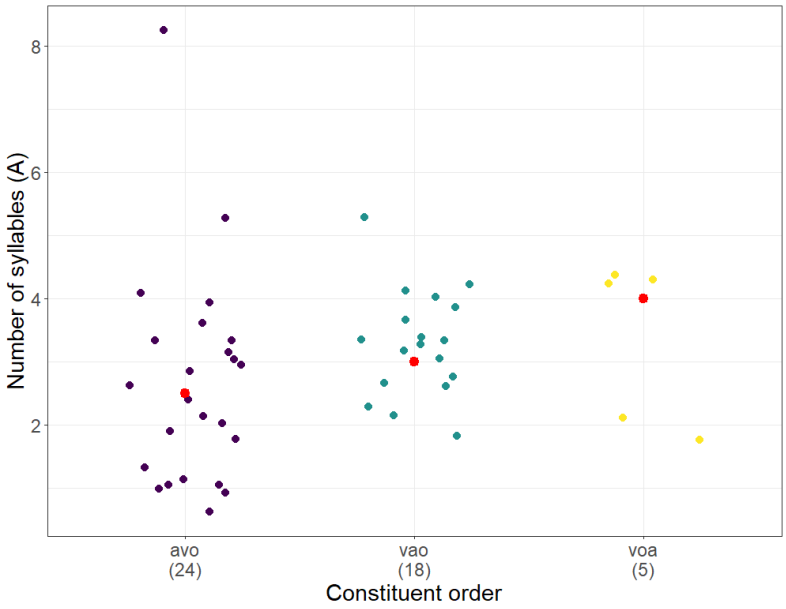

Figure 5.21: Syllable lengths of A in different constituent orders.

The distribution of syllable lengths of O within the fully realized transitive clauses of the data is shown in Figure 5.22. In both AVO and VAO order O is placed towards the end. O in VAO order has the highest median of all three groups with 2.5 and also shows the greatest distribution with one to six syllables. By contrast, O in AVO order has the same median as in VOA, where O directly follows the verb and is not the last element. However, O in AVO order shows a greater range in syllable length than in VOA. Thus, O also tends to occur towards the end when it is a heavy constituent.

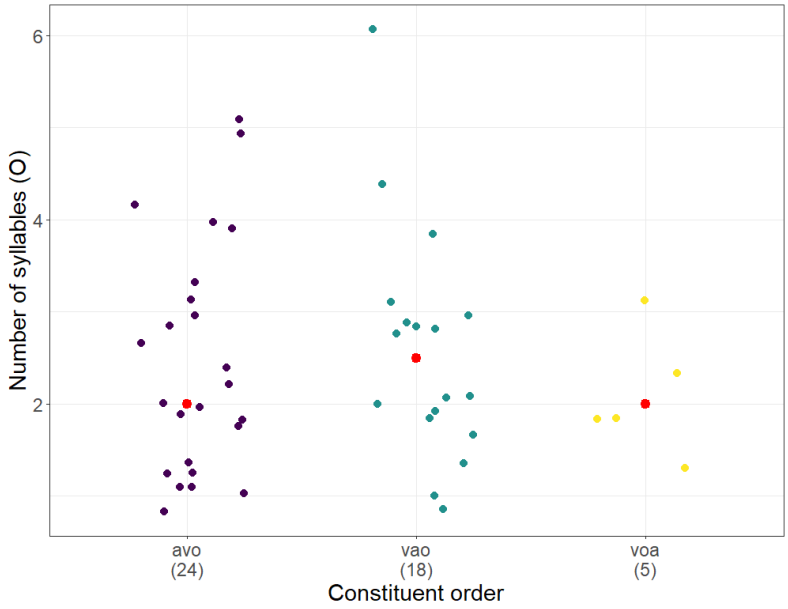

Figure 5.22: Syllable lengths of O in different constituent orders.

Figure 5.23 shows the interaction of syllable lengths of A and O in fully realized transitive clauses. In AVO and VAO order all three possible interactions in terms of syllable length occur. In AVO order, O being longer than A occurs slightly more often than A being longer. The majority of clauses with VAO order have an A argument that is longer than O, followed by both arguments sharing a length, followed by O being longer than A. In the five clauses with VOA order, three of them have a longer A than O while the other two have both arguments with the same length. In VOA order clauses within the data, there is no clause where O is longer than A. Overall, this figure shows that although the longest argument may be the last, as it is the case for some clauses within all three orders, this is not always the case. In fact, there is also a greater amount of clauses where the longest argument is first in AVO clauses.

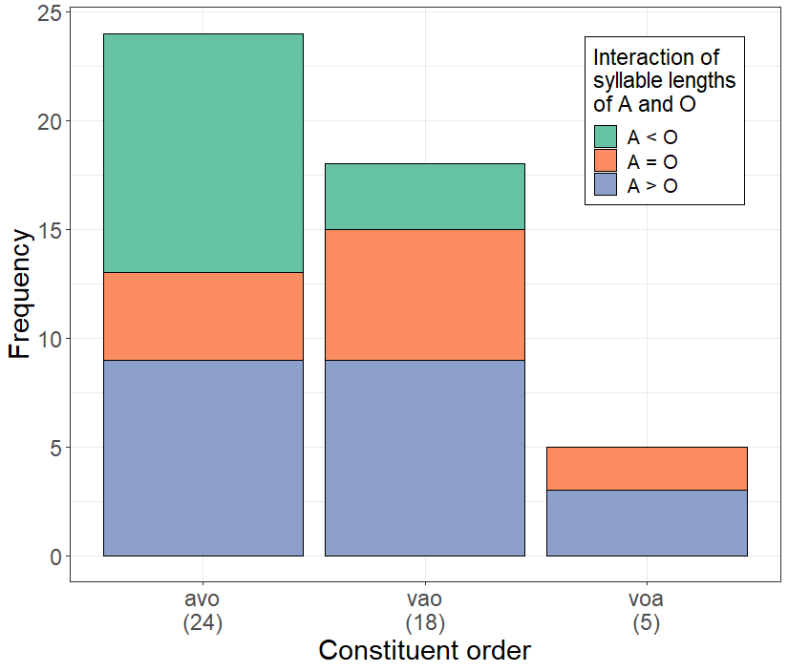

Figure 5.23: Interaction of syllable lengths of A and O in different constituent orders.

These findings lead to the conclusion that the syllable length of an argument may result in a position near the end of a sentence. This is especially prevalent with S arguments although the amount of longer S arguments is low. However, as the findings for both the intransitive and transitive clauses demonstrate, heavier elements are not required to be placed toward the end. This implies that syllable length is not an important condition for constituent order variation, if at all.

5.4 Evaluation of the stimuli data

This section shows the results of the evaluation of the video-stimuli data. As explained in Section 5.1.2, each of the 18 stimuli were shown six different speakers, resulting in six different elicitation sets of which three have been annotated so far. One of the 18 stimuli was excluded from this analysis, precisely stimulus number five, as it shows an intransitive setting though the analysis focuses on transitive settings.

Our aim with the stimuli-based elicitation was to investigate whether certain discourse properties have an influence on constituent order in Datooga. In particular, we were interested in whether animacy, givenness, and/or the identifiability of referents in A and/or O function would lead to a constituent order that deviates from the basic verb-initial order. Although the videos (excluding stimulus five) show transitive scenarios, speakers typically described them in more than one sentence, also including different clause types, such as non-verbal clauses or imperative constructions, as well as intransitive clauses.

Before discussing these properties in more detail, Table 5.13 provides an overview on the distribution of clause types within the three elicitation sets that have been annotated so far. As can be seen, the majority of clauses in the elicited data are declarative main clauses (68.85%), followed by non-verbal (16.94%) and subordinate clauses (41%). Impersonal constructions are distinguished from declarative main clauses, as they do not have a referent in agent function. The following analysis focuses on the 252 declarative main clauses and further includes adverbial and complement clauses, resulting in a total of 262 clauses.

	Frequency	**%**
Declarative main clause	252	68.85
Subordinate	41	11.20
Non-verbal	62	16.94
Question	1	0.27
Imperative	4	1.09
Impersonal	6	1.64
Total	**366**	**100**

Table 5.13: Clause types in the stimuli data.

Table 5.14 shows the distribution of animacy combinations of agent and patient, as well as the resulting constituent orders and the number of stimuli with the respective conditions. As mentioned above, speakers used to describe the videos in more than one sentence and they also use intransitive clauses. They are therefore included in the analysis.

In the conditions where the agent has a lower animacy status or is on the same level as the patient, the majority of clauses are intransitive with a zero-realized S argument. When the agent has a higher animacy status than the patient, i.e. human agent vs. animate or inanimate patient, VO is more frequent or as frequent as V (intransitive). With an exception of VO which occurs with every condition, transitive orders, i.e. AV, AVO, VAO, VOA, V, are further exclusive to the constructions in which the agent has a higher animacy status. Concerning the preverbal arguments, only SV, AV, and AVO are found in the elicited data, though their frequency is low.

	Hum agent Hum patient		**Hum agent An patient**		**Hum agent Inan patient**		**Inan agent Hum patient**	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**
SV	7	16.28%	2	8.33%	2	1.19%	1	3.7%
AV	0	0%	0	0%	1	0.6%	0	0%
AVO	0	0%	0	0%	3	1.79%	0	0%
VS	3	6.98%	5	20.83%	10	5.95%	4	14.81%
VAO	0	0%	0	0%	1	0.6%	0	0%
VOA	0	0%	1	4.17%	0	0%	0	0%
V (intr.)	23	53.49%	5	20.83%	65	38.69%	15	55.56%
V (tr.)	0	0%	3	12.5%	21	12.5%	0	0%
VO	10	23.69%	8	33.33%	65	38.69%	7	25.93%
Total	**43**	**100%**	**24**	**100%**	**168**	**100%**	**27**	**100%**
No. of stimuli with condition	2		2		11		2	

Table 5.14: Animacy of agent and patient combinations and the resulting constituent orders in the stimuli data.

However, what the table also shows is the number of stimuli with each condition. Thus, the majority of stimuli had a human agent acting on an inanimate patient, resulting in a significantly higher number of analyzable clauses compared to the other conditions. Given the circumstances, it was also not possible to create stimuli with animate non-human agents. For future research, it would be advisable to consider more variety in terms of animacy.

The next parameter we altered in the stimuli was the order of appearance of agent and patient. In particular, we created five stimuli which showed the agent first, six were the patient occurred first, and another six were both were in frame at the same time. The results are given in Table 5.15. As can be deduced from the table, the stimuli where the agent appeared first and where both were in frame mostly resulted in intransitive constructions without an overt S argument. When the patient was seen first, the majority of clauses had a VO order. Preverbal arguments are rare in all three conditions with SV occurring the most in the stimuli where both referents were seen at the same time, though the total numbers are low.

	Agent first		Patient first		Both in frame	
	N	**%**	**N**	**%**	**N**	**%**
SV	1	1.28%	2	2.08%	9	10.23%
AV	0	0%	1	1.04%	0	0%
AVO	0	0%	1	1.04%	2	2.27%
VS	8	10.26%	8	8.33%	6	6.82%
VAO	0	0%	0	0%	1	1.14%
VOA	0	0%	1	1.04%	0	0%
V (intr.)	41	52.56%	30	31.25%	37	42.05%
V (tr.)	5	6.41%	13	13.54%	6	6.82%
VO	23	29.49%	40	41.67%	27	30.68%
Total	**78**	**100%**	**96**	**100%**	**88**	**100%**
No. of stimuli with condition	5		6		6	

Table 5.15: Order of appearance of agent and patient and the resulting constituent orders in the stimuli data.

Our original goal was to determine if givenness affects constituent order variation in Datooga under these settings. However, we designed the stimuli with just two people visible in the videos, and all of the stimuli were presented throughout the elicitation. Thus, they could only be identified as new referents when the speakers viewed the initial stimulus, but for the following stimuli they were already given referents as soon as the speakers noticed them. Therefore, future studies should include videos with more human referents.

For the six stimuli in which the patient was presented first, we further altered whether the agent was fully visible or only its type, i.e. in this case human. Table 5.16 shows the results. For both conditions, VO order is the most frequent, followed by intransitive clauses with a zero-realized S argument. As it was the case in the discussions of the previous tables, preverbal arguments are infrequent in both conditions. There are slightly more preverbal arguments when only the type of the agent could be identified. However, given the low numbers, one can hardly speak of a pattern here.

	Token-identifiable agent		Type-identifiable agent	
	N	**%**	**N**	**%**
SV	0	0%	2	5.13%
AV	0	0%	1	2.56%
AVO	1	1.89%	0	0%
VS	2	3.77%	6	15.38%
VAO	0	0%	0	0%
VOA	0	0%	1	2.56%
V (intr.)	16	30.19%	12	30.77%
V (tr.)	8	15.09%	3	7.69%
VO	26	49.06%	14	35.90%
Total	**53**	**100%**	**39**	**100%**
No. of stimuli with condition	3		3	

Table 5.16: Identifiability of the agent and its resulting constituent orders in the stimuli data.

In summary, the evaluation of the stimuli suggests that the tested conditions have little to no impact on constituent order in Datooga. However, some stimuli, particularly those affecting the animacy of referents, are not fully balanced, which could influence the results. Additionally, it was challenging to assess the impact of givenness on constituent order, as the stimuli involved only two human referents consistently recognized by the speakers throughout the elicitation process. It would also be interesting to see in future studies whether the identifiability of the agent would have an impact on constituent order variation regardless of the givenness of agent and patient. So far, we have only tested it with the patient being presented first. However, presenting a type-identifiable agent and the patient at the same time would be challenging with the video format. To achieve more reliable conclusions, future studies should consider using more balanced data and potentially a larger dataset.

5.5 Discourse pragmatics and constituent order variation

As demonstrated in the previous sections, semantic and syntactic characteristics contribute to constituent order variation to some degree. However, these factors alone cannot account for the full range of variation observed in the dataset.

The following sections explore the discourse-pragmatic features that may influence constituent order variation in Datooga. While Section 5.5.1 adopts a quantitative approach based on a subset of the data, the subsequent analysis in Section 5.5.2 shifts to a qualitative, discourse-oriented perspective.

Firstly, Section 5.5.1 examines the concept of givenness. This analysis is based on a subset of the data and were conducted using INCEpTION (Klie et al. 2018) and R, as detailed in Section 5.1.3. Section 5.5.2 offers a qualitative analysis of the information structural categories of topic, focus, and contrast within the fully realized clauses from the entire analyzed dataset, i.e. the same subset that was used in Section 5.3.4. It additionally includes the stimuli-based texts.

5.5.1 Givenness

Figure 5.24 shows the findings of the givenness study. This analysis was carried out on a subset of data, including the three stimuli-based texts.

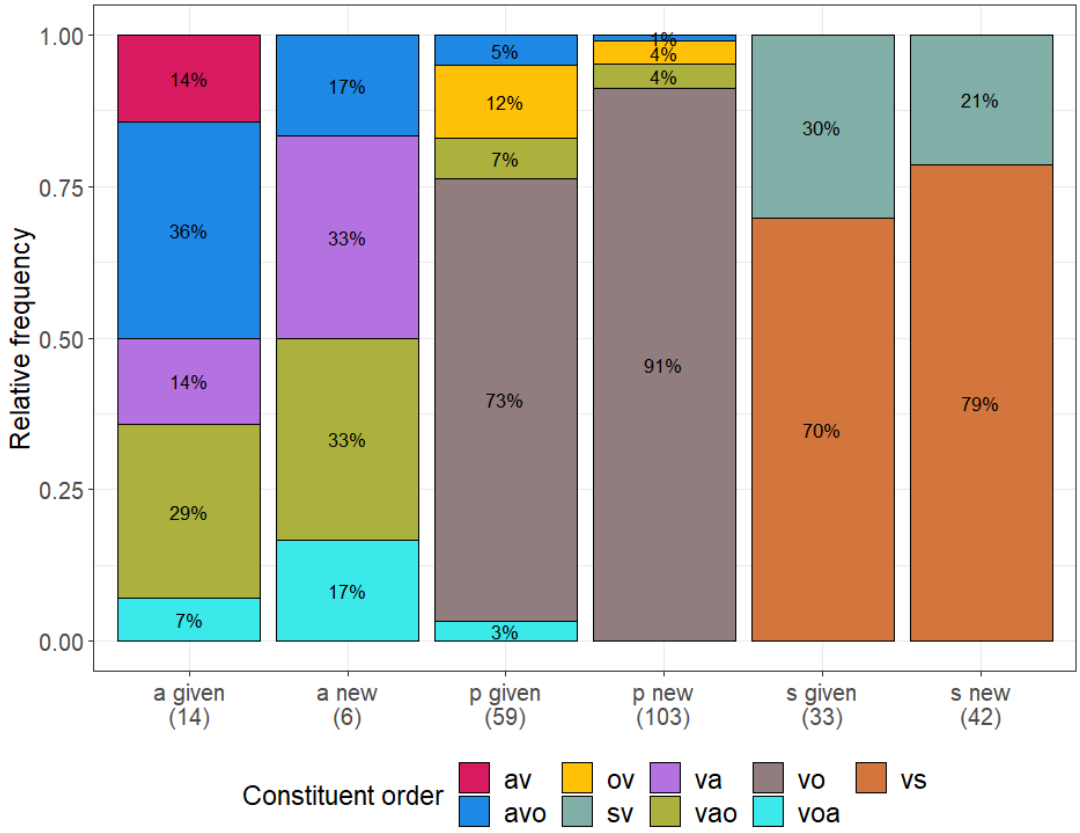

Figure 5.24: Givenness of the argument roles and the constituent order they occur in.

The figure indicates the distribution of constituent orders for each givenness status of each argument role (S, A, and O). For example, one bar represents the argument and its givenness status, such as "A given", as well as the proportion of the various constituent orders where a given A argument appears. The figure displays the total number of argument roles and their givenness status. For example, there are 14 given A arguments in the studied data.

Overall, O, out of the three argument roles, is the most often fully realized, supporting the findings described in Section 5.2 and the typological viewpoint on the frequency of fully realized O arguments relative to the other roles. Interestingly, A is the sole argument role that appears more frequently as given than as new, implying that referents are more often introduced to the

discourse in O or S function. This aligns with findings in other languages on referent introduction (Du Bois 1987; Haig & Schnell 2016).

Furthermore, this image highlights the effect of givenness on the constituent order. Its influence is particularly noticeable when comparing the given and new A. When A is new, it is mainly found in postverbal position, i.e., in VA, VAO, or VOA order, which adds up to a total of 83% of all occurrences of a new A referent. In contrast, a preverbal position for a new A is only found in 17% of the cases. Yet, if A is given, the number of occurrences where A is in a preverbal position, i.e. in AVO or AV order, increases to 50%. This tendency is also observed with the other two argument roles, O and S, albeit it is not as strongly as with A.

5.5.2 Topic, focus, and contrast

As shown in Section 5.2, clauses with preverbal arguments are relatively rare compared to those with verb-initial structures, and, in the case of transitive clauses, even rarer when all arguments are fully realized. This suggests that preverbal arguments may serve specific communicative and discourse-related functions.

In contrast to the preceding quantitative analyses, the following section adopts a qualitative approach. It examines individual clauses in their discourse context in order to analyze the information-structural categories of topic (topicalization), focus, and contrast, as outlined in Section 2.2.3.2. The analysis is based on fully realized clauses and includes both corpus data and stimuli-based texts.

Datooga is a verb-initial language, so the sentence-initial or preverbal position is not inherently reserved for the topic as it is in some other languages. However, topics can appear in this position. Example (59) shows a context where the topic occurs preverbally. This extract comes from a description of rituals for a newborn baby. In the first clause in Example (59a), the topic is *qeamata* 'mother' in the sense of aboutness, i.e. the utterance is about her. The noun *qeamata* is given, as she has been introduced before and is picked up again as a topic in this clause with 17 clauses since her last mention. Prior to this extent, mostly impersonal constructions occurred, which was also found in the context of other clauses in the data with postverbal topics. After her reintroduction to the discourse, she is non-overtly expressed as can be seen in Example (59b).

(59) a. *gageanu* ***qeamata*** *deareeda*
g-a-gean-u qeamad-da dear-ee-da
AFF-3-get-CP mother-UR ash-PS-UR
'The mother gets some ash.'

b. *qawada* *qeamamda* *jeefta*
g-a-wad-a qeamam-da jeep-da
AFF-3-rub-IS face-UR child-UR
'She rubs it onto the child's face.'

[2017-11-22_Udabuunga3_058–059]

Nonetheless, topics can also occur preverbally under certain circumstances. Following Kießling (2007: 171) the topicalization of arguments, i.e. their fronting to a preverbal position, is due to "pragmatic purposes". He further notes that "preverbal phrases are used in citation, address and fronted topics", where the phrase occurs in the absolute case (Kießling 2007: 171). The following extract (60) is from one of the conversational texts of the data. Prior to the extract in the following example, the speaker talks about the attack on a group of people by others. The speaker specifically mentions two woman, a child, and two men. All of these have been introduced in the discourse preceding the extract below. In the first clause in Example (60a), *gatmooda sifu* 'woman who had been beaten' is a dislocated topic. It is the same referent as *diida* 'this one' in Example (60b) which is also dislocated. The following clause in Example (60d) provides more details about *gatmooda*. However, in this case the argument is only referenced on the verb with the prefix *-a*. The clause in Example (60e) follows after a short pause and has a change in topic as it is about one of the men. In the last Example (60f) following the short discourse about the man, not shown in the examples, the other one of the two women becomes the topic.

(60) a. *ah* ***gatmooda*** *sifu*
ah gadeem-oo-da si-fu
INTERJ woman-PS-UR IMPRS.PRF-beat
'No, that woman who'd been beaten.'

b. ***diida*** *sifu*
diida si-fu
DEM.DIST IMPRS.PRF-beat
'The one they'd beaten.'

c. *ah*
ah
INTERJ
'Ah.'

d. *qam*
g-**a**-mi
AFF-3-die
'She died.'

e. *ng'eedeeda qafaj há ng'eedeeda muuchi*
ng'eed-ee-da g-a-faj ha ng'eed-ee-da muuchi
man-PS-UR AFF-3-run DSC man-PS-UR very
'The man really ran.'

f. ***gatmoodiideanji*** *qahabaas ng'eanyi*
gadeem-oo-da-ida-eani g-a-habaas n'geany
woman-PS-UR-DEM.DIST-DEM.PROX AFF-3-lie.low ground-PS
'That woman laid flat.'

[2014-09-13_YonasHulanda1_261–268]

The preverbal position of *gatmooda* in the last clause within this discourse extract puts some emphasis on the argument which seems to be related to the change in topic. Both of the other topics preceding *gatmoodiideanji* 'woman' (second woman) in Example (60f), i.e. *gatmooda* 'woman' (first woman) in Example (60a) and *ng'eedeeda* 'man' in Example (60e) are reintroduced as dislocated topics and not as arguments of the verb. Following Kießling (2007: 171), NPs that have a demonstrative suffix, such as *gatmoodiideanji* in Example (60f), are preferred in preverbal position. However, the use of the demonstrative suffix here seems odd given that the previous clauses were about the man and the other woman. Thus, the demonstrative suffix might add additional emphasis on *gatmoodiideanji*. As mentioned above, this might be related to a change in topic which also occurred in the previous example with a postverbal topic in Example (59). In contrast to this example which was preceded by impersonal constructions, the clause in (60f) was preceded by expressed referents.

A similar example comes from the other conversational text within the sample. This text is also the text in the sample that has the highest frequency of clauses with preverbal arguments in fully realized clauses. The following extract in Example (61) shows not only a change of the topic but also a change of the speaker. The first two clauses in the Examples (61a) and (61b) are both uttered by the same person and they both share the same referent as the topic, namely *bareandida* 'group of travelers'. In the third clause of this example, (61c), a change in turn happens as a different speaker starts speaking. This further shows a change of speech act participant. Following Kießling (2007), speech act participants, i.e. first and second person pronouns, typically do not occur independently with overt forms. When they do appear, it is due to "emphasis or pragmatic prominence", whereby the latter leads to a fronting to the preverbal position (Kießling 2007: 160).

(61) a. *nea* *gwarukta* *bareandida*
nea g-wa-rug-d-a barean-d-i-da
CONJ AFF-3-tell-TERM-IS group.of.travellers-?-PS-UR
bea *badahuutkeaba* *súu*
bea badahuutkeaba suu
ASSOC ? DEM.PROX.PL
'And those people who were behind the others, they said.'

b. *súu* *duukseani*
suu duuks-aan-i
DEM.PROX.PL be.close.to-AM-IS
'Those ones who were close.'

c. ***ániini*** *qáahiti* *nea* *gwacha* *gabay* *ea* *nea*
aniini q-aa-hid-i nea gwacha g-a-bay ea nea
1SG AFF-1SG-come-IS CONJ recent AFF-3-do.already CONJ CONJ
'I arrived and [...]'

[2017-04-19_Maandazi_481–483]

Similar examples to this where a topic change is accompanied by a change of the speech act participant can be found throughout this conversational text. It is also found to some extent in the other conversational text of the previous example although it does not occur as often there. The higher frequency of this pattern in this specific conversational text might be related to the nature of both texts: In the first one, i.e. the one Example (60) stems from, multiple people are speaking but there is one main speaker who is telling a story and gets input from the other people throughout the discourse. However, in contrast to the other multi-party conversation of the data, i.e. the one illustrated in the previous example, there are not a lot of changes of speech act participants accompanied by topic changes. If the topic changes, it mostly happens to be a third person referent.

Another function of topics in the preverbal position seems to be the establishing of the context in storytelling. An illustration of this is provided in Example (62), taken from one of the narrative texts. This particular story is a folk tale about the encounter between a donkey and a hyena, which are able to talk to each other. The initial sentence of the story places *dìiyángà* 'animals' in the preverbal position.

(62) *qáyi gàráy* ***dìiyángà*** *gábàlóolí* *gáwùrjì* *búunèeda*
qayi garay diiy-an-ga g-a-balool-i g-a-urj-i buun-ee-da
PST PST animal-PS-PL AFF-3-talk-IS AFF-3-be.like-IS folk-PS-UR
'Once upon a time, animals talked like people.'

[2013-06-06_HM_digeeda_01]

The sentence sets the stage for the entire story. While *dìiyángà* is a new entity, it serves as the topic that the subsequent discourse will revolve around, namely a donkey and a hyena. In contrast, the focus is on the action *gábàlóolí gáwùrjì búunèeda* 'talked like people', which emphasizes a significant event, considering that animals typically cannot speak except in folk tales. This analysis aligns with how topics and focuses function in storytelling, where the topic sets up the context or characters and the focus highlights the notable events or actions (see Lambrecht 1994; Krifka 2008).

It is vital to remember that topics do not always occur preverbally in the Datooga data. Arguments in the topic function appear considerably more frequently in postverbal position, as seen by the proportion of clauses with preverbal arguments against those with postverbal arguments. If an argument is required preverbally when it serves as the topic, we would expect more preverbal arguments. As this is not the case in the study's data, simply being assigned the topic may not be sufficient to establish a preverbal position for an argument. The examples presented above have in common that the argument functioning as the topic precedes the verb. All of these, however, have additional functions within the discourse; for instance, in the first example, the topic is picked up again; in the second, the speaker switches along with the topic; and in the third, it appears as though the setting for the forthcoming story is established.

In addition to having topics in the preverbal position, arguments can also occupy this position when they serve a focus function. Unlike topics, these arguments typically introduce new

information. However, there is no information on non-contrastive focus constructions in the literature on Datooga.

The initial example in (63) is an excerpt from a multi-party conversation. Before this excerpt, the first speaker in Example (63a) inquires about a rope, asking if anyone has seen it. The subsequent utterance, shown in Example (63b), comes from a different speaker (a three-year-old child) who indicates that the rope is theirs. In the final Example (63c), another speaker remarks that Madootay believes the rope resembles the one they had the previous day. Here, both the proper noun *Madootay* and the third-person pronoun *ninyi* appear preverbally. This instance marks the first mention of *Madootay* in the discourse, thus introducing a new referent.

(63) a. *gwayii* *gaba* *eenu*
g-wa-yii g-a-bea eenu
AFF-3-say AFF-3-belong.to where
'Did you see where is it from?'

b. *ánì*
anii
1SG
'Me!'

c. ***Madootay*** ***ninyi*** *gwayeena* *gawurji*
Madootay ninyi g-wa-yee-n-a g-a-urj-i
Madootay 3SG AFF-3-say-CP-IS AFF-3-be.like-IS
dea *heeta* *gida* *Datoo*
dea heeta gida dat-oo
ANAPH.PRO yesterday thing Datooga-PS
'Madootay thinks it's like that one of yesterday.'

[2017-04-19_Maandazi_294–296]

The placement of the proper noun in the preverbal position, followed by the third person pronoun, brings the referent into the focus of the entire utterance. This example might illustrate reported speech, where the speaker aims to emphasize that the thought is not their own. The speaker may want to clarify that it is Madootay who holds this belief, which possibly contradicts their own views.

The following Example (64) represents another context in which an argument serving as the focus can precede the verb.

(64) a. *gwanda* *híjí* *eara* *ninyi*
g-a-nda hiji eara ninyi
AFF-3-be here even 3SG

b. ***ninyi*** *há* *goodeeyi* *buuneeda*
ninyi ha g-a-deey-i buun-ee-da
3SG DSC AFF-3-accompany-IS people-PS-UR

c. *mudugaba* *há* *gídêabá* *gooptead* *aba* *qwa*
m-udu-a-bea ha gideaba g-a-bidead aba qwa
NEG-CONT-3-belong.to DSC CONJ AFF-3-sit PREP house
'He was there too, even him, he accompanied the people, it wasn't that he stayed at home.'

[Samoota_058]

In this instance, *ninyi* in Example (64b) seems to be the focus. In contrast to the previous example, there is no rising pitch on the second *ninyi* that could indicate emphasis, as Figure (5.25) shows. However, there are other reasons supporting the interpretation of focus or emphasis on *ninyi*, such as the use of the discourse particle *há* following *ninyi*. Moreover, the emphasis on this particular individual is reinforced by the inclusion of *eara* 'even' in the preceding clause (64a) which also has a kind of counterexpectational quality in this context. That is, it seems like the speaker wants to point out that the presence of *ninyi* was unexpected here. The following clause (64c) puts additional emphasis on *ninyi* stating that he did not remain inactive by staying at home. This emphasis likely aims to highlight that it was specifically him who undertook the action of accompanying the people, potentially contrasting with other possible actors.

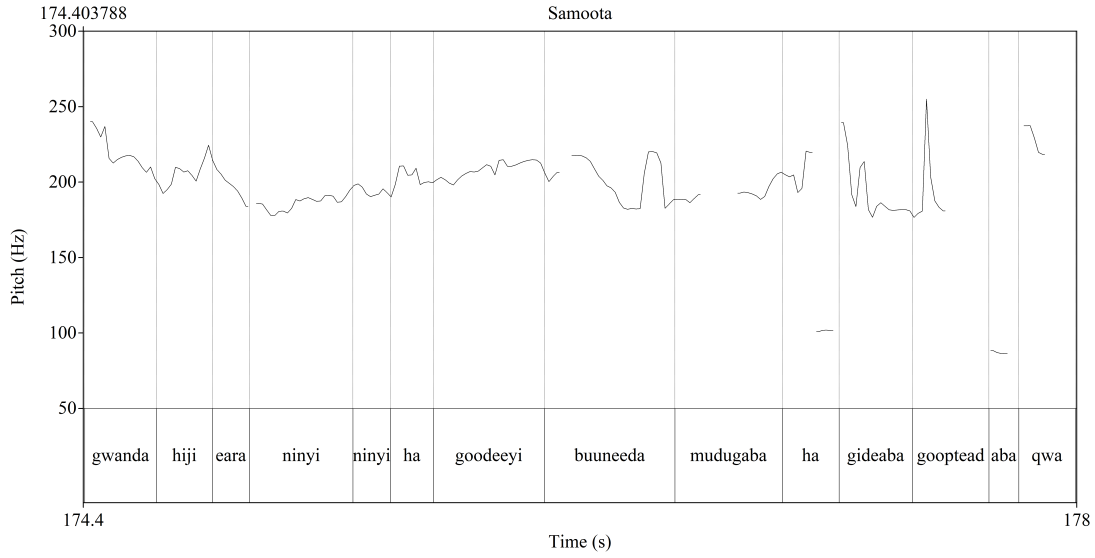

Figure 5.25: Pitch contour of Example (64).

Kießling (2007: 158) states that contrastive focus leads to the movement of an argument to the "immediate postverbal position". In the following example provided by Kießling (2007), this results in a tonal change of the A argument from *qáarèemáŋgá* 'youths' in Example (65a) to *qáarèemàŋgə̰* in Example (65b) as A is now in the final position. Following this, *bùŋéedà* 'grave' in Example (65b) is in contrastive focus, as indicated by the italicization, which has been added by me and is not present in (Kießling 2007). However, Kießling (2007) does not provide further context for these clauses. As a result, it remains unclear which entity *bùŋéedà* 'grave' is contrasting with, or what the specific properties of contrastive focus are in this context.

(65) a. *gày-qá-bár qáarèemáŋgá bùɲêedə̰*
FUT-3SBJ-dig youths.NOM grave.ABS
'The youths will dig the grave.'

b. *gày-qá-bár bùɲéedà qáarèemàŋgə̰*
FUT-3SBJ-dig grave.ABS youths.NOM
'The youths will dig *the grave*.'

(Kießling 2007: 158)

In the data observed in this study, there are also clauses with a contrasting entity in the preverbal position. The following Example (66) is taken from a narrative text about a woman who survived an attack by a group of people by hiding herself. The woman, *gatmooda*, was introduced to the story by means of a postverbal NP a few clauses prior to this extract. After being picked up again as the topic in the postverbal position in a clause preceding this extract, the argument is only referenced in the verb, as shown in Examples (66a) and (66b). In the subsequent direct speech, where she recounts what happened, the first-person pronoun appears preverbally.

(66) a. *qwahita*
q-wa-hit-a
AFF-3-come-IS
'She got there.'

b. *gwayeesha há báabà* ***ánìini*** *nɛ́ɛhidu*
g-wa-yee-sh-a ha baaba aniini n-aa-hid-u
AFF-3-say-AP-IS DSC father 1SG PRF-1SG-come-CP
'She said, "Father, I've come...'

c. *nea buuneeda ea sibar seani*
nea buun-ee-da ea si-bar seani
CONJ people-PS-UR CONJ IMPRS.PRF-kill all
'..but all the people have been killed".'

[2014-11-14_YG_story1_096–098]

As mentioned earlier, speech act participants typically occur preverbally for pragmatic reasons. In this example, *ánìini* '1SG' functions as the contrastive focus. It is highlighted and emphasized in contrast to *buuneeda*'people' in Example (66c), who have experienced a different fate. This is further indicated by the use of the conjunction *nea* 'but'. This example deviates from the Example (65) by Kießling (2007) in that the contrasting entities, i.e. *ánìini* and *buuneeda* 'people' each precede the verb.

The following Example (67) is another example for the role of contrast in accounting for constituent order variation. The clause in Example (67a)–(67b) sets up a conditional context, framing the actions of Ng'araajan based on the direction she comes from. The women in this conversation are discussing bartering and prices at the market. The following clauses (67c)–(67f)

each start with the proximal demonstrative *níi*, which refers to different market traders, followed by the same verb *gwayii* 'say', followed by a number.

(67) a. *mɛ́ɛyii* *iileabu* *murdeeni*
m-ea-yii ii-leab-u murd-ee-ni
NEG-1SG-say 3-pass-CP side-PS-DEM.PROX.SG
'If he comes from this side'

b. *akkwaleabu* *murdeeni* *Ng'arajaan*
ag=g-wa-leab-u murd-ee-ni Ng'arajaan
SEQ=AFF-3-pass-CP side-PS-DEM.PROX.SG Ng'arajaan
'and he comes from this side Ng'araajan'

c. *níi* *gwayii* *iiyèenya*
nii g-wa-yii iiyeenya
DEM.PROX.SG AFF-3-say two
'and this one says two'

d. *níi* *gwayii* *ági*
nii g-wa-yii agi
DEM.PROX.SG AFF-3-say one
'and this one says one'

e. *níi* *gwayii* *iiyèenya* *sey*
nii g-wa-yii iiyeenya sey
DEM.PROX.SG AFF-3-say two VOC
'and this one says two'

f. *níi* *gwayii* *múut*
nii g-wa-yii muut
DEM.PROX.SG AFF-3-say five
'and this one says five.'

[2017-04-19_Maandazi_418–419]

The clauses (67c)–(67f) show contrasting actions by different individuals which is also supported by the same clause structure. However, the focus in these clauses lies on the actions of these individuals which serve as the new and relevant information in this context, i.e. they are the primary points of interest. This example thus aligns with the findings by Kießling (2007) as the contrastive entities immediately follow the verb. However, in contrast to Kießling's (2007) examples, A in each clause *níi* 'this one' is shifted to the preverbal position. Nonetheless, it is challenging to compare this example with the one by Kießling (2007) where no context was provided.

5.6 Discussion

This chapter has aimed to give an explanation about constituent order variation in Datooga and about its conditions. Constituent order variation was analyzed in terms of semantic and

syntactical features like animacy and negation which played a role for the variation in other related Nilotic languages, as shown in Chapter 3.2. It additionally included an analysis of givenness, as well as a qualitative study on the impact of information structure.

The semantic and syntactical features observed in this study seem to have an impact on constituent order variation in that they show tendencies, such as personal pronouns occurring more frequently preverbally in S and A function. However, it is also important to note that conditions like animacy and argument realization are usually related, such as that first and second person referents are realized as personal pronouns in the discourse and information structure. Further, first and second person referents are typically only indexed on the verb and when they do occur with pronouns, these are mostly found preverbally. That is, they appear in a position which deviates from the dominant verb-initial order.

What the results further showed is that within the conditions tested, no single factor can explain or predict the choice of constituent order in Datooga. This was for example the case with negation in other languages of the typological survey where a specific order for negative clauses is applied but this does not held true for the Datooga data observed.

This leads to the conclusion that constituent order in Datooga might be more tied to the discourse pragmatics which formed the topic of the third research question. The results showed that preverbal arguments are emphasized in contrast to postverbal arguments as they are moved out of their canonical postverbal position. From an information structural perspective, preverbal arguments either occur in the function of topics, focus, or contrast. However, topics have been shown to also occur postverbally highlighting the fact that the preverbal position is reserved for this concept in Datooga. This is further supported by the numbers of verb-initial constituent orders in Section 5.2 compared to other constituent orders where core arguments precede the verb. As for contrast, according to Kießling (2007) it is even the immediate postverbal position that is restricted for contrastive focus, though the first Example for contrast (66) had the contrastive entity in preverbal position. Thus, when these categories appear preverbally, it puts a special emphasis on them.

The findings on the givenness of the arguments in the subset of data observed in INCEpTION showed that especially for A being given to the discourse is more commonly related to a preverbal position whereas new A arguments are rarely found in preverbal position. Givenness is often associated with the notion of topic which was supported by the qualitative analysis which showed examples of arguments in the preverbal position serving as the topic.

Cross-linguistically a preverbal position is often associated with the notion of topic. However, this raises the question of topics in verb-initial languages, such as Datooga, as it implies that an argument only serves as the topic when in preverbal position. Therefore, preverbal position might not per-se be restricted solely to the notion of topic in such languages. The results rather showed that more context is needed and that there are more discourse pragmatics involved, i.e. the context of the utterances with preverbal arguments plays a great role.

In terms of genres, the findings of the qualitative analysis also showed that clauses with preverbal arguments are more prevalent in conversations than in narratives or in the stimuli

included. Within the two conversations of the study, the difference in the type of conversation also seemed to play a role for the frequency of preverbal arguments. The conversation where people interacted while doing daily chores has more clauses with preverbal arguments compared to the conversation which had one main speaker that told a story to his audience. This suggests that conversational properties, such as turn-taking, might also be relevant for constituent order variation.

As mentioned in the section of methods, the number of occurrences for certain values is very low compared to other values, and thus restrict the strength of the conclusions. Future research with larger sample sizes is necessary to validate these preliminary observations and provide more robust evidence.

Chapter 6

Conclusions

The overall aim of this thesis was to investigate constituent order variation in Nilo-Saharan languages of East Africa. This research was conducted in two parts.

The first part consisted of a typological survey of the entire Nilo-Saharan phylum using grammars and other language descriptions to provide a general overview of constituent order variation and its interaction with case marking, as well as the conditions influencing this variation. The survey extended the findings of König (2008a) on case marking and constituent order variation in Nilo-Saharan languages. These languages exhibit a pattern of case marking restricted to postverbal agents or subjects and a lack or reduction of case marking when agents or subjects are in preverbal positions. This phenomenon is sometimes referred to as the 'no case before the verb' rule (König 2008a: 240). Nilo-Saharan languages were selected for this study due to the significant number of case-marking languages they include.

The first research question addressed the distribution of the interaction between case marking and constituent order variation, specifically the prevalence of the 'no case before the verb' rule. The analysis revealed that, although all Nilo-Saharan language families have languages with case distinctions, interaction was found only in three families: Eastern Sudanic, Gumuz, and Koman. Additionally, the results indicated that languages with such interaction either use a marked nominative or an ergative case marking system. Marked nominative languages predominantly employ tonal inflection for case marking, while ergative languages mainly use morphological marking. Regarding dominant constituent order, marked nominative languages typically exhibit a verb-initial order in both intransitive (VS) and transitive (VAO or VOA) clauses, whereas ergative languages generally show a verb-medial order (AVO or OVA).

The second research question examined the conditions relevant for case marking split or constituent order variation in languages with case marking and constituent order interaction. The results identified animacy, argument realization, definiteness, verb semantics, clause type, and negation as relevant conditions in some languages. However, the majority of languages in the survey highlighted information structure as the most significant condition for constituent order variation. In many cases, information structure was the sole explanation for changes in constituent order.

The impact of information structure on constituent order was the third research aim of the typological survey. The descriptions of information structure and its relevance for constituent order variation varied in detail. In most languages, topicalization or emphasis was cited as the primary reason for constituent order variation, while some languages associated a preverbal position of arguments with focus.

It is important to note that studying information structure is complex and requires a detailed consideration of a language's discourse pragmatics. Consequently, these topics are often not extensively covered in language descriptions. For future research, including existing corpora of these languages would facilitate qualitative analyses of discourse properties based on actual data rather than descriptions alone. In addition, the types of data that descriptions are typically based on would also need to be overhauled.

Given the significance of constituent order variation for case marking splits in languages of East Africa, the second part of this dissertation involved an in-depth study of constituent order variation in Datooga, a Southern Nilotic dialect cluster spoken in Tanzania. Previous research suggested that Datooga has a relatively flexible constituent order (Kießling 2007: 171; Griscom 2019: 31), but the reasons for choosing one order over another were unclear.

The first research question of the in-depth study on Datooga explored differences between genres, specifically between multi-party conversations and monologic narratives. The findings indicated differences in constituent order frequency of overt S and A arguments, with conversations showing a higher frequency of preverbal S and A arguments, while narratives had more postverbal occurrences.

The second research question aimed to identify the conditions relevant for constituent order variation in Datooga. With the use of the GRAID annotation scheme, conditions that played a role in constituent order variation in other Nilo-Saharan languages could be annotated. The findings showed that, although some conditions influenced argument placement relative to the verb, such as the preference for a preverbal position of animate S and A arguments, none mandated a specific and obligatory constituent order.

The third research question examined the role of discourse pragmatics and information structure in constituent order variation in Datooga. The findings revealed that a preverbal position in Datooga is not restricted to arguments serving as the topic or focus of an utterance. The qualitative analysis highlighted the importance of context in determining the choice of one constituent order over another. Overall, the findings from both the typological survey and the in-depth study on Datooga underscore the importance of discourse pragmatics and information structure in understanding constituent order variation in Nilo-Saharan languages. While other conditions such as animacy play a role, information structure was identified as the main factor. However, without further information about the context of an utterance, it is hard, if not impossible, to make assumption about the discourse roles and information structural properties of the referents which information in turn is necessary for discussing constituent order and especially the function of constituent orders deviating from a given basic order.

The Datooga study also demonstrated the importance of including various genres, such as conversational data, in analyses. Conversations exhibited more instances of preverbal arguments compared to monologic narratives. The difference between conversational texts, such as those recorded during daily chores versus story-telling sessions, suggests that a range of sociolinguistic contexts may provide additional insights into constituent order preferences. Future studies should analyze data from different social settings and groups to gain a deeper understanding.

Multi-party conversations introduce properties largely absent in monologic narratives, such as turn-taking. In some instances, the qualitative analysis suggested that speaker changes might influence constituent order variation. Although investigating the role of turn-taking was beyond the scope of this study, it presents an avenue for future research. Therefore, future studies should prioritize conversational data to further explore these dynamics.

References

Abdel-Hafiz, Ahmed Sokarno. 1988. *A reference grammar of Kunuz Nubian*. New York: State University of New York PhD dissertation.

Adelaar, Willem F. H. & Pieter C. Muysken. 2004. *The languages of the Andes*. Cambridge: Cambridge University Press.

Ahland, Colleen. 2016. Daats'íin, a newly identified undocumented language of western Ethiopia: A preliminary examination. In Doris L. Payne, Sara Pacchiarotti & Mokaya Bosire (eds.), *Diversity in African languages: Selected papers from the 46th Annual Conference on African Linguistics*, vol. 1 Contemporary African Linguistics, 417–449. Berlin: Language Science Press.

Ahland, Colleen Anne. 2012. *A grammar of Northern and Southern Gumuz*. Eugene: University of Oregon PhD dissertation.

Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory* 21(3). 435–483.

Alamin Mubarak, Suzan. 2006. An initial description of Laggori noun morphology and noun phrase. In Al-Amin Abu-Manga, Leoma G. Gilley & Anne Storch (eds.), *Insights into Nilo-Saharan language, history and culture: Proceedings of the 9th Nilo-Saharan Linguistic Colloquium, Institute of African and Asian Studies, University of Khartoum, 16-19 February 2004*, 9–24. Cologne: Rüdiger Köppe.

Alfira, David Abbi, Timothy Kuku Kafi, Hassan Kuwa Kaki, Ali Alaliim Hasan, Anjo Kuku Anjo, Dayan Kuku Jas & Sadik Kafi Sarukh. 2016. *Caning grammar book*. Juba: Sudan Workshop Program.

Amha, Azeb. 2017. The Omotic language family. In Alexandra Y. Aikhenvald & Robert M. W. Dixon (eds.), *The Cambridge handbook of linguistic typology* Cambridge handbooks in language and linguistics, 815–853. Cambridge: Cambridge University Press.

Andersen, Torben. 1981. *A grammar of Modo: A preliminary sketch*. Aalborg: University Centre of Aalborg.

Andersen, Torben. 1984. Aspect and word order in Moru. *Journal of African Languages and Linguistics* 6(1). 19–34.

Andersen, Torben. 1988. Ergativity in Päri, a Nilotic OVS language. *Lingua* 75(4). 289–324.

Andersen, Torben. 1991. Subject and topic in Dinka. *Studies in Language* 15(2). 265–294.

Andersen, Torben. 1995. Absolutive and nominative in Berta. In Robert Nikolaï & Franz Rottland (eds.), *Actes du cinquième colloque de linguistique Nilo-Saharienne: Proceedings of the fifth Nilo-Saharan linguistics colloquium, Nice, 24-29 August 1992, Nilo-Saharan language studies* Nilo-Saharan: Linguistic analyses and documentation, 39–69. Cologne: Rüdiger Köppe.

Andersen, Torben. 1999. Anti-logophoricity and indirect mode in Mabaan. *Studies in Language* 23(3). 499–530.

Andersen, Torben. 2002. Case inflection and nominal head marking in Dinka. *Journal of African Languages and Linguistics* 23(1). 1–30.

Andersen, Torben. 2015. Syntacticized topics in Kurmuk: A ternary voice-like system in Nilotic. *Studies in Language* 39(3). 508–554.

Andersen, Torben. 2018. The encoding of subjects and objects in Jumjum, a Nilotic OV language. *Lingua* 204. 78–116.

Arensen, Jon. 1982. *Murle grammar* Occasional papers in the study of Sudanese languages. Juba: Summer Institute of Linguistics.

Armbruster, Charles H. 1960. *Dongolese Nubian: A grammar.* Cambridge: Cambridge University Press.

Arnold, Jennifer E., Anthony Losongco, Thomas Wasow & Ryan Ginstrom. 2000. Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language* 76(1). 28–55.

Banga, Yey Dokery. 1978. *Éléments de la grammaire générative et transformationnelle du Membitu (Ndo).* Kisangani Institut Supérieur Pédagogique MA thesis.

Barasa, David. 2017. *Ateso grammar: A descriptive account of an Eastern Nilotic language.* Cape Town: University of Cape Town PhD dissertation.

Baroja, Tomás Herreros, Peter Sikamoy & Daniel Partany. 1989. *Analytical grammar of the Pokot language: Kitapu ngala pökot nyo kikir* Bibliotheca Africana. Trieste: Università di Trieste.

Bayou, Tesfaye Negash. 2014. *Documentation and grammatical description of Kwom.* Addis Ababa: Addis Ababa University PhD dissertation.

Beer, Samuel James. 2017. *Grammatical contraction in Nyang'i: A descriptive and comparative study.* Boulder: University of Colorado Boulder PhD dissertation.

Beltrame, Giovanni. 1876, 1877, 1877. Saggio di grammatica della lingua degli Akka' [sic!], Vocabolario Italiano- Akkà. *Bollettino della Società Geografica Italiana* XIII, XIV, XIV. 622–629, 13–41, 128–136.

Bender, M. Lionel. 1996. *Kunama* Languages of the world/materials. München: Lincom.

van den Berg, Helma. 1995. *A grammar of Hunzib.* München: Lincom Europa.

Bickel, Balthasar, Alena Witzlack-Makarevich & Taras Zakharko. 2015. Typological evidence against universal effects of referential scales on case alignment. In Ina Bornkessel-Schlesewsky, Andrej L. Malchukov & Marc Richards (eds.), *Scales and hierarchies*, 7–44. Berlin: Mouton de Gruyter.

Bickel, Balthasar & Yogendra P. Yādava. 2000. A fresh look at grammatical relations in Indo-Aryan. *Lingua* 110(5). 343–373.

Blackings, Mairi & Nigel Fabb. 2003. *A grammar of Ma'di.* Berlin/New York: Mouton de Gruyter.

Blake, Barry J. 2004. *Case* Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Blystad, Astrid & Ole Bjørn Rekdal. 2004. Datoga. In Carol R. Ember & Melvin Ember (eds.), *Encyclopedia of medical anthropology: Health and illness in the world's cultures*, 629–638. New York: Kluwer Academic Publishers.

Bock, J. Kathryn & Richard K. Warren. 1985. Conceptual accessibility and syntactic structure in sentence formulation. *Cognition* 21(1). 47–67.

Borgman, Donald M. 1990. Sanuma. In Desmond C. Derbyshire & Geoffrey K. Pullum (eds.), *Handbook of Amazonian languages*, 15–248. Berlin: Mouton de Gruyter.

Bossong, Georg. 1982. Der präpositionale Akkusativ im Sardischen. *Festschrift für Johannes Hubschmid zum 65.* 579–599.

Bossong, Georg. 1985. *Differentielle Objektmarkierung in den neuiranischen Sprachen.* Tübingen: Narr.

Bossong, Georg. 1991. Differential object marking in Romance and beyond. In Dieter Wanner & Douglas A. Kibbee (eds.), *New analyses in Romance linguistics. Selected papers from the XVIII Linguistic Symposium on Romance Languages, Urbana-Champaign, April 7–9, 1988*, 143–170. Amsterdam: John Benjamins.

Branigan, Holly P. & Eleonora Feleki. 1999. Conceptual accessibility and serial order in Greek speech production. In *Proceedings of the 21st Annual Conference of the Cognitive Science Society*, 96–101. Lawrence Erlbaum Associates.

Branigan, Holly P., Martin J. Pickering & Mikihiro Tanaka. 2008. Contributions of animacy to grammatical function assignment and word order during production. *Lingua* 118(2). 172–189.

Bruckhaus, Stefan. 2015. Locational nouns in Datooga. In Angelika Mietzner & Anne Storch (eds.), *Nilo-Saharan: Models and descriptions*, 73–92. Cologne: Rüdiger Köppe.

Bruckhaus, Stefan. 2021. *Motion in Datooga: A Southern Nilotic language of Tanzania* Nilo-Saharan: Studies in Language and Context. Cologne: Rüdiger Köppe.

Bryant, Michael G. 1999. *Aspects of Tirmaga grammar.* Ann Arbor University of Michigan MA thesis.

Carlin, Eithne. 1993. *The So language* Afrikanistische Monografien (AMO). Cologne: Institut fur Afrikanistik, Universität zu Köln.

Casaretto, Antje, Gerrit J. Dimmendaal, Birgit Hellwig, Uta Reinöl & Gertrud Schneider-Blum. 2020. Roots of ergativity in Africa (and beyond). *Studies in African Linguistics* 40(1). 185–215.

Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (ed.), *Subject and topic: A new typology of language*, 25–55. New York: Academic Press.

Chafe, Wallace L. 1979. The flow of thought and the flow of language. In Talmy Givón (ed.), *Discourse and syntax* Syntax and semantics, 159–181. Leiden: Brill.

Chappell, Hilary & Jean-Christophe Verstraete. 2019. Optional and alternating case marking: Typology and diachrony. *Language and Linguistics Compass* 13(3). e12311.

Cien, Mark Akec, Andrew Agok Igai Piiny, David Mabor Makuei Angong, John Malak Balang Kerjok, Daniel Majier Johou & David Kuacreng Nhial. 2016. *Reel grammar book.* Juba: SIL South Sudan.

Cohen, Kevin Bretonnel. 2000. *Aspects of the grammar of Kukú* LINCOM studies in African linguistics. München: Lincom.

Comrie, Bernard. 1989. *Language universals and linguistic typology: Syntax and morphology.* Chicago: University of Chicago Press.

Comrie, Bernard. 2013a. Alignment of case marking of full noun phrases (v2020.3). In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*, Online: Zenodo.

Comrie, Bernard. 2013b. Alignment of case marking of pronouns (v2020.3). In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*, Online: Zenodo.

Crazzolara, Pasquale J. 1960. *A study of the Logbara (Ma'di) language: Grammar and vocabulary.* London: Oxford University Press.

Creider, Chet A. & Jane Tapsubei Creider. 1989. *A Grammar of Nandi* Nilo-Saharan: Linguistic analyses and documentation. Hamburg: Helmut Buske.

Creider, Chet A. & Franz Rottland. 1997. Noun classification in Southern Nilotic: Datooga. *Afrika und Übersee* 80(1). 71–93.

Creissels, Denis, Gerrit J. Dimmendaal, Zygmunt Frajzyngier & Christa König. 2007. Africa as a morphosyntactic area. In Bernd Heine & Derek Nurse (eds.), *A linguistic geography of Africa*, 86–150. Cambridge: Cambridge University Press.

Croft, William. 2003. *Typology and universals.* Cambridge: Cambridge University Press.

DeLancey, Scott. 1981. An interpretation of split ergativity and related patterns. *Language* 57(3). 626–657.

Dimmendaal, Gerrit J. 1983. *The Turkana language* Publications in African languages and linguistics. Dordrecht: Foris Publications.

Dimmendaal, Gerrit J. 1985. Prominence hierarchies and Turkana syntax. In *Current approaches to African linguistics 3: proceedings of the 14th annual conference on African linguistics, University of Wisconsin, Madison* Publications in African languages and linguistics, 127–148. Dordrecht: Foris.

Dimmendaal, Gerrit J. 1986. Notes on Avokaya. *Afrikanistische Arbeitspapiere* 6. 5–33.

Dimmendaal, Gerrit J. 2009. Tama. In Gerrit J. Dimmendaal (ed.), *Coding participant marking: Construction types in twelve African languages* (Studies in language companion series 110), 305–330. Amsterdam/Philadelphia: John Benjamins.

Dimmendaal, Gerrit J. 2012. Metrical structures: A neglected property of Nilotic languages. *Studies in Nilotic Linguistics* 5. 1–26.

Dimmendaal, Gerrit J. 2017. Areal contact in Nilo-Saharan. In Raymond Hickey (ed.), *Japanese, Korean and the Transeurasian languages*, 446–470. Cambridge: Cambridge University Press.

Dimmendaal, Gerrit J. 2020. Nilo-Saharan and its limits. In Rainer Voßen & Gerrit J. Dimmendaal (eds.), *The Oxford handbook of African languages*, 364–382. Oxford: Oxford University Press.

Dimmendaal, Gerrit J., Colleen Ahland, Angelika Jakobi & Constance Kutsch Lojenga. 2019. Linguistic features and typologies in languages commonly referred to as 'Nilo-Saharan'. In H. Ekkehard Wolff (ed.), *Cambridge handbook of African languages*, 326–381. Cambridge: Cambridge University Press.

Dixon, Robert M. W. 1979. Ergativity. *Language* 55(1). 59–138.

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

Dowty, David. 1991. Thematic Proto-Roles and argument selection. *Language* 67(3). 547–619.

Dryer, Matthew S. 1995. Frequency and pragmatically unmarked word order. In Pamela A. Downing & Michael Noonan (eds.), *Word order in discourse*, 105–135. Amsterdam/Philadelphia: John Benjamins.

Dryer, Matthew S. 1997. On the six-way word order typology. *Studies in Language* 21(1). 69–103.

Dryer, Matthew S. 2007. Word order. In Timothy Shopen (ed.), *Clause structure*, 61–131. Cambridge: Cambridge University Press.

Du Bois, John W. 1987. The discourse basis of ergativity. *Language* 63. 805–855.

Edgar, John. 1989. *A Masalit grammar: With notes on other languages of Darfur and Wadai*, vol. 3 Sprache und Oralität in Afrika: Frankfurter Studien zur Afrikanistik. Berlin: Dietrich Reimer.

Ehret, Christopher. 1995. *Reconstructing Proto-Afroasiatic (Proto-Afrasian): Vowels, tone, consonants, and vocabulary.* Berkeley: University of California Press.

ELAN. 2023. ELAN (Version 6.6) [Computer Software]. Max Planck Institute for Psycholinguistics, The Language Archive.

Elnur, Elsadig Omda Ibrahim. 2016. *Major word categories in Nara.* Universität zu Köln MA thesis.

England, Nora C. 1983. *A grammar of Mam, a Mayan language* Texas Linguistics Series. Austin: University of Texas Press.

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

Fauconnier, Stefanie. 2011. Differential agent marking and animacy. *Lingua* 121(3). 533–547.

Fauconnier, Stefanie. 2012. *Constructional effects of involuntary and inanimate agents: A cross-linguistic study.* Leuven: Katholieke Universiteit Leuven PhD dissertation.

Flores, Joseph. 1971. *A simple introduction to the Ketebo or Mening language with vocabulary.* Naporo: Catholic Church Karonga.

Frajzyngier, Zygmunt. 2018. Afroasiatic languages. *Oxford Research Encyclopedia of Linguistics* 1–42.

Frajzyngier, Zygmunt & Erin Shay. 2020. Contact and Afroasiatic languages. In Raymond Hickey (ed.), *The handbook of language contact*, 571–591. Hoboken, NJ: Wiley-Blackwell.

Gerven Oei, Vincent W. J. van Gerven. 2021. *A reference grammar of Old Nubian*. Louvain: Peeters.

Good, Jeff. 2017. Niger-Congo languages. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 471–499. Cambridge: Cambridge University Press.

Goyvaerts, D. L. 1986. Kaliko: A first encounter. In *Language and history in Central Africa* (Antwerp papers in linguistics 44), 57–69. Antwerp: University of Antwerpen.

Green, Christopher R. 2021. *Somali grammar*. Berlin: Mouton de Gruyter.

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

Greenberg, Joseph H. 2005. *Language universals: With special reference to feature hierarchies*. Berlin/New York: De Gruyter Mouton.

Griscom, Richard. 2019. *Topics in Asimjeeg Datooga verbal morphosyntax*. Eugene: University of Oregon PhD dissertation.

Grossman, Eitan & Noam Faust. 2015. Nuer.

Guillaume, Antoine & Harold Koch. 2021. Introduction: Associated Motion as a grammatical category in linguistic typology. In Antoine Guillaume & Harold Koch (eds.), *Associated motion*, 3–30. De Gruyter Mouton.

Güldemann, Tom. 2018. Historical linguistics and genealogical language classification in Africa. In Tom Güldemann (ed.), *The Languages and Linguistics of Africa* The World of Linguistics, 58–444. Berlin: DeGruyter Mouton.

Gundel, Jeanette K & Thorstein Fretheim. 2006. Topic and focus. In Laurence R. Horn & Gregory Ward (eds.), *The handbook of pragmatics*, 175—196. Blackwell.

Haig, Geoffrey & Stefan Schnell. 2014. *Annotations using GRAID. (Grammatical Relations and Animacy in Discourse). Version 7.0.*

Haig, Geoffrey & Stefan Schnell. 2016. The discourse basis of ergativity revisited. *Language* 92(3). 591–618.

Hall, Beatrice, R. M. R. Hall, Martin D. Pam, Amy Myers, Stephen A. Antell & Godfrey K. Cherono. 1974. African vowel harmony systems from the vantage point of Kalenjin. *Afrika und Übersee* 57. 241–267.

Halliday, Michael A. K. 1967. Notes on transitivity and theme in English: Part 2. *Journal of Linguistics* 3(2). 199–244.

Hammarström, Harald. 2016. Linguistic diversity and language evolution. *Journal of Language Evolution* 1(1). 19–29.

Hammarström, Harald. 2018. A survey of African languages. In Tom Güldemann (ed.), *The Languages and Linguistics of Africa*, 1–57. Berlin: DeGruyter Mouton.

Harold, Bruce B. 1995. Subject-verb word order and the function of early position. In Pamela A. Downing & Michael Noonan (eds.), *Word order in discourse* (Typological studies in language 30), 137–161. Amsterdam/Philadelphia: John Benjamins.

Harris, Alice C. 2009. *Georgian syntax: A study in relational grammar* Cambridge studies in linguistics. Cambridge: Cambridge University Press.

Haspelmath, Martin. 2007. Ditransitive alignment splits and inverse alignment. *Functions of Language* 14(1). 79–102.

Haude, Katharina & Alena Witzlack-Makarevich. 2016. Referential hierarchies and alignment: An overview. *Linguistics* 54(3). 433–441.

Hayward, Dick. 1984. *The Arbore language: A first investigation including a vocabulary* (Kuschitische Sprachstudien 2). Hamburg: Helmut Buske.

Heine, Bernd. 1980. Camus, a Maa dialect. In Bernd Heine (ed.), *The Non-Bantu languages of Kenya* (Language and dialect atlas of Kenya 2), 99–140. Berlin: Dietrich Reimer.

Heine, Bernd & Christa König. 2010. *The Labwor language of northeastern Uganda: A grammatical sketch*. Tokyo: Research Institute for Languages and Cultures of Asia and Africa.

Heine, Bernd, Christa König & Karsten Legère. 2015. *The Akie language of Tanzania: A sketch of discourse grammar*. Tokyo: Research Institute for Languages and Cultures of Asia and Africa.

Heine, Bernd & Rainer Voßen. 1976. Zur Stellung der Ongamo-Sprache (Kilimandscharo). *Afrika und Übersee* 59(2). 81–104.

von Heyking, Beatrix. 2013. *A grammar of Belanda Boor: Phonology and morphology* (Nilo-Saharan: Linguistic analyses and documentation 27). Cologne: Rüdiger Köppe.

Hieda, Osamu. 1998. A sketch of Koegu grammar: Towards reconstructing Proto-Southeastern Surmic. In Gerrit J. Dimmendaal & Marco Last (eds.), *Surmic languages and cultures*, vol. 13 Nilo-Saharan: linguistic analyses and documentation (NISA), 345–373. Cologne: Rüdiger Köppe.

Hieda, Osamu. 2000. Some problems in Datooga (Bajuuta dialect) consonantal system. *'Mehr als nur Worte...': Afrikanistische Beiträge zum 65. Geburtstag von Franz Rottland* 285–293.

Hieda, Osamu. 2013. *A grammar of Kumam: the interaction between syntax and pragmatics* (Studies in Nilotic Linguistics 7). Tokyo: Research Institute for Languages and Cultures of Asia and Africa.

Holisky, Dee Ann. 1987. The case of the intransitive subject in Tsova-Tush (Batsbi). *Lingua* 71. 103–132.

de Hoop, Helen & Peter de Swart (eds.). 2009. *Differential subject marking.* Dordrecht: Springer.

de Hoop, Helen & Andrej L. Malchukov. 2008. Case-marking strategies. *Linguistic Inquiry* 39(4). 565–587.

Iemmolo, Giorgio. 2013. Symmetric and asymmetric alternations in direct object encoding. *STUF - Language Typology and Universals* 66(4). 378–403.

Irumu, Agozia-Kario. 1986. *Description du Logoti (Nord-Est du Zaïre): Phonologie, esquisse grammaticale et lexique logoti-français.* Lille: Université de la Sorbonne Nouvelle (Paris 3) PhD dissertation.

Ismail, Muhammad Abbaker. 2000. *Linguistic analysis of the Daju language.* Khartoum University of Khartoum MA thesis.

Jakobi, Angelika. 1990. *A Fur grammar: Phonology, morphophonology, and morphology.* Hamburg: Helmut Buske.

Jakobi, Angelika & Joachim Crass. 2004. *Grammaire du Beria (langue saharienne)* (Nilo-Saharan: Linguistic analyses and documentation 18). Cologne: Rüdiger Köppe.

Jakobi, Angelika & Gerrit J. Dimmendaal. 2022. Number marking in Karko and Nilo-Saharan. In Paolo Acquaviva & Michael Daniel (eds.), *Number in the world's languages: A comparative handbook*, 63–106. Berlin: De Gruyter Mouton.

Jerono, Prisca. 2011. *Tugen word order - A minimalist perspective.* Nairobi: University of Nairobi PhD dissertation.

Joseph, Clement Lopeyok, Michael Vorgol Lochule, Joseph Celestine Lolar, Lolar John Hillary & Davĩd Nyangamoi Nayoro. 2013. *Laarim grammar book.* Juba: SIL South Sudan.

Joswig, Andreas. 2019. *The Majang language.* Leiden: Rijksuniversiteit te Leiden PhD dissertation.

Joswig, Andreas, Fajio Mokonne, Tulua Amaro & Achim Diehl. 2011. *A brief grammar of the Me'en language.* Addis Ababa: SIL Ethiopia.

Kauczor, Daniel. 1920. *Die bergnubische Sprache (Dialekt von Gebel Delen)* (Schriften der Sprachenkommission 7). Wien: Alfred Hölder.

Kellermann, Petra. 2000. *Eine grammatische Skizze des Tama auf der Basis der Daten von R.C.~Stevenson.* Mainz Johannes Gutenberg Universität MA thesis.

Kenesei, István, Robert M. Vago & Anna Fenyvesi. 1998. *Hungarian.* London/New York: Routledge.

Kießling, Roland. 1998. Der Pluraktionalis im Datooga. In Ines Fiedler, Catherine Griefenow-Mewis & Brigitte Reineke (eds.), *Afrikanische Sprachen im Brennpunkt der Forschung: Linguistische Beiträge zum 12. Afrikanistentag, Berlin, 3.-6. Oktober 1996*, 179–196. Cologne: Rüdiger Köppe.

Kießling, Roland. 2000a. Number marking in Datooga nouns. In *'Mehr als nur Worte...': Afrikanistische Beiträge zum 65. Geburtstag von Franz Rottland*, 349–366. Cologne: Rüdiger Köppe.

Kießling, Roland. 2000b. Verb classes in Nilotic: Evidence from Datooga (Southern Nilotic). In *Proceedings of the 2nd world congress of African linguistics, Leipzig 1997*, 603–616. Cologne: Rüdiger Köppe.

Kießling, Roland. 2001. Space and reference in Datooga verbal morphosyntax, vol. 22 Nilo-Saharan: linguistic analyses and documentation, 123–142. Cologne: Rüdiger Köppe.

Kießling, Roland. 2007. The 'marked nominative' in Datooga. *Journal of African Languages and Linguistics* 28(2). 149–191.

Kießling, Roland. 2022. Future, persistive, and negative perfect in Datooga verbal inflection. In Roger M. Blench, Petra Weschenfelder & Georg Ziegelmeyer (eds.), *Current research in Nilo-Saharan: 14th Nilo-Saharan Linguistics Colloquium, Vienna, 30th May–1st June 2019 (Nilo-Saharan volume 32)*, 141–152. Cologne: Rüdiger Köppe.

Kießling, Roland & Stefan Bruckhaus. 2017. Associated locomotion in Datooga (Southern Nilotic). *Proceedings of the 8th World Congress of African Linguistics* 243–258.

Kievit, Dirk & Erika Robertson. 2012. Notes on Gwama grammar. *Studies in African Linguistics* 41(1). 39–97.

Killian, Don. 2015. *Aspects of Uduk morphology and phonology.* Helsinki: University of Helsinki PhD dissertation.

Kimball, Geoffrey D. 1991. *Koasati grammar.* London: Lincoln.

Kittilä, Seppo. 2008. Animacy effects on differential Goal marking. *Linguistic Typology* 12(2). 245–268.

Kittilä, Seppo & Andrej L. Malchukov. 2008. Varieties of accusative. In Andrej L. Malchukov & Andrew Spencer (eds.), *The Oxford handbook of case*, 549–561. Oxford: Oxford University Press.

Klie, Jan-Christoph, Michael Bugert, Beto Boullosa, Richard Eckart De Castilho & Iryna Gurevych. 2018. The INCEpTION platform: Machine-assisted and knowledge-oriented interactive annotation. In *Proceedings of the 27th international conference on computational linguistics: system demonstrations*, 5–9. Santa Fe.

Koelle, Sigismund W. 1854. *A grammar of the Bo̱rnu or Kānurī language*. London: Church Missionary Society.

Kol, Deng Gideon, John Majok Matuek, Zechariah Manyiel Mawuor, Peter Tiu Mberu, Mbio Nelson Moli & Charles Mayen Manyang. 2013. *'Bëlï grammar book*. Juba: SIL South Sudan.

König, Christa. 2006. Marked nominative in Africa:. *Studies in Language* 30(4). 655–732.

König, Christa. 2008a. *Case in Africa*. Oxford: Oxford University Press.

König, Christa. 2008b. Khoisan languages. *Language and Linguistics Compass* 2(5). 996—1012.

Koshal, Sanyukta. 1979. *Ladakhi grammar*. Delhi: Motilal Banarsidass.

Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica* 55(3–4). 243–276.

Kutsch Lojenga, Constance. 1994. *Ngiti: A Central-Sudanic language of Zaire* (Nilo-Saharan: Linguistic analyses and documentation 9). Cologne: Rüdiger Köppe.

Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Lambrecht, Knud. 2001a. A framework for the analysis of cleft constructions. *Linguistics* 39(3). 463–516.

Lambrecht, Knud. 2001b. Dislocation. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals* Handbücher zur Sprach- und Kommunikationswissenschaft, 1050–1078. New York: De Gruyter Mouton.

Larochette, J. 1958. *Grammaire des dialectes mangbetu et medje: Suivie d'un manuel de conversation et d'un lexique* (Sciences de L'homme 18). Tervuren: Commission de Linguistique Africaine.

Larsen, Iver A. 1991. Sabaot noun classification. In *Proceedings of the 3rd Nilo-Saharan linguistics colloquium, Kisumu, Kenya, August 4-9, 1986* (Nilo-Saharan: Linguistic analyses and documentation 6), 143–163. Cologne: Rüdiger Köppe.

Last, Marco & Deborah Lucassen. 1998. A grammatical sketch of Chai, a Southeastern Surmic language. In Gerrit J. Dimmendaal & Marco Last (eds.), *Surmic languages and cultures* (Nilo-Saharan: Linguistic analyses and documentation 13), 375–436. Cologne: Rüdiger Köppe.

Levergood, Barbara Jo. 1987. *Topics in Arusa phonology and morphology (Tanzania).* Ann Arbor: UMI PhD dissertation.

Leyew, Zelealem. 2005. *Gwama, a little-known endangered language of Ethiopia: A sketch of its grammar and lexicon.* Addis Ababa: Addis Ababa University.

Lohitare, Loki Dominic, Darius Lokure Beato Lohammarimoi, Dominic Timan Peter & Peter Lopeyok Joseph. 2012. *Didinga grammar book.* Juba: SIL South Sudan.

Lukas, Johannes. 1933. Beiträge zur Kenntnis der Sprachen von Wadái (Mararēt, Mába). *Journal de la Société des Africanistes* 3(1). 25–55.

Lukas, Johannes. 1938. Die Sprache der Sungor in Wadai. *Mitteilungen der Ausland-Hochschule an der Universität Berlin* 41(3). 171–246.

Lutwori, Allen Pitya, Enike Amina Wani, Lodu Philip Jembeke, Robert Gajuk Paul Wanit, Martin Lomu Goke & Augustino Laku Buli. 2013. *Mundari grammar.* Juba: SIL South Sudan.

Malchukov, Andrej L. 2005. *Even* (Languages of the world/materials 12). München: Lincom.

McGregor, William B. 2006. Focal and optional ergative marking in Warrwa (Kimberley, Western Australia). *Lingua* 116(4). 393–423.

McGregor, William B. 2010. Optional ergative case marking systems in a typological-semiotic perspective. *Lingua* 120(7). 1610–1636.

McGregor, William B. 2018. Emergence of optional accusative case marking in Khoe languages. In Ilja A. Seržant & Alena Witzlack-Makarevich (eds.), *Diachrony of differential argument marking*, 243–279. Berlin: Language Science Press.

Mdoe, Kuria Michael. 2024. Gender assignment in Datooga nouns. *South African Journal of African Languages* 44(2). 188–193.

Meakins, Felicity & Carmel O'Shannessy. 2010. Ordering arguments about: Word order and discourse motivations in the development and use of the ergative marker in two Australian mixed languages. *Lingua* 120(7). 1693–1713.

Mellese, Gelaneh Alemu. 2017. *Documentation and grammatical description of Tapo.* Addis Ababa: Addis Ababa University PhD dissertation.

Micheli, Ilaria. 2018. *Grammatical sketch and short vocabulary of the Ogiek language of Mariashoni.* Trieste: Università di Trieste.

Mietzner, Angelika. 2016. *Cherang'any: A Kalenjin language of Kenya.* Cologne: Rüdiger Köppe.

Miller, Cynthia L & Leoma G. Gilley. 2001. Evidence for ergativity in Shilluk. *Journal of African Languages and Linguistics* 22(1). 33–68.

Mitchell, Alice. 2015a. *Linguistic avoidance and social relations in Datooga.* New York: State University of New York at Buffalo PhD dissertation.

Mitchell, Alice. 2015b. Words that smell like father-in-law: A linguistic description of the Datooga avoidance register. *Anthropological Linguistics* 57(2). 195–217.

Mitchell, Alice. 2017. The pragmatics of a kinship term: The meaning and use of íiyá 'mother' in Datooga (Nilotic). In Raija Kramer & Roland Kießling (eds.), *Mechthildian Approaches to Afrikanistik*, 287–301. Cologne: Rüdiger Köppe.

Mitchell, Alice. 2021. Phasal polarity in Barabaiga and Gisamjanga Datooga (Nilotic): Interactions with tense, aspect, and participant expectation. In Raija L. Kramer (ed.), *The expression of phasal polarity in African languages* (Empirical approaches to language typology 63), 419–442. Berlin/Boston: De Gruyter Mouton.

Mitchell, Alice. 2022. "Hey, bonikala" - language contact and experiences of Swahili among rural Datooga children. In Roger M. Blench, Petra Weschenfelder & Georg Ziegelmeyer (eds.), *Current research in Nilo-Saharan: 14th Nilo-Saharan Linguistics Colloquium, Vienna, 30th May–1st June 2019 (Nilo-Saharan volume 32)*, 153–162. Cologne: Rüdiger Köppe.

Mitchell, Alice & Péter Marton Rácz. 2021. Children's knowledge of a name-based avoidance register: A quantitative study among Datooga of Tanzania. *American Anthropologist* 123(2). 389–400.

Montgomery, Christine Anne. 1966. *The morphology of Sebei.* Los Angeles: University of California at Los Angeles PhD dissertation.

Moodie, Jonathan & Rosey Billington. 2020. *A grammar of Lopit: An Eastern Nilotic language of South Sudan.* Leiden: Brill.

Morolong, Malillo & Larry M. Hyman. 1977. Animacy, objects and clitics in Sesotho. *Studies in African Linguistics* 8(3). 199–218.

Mukarovsky, Hans. 1956. Zur Sprache der Ju=p=Adhola. *Archiv für Völkerkunde* 11. 83–103.

Mulder, Monique Borgerhoff. 1992. Demography of pastoralists: Preliminary data on the Datoga of Tanzania. *Human Ecology* 20(4). 383–405.

Muratori, Carlo P. 1938. *Grammatica Lotuxo.* Verona: Missioni Africane.

Muzale, Henry R. T. & Josephat M. Rugemalira. 2008. Researching and documenting the languages of Tanzania. *Conservation* 2(1). 68–108.

Næss, Åshild. 2004. What markedness marks: The markedness problem with direct objects. *Lingua* 114(9). 1186–1212.

Nebel, Arturo. 1978. *Grammatica e Dizionario Dinka* Museum Combonianum. Bologna: Editrice Missionaria Italiana.

Neeleman, Ad, Elena Titov, Hans Van De Koot & Reiko Vermeulen. 2009. A syntactic typology of topic, focus and contrast. In Jeroen Van Craenenbroeck (ed.), *Alternatives to cartography*, 15–52. De Gruyter Mouton.

Nichols, Johanna. 1986. Head-marking and dependent-marking grammar. *Language* 62(1). 56–119.

Noonan, Michael. 1992. *A grammar of Lango* (Mouton grammar library 7). Berlin/New York: De Gruyter Mouton.

Nougayrol, Pierre. 1986. Note sur la langue Kibet (Tchad). *Africana Marburgensia* 19(2). 38–55.

Nougayrol, Pierre. 1990. *La langue des Aiki dits Rounga (Tchad, République centrafricaine): Esquisse descriptive et lexique* Publications du Département Langues et parole en Afrique centrale. Paris: Geuthner.

Novelli, Bruno. 1985. *A grammar of the Karimojong language* (Language and Dialect Studies in East Africa 7). Berlin: Dietrich Reimer.

Odden, David. 1983. Aspects of Didinga phonology and morphology. In *Nilo-Saharan language studies* (Monograph / committee on Northeast African studies 13), 148–176. East Lansing: African Studies Center, Michigan State University.

Overall, Simon. 2007. *A grammar of Aguaruna.* Melbourne: La Trobe University PhD dissertation.

Palancar, Enrique L. 2008. Varieties of ergative. In Andrej L. Malchukov & Andrew Spencer (eds.), *The Oxford handbook of case*, 562–571. Oxford: Oxford University Press.

Payne, Doris L. 1987. Information structuring in Papago narrative discourse. *Language* 63(4). 783–804.

Payne, Doris L. 1995. Verb initial languages and information order. In Pamela A. Downing & Michael Noonan (eds.), *Word order in discourse* (Typological studies in language 30), 449–485. John Benjamins.

Payne, John R. 1980. The decay of ergativity in Pamir languages. *Lingua* 51(2–3). 147–186.

Petráček, Karel. 1966. Morphologie (Nomen, Pronomen) der Berti-(Siga) Sprache in Dār Fūr (Sūdān). *Archiv Orientální* 34. 295–319.

Pike, Ivy L. & Crystal L. Patil. 2006. Understanding women's burdens: Preliminary findings on psychosocial health among Datoga and Iraqw women of northern Tanzania. *Culture, medicine and psychiatry* 30. 299–330.

R Core Team. 2022. *R: A language and environment for statistical computing*. Vienna.

Randal, Scott. 1998. A grammatical sketch of Tennet. In *Surmic languages and cultures* (Nilo-Saharan: Linguistic analyses and documentation 13), 219–272. Cologne: Rüdiger Köppe.

Reh, Mechthild. 1996. *Anywa language: Description and internal reconstructions* (Nilo-Saharan: Linguistic analyses and documentation 11). Köln: Rüdiger Köppe.

Ringe, Peter C. 1953. *A simple Alur grammar and Alur-English-Alur vocabularies* Eagle language study series. Kampala: Eagle Press.

Rottland, Franz. 1982. *Die südnilotischen Sprachen: Beschreibung, Vergleichung und Rekonstruktion*. Berlin: Dietrich Reimer.

Ryan, Kevin M. 2016. Phonological weight. *Language and Linguistics Compass* 10(12). 720–733.

Santandrea, Stefano. 1963. *A concise grammar outline of the Bongo language (Sudan, Bahr el Ghazal province)* (Museum Combonianum 14). Rome: Sodality of St. Peter Claver.

Santandrea, Stefano. 1966. *The Birri language: Brief elementary notes* (Museum Combonianum 20). Bologna: Editrice Nigrizia.

Santandrea, Stefano. 1976. *The Kresh group, Aja and Baka languages (Sudan): A linguistic contribution*. Naples: Istituto Universitario Orientale.

Schneider-Blum, Gertrud. 2023. The impact of attentional centering on ergative marking in Tima. *Language Typology and Universals* 76(1). 87–112.

Schrock, Terrill. 2014. *A grammar of Ik (Icé-tód): Northeast Uganda's last thriving Kuliak language*. Utrecht: LOT PhD dissertation.

Schröder, Helga. 2008. *Word order in Toposa: An aspect of multiple feature-checking*, vol. 142 Publications in linguistics. Dallas: SIL International.

Schröder, Martin & James Lokuuda Kadanya. 2011. *A brief grammar of the Nyangatom language*. Addis Ababa: SIL Ethiopia.

Sellen, Daniel W. 2003. Nutritional consequences of wealth differentials in East African pastoralists: The case of the Datoga of Northern Tanzania. *Human Ecology* 31(4). 529–570.

Sieff, Daniela F. 1995. *The effects of resource availbility on the subsistence strategies of Datoga pastoralists of north west Tanzania.* Oxford: University of Oxford PhD dissertation.

Sieff, Daniela F. 1999. The effects of wealth on livestock dynamics among the Datoga pastoralists of Tanzania. *Agricultural Systems* 59(1). 1–25.

Siewierska, Anna. 1993. On the interplay of factors in the determination of word order. In Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld & Theo Vennemann (eds.), *Syntax: An international handbook of contemporary research*, 826–846. Berlin: De Gruyter Mouton.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In Robert M. W. Dixon (ed.), *Grammatical categories in Australian languages*, 112–171. Canberra: Australian Institute of Aboriginal Studies.

Skopeteas, Stavros, Ines Fiedler, Samantha Hellmuth, Anne Schwarz, Ruben Stoel, Gisbert Fanselow, Caroline Féry & Manfred Krifka. 2006. *Questionnaire on information structure (QUIS): Reference manual*, vol. 4. Potsdam: Universitätsverlag Potsdam.

Smith, Edwin W. 1938. *A tentative grammar of the Efe or Mbuti language: The reputed language of the Pygmies of the Ituri Forest, Belgian Congo.* London: The Bible House.

Spagnolo, Lorenzo M. 1933. *Bari grammar.* Verona: Missioni Africane.

Spire, F. 1905. Notes on the Madi N****** (Equatorial Nile). *Journal of the African Society* 4(15). 301.

Stevenson, Roland C. 1938. A grammar of the Nyimang language (Nuba Mountains).

Stirtz, Timothy M. 2011. *Laarim (loh) tone.* SIL International.

Stirtz, Timothy M. 2014. Ergative, antipassive and other verb derivational morphemes in Gaahmg. *Journal of African Languages and Linguistics* 35(2). 243–272.

Storch, Anne. 2014. *A grammar of Luwo: An anthropological approach* Culture and Language Use. Amsterdam/Philadelphia: John Benjamins.

Sulkala, Helena & Merja Karjalainen. 1992. *Finnish.* London: Routledge.

Tanaka, Mikihiro N., Holly P. Branigan, Janet F. McLean & Martin J. Pickering. 2011. Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language* 65(3). 318–330.

Tornay, Serge. 1970. *Une contribution à l'étude du kalenjin: Recherches sur le parler des Keyo du Kenya.* Paris: Université de la Sorbonne (Paris IV) PhD dissertation.

Toweett, Taaitta. 1979. *A study of Kalenjin linguistics.* Nairobi: Kenya Literature Bureau.

Tucker, Archibald N. 1940a. Lendu. In Archibald N. Tucker (ed.), *The Eastern Sudanic languages*, vol. 1, 380–418. Oxford: Oxford University Press.

Tucker, Archibald N. 1940b. The Moru-Ma'di language group. In Archibald N. Tucker (ed.), *The Eastern Sudanic languages*, vol. 1, 87–379. Oxford: Oxford University Press.

Tucker, Archibald N. 1994. *A grammar of Kenya Luo (Dholuo)* (Nilo-Saharan: Linguistic analyses and documentation 8). Cologne: Rüdiger Köppe.

Tucker, Archibald N. & Margaret A. Bryan. 1966. *Linguistic analyses: The Non-Bantu languages of North-Eastern Africa* Handbook of African languages. London: Oxford University Press.

Tucker, Archibald N. & J. Tompo Ole Mpaayei. 1955. *A Maasai grammar with vocabulary* (Publications of the African Institute, Leyden II). London: Longmans, Green and Co.

Unyutha, Wokocan. 1975. *Aspects de la grammaire générative et transformationnelle de la langue Kakwa.* Lubumbashi Université de Lubumbashi (UNILU) MA thesis.

van Urk, Coppe. 2015. *A uniform syntax for phrasal movement: A case study of Dinka Bor.* Cambridge: Massachusetts Institute of Technology PhD dissertation.

de Voogt, Alex. 2011. Dual marking and kinship terms in Afitti. *Studies in Language* 35(4). 898–911.

Vorbichler, Anton. 1965. *Die Phonologie und Morphologie des Balese (Ituri-Urwald, Kongo)* (Afrikanistische Forschungen 2). Glückstadt: J. J. Augustin.

Vorbichler, Anton. 1971. *Die Sprache der Mamvu*, vol. 5 Afrikanistische Forschungen. Glückstadt: J. J. Augustin.

Waag, Christine. 2010. *The Fur verb and its context*, vol. 26 Nilo-Saharan: Linguistic analyses and documentation. Cologne: Rüdiger Köppe.

Waag, Christine. 2018. The grammar of space in Keliko, a Central Sudanic language of South Sudan. In Helga Schröder & Prisca Jerono (eds.), *Nilo-Saharan issues and perspectives*, 193–209. Cologne: Rüdiger Köppe.

Walters, Josiah Keith. 2016. *A grammar of Dazaga* Grammars and sketches of the world's languages: Africa. Leiden/Boston: Brill.

Ward, Gregory & Betty Birner. 2006. Information structure and non-canonical syntax. In Laurence R Horn & Gregory Ward (eds.), *The handbook of pragmatics*, 152–174. Oxford: Wiley-Blackwell.

Wegener, Claudia U. 2012. *A grammar of Savosavo*, vol. 61 Mouton grammar library. Berlin/Boston: De Gruyter Mouton.

Weiss, Doris. 2009. *Phonologie et morphosyntaxe du Maba*. Lyon: Université Lumière-Lyon II PhD dissertation.

Wellens, Inneke. 2003. *The Nubi language of Uganda: An Arabic Creole in Africa*. Nijmegen: Katholieke Universiteit Nijmegen PhD dissertation.

Werner, Roland. 1987. *Grammatik des Nobiin (Nilnubisch): Phonologie, Tonologie und Morphologie* (Nilo-Saharan: Linguistic analyses and documentation 1). Hamburg: Helmut Buske.

Werner, Roland. 1993. *Tìdn-Áal: A study of Midob (Darfur-Nubian)* (Sprache und Oralität in Afrika: Frankfurter Studien zur Afrikanistik 17). Berlin: Dietrich Reimer.

Westermann, Diedrich. 1944. Das Oxoriok: eine sprachliche Skizze. *Studien zur Auslandskunde: Afrika* 3. 19–46.

Williams, Robert S. & Jade Comfort. no date. Ghulfan grammar sketch.

Williamson, Kay. 1989. Niger-Congo overview. In J. T. Bendor-Samuel (ed.), *The Niger-Congo languages: a classification and description of Africa's largest language family*, 3–45. Lanham: University Press of America.

Witzlack-Makarevich, Alena & Ilja A. Seržant. 2018. Differential argument marking: Patterns of variation. In Ilja A. Seržant & Alena Witzlack-Makarevich (eds.), *Diachrony of differential argument marking*, 1–40. Berlin: Language Science Press.

Worku, Firew Girma. 2020. *A grammar of Mursi, a Nilo-Saharan language*. Cairns: James Cook University PhD dissertation.

Yigezu, Moges & Gerrit J. Dimmendaal. 1998. Notes on Baale. In Gerrit J. Dimmendaal & Marco Last (eds.), *Surmic languages and cultures*, 273–317. Cologne: Rüdiger Köppe.

Yokwe, Eluzai Moga. 1987. *The tonal grammar of Bari*. Ann Arbor: UMI PhD dissertation.

Zavoni, Ntondo. 2003. *Éléments de description du Ngangela*. Lyon: Université Lumière Lyon 2 PhD dissertation.

Zwicky, Arnold. 1977. Hierarchies of person. *Papers from the Thirteenth Annual Meeting of the Chicago Linguistic Society* 712–733.

Appendix A

Tables

Table A.1: Languages with case marking and interaction, including the three languages with interaction different from the 'no case before the verb' rule.

#	Language family	Branch	Language	Geographical area	Case marking system	Marking strategy	Default order (TR)	Default order (INTR)	Flexible	Source
1	Eastern Sudanic	Nilotic (Eastern)	Maasai [masa1300]	Kenya, Tanzania	MN	Tone	VAO	VS	yes	(Tucker & Mpaayei 1955)
2	Eastern Sudanic	Nilotic (Eastern)	Maasai (Arusa) [arus1243]	Tanzania	MN	Tone	VAO	unclear	yes	(Levergood 1987)
3	Eastern Sudanic	Nilotic (Eastern)	Turkana [turk1308]	Kenya	MN	Tone	VAO\ VOA	VS	yes	(Dimmendaal 1983, 1985)
4	Eastern Sudanic	Nilotic (Eastern)	Toposa [topo1242]	South Sudan	MN	Tone	VAO	VS	yes	(Schröder 2008)
5	Eastern Sudanic	Nilotic (Eastern)	Teso [testo1249]	Uganda, Kenya	MN	Tone	VAO	VS	yes	(Barasa 2017)
6	Eastern Sudanic	Nilotic (Eastern)	Samburu [samb1315]	Kenya	MN	Tone	VAO	VS	yes	(Heine 1980)
7	Eastern Sudanic	Nilotic (Eastern)	Lopit [lopi1242]	South Sudan	MN	Tone	VAO	VS	yes	(Moodie & Billington 2020)

Table A.1 continued from previous page

#	Language family	Branch	Language	Geographical area	Case marking system	Marking strategy	Default order (TR)	Default order (INTR)	Flexible	Source
8	Eastern Sudanic	Nilotic (Southern)	Cherang'any [mark1255]	Kenya	MN	Tone	VAO	VS	yes	(Mietzner 2016)
9	Eastern Sudanic	Nilotic (Southern)	Datooga (Gisamjanga) [dato1239]	Tanzania	MN	Tone	VAO	VS	yes	(Kießling 2007)
10	Eastern Sudanic	Nilotic (Southern)	Datooga (Asimjeega) [isim1234]	Tanzania	MN	Tone	AVO	VS\SV	yes	(Griscom 2019)
11	Eastern Sudanic	Nilotic (Southern)	Nandi [nand1266]	Kenya	MN	Tone	VAO	VS	yes	(Creider & Creider 1989)
12	Eastern Sudanic	Nilotic (Southern)	Tugen [tuge1241]	Kenya	MN	Tone	VAO\ VOA	VS	yes	(Jerono 2011)
13	Eastern Sudanic	Nilotic (Southern)	Akie [mosi1247]	Tanzania	MN	Tone	VAO	VS	yes	(Heine et al. 2015)
14	Eastern Sudanic	Nilotic (Western)	Päri [pari1256]	South Sudan	ERG	Morph	OVA	SV	partly	(Andersen 1988)

Table A.1 continued from previous page

#	Language family	Branch	Language	Geographical area	Case marking system	Marking strategy	Default order (TR)	Default order (INTR)	Flexible	Source
15	Eastern Sudanic	Nilotic (Western)	Anywa [anua1242]	Ethiopia, South Sudan	ERG	Morph	OVA	SV	yes	(Reh 1996)
16	Eastern Sudanic	Nilotic (Western)	Shilluk [shil1265]	South Sudan	ERG	Morph	OVA	SV	yes	(Miller & Gilley 2001)
17	Eastern Sudanic	Nilotic (Western)	Jur-Luwo [luwo1239]	South Sudan	ERG	Morph	OVA	SV	yes	(Storch 2014)
18	Eastern Sudanic	Nilotic (Western)	Burun [buru1301]	Sudan, South Sudan	MN	Synt.	AVO	SV	yes	(Andersen 2015)
19	Eastern Sudanic	Nilotic (Western)	Dinka (Agar) [agar1249]	South Sudan	MN or ERG	Tone	AVO\ OVA	SV	partly	(Andersen 1991, 2002)
20	Eastern Sudanic	Nilotic (Western)	Dinka (Southeastern) [sout2834]	South Sudan	MN or ERG	Tone	AVO	SV	partly	(van Urk 2015)
21	Eastern Sudanic	Surmic (South)	Murle [murl1244]	South Sudan	MN	Morph	VAO	VS	partly	(Arensen 1982)

Table A.1 continued from previous page

#	Language family	Branch	Language	Geographical area	Case marking system	Marking strategy	Default order (TR)	Default order (INTR)	Flexible	Source
22	Eastern Sudanic	Surmic (South)	Tirma-Chai (Tirma) [tirm1240]	Ethiopia	MN	Comb.	AVO	SV	yes	(Bryant 1999)
23	Eastern Sudanic	Surmic (South)	Tirma-Chai (Chai) [chai1252]	Ethiopia	ERG	Morph.	AVO	SV	yes	(Last & Lucassen 1998)
24	Eastern Sudanic	Surmic (South)	Mursi [murs1242]	Ethiopia	MN	Morph.	AVO	SV	yes	(Worku 2020)
25	Eastern Sudanic	Surmic (South)	Baale [kaci1244]	Ethiopia	MN	Comb.	AVO	unclear	yes	(Yigezu & Dimmendaal 1998)
26	Eastern Sudanic	Surmic (South)	Tennet [tenn1246]	South Sudan	MN	Comb.	VAO	VS	partly	(Randal 1998)
27	Eastern Sudanic	Surmic (North)	Majang [maja1242]	Ethiopia	ERG	Tone	VAO	VS	partly	(Joswig 2019)
28	Eastern Sudanic	Jebel	Gaahmg [gaam1241]	Sudan	ERG	Tone	AVO	SV	yes	(Stirtz 2014)

Table A.1 continued from previous page

#	Language family	Branch	Language	Geographical area	Case marking system	Marking strategy	Default order (TR)	Default order (INTR)	Flexible	Source
29	Koman	Central Koman	Uduk [uduk1239]	Sudan, Ethiopia, South Sudan	ERG	Synt.	AVO\ OVA	SV	partly	(Killian 2015)
30	Eastern Sudanic	Berta	Berta [bert1248]	Ethiopia, Sudan	MN	Tone	AVO\ OVA	SV	partly	(Andersen 1995)
31	Gumuz	Gumuz	Gumuz (Southern) [sout3236]	Ethiopia, Sudan	MN	Morph.	AVO	SV	yes	(Ahland 2012)
32	Gumuz	Gumuz	Gumuz (Northern) [gumu1244]	Ethiopia, Sudan	MN	Morph.	AVO	SV	yes	(Ahland 2012)
33	Kuliak		Ik [ikkk1242]	Kenya, Uganda	ACC	Morph.	VAO	VS	partly	(Schrock 2014)

Table A.2: Languages with no case marking.

#	Language family	Branch	Language	Geographical area	Default order (TR)	Default order (INTR)	Flexible	Source
1	Eastern Sudanic	Nilotic (Eastern)	Kuku [kuku1285]	South Sudan, Uganda	AVO	unclear	unclear	(Cohen 2000)
2	Eastern Sudanic	Nilotic (Western)	Jumjum [jumj1238]	South Sudan	OVA\ AOV	SV	partly	(Andersen 2018)
3	Eastern Sudanic	Nilotic (Western)	Luo [luok1236]	Kenya, Tanzania	AVO	SV	unclear	(Tucker 1994)
4	Eastern Sudanic	Nilotic (Western)	Lango [lang1324]	Uganda	AVO	SV	partly	(Noonan 1992)
5	Eastern Sudanic	Nilotic (Western)	Nuer [nuer1246]	Ethiopia, South Sudan	AOV	SV	unclear	(Grossman & Faust 2015)
6	Eastern Sudanic	Nilotic (Western)	Reel [reel1238]	South Sudan	AOV	SV	yes	(Cien et al. 2016)
7	Eastern Sudanic	Nilotic (Western)	Belanda Bor [bela1256]	South Sudan	AVO	SV	no	(von Heyking 2013)
8	Eastern Sudanic	Nilotic (Western)	Acoli [acol1236]	South Sudan, Uganda	AVO	SV	no	(Heine & König 2010)

Table A.2 continued from previous page

#	Language family	Branch	Language	Geographical area	Default order (TR)	Default order (INTR)	Flexible	Source
9	Eastern Sudanic	Nilotic (Western)	Alur [alur1250]	Congo, Uganda	AVO	unclear	unclear	(Ringe 1953)
10	Central Sudanic	Moru-Madi (Central)	Keliko [keli1248]	Congo, South Sudan, Uganda	AVO	unclear	partly	(Waag 2018)
11	Central Sudanic	Moru-Madi (Central)	Lugbara [lugb1240]	Congo, Uganda	AVO	SV	partly	(Crazzolara 1960)
12	Central Sudanic	Moru-Madi	Moru [moru1253]	South Sudan	AVO	SV	partly	(Andersen 1984)
13	Central Sudanic	Moru-Madi (Southern)	Ma'di [madi1260]	South Sudan, Uganda	AVO	SV	partly	(Blackings & Fabb 2003)
14	Central Sudanic	Moru-Madi (Southern)	Olu'bo [olub1238]	South Sudan	AVO	SV	unclear	(Tucker 1940b)
15	Eastern Sudanic	Surmic (South)	Kwegu [kweg1241]	Ethiopia	AVO	SV	no	(Hieda 1998)
16	Eastern Sudanic	Surmic (South)	Me'en [meen1242]	Ethiopia	AVO	SV	partly	(Joswig et al. 2011)

Table A.2 continued from previous page

#	Language family	Branch	Language	Geographical area	Default order (TR)	Default order (INTR)	Flexible	Source
17	Kuliak	Ngangea-So	Nyang'i [nyan1313]	Uganda	VAO	VS	unclear	(Beer 2017)
18	Kuliak	Ngangea-So	Soo [sooo1256]	Uganda	VAO	VS	unclear	(Carlin 1993)
19	Eastern Sudanic	Nara	Nara [nara1262]	Eritrea	AOV	SV	unclear	(Elnur 2016)
20	Eastern Sudanic	Temeinian	Temein [nucl1339]	Sudan	AVO	unclear	unclear	(Tucker & Bryan 1966)
21	Central Sudanic	Lenduic	Ngiti [ngit1239]	Congo, Uganda	AVO	SV	partly	(Kutsch Lojenga 1994)
22	Central Sudanic	Mangbetu-Asua (Mangbetuic)	Asoa [asoa1238]	Congo	AVO	unclear	unclear	(Beltrame 1876, 1877, 1877)
23	Central Sudanic	Sara-Bongo-Bagirmi (Baka-Beli)	Beli [beli1257]	South Sudan	AVO	SV	unclear	(Kol et al. 2013)
24	Central Sudanic	Sara-Bongo-Bagirmi (Baka-Beli)	Jur Modo [jurm1239]	South Sudan	AVO	SV	unclear	(Andersen 1981)

Table A.2 continued from previous page

#	Language family	Branch	Language	Geographical area	Default order (TR)	Default order (INTR)	Flexible	Source
25	Koman	Central Koman	Opo [opuu1239]	Ethiopia, South Sudan	AVO	SV	yes	(Mellese 2017)
26	Saharan	Eastern Saharan	Beria [zagn1240]	Tchad	AOV	SV	partly	(Jakobi & Crass 2004)

Table A.3: Languages with case marking but no interaction with constituent order.

#	Language family	Branch	Language	Geographical area	Case marking system	Default order (TR)	Default order (INTR)	Flexible	Source
1	Eastern Sudanic	Taman (Tamam-Sungor)	Tama [tama1331]	Sudan, Chad	ACC	AOV	SV	partly	(Dimmendaal 2009), (Kellermann 2000)
2	Eastern Sudanic	Nubian (Nobiin Nubian)	Nobiin [nobi1240]	Egypt, Sudan	ACC	AOV	SV	no	(Werner 1987)
3	Eastern Sudanic	Nubian (Central)	Dongola [dong1288]	Sudan	ACC	AOV	SV	no	(Armbruster 1960)
4	Eastern Sudanic	Nubian (Central)	Kenuzi [kenu1243]	Egypt	ACC	AOV	SV	no	(Abdel-Hafiz 1988)
5	Eastern Sudanic	Nubian (Kordofan)	Ucunwee [ghul1238]	Sudan	ACC	AOV	unclear	unclear	(Williams & Comfort no date)
6	Eastern Sudanic	Nubian (Kordofan)	Dilling [dill1242]	Sudan	ACC	unclear	unclear	unclear	(Kauczor 1920)
7	Eastern Sudanic	Nubian (West Central)	Midob [mido1240]	Sudan	ACC	unclear	unclear	unclear	(Werner 1993)

Table A.3 continued from previous page

#	Language family	Branch	Language	Geographical area	Case marking system	Default order (TR)	Default order (INTR)	Flexible	Source
8	Central Sudanic	Membi-Mangbutu-Efe (Mangbutu-Efe)	Lese [lese1243]	Congo	ACC	AVO	SV	unclear	(Vorbichler 1965)
9	Central Sudanic	Membi-Mangbutu-Efe (Mangbutu-Efe)	Mamvu [mamv1243]	Congo	ACC	AVO	SV	yes	(Vorbichler 1971)
10	Central Sudanic	Sara-Bongo-Bagirmi (Baka-Beli)	Baka [baka1274]	Congo, South Sudan	ACC	AVO	SV	unclear	(Santandrea 1976)
11	Maban	Mabang (Maba-Masalit)	Masalit [nucl1440]	Sudan	ACC	AOV	unclear	unclear	(Edgar 1989)

Table A.3 continued from previous page

#	Language family	Branch	Language	Geographical area	Case marking system	Default order (TR)	Default order (INTR)	Flexible	Source
12	Maban	Mabang (Maba-Masalit)	Maba [maba1277]	Chad	ACC	AOV	SV	partly	(Weiss 2009)
13	Maban	Mabang (Runga-Kibet)	Runga [rung1258]	Sudan, Chad, Central African Republic	ACC	AOV	unclear	unclear	(Nougayrol 1990)
14	Saharan	Western Saharan (Tebu)	Dazaga [daza1242]	Tchad	ACC	AOV	SV	partly	(Walters 2016)
15	Furan		Fur [furr1244]	Sudan, Chad, Central African Republic	ACC	AOV	unclear	unclear	(Waag 2010)
16	Kunama		Kunama [kuna1268]	Eritrea, Ethiopia, Sudan	ACC	AOV	unclear	unclear	(Bender 1996)

Table A.4: Languages with case marking, but the type of system and/or interaction with constituent order is unclear.

#	Language family	Branch	Language	Geographical area	Unclearity	Default order (TR)	Default order (INTR)	Flexible	Source
1	Eastern Sudanic	Nilotic (Eastern)	Karamojong [kara1483]	Uganda, Kenya, South Sudan	Case system/ Interaction	VAO	unclear	unclear	(Novelli 1985)
2	Eastern Sudanic	Nilotic (Eastern)	Ongamo [ngas1238]	Tanzania	Interaction	VAO	unclear	partly	(Heine & Voßen 1976)
3	Eastern Sudanic	Nilotic (Eastern)	Nyangatom [nyan1315]	Ethiopia, South Sudan	Case system/ Interaction	VAO	VS	unclear	(Schröder & Kadanya 2011)
4	Eastern Sudanic	Nilotic (Southern)	Omotik [omot1239]	Kenya, Tanzania	Case system/ Interaction	unclear	unclear	unclear	(Rottland 1982)
5	Eastern Sudanic	Nilotic (Southern)	Kipsigis [kips1239]	Kenya	Interaction	unclear	unclear	unclear	(Toweett 1979)
6	Eastern Sudanic	Nilotic (Southern)	Pökoot [poko1263]	Kenya, Uganda	Interaction	VAO	VS	unclear	(Baroja et al. 1989)
7	Eastern Sudanic	Nilotic (Western)	Dinka (Agar) [agar1249]	South Sudan	Case system	AVO/ OVA	SV	partly	(Andersen 1991, 2002)

Table A.4 continued from previous page

#	Language family	Branch	Language	Geographical area	Unclearity	Default order (TR)	Default order (INTR)	Flexible	Source
8	Eastern Sudanic	Nilotic (Western)	Dinka (Southeastern) [sout2834]	South Sudan	Case system	AVO	SV	partly	(van Urk 2015)
9	Eastern Sudanic	Surmic (South)	Narim [nari1240]	South Sudan	Interaction	VAO	unclear	unclear	(Stirtz 2011), (Joseph et al. 2013)
10	Eastern Sudanic	Surmic (South)	Didinga [didi1258]	South Sudan	Interaction	VAO	unclear	unclear	(Odden 1983), (Lohitare et al. 2012)
11	Eastern Sudanic	Nyima	Afitti [afit1238]	Sudan	Case system/ Interaction	unclear	unclear	unclear	(de Voogt 2011)
12	Eastern Sudanic	Nyima	Ama [amas1236]	Sudan	Interaction	unclear	unclear	unclear	(Stevenson 1938)
13	Eastern Sudanic	Nubian (Kordofan)	Karko [kark1256]	Sudan	Interaction	AOV	SV	unclear	(Jakobi & Dimmendaal 2022)
14	Eastern Sudanic	Nubian (Nobiin)	Old Nubian [oldn1245]	Sudan	Interaction	AOV	unclear	unclear	(Gerven Oei 2021)
15	Eastern Sudanic	Taman	Mararit [mara1396]	Chad	Interaction	AOV	SV	unclear	(Lukas 1933)

Table A.4 continued from previous page

#	Language family	Branch	Language	Geographical area	Unclearity	Default order (TR)	Default order (INTR)	Flexible	Source
16	Saharan	Western Saharan (Kanuri-Kanembu)	Kanembu [kane1243]	Tchad, Niger	Interaction	unclear	unclear	unclear	(Koelle 1854)
17	Saharan	Eastern Saharan	Berti [bert1249]	Sudan	Interaction	unclear	unclear	unclear	(Petráček 1966)

Table A.5: Languages with unclear case marking.

#	Language family	Branch	Language	Geographical area	Default order (TR)	Default order (INTR)	Flexible	Source
1	Eastern Sudanic	Nilotic (Eastern)	Bari [bari1284]	South Sudan, Uganda	unclear	unclear	unclear	(Spagnolo 1933), (Yokwe 1987)
2	Eastern Sudanic	Nilotic (Eastern)	Mandari [mand1425]	South Sudan	unclear	unclear	unclear	(Lutwori et al. 2013)
3	Eastern Sudanic	Nilotic (Eastern)	Okolie [okol1243]	South Sudan, Uganda	unclear	unclear	unclear	(Flores 1971)
4	Eastern Sudanic	Nilotic (Eastern)	Lokoya [loko1254]	South Sudan	VAO	unclear	partly	(Westermann 1944)
5	Eastern Sudanic	Nilotic (Eastern)	Otuho [otuh1238]	South Sudan, Uganda	VAO	unclear	partly	(Muratori 1938)
6	Eastern Sudanic	Nilotic (Western)	Adhola [adho1243]	Kenya, Uganda	AVO	unclear	unclear	(Mukarovsky 1956)
7	Eastern Sudanic	Nilotic (Southern)	Kupsabiny [kups1238]	Kenya, Uganda	unclear	unclear	unclear	(Montgomery 1966)
8	Eastern Sudanic	Nilotic (Southern)	Sabaot [saba1262]	Kenya, Uganda	unclear	unclear	unclear	(Larsen 1991)

Table A.5 continued from previous page

#	Language family	Branch	Language	Geographical area	Default order (TR)	Default order (INTR)	Flexible	Source
9	Eastern Sudanic	Nilotic (Southern)	Okiek [okie1245]	Kenya, Tanzania	unclear	unclear	unclear	(Micheli 2018)
10	Eastern Sudanic	Dajuic (Eastern)	Logorik [logo1261]	Sudan	unclear	unclear	unclear	(Alamin Mubarak 2006)
11	Eastern Sudanic	Dajuic (Eastern)	Shatt [shat1244]	Sudan	unclear	unclear	unclear	(Alfira et al. 2016)
12	Eastern Sudanic	Dajuic (Western)	Dar Fu Daju [darf1239]	Sudan	unclear	unclear	unclear	(Ismail 2000)
13	Eastern Sudanic	Tama (Sungor)	Assangori [assa1269]	Chad	unclear	unclear	unclear	(Lukas 1938)
14	Central Sudanic	Lenduic (Bale)	Lendu [lend1245]	Congo, Uganda	unclear	unclear	unclear	(Tucker 1940a)
15	Central Sudanic	Mangbetu-Asua (Mangbetuic)	Mangbetu [mang1394]	Congo	unclear	unclear	unclear	(Larochette 1958)

Table A.5 continued from previous page

#	Language family	Branch	Language	Geographical area	Default order (TR)	Default order (INTR)	Flexible	Source
16	Central Sudanic	Membi-Mangbutu-Efe (Mangbutu-Efe)	Efe [efee1239]	Congo	unclear	unclear	unclear	(Smith 1938)
17	Central Sudanic	Membi-Mangbutu-Efe	Birri [birr1240]	Central African Republic	unclear	unclear	unclear	(Santandrea 1966)
18	Central Sudanic	Moru-Madi (Central)	Avokaya [avok1242]	South Sudan, Congo	AVO	unclear	partly	(Dimmendaal 1986)
19	Central Sudanic	Moru-Madi (Central)	Omi [omii1238]	Congo	unclear	unclear	unclear	(Goyvaerts 1986)
20	Central Sudanic	Moru-Madi (Central)	Logo [logo1259]	Congo, South Sudan	unclear	unclear	unclear	(Irumu 1986)
21	Central Sudanic	Sara-Bongo-Bagirmi	Bongo [bong1285]	South Sudan	AVO	unclear	unclear	(Santandrea 1963)

Table A.5 continued from previous page

#	Language family	Branch	Language	Geographical area	Default order (TR)	Default order (INTR)	Flexible	Source
22	Koman	Central Koman	Komo [komo1258]	Ethiopia, South Sudan, Sudan	AVO	unclear	partly	(Bayou 2014)
23	Maban	Mabang (Runga-Kibet)	Kibet [kibe1241]	Chad	unclear	unclear	unclear	(Nougayrol 1986)
24	Koman		Gwama [kwam1249]	Ethiopia	AVO	unclear	partly	(Leyew 2005), (Kievit & Robertson 2012)
25	Gumuz		Daats'iin [daat1234]	Ethiopia	AVO	unclear	unclear	(Ahland 2016)

#	**Language family**	**Branch**	**Language**	**Geographical area**	**Unavailable references**
1	Eastern Sudanic	Nilotic (Eastern)	Kakwa [kakw1240]	South Sudan, Uganda, Congo	(Unyutha 1975)
2	Eastern Sudanic	Nilotic (Southern)	Keiyo [keiy1238]	Kenya	(Tornay 1970)
3	Eastern Sudanic	Nilotic (Western)	Mabaan [maba1273]	Sudan, South Sudan	(Andersen 1999)
4	Eastern Sudanic	Nilotic (Western)	Kumam [kuma1275]	Uganda	(Hieda 2013)
5	Eastern Sudanic	Nilotic (Western)	Dinka (Southwestern) [sout2832]	Sudan, South Sudan	(Nebel 1978)
6	Central Sudanic	Membi-Mangbutu-Efe	Ndo [ndoo1242]	Congo, Uganda	(Banga 1978)
7	Central Sudanic	Moru-Madi (Southern)	Southern Ma'di [sout2828]	Uganda	(Spire 1905)

Table A.6: Unavailable grammars or language descriptions at the time of the survey.