A functional approach to differential indexing: combining perspectives from typology and corpus linguistics

Just, E.C.

Chapter 5

Variable index placement in Gutob from a typological perspective[1]

Abstract

Gutob (Munda, India) displays a special kind of differential indexing in that S/A indexes can attach to other hosts apart from the verb, unconstrained by syntax. Previous studies have described non-verbal index placement in Gutob as exceptional, establishing verbal indexes as the default; however, a corpus based analysis has still been owing until now. Comparative studies on variation in the placement of indexes show that there is not only inter-linguistic variation with regard to index placement, but in some cases also intralinguistic variation. Against this background, we present a case study on index placement in Gutob based on quantitative corpus data. Our analysis shows that although index placement in Gutob is in fact conditioned by discourse effects, non-verbal clitics cannot be considered particularly exceptional. Strikingly, we observe that indexing does not succumb to discourse, but can itself be used to structure it, marking the hosts as particularly noteworthy.

[1]This chapter is submitted as: Just, Erika & Voß, Judith. *Variable index placement in Gutob from a typological perspective.* Author contributions: EJ and JV worked out the annotation scheme together; EJ wrote the paper; JV provided the expertise on Gutob, as well as the corpus; annotation was largely carried out by student assistant Luna Hemmerling under the supervision of EJ and JV.

5.1 Introduction

In this study we investigate a special type of differential indexing, i.e. variation in the encoding of referents through bound person marking (traditionally referred to as agreement). Whereas differential indexing as described in previous studies (e.g. Iemmolo 2011) typically refers to conditions under which an index is present or absent, we are concerned with the position of indexes, in the following referred to as variable index placement.

In this section we will first elaborate on the concepts relevant for our analysis, including information-structural concepts. The remainder of the chapter is structured as follows: Section 5.2 will give an overview over variable index placement in various languages, first providing some theoretical background in Section 5.2.1, then illustrating these findings with examples from some languages where indexing is not confined to one position only but still syntactically determined (Section 5.2.2). We will then turn to syntactically unconstrained index placement in Section 5.2.3. Section 5.3 will start off with a summary of referent indexing in Munda languages, followed by our Gutob case study in Section 5.4. After a brief introduction to the language, we present the formal properties of S/A indexes in the language in Section 5.4.2, before presenting our corpus-based findings in Sections 5.4.3 and 5.4.4. Section 5.5 will elaborate on the discourse effects of index placement in Gutob, and Section 5.6 concludes the study with some final remarks.

5.1.1 Differential Indexing

Indexes are defined as bound markers expressing argument features, most commonly person and number, and most commonly attached to the verbal predicate. Indexing (Haspelmath 2013) is a more neutral term than agreement, as it does not presuppose any syntactic relationship between the marker and the co-referential NP, nor whether the latter is obligatorily expressed (Haig & Forker 2018). Also, the morphological status of the index, as a clitic or an affix, is irrelevant, as the latter is often unjustifiably equated with obligatoriness of marking when it comes to agreement or indexing (Haig & Forker 2018: 720). This chapter discusses differential indexing as a type of differential marking, which in turn refers to any situation where an argument of a predicate bearing the same generalized semantic role may be coded in differ-

ent ways, depending on factors other than the argument role itself (Witzlack-Makarevich & Seržant 2018: 16).

This definition of differential marking captures changes in marking patterns, but it does not imply any conditions on the differences in coding. Thus, the term differential marking can refer to differential flagging, i.e. case marking and adpositions, as well as to differential indexing. Also, it can involve the uses of different markers, the general presence of marking, or, as in the present case, the position of a marker in a clause. The definition also includes both differential marking due to predicate properties (such as TAM, polarity or clause type) as well as argument properties (both inherent as well as non-inherent).

The study of differential indexing has largely been focusing on the presence vs. the absence of indexes, both in language or family specific studies as well as in typological ones (e.g. Arkadiev 2010 or Iemmolo 2011). There has not been much cross-linguistic work on variable index placement (but see Cysouw 2003 and Forker 2016). This type of differential indexing that is not characterized by whether there *is* an index or not, but *where* the respective index is placed, is illustrated by (20). It is a minimal pair from a conversation, showing how the S/A index can attach to different constituents: the noun specifying what was brought by the guests in the first sentence (palm wine), and the amount of it (one goria) in the third sentence.

(20) Gutob

Speaker A: *iŋɋi? solop=nen gor-ek riŋ-o? ɖugu*
HES palm.wine=3PL goria-one bring-CVB AUX.PST
'Eh, they had brought one goria of palm wine.'

Speaker B: *riŋ-o?=nen ɖugu*
bring-CVB=3PL AUX.PST
'Had they brought it?'

Speaker A: *ũ solop gor-ek=nen riŋ-o? ɖugu*
yes palm.wine goria-one=3PL bring-CVB AUX.PST
'Yes, they had brought one goria full of palm wine [...].'

Examining different studies on the placement of indexes (e.g. Capell 1972, Barbosa 1996, Harris 2000, Baker 2002 or Dixon 2002) shows that there is both interlinguistic and intralinguistic variation with regard to index placement. Not only do different languages prefer different positions for referent

indexing, but very often a given language has several potential positions for an index. And although Cysouw's (2003) study already provides an insightful description of the cross-linguistic variation with regard to the placement of indexes, there is to our knowledge no corpus-driven study to account for the internal variation in a language in which the placement of the index is not predictable by grammatical rules but is sensitive to pragmatic and/or semantic factors.

Such lack of hard coding for differential marking phenomena makes them especially liable to what is often called "optionality" in reference grammars and other language specific studies. This usage of this term somewhat blurs the fact that the choice of one marking strategy over another might be well motivated, albeit not necessarily syntactically. Usage based studies have shown that even though certain constructions might not be put down to a single factor which is easy to discern, the respective form serves an intentional communicative goal on the part of the speaker (see Schikowski 2013 on DOM in Nepali, Erika & Witzlack-Makarevich (accepted) on differential P indexing in Ruuli, or Just & Čéplö (to appear) on the same phenomenon in Maltese). Section 5.1.2 now briefly elaborates on information-structural categories, especially *topic* and *focus*, and how they are dealt with in the present study.

5.1.2 A note on information-structural categories

Unlike morphological or semantic referential properties (such as gender or animacy, respectively), information-structural phenomena are often difficult to identify as the cause of differential marking patterns (Witzlack-Makarevich & Seržant 2018: 10–11). The reason for that is the variety of discourse phenomena associated with the traditional categories topic and focus, and the linguistic diversity both in form and function of the features involved. It is therefore questionable to conceive of topic and focus as language external universal categories reflected in cross-linguistically stable categories (Matić & Wedgwood 2013, Ozerov 2018, Ozerov 2021).

It has been generally accepted that topic is associated with givenness, a high degree of identifiability, and is assumed to relate to the hearer's knowledge. Focus, on the the other hand, brings about an information update, and is thus associated with notions such as newness or contrast. And although there are comparable constructions in different languages which are ascribed to the

concepts of topic or focus (such as left-dislocation or clefts), they are used to map different types of interactional management (Ozerov 2018) and there is no one-to-one relation between recurrent structures and their pragmatic effects in the respective languages (e.g. Gómez González 1997 or Skopeteas & Fanselow 2010).

Following Matić & Wedgwood (2013), Ozerov (2018) convincingly argues that topic and focus are not universal categories but rather constitute umbrella terms for a pool of different features such as – in the case of focus – contrast, correction, or an answer to a content question (also see Mushin 2006: 292–293). This clustering of features under a single term like focus has lead to the application of testing methods which constrain the interpretation of a linguistic form: once a form has been ascribed to expressing one of the prototypical features (e.g. if it marks contrast or the new piece of information in an answer), some of its other functions might be overlooked, resulting in a biased or incomplete picture of its actual contribution to information management[2].

Differential indexing entailing the absence of an otherwise present index has often been considered to be such a structure: referent indexing (often, but not exclusively, of S/A) can be suspended if the respective referent is in focus (see e.g. Ouhalla 1993, Lambrecht & Polinsky 1997, Mereu 1999 and Siewierska 2004: 159–162). But differential marking manifested in the *placement* of an index rather than its absence has been ascribed to focus of the *host* constituent (Cysouw 2003, Forker 2016).

Considering example (20) from Gutob above, it seems tempting to assume the index marks focus, as it can be interpreted as emphasizing its host constituent: *what* was brought by the guests in the first sentence in (1) and, upon further request on the part of the dialogue partner, *the amount* of what has been brought in the last one. However, due to the reasons just outlined, we avoid an *a priori* establishment of a focus category for Gutob in the present study, and instead give a bottom-up description of the motivations for the shift of the index from the verbal predicate. Nevertheless, the notions of focus and topic have played a central role in the description of differential indexing phenomena, and especially focus has been considered very important in accounts

[2]The term information management is used as an alternative to information structure, bypassing the challenges of the traditional notions associated with the latter (Ozerov 2018, 2021)

of index placement. Thus, the terms are used in Section 5.2, whenever we adopt them, as they have been used in the respective studies.

5.2 Variable index placement

In the following section we consider the distinction between indexes which have a dedicated host and those which have variable hosts. As for the latter, one can further differentiate between those which occupy a fixed position syntactically and those which do not, as is the case in Gutob.

5.2.1 Typological overview

Siewierska (2004: 26–32) gives an extensive description of indexes in various languages which are not always attached to a particular type of stem (and which are therefore not "bound" in her terminology) but which have a designated syntactic position. Designated syntactic position means that there is some variation with regard to the part-of-speech of the host word, however the position of the index within the clause is nevertheless grammatically determined and not flexible. Following Anderson (1993: 74), she uses the typology of specialized positions listed in (21) for indexes which do not always attach to the same stem. She also mentions languages in which two of the positions in Anderson's (1993: 74) typology are possible (Siewierska 2004: 26-32).

(21) a. verb phrase initial position
b. verb phrase final position
c. second position
d. penultimate position
e. pre-head position
f. post-head position

Example (22) from Kharia and (23) from Vera'a illustrate indexes showing variability with regard to their host, but occupying syntactically fixed positions. In Kharia, the S/A index is an enclitic to the verb in affirmative clauses, as in (22a), whereas it attaches to the negative particle in negated clauses, as in (22b). According to Anderson's (1993) typology, indexes in Kharia are therefore either in a pre-head or in a post-head position within the

verb phrase,[3] conditioned by polarity, and no other position is possible. In Vera'a, S/A referents are only indexed in the aorist, and not all person markers in this paradigm are phonologically bound. However, the ones which are bound (as the non-singular marker *=k* in (23b)) attach to whatever element precedes the predicate, which in most cases is the last word of the S/A NP, or an adverb of a closed class, which can intervene between the S/A NP and the predicate (Schnell 2018: 750–752).

(22) Kharia (Munda, Peterson 2011: 58)

a. *kayom=ta=ɲ*
speak=PRS=1SG
'I speak.'

b. *um=iɲ kayom=ta*
NEG=1SG speak=PRS
'I don't speak.'

(23) Vera'a (Oceanic, Schnell 2018: 759)

a. *dē=k van 'ō' di mē=n sisidiñ*
1PL.INCL=AOR:NSG go carry 3SG DAT=ART RDPL.bird.trap
'and we will go bird catching with him.'

b. *gidu=k van=ēk traem*
1DU.INCL=AOR:NSG go=AOR:NSG try
'"Let's go try!"' [lit: "We two will/shall go, (we) will/shall try."]

Example (23) from Vera'a shows that the typology of index positions in (21) is not exhaustive: the index is not part of the verb phrase (therefore neither in position a., b., e. or f.), nor is it in second or penultimate position with regard to the whole clause. It is "detached" (Bickel & Nichols 2007: 176) from its predicate, but unlike positions c. and d., its position is fixed in relation to the predicate, always directly preceding it.

5.2.2 More than one option for the position of an index

As outlined in the previous section, even in systems where indexing has grammaticalized in a sense that it becomes obligatory (even if sometimes only in parts of the paradigm), the index can still be flexible in selecting a host, which

[3] These indexes have also been called anticipatory clitics (Peterson 2011: 61–62, see also Dixon & Aikhenvald 2002: 46).

once again demonstrates that there is no equation of obligatoriness with the status (as an affix or clitic) of marking (see Haig & Forker 2018: 720).

There are, however, languages which display an intermediate position between a grammatically conditioned, fixed syntactical position for an index on the one hand, and great freedom of position on the other hand: in such languages, the index has a default position, often the head of the predicate, but can leave it and appear in an alternative, syntactically fixed position. One of these languages, also reported by Siewierska (2004: 27–29), is Nganhcara, where indexes can occur in two syntactic positions. Smith & Johnson (1985: 104) state that the favored position is that encliticized to the last element before the verb (as in 24a), though indexes also occur encliticized to the verb itself (as in 24b).

(24) Nganhcara (Pama-Nyungan, Smith & Johnson 1985: 102, 106)

a. *Nganhca nga'a-nhca yenta*
1PL.EXCL.NOM fish-1PL.EXCL.NOM spear
'We speared the fish.'

b. *Nganhca nhingu nga'a waa-ngu-nhca*
1PL.EXCL.NOM 3SG.DAT fish give-3SG.DAT-1PL.EXCL.NOM
'We gave him a fish.'

As the placement of the index in these examples is not determined by phonological, morphological or syntactic factors, it probably serves a communicative goal of the speaker. The discourse function of index placement is somewhat under-studied, but Cysouw (2003) investigates the attraction of indexes to various positions of discourse prominence in a sample of 40-odd languages. He aligns the different positions of index placement with the status of its host with regard to focus. Indexes which are not confined to the head of the predicate most often attach to elements which are considered to be inherently focused, such as question words and negation markers. Next in the focus hierarchy are constituents with intended focus, i.e. their focus status arises out of a particular situation, as, for example, NPs in contrastive or emphatic contexts. The two other focused contexts that play a role in this hierarchy are stage setting (clause linkers and adverbs), and modified (indefinite and quantified) NPs.

As an explanation for this attraction of indexes towards focal elements, Cysouw (2003) proposes that as indexes themselves are highly topical, and

therefore non-focal, this combination is a "juncture of opposites", the less focal element binding itself on to the most focal one. He also demonstrates that clause-second position is very frequently used for indexes, either as default (such as in some Pama-Nyungan languages like Yingkarta, Wajarri, Ngiyambaa or Warlpiri as well as in the Uto-Aztecan language Yaqui[4]) or as an alternative to a position within the verb phrase (e.g. in Suleimaniye Kurdish or Cypriot Greek).

An example for clause-second position as an alternative to verbal position comes from Kuuk Thaayorre. In this language, there is differential indexing in two ways: i) the index is not (yet) grammatically obligatory (Gaby 2006: 342–343)[5], as exemplified by (25a), which would be equally grammatical without *=ay*, and in (25b) where the third person singular accusative *=unh* features twice; and ii) the position of the index alternates between clause-second position, as in (25c) and (25d), and verb-final position, as in (25a), which is preferred (Gaby 2006: 216–217).

(25) Kuuk Thaayorre (Pama-Nyungan, Gaby 2006: 217)

a. *ngay ii-rr-kuw Darwin-ak yat=ay*
1SG.NOM there-towards-west Darwin-DAT go.PFV=1SG.NOM
'I went west to Darwin.'

b. *thil=unh koow rathirr=eln=unh*
again=3SG.ACC nose.ACC chop-PFV=3PL.ERG=3SG.ACC
'They slashed his nose once more.'

c. *inh'nhul=ay yik, kuuk inh'nhul*
this.one=1SG.NOM say-NPST word this.on
'I'm telling this story.'

d. *ngul=ul=unh man.pert-e theerka-n-r*
then=3SG.ERG=3SG.ACC shoulder-ERG return-TR-NPST
nhaknkath-an
camp-DAT
'And he carried it back home on his shoulder to camp.'

Second position clitics, also called Wackernagel-clitics, are cross-linguistically quite common and person indexes are not the only elements which are

[4]For more information on some Uto-Aztecan languages where the index can shift away from the verb to the clause-second position see e.g. Press (1979: 77) on Chemehuevi or Wistrand Robinson & Armagost (1990: 250–252) on Comanche.

[5]Gaby (2006) argues that the index is in the early stages of grammaticalizing from the respective free pronouns into indexes.

prone to this position, but also other inflectional material, such as TAM or evidentiality markers, or inflected verbs (Anderson 1993, Mushin 2006).[6] There is a considerable amount of literature addressing the syntactic and phonological properties of Wackernagel-clitics, but less on the functions of this position (but see Anderson 1993, Mushin 2006, Mushin & Simpson 2008). Two rather contrary motivations that draw elements to the clause-second position have been proposed. On the one hand, it was suggested that the clause-second position elements (indexes or other inflection) are actually targeting a clause initial position, but are blocked from occurring there due to language-specific phonological or morphosyntactic constraints, and therefore shift to clause-second position (Anderson 2005: 142–152). On the other hand, it was suggested that the elements in question are "bare-bones grammatical information" (Mushin 2006: 296) and thus are attracted to elements in the first position, which, in turn, is generally recognized to be associated with focal effects (e.g. Mithun 1992, McConvell 1996 or Cysouw 2003).

The latter idea of a syntactic "beacon" (Mushin 2006: 296) attracting markers or constituents with low pragmatic impact resulting in Wackernagel-clitics or clause-second position verbs in many languages is compliant with Cysouw's (2003) observation that indexes as "less focal elements" bind themselves "on the most focal element". It also goes well with the fact that there are languages where markers of modality or evidentiality attach either to the verb or to any other focused constituent (Facundes 2000, Aikhenvald 2003).[7]

In a nutshell, there is evidence that indexes (as well as other grammatical markers) are often attracted to constituents that the speaker wants to highlight – and these constituents are frequently found clause initially. This can lead to indexes being found in clause-second position, either exclusively or as an alternative to attaching to the verb. We will now turn to languages where indexes are not confined to one or two syntactic positions but can be attracted to any constituent in the clause.

[6] Clause-second position sounds more straight forward than it is, as there is variation with regard to the rules of attachment and whether the clitics attach to the first (prosodic) word or first (prosodic) constituent of a clause.

[7] As outlined in Section 5.1.1 no two languages have identical categories of focus, i.e. focus marking in language A does not have the exact same pragmatic effect as focus marking in language B; however, there is undoubtedly a set of communicative functions which can be ascribed to this traditional notion, and which overlap to various degrees from language to language (see Ozerov 2018 for a discussion and overview of these features).

5.2.3 Syntactically unconstrained index placement

The placement of indexes may be even less constrained by morphosyntactic criteria than having several alternative syntactic slots. Some language descriptions suggest that the position of an index cannot always be lead back to a hard and fast grammatical rule, and is therefore sensitive to information management. For Mutsun, an extinct Utian language of California, Okrand (1977: 171) reports that indexes are usually second position clitics, following the first word of a sentence, whatever this word may be. However, there are a few exceptions to this rule where the index attaches to other constituents. The motivation for this cannot be discerned based on his data. Also, the distribution of the indexes and respective independent pronouns, which cannot be used together, remains unclear.

The situation seems to be even more complex in Ute (Uto-Aztecan, Givón 2011: 170–192). Here, the same set of indexes can in principle be used for either S, A or P, and compete with free pronouns as well as zero anaphora. Zero anaphora has been identified as means of tracking of S/A if it persists as "agentive" subject; if participants start interacting, indexes are used for the absolutive (S/P) argument. As for the position of the index, Givón (2011: 170) states that it can attach "not only to the verb, but to any first word in the clause". However, in his count of host positions in the clause, he finds that although 81.9% of all non-verbal indexes are in fact in clause-second position, nearly 20% of non-verbal indexes are not clause-second; unfortunately he does not enlarge upon these cases.

The conditions of index placement are described more transparently for Sanzhi Dargwa (Nakh-Daghestanian).[8] The default position of the index is postverbal, but it can also be conditioned by the focal status of the host of the index, a phenomenon called "floating agreement" by Forker (2016): the index leaves its postverbal position (exemplified in 26a) and floats off to constituents which are focal or contribute new information, thus serving to emphasize its host, as e.g. 'the dishes' in (26b) or 'I' in (26c). The emphasized constituents are underlined in the translation. According to Forker, the host can be any other constituent without fixed syntactic position. However, these examples are restricted to elicited sentences (Forker 2016: 1). Which role is indexed

[8] See (Bickel & Nichols 2007: 176–177) quoting (Kibrik 1997) for discussing a similar phenomenon in Tsakhur, another Nakh-Daghestanian language.

in Sanzhi-Dargwa is governed by the person hierarchy 2 > 1 > 3 (Forker 2016: 4).

(26) Sanzhi Dargwa (Nakh-Daghestanian, Forker 2016: 2)

a. *du-l hana t'alaʕħ-ne ic-an = da*
1SG-ERG now dishes-PL wash.IPFV-PTCP = 1
'Now I will / have to wash the dishes.'

b. *du-l hana t'alaʕħ-ne = da ic-an, c'il t'ult'-e*
1SG-ERG now dishes-PL = 1 wash.IPFV-PTCP then bread-PL
d-uc'-an = da
NPL-bake.IPFV-PTCP = 1
'Now I will / have to wash the dishes, later I will make bread.'

c. *du = da Sanijat-li-j χabar b-urs-an*
1SG = 1 Sanijat-OBL-DAT story N-tell.PFV-PTCP
'I will / have to tell Sanijat the story.'

Forker (2016: 20) also mentions Polish, Paez (isolate, Colombia) and Zargulla (Omotic) as further examples for index placement conditioned by information management. In Zargulla, the situation is quite intriguing and deserves to be elaborated on: first of all, S/A indexing is described as "optional", which is in itself an interesting fact worth to be studied in further detail;[9] Amha (2007) mentions that identifiability and animacy play a role to some extent. However, a prerequisite for indexing is the presence of the focus marker *-tte*, i.e. indexing cannot occur on its own, at least not in declarative clauses. The focus marker, on the other hand, can be used on its own without an index, as in (27a), and it can shift to various constituents to mark them for focus, and the index (if present) always moves along, as exemplified in (27b)-(27d). The index can, however, be attached to question words without the focus marker, as in (27e) (Amha 2007: 200–202).

(27) Zargulla (Omotic, Amha 2007: 201–202)

a. *s'úho ʔíndó-y ʔúkkó-tte-ínne*
Tsuho:GEN mother-NOM be.close-FOC-PST
'Tsuho's mother moved closer.'

b. *naʔá-z-í kátsa bays-í, maaʔó = tte-s sang-í,*
child-M-NOM grain.ABS sell-CVB cloth.abs = FOC-3SG.M buy-CVB

[9] In contrast to the indexes in declarative clauses, which can be used only in focused constructions, S/A indexes in the negative interrogative and imperative/optative are obligatory and seem to be well entrenched and older (Amha 2009: 215).

ɗum-us-í *yeénne*
be.dark-CAUS-CVB come.PFV

'The boy sold grain, bought cloth, and came late.'

c. *naʔá-z-í* *kátsa* *bays-í,* *maaʔó* *sang-í=tte-s,*
child-M-NOM grain.abs sell-CVB cloth.ABS buy-CVB=FOC-3SG.M
ɗum-us-í *yeénne*
be.dark-CAUS-CVB come.PFV

'The boy sold grain, bought cloth, and came late.'

d. *naʔá-z-í* *kátsa* *bays-í,* *maaʔó* *sang-í,*
child-M_NOM grain.ABS sell-CVB cloth.ABS buy-CVB
ɗum-us-í=tte-s *yeénne*
be.dark-CAUS-CVB=FOC-3SG.M come.PFV

'The boy sold grain, bought cloth, and came late.'

e. *ʔas'o-y* *ʔánna-s* *yene*
man-NOM where-3SG.M exist.NPST

'Where is the man?'

Question words can be considered inherently focal, and this seems to be the reason why additional focus marking is blocked from them in Zargulla, but indexing is nevertheless possible. Before discussing index placement in Gutob, which is also syntactically unconstrained but sensitive to information management, we will first provide a short overview of referent indexing in Munda languages more generally in the following section. This is worthwile as S/A indexing in Gutob is quite exceptional compared to the other members of the family.

5.3 Referent indexing in Munda

The Munda languages belong to the Austroasiatic phylum and are spoken in Central and Eastern India, surrounded by Indo-Aryan and Dravidian languages. The internal classification of Munda languages is still a matter of debate, but there is some consensus that Gutob is most closely related to Remo and Gta' (Anderson 2008: 1–4; Sidwell 2015: 194–197).

The verbal complex in Munda languages exhibits a range of inflectional categories, including indexing for person and number. However, indexing in the languages of the family differs with regard to three aspects. There is variation among the languages concerning i) the morphological form of the indexes;

ii) the argument roles which can be indexed (see Cysouw (2004) and Anderson (2007: 64) for an overview); and iii) the position of the indexes. While object indexes are mostly suffixes, S/A indexes are either prefixes or enclitics/suffixes[10]. Table 5.1 shows an overview of Munda S/A indexing. Korku is an exception among the Munda languages, and is not listed in the table, as it lacks indexing altogether, except for of some locational copular expressions (Anderson 2007: 64). The variation in form and function of indexes in Munda languages has caused considerable debate with regard to their historical development (cf. Pinnow 1966, Anderson 2001, Anderson & Zide 2001 and 2007, and Donegan & Stampe 2004). In this section we will mainly pay attention to enclitic S/A indexing in Munda languages, as this is the only indexing that is found in Gutob, the language of our primary focus.

[10]Even for individual languages there is sometimes no consistency in labelling the indexes, cf. e.g Osada (2008) (suffixes) and Anderson (2007) (clitics) on the Mundari indexes.

	1SG	1DU.INCL	1DU.EXCL	1PL.INCL	1PL.EXCL	2SG	2DU	2PL	3SG	3DU	3PL
Santali	=ɲ	=laŋ	=liɲ	=bon	=le	=m	=ben	=pe	=e	=kin	=ko
Mundari	=ñ	=laŋ	=liŋ	=bu	=le	=m	=ben	=pe	=e(ʔ)/ =i(ʔ)	=kin	=ko
Kharia	=[i]ɲ/ŋ	=naŋ	=dʒar	=niŋ	=le	=[e]m	=bar	=pe		=kijar	=ki, moj
Remo	=niŋ	=naŋ		=naj		=no	=pa	=pe			
Gutob	=niŋ			=nei		=nom		=pen			=nen
Juang	-V_1-	ba-		nV_1-		mV_1-	ha-	(h)V_1-		-kia	-ki
Gorum	ne-			le-		mo-		bo-			gi-
Sora	-aj			-be	ə-…-aj			ə-…-ɛ			-dʒi
Gta’	n-			ni-	næ-/nɛ-	na-	pa-	pe-			-har-

Table 5.1: S/A indexes in Munda languages (adapted from Anderson (2007: 76), supplemented by data from Neukom (2001) on Santali and Osada (2008) on Mundari

)

Santali and Mundari, as well as Kharia, Remo and Gutob display S/A indexing as enclitics. In most of these languages, S/A indexes either attach to the main verb or to the constituent immediately preceding it, though the factors which determine the placement can vary. In Mundari, for instance, indexes obligatorily attach to the element immediately preceding the predicate, except if the clause consists of a predicate only in which case they attach to the verb (Hoffmann 1903: 12–13). The situation is similar in Santali where the S/A indexes (which are obligatory for animates, and possible for inanimates) either attach to the verb itself, or, more commonly, to the immediately preceding element if there is one. However, the shift of the S/A index away from the verb does not seem to be obligatory, but more of a strong tendency as there are exceptions (which are not elaborated on, see Neukom 2001: 113–115). The following examples illustrate the acceptability of the index attached to the constituent preceding the verb, to an independent pronoun in (28-a), or to the affirmative particle in (28-b), as well as attached to the verb itself, as in (28-c). Example (28-d) is ungrammatical, as animates always have to be indexed:[11]

(28) Santali

Q *cala-k'a-m?*
go-IND-2SG
'Will you go?.'

A a. *hɛ̃, iɲ-iɲ cala-k'a.*
yes 1SG-1SG go-IND
'Yes, I shall go.'

b. *hɛ̃-ɲ cala-k'a.*
yes-1SG go-IND
'Yes, I shall go.'

c. *hɛ̃, cala-k'a-ɲ.*
yes go-IND-1SG
'Yes, I shall go.'

d. * *hɛ̃, iɲ cala-k'a.*
yes 1SG go-IND
'Yes, I shall go.'

[11] Also P and G, as well as possessors can be indexed if they are animate. However, the position of these indexes is confined to the verb, i.e. non-S/A indexes cannot attach to other constituents (Neukom 2001: 115–117), although they are morphologically identical.

Gutob, as will be shown in the following sections, seems to be the only language of the family where the position of the index is not determined on purely syntactic grounds.

5.4 Case study: S/A indexing in Gutob

Section 5.3 shows that S/A indexing in Munda languages is not uniform. The South Munda languages Kharia, Gutob and Remo all have enclitic indexes which are formally similar. Gutob and Remo clearly form a subgroup of the family, but with regard to indexing, Gutob shows more similarity with Kharia and the North Munda languages Mundari and Santali.

The Remo S/A indexes are identified as clitics, however, nothing about hosts other than verbs is mentioned in the descriptions (Fernandez 1983, Swain 1997 and Anderson & Harrison 2008). In Kharia, exemplified in (22) above, the subject index is enclitic to the verb in affirmative clauses and attaches to the negative particle in negative clauses, and its placement is therefore rule governed. The situation is different in Gutob, where the position of the S/A index is syntactically unconstrained. This will be described further in Section 5.4.2, after a brief introduction of the language, including some relevant information on its morphology and syntax.

5.4.1 Language and data

The Gutob language (ISO gbj, sometimes referred to as Bodo Gadaba) is mainly spoken in the Koraput District in the highlands of the state of Odisha and in neighbouring districts in the state of Andhra Pradesh in Eastern India. In the present study, the name Gutob is used as it is the name the speakers themselves use both for their language and for themselves as a social group. There is little reliable information on the number of speakers of Gutob. The census of 1991 counts around 28,000 Gadabas, but does not distinguish between Gutob Gadaba and the Dravidian Ollari Gadaba. Estimates range from 5.000 to 20.000 speakers (Rajan & Rajan 2001, Griffiths 2008, Berger 2015). Our study is based on a corpus collected during a recent language documentation project (Voß 2018) between 2016 and 2018. The whole corpus contains 18.5 hours of transcribed audio and video recordings consisting of fictional stories, conversations and interviews, personal narratives and elicitations.

The Gutob people in the Koraput district live in a multilingual setting and Desia, the regional Indo-Aryan *lingua franca*, is present on a daily basis. In the village of Jalarhanzar, where the data for this study were obtained, most younger speakers have shifted to Desia. According to Griffiths (2008: 635–636), there are at least two dialects of Gutob, which he calls Koraput Gutob and Andhra Gutob. Most of the previous research as well as the present analysis is based on Koraput Gutob.

The prevalent constituent order in Gutob is APV, although A and P are sometimes reversed. The clause-final position of the verb is more fixed. In afterthought constructions A and more commonly P may follow the verb, but are clearly set off prosodically. Adjectives, adverbs, demonstratives and quantifiers usually precede their head. With regard to morphosyntactic alignment, Gutob is a nominative-accusative language.

As is typical for Munda languages, Gutob displays complex verbal morphology. Gutob has basic voice, which is marked by TAM/voice portemanteau suffixes. Voice marking closely correlates with transitivity and most verbs are either always in the middle or always in the active voice. A small set of voice-alternating verbs exists, such as verbs with a causative alternation, e.g. 'break' or 'tear', or motion verbs in which voice marks directionality. A change of voice from active to middle can be employed to reduce transitivity of normally transitive verbs, e.g. in reflexive constructions, although this is rare. To increase the transitivity of an otherwise intransitive verb it has to take the causative marker, a change of voice alone is not sufficient in this case. Further categories marked on the verb are negation, reality status and honorifics. The following template illustrates the morphological structure of a finite verb in Gutob. It indicates that verbal indexes are not obligatory on the verb, but if they do occur, they have a fixed slot within the verbal morphology.

NEG-	CAUS-/<CAUS>	ROOT	-TAM.VOICE	(=S/A)	-PRS	-HON

Table 5.2: The morphological structure of the Gutob verb

For our investigation of the behavior of S/A indexes in Gutob, we annotated 2318 finite main clauses for overt S/A reference through an index and/or a pronoun in a subcorpus of 32669 words, comprised of 12 narratives and stories from everyday life (approx. 360 min) by 7 speakers (20-70 years,

all female). Clauses were annotated for person and number, the position of the index, and the part of speech of the host of preverbal indexes; with verbal indexes, we also annotated whether non-verbal placement would have been syntactically possible (see Appendix C for details).

5.4.2 Formal properties of indexes in Gutob

As for S/A reference in general, third person NP arguments may be omitted if the referents can be inferred from the context. First and second referents, however, are usually expressed by a full pronoun and/or a bound index. The person indexes are formally identical to the free personal pronouns, except for the third person, which is zero marked. The clitic used for third person plural *=nen* is a general plural marker, which, apart from marking reference to a third person plural S/A argument, also attaches to NPs to mark them for plural, as well as to imperatives with second plural reference. Table 5.3 illustrates the identity in form of the free pronouns and the indexes for all but the third persons.

	SG		PL	
	free	bound	free	bound
1	niŋ	=niŋ	nei/naj	=nei/=naj
2	nom	=nom	pen	=pen
3	maj	∅	maj=nen	=nen

Table 5.3: Free pronouns and bound indexes

Despite the fact that person indexes are crosslinguistically commonly derived from pronouns, and that similarity between indexes and their free counterparts is more common in first and second person than in third person, identity of the two paradigms as found in Gutob has to be viewed as exceptional (Siewierska 2004: 251). The formal identity between the two paradigms as well as the fact that referents can be marked by either free forms or bound forms or both at the same time (see examples in 32 below) has induced Voß (2015) to deal with the issue of whether the two paradigms can actually be

distinguished. She finds that they can: clause-initial person markers like in (29) are unambiguously pronouns, as they can host clitics which are reserved for nominals, such as the additive marker *=sa* in (29b). Also, they are not repeated in coordinated clauses, as in (29c). As for indexes as part of the verbal predicate, it is clear that they are part of the morphology, as they have their fixed slot within the verbal template (see Table 5.2 above). Also, they are often repeated in coordinated clauses with the same S/A referent (29c).

(29) a. *uraʔ kunig, nom maŋɖem piŋ-loŋ*
NEG old.woman 2SG why come-FUT
'No old woman, why should you come.'

b. *niŋ=sa ɖoɲ-tu=niŋ*
1SG=too cook-FUT=1SG
'I, too, will cook.'

c. *nom ca iɖ-tu=nom, lai som-tu=nom maʔ som-tu=nom*
2SG tea drink-FUT=2SG rice eat-FUT=2SG curry eat-FUT=2SG
'You drink tea, eat rice and eat curry.'

Referent expression can either be in the form of a free pronoun only, as in (29a), or an index, as in (29c), or both, as in (29b), and also in (32b) below, where there are even three realizations of the same referent in one clause.

If indexes are part of the verbal complex, a distinction has to be made with regard to the form of the predicate. Complex predicates in Gutob distinguish between explicator verb constructions and auxiliary constructions (Voß 2015: 225–226). Explicator verb constructions consist of a main verb carrying the semantic load, plus a second inflected verb. Explicator verbs are homophonous with lexical verbs but have undergone extensive semantic bleaching and express aktionsart distinctions (cf. Butt & Geuder 2003: 330–331). In these constructions, the clitic tends to attach to the last element, the explicator verb, as (30a) shows, although there are exceptions. In auxiliary constructions conveying TAM distinctions, however, the index is more likely to attach to the main verb, as shown by example (30b).

(30) a. *nom dapre moɽ-gu piŋ-gi=nom*
2SG quickly rise-MID.CVB come-MID.PST=2SG
'You got up quickly.'

b. *bezri aɽo=bo? ui-gi=pen ɖugu*
tomato garden=LOC go-MID.CVB=2PL AUX.PST
'You went to the tomato garden.'

There can be further variation in complex predicates when it comes to negation. Whereas in simple predicates, standard negation is marked by a prefix (see Table 5.2 above) which does not affect index placement, auxiliary constructions are negated by means of the negative particle *ura?*. In present tense, *ura?* usually replaces the auxiliary, whereas in the past tense, the auxiliary follows the negative particle. In our subcorpus, out of 62 clauses which are negated by means of *ura?*, there are 19 clauses where S/A reference is either expressed by a pronoun or a preverbal index (which will be further elaborated on in 5.4.4). In the remaining 43 clauses which have an index in the negated complex predicate, we find six instances (14%) where the index attaches to the negative particle *ura?*, while it attaches to the verb in the remaining clauses. So while this particle may host the index, as in (31a), this is by no means the mandatory position. More commonly, the index attaches to the lexical verb, as in (31b). This is different from the situation in Kharia, where a negative particle becomes the mandatory host of the person index.

(31) a. *buzei ura?=niŋ*
INF.understand NEG=1SG
'I didn't understand.'

b. *sa?mel ri~riŋ=niŋ ura?*
millet INF~bring=1SG NEG
'I didn't bring millet.'

The verbal position has been considered default in previous studies, although the placement of the indexes varies considerably. What has caused some debate in the formal analyses of the Gutob person marking system are the indexes which are not part of the verb complex, in the following called preverbal indexes. Earlier accounts have ascribed the non-verbal placement of indexes to the inherent features of the hosts, with certain adverbs, adverbials and interrogatives being preferred as hosts (Zide 1997). Like pronouns and verbal indexes, preverbal indexes can often be the only realization of a referent in a clause (see (20) and (36)). Voß (2015) found that preverbal indexes frequently co-occur with verbal indexes, as in (32b), which suggests a

functional similarity to free pronouns. On the other hand they usually do not host nominal morphology like the additive clitic and are often repeated in coordinated clauses, as in (32a), which makes them more similar to the verbal indexes than to the pronouns.

(32) a. *o-maj lai=niŋ beɖ-o? ma?=niŋ beɖ-o?*
OBJ-3SG rice=1SG give-PST curry=1SG give-PST
'I gave him/her rice and I gave (him/her) curry.'

b. *naj maŋɖem=naj gisiŋ=nen bon-o?=naj ɖutu*
1PL why=1PL chicken=PL raise-CVB=1PL.S/A AUX.PRS
'Why have we raised chicken?.'

In Section 5.4.3, we present the analysis of our corpus annotation regarding index placement in Gutob, before proposing an account of possible contexts in which indexes shift to preverbal hosts in Section 5.4.4.

5.4.3 First and second vs. third person reference

As has been mentioned in Section 5.4.2, there are no indexes for third person singular in Gutob and the referents are very often not overtly expressed at all. In our annotated subcorpus, there were only five instances of a third person singular S/A referent expressed by a pronoun. Third person referents can only be indexed in the plural by the use of the general plural marker *=nen*. This marker, however, behaves quite differently from the indexes expressing person and number of first and second person referents. As with the non-third person indexes, also *=nen* in its indexing function does not have to attach to the verb, as in (33) where it attaches to a preverbal adverb. However, our data show that third plural indexes attach to the verb more often than non-third person indexes.

(33) *a?=nen bana-gu beɖ-o?*
now=3PL forget-MID.PST give-ACT.PST
'Now they forgot.'

Table 5.4 shows the distribution (with absolute numbers in brackets) of indexes in affirmative clauses with third person plural vs. non-third persons, comparing the numbers for clauses with only verbal indexes, only preverbal indexes, as well as both options within a single clause. It shows that for third plural referents, preverbal indexes are uncommon, 96.7% of the clauses have

a single index in the predicate. In contrast, for non-third persons, only 59% of the clauses have verbal indexes only. In both cases a small amount of clauses also has indexes in both verbal and preverbal position. These numbers already show that the preverbal placement of an index is much more common for non-third persons, whereas for the marker *=nen* '3PL' the verbal position appears to be a default.

	verbal	preverbal	both	total
1st & 2nd	59.0% (646)	32.4% (354)	8.6% (94)	1094
3PL	96.7% (1089)	2.00% (23)	1.2% (14)	1126

Table 5.4: Verbal and non-verbal indexing for non-3rd person vs. 3PL referents

The difference between third and non-third person indexes is even more striking if one excludes those clauses that consist of a predicate only, thus making a preverbal placement impossible. In this subset, given in Table 5.5 the picture does not change much for third person: the verbal placement remains by far the most common. For non-third persons, however, the preverbal position is now much more prevalent than the verbal position. The majority of clauses, namely 54.1%, have a single index in preverbal position whereas only 31.5% have a single index in the predicate. In the remaining 14.4% of clauses there are indexes in both positions.

	verbal	preverbal	both	total
1st & 2nd	31.5% (206)	54.1% (354)	14.4% (94)	654
3PL	93.7% (552)	3.9% (23)	2.4% (14)	589

Table 5.5: Verbal and non-verbal indexing for non-3rd person vs. 3PL referents, excluding clauses comprised of verbs only

Not only do indexes for third person plural and non-third persons display very distinct distributions, but there are functional differences as well. The enclitic *=nen* can also mark plurality in nouns and functions as an associative plural marker even with uncountable nouns, as in (34). At the same time

plural NPs in Gutob do not have to be overtly marked for plural. Therefore, the use of *=nen* as an index for third person plural referents attaching to a preverbal (object) NP can cause ambiguity. Furthermore, object NPs do not have to be marked overtly by the non-pronominal object enclitic *=lai*, which would follow the nominal plural marker and eleviate the ambiguity. Consider example (35), where, similarly to (32a) above, the index could in principle also attach to *paʈai*, but this would result in the plural reference being unclear. Due to these circumstances, it was hard for several cases in our subcorpus to decide whether *=nen* is in fact an index or a marker of a nominal plural. Therefore, in the analysis of the peculiarities of preverbal indexes in Section 5.4.4 we focus on first and second person referent indexes only.[12]

(34) *tonda=nen baŋ-to=nei*
lemonade=PL send-HAB=1SG
'We send lemonade and such things.'

(35) *o-maj paʈai beḍ-tu=nen*
OBJ-3SG dress give-ACT.FUT=PL
'They will give her a dress/dresses.'

5.4.4 Preverbal indexes in Gutob

This section deals with a more detailed analysis of indexes in preverbal position. We have already shown in Tables 5.4 and 5.5 above that the assumption that the verbal position for indexes in Gutob is the default one (Zide 1997, Anderson 2007: 70–71 and Griffiths 2008: 643, 653), and that anything else is an exception, has to be reconsidered at least for non-third person indexes. This assumption might be issuing from three facts: firstly, many clauses consist of a predicate only, and do thus not provide an alternative for verbal placement of the index. Secondly, previous investigations have not differentiated between third and non-third person indexes. Finally the indexing behavior in other Munda languages might have biased previous analyses for Gutob.

Table 5.6 shows the kinds and positions of (non-third) reference in more detail. The most frequent position is, indeed, verbal, but not by a great deal: taken together, clauses with preverbal indexes (either as sole reference or

[12]For a discussion of conceptual characteristics of first and second vs. third person forms see e.g. Benveniste (1971: 195–205) or Dahl (2000).

in combination with another index or a pronoun) make up 41% of all the clauses.

verbal only	44,70% (523)
preverbal only	26,84% (314)
pronoun & verbal	10,43% (122)
preverbal & verbal	7,18% (84)
pronoun only	6,50% (76)
pronoun & preverbal	3,50% (41)
pronoun & preverbal & verbal	0,85% (10)
total	1170

Table 5.6: Type and position of reference in non-3rd persons

The preverbal elements indexes can attach to can be locative (as in 36a) or temporal (as in 36b) adverbials, object NPs (as in 36c), interrogatives (as in 32b) above, but also adjectives or demonstrative pronouns (as in 36d).

(36) a. *pen=nu=bo?=nom ui-a=be*
2PL=ATTR=LOC=2SG go-MID.IMP=HON
'You (SG) go to your (PL) place.'

b. *usoŋ muiro? gisiŋ=naj sir-o? som=be kina*
today one chicken=1PL roast-ACT.CVB eat=HON please
'Let's roast and eat one chicken today, please.'

c. *maŋ ma?=nom ɖoɲ-tu=be*
why sauce=2SG cook-ACT.FUT=HON
'Why did you cook curry sauce?'

d. *ito?=o? dinke=niŋ olai-o? ɖuloŋ*
like.this=EMPH daily=1SG hang-CVB be.FUT
'Like this I will daily hang it [the calabash] up.'

Zide (1997) in his analysis of Gutob person reference finds himself quite puzzled in view of this variation of positions and hosts, and ascribes it to "extreme (rhetorical) conditions" and something which would not come up in

"ordinary sentences" (Zide 1997: 326); he admits, however, that his data is too scarce to come up with a satisfying analysis. In his data, verbs are also the most preferred hosts, followed by certain adverbials as well as wh-words; among this class of constituents there are some members that are even more favored, namely *eke?* 'here', *a?* 'now', *begi* 'quickly', *dapre* 'afterwards', *ũdoj* 'when', *mono?* 'where' and *maŋ* 'why' (Zide 1997:313-327).

However, even though these words have a strong tendency to be the host of the S/A index, it would be likewise grammatical if the verb carried the index in their presence, as exemplified in (37). And it should be remembered that Table 5.5 above also provided quite a substantial number of cases where the verb was preceded by one or more suitable candidates for hosts but nevertheless carried the index.

(37) *a? keʈei-gu=nom*
now arrive-MID.PST=2SG
'Did you arrive just now?'

Table 5.7 provides the numbers for the different preverbal types of hosts; we can see that adverbials (including adverbial phrases like locative phrases) are indeed very frequent hosts, followed by NPs, interrogatives and object pronouns.

ADV	39% (189)
OBJ.NP	37% (176)
interrogatives	17% (83)
OBJ.PRO	3% (14)
NUM	2% (8)
DEM	2%(8)
total	480

Table 5.7: Different hosts for non-3rd person preverbal indexes

Thus, no lexical item or class in Gutob can be said to serve as the default host of an index. As a consequence, the index placement regarding the position

in the clause is variable as well and several examples (20, 32b and 38) show that the idea that preverbal indexing is limited "to the word immediately preceding the verbal complex" (Anderson 2007: 70) has to be re-evaluated. What is more, there are also a few tokens in the corpus used for this study where there are two preverbal indexes in one clause. In (38), there are not only two non-verbal indexes attaching to preverbal constituents (an interrogative and a manner demonstrative), but also a clause initial pronoun and a verbal index. Examples like this suggest that person indexes are not merely referential in Gutob and that their placement is not merely sensitive to discourse structure, but that indexes themselves have a discourse structuring function, which we will come back to in Section 5.5.

(38) *nom maŋɖem=nom ito?=nom ɖe~ɖem piŋ-gu=nom*
2SG why=2SG like.this=2SG INF~do come-MID.PST=2SG
'Why did you come this way?'

Finally, it should be mentioned that although index placement is variable in Gutob, it is not completely without constraints. For one thing, an S/A pronoun cannot host an index, see (39). Also, postverbal constituents in afterthought constructions do not host indexes, whereas free pronouns can appear there.

(39) **niŋ=niŋ sun-tu*
1SG=1SG speak-FUT
intended: 'I will speak.'

5.5 The discourse effect of index placement in Gutob

Having described the properties and frequency of preverbal indexing in Gutob, we now turn to illustrate how it contributes to information management. We argue that the placement of S/A indexes in Gutob is not merely sensitive to discourse, but that it is deployed judiciously as an information management marker.

This becomes especially evident in the light of examples where an index attaches to a preverbal constituent although there is already S/A reference in the clause, either by a verbal index and/or a pronoun. Thus, an index

would not be required if it would simply be needed for the sake of referencing (examples 32b and 38, also see Table 5.6). In these cases, the index is used as a device for marking a piece of information (or several pieces of information) as particularly noteworthy.

Marking a constituent as particularly noteworthy means that the speaker assumes that the interlocutor might not meet it with the right level of attention. Answers to content questions, question words or negators, which have been considered as being inherently focal in the literature, can host an index in Gutob, but they often don't. So they do not attract them by default, but only in contexts where the constituent needs more attention than the speaker expects it to receive purely on the grounds of it being new or unexpected information.

Consequently, there is also variation in sentences which contribute several pieces of new information: the speaker can use indexing to mark one or several constituents as being particularly noteworthy. In (40), where the speaker is telling about preparations that are done for the rituals for the dead, they want to highlight the objects of the actions. The predicates contribute new information, too, and could have been just as well the hosts of the indexes, but to a different effect. Indexing the objects as well as the verbs would have equated them in terms of how much attention they are to receive.

(40) *suŋol=nei goi-tu peɳɖom=nei ɖoɲ-tu*
firewood=1PL cut.down-ACT.FUT rice.liquor=1PL cook-ACT.FUT
'We will cut down firewood, we will cook rice liquor.'

Indexing can coincide with contrast marking, but it rarely does so. In (41), an index attaches to a constituent marked for contrast by the clitic *=o?*. In such a case, an index can increase the effect of this marker and draws even more attention to the host than enhanced by *=o?* alone.

(41) *eke=o?=niŋ lai ma? som-tu=niŋ*
here=EMPH=1SG rice curry eat-ACT.FUT=1SG
'I will eat rice and curry here. (i.e. not at home)'.

That indexing is not confined to contrastive or new information is also reflected in the fact that also discourse-given information like object pronouns can be host of an index (see Table 5.7). This is in line with the observation that the allocation of attention is logically independent from other discourse

effects which have been ascribed to the notions of topic or focus (Ozerov 2021).

The commitment towards a certain constituent on the part of the speaker can also shift within the scope of only a few utterances. This can be shown quite nicely in examples like (20) and (42) which constitute clausal minimal pairs, of which there are actually quite a few in our data. In each consecutive utterance, a different element is considered of more importance by the speaker and thus hosts the index. Examples (42a) and (42b), produced by the same speaker within the same text, illustrate this nicely. A delegation from one village had gone to visit another village for wedding negotiations and stayed over night. In the morning they decide to leave, uttering (42a) with the index attached to the predicate 'stay'. When asked by the people from their own village why they had come back so early, they reply with the sentence in (42b), but this time the index does not attach to the verb, but to the interrogative, thus drawig the attention towards the 'why'.

(42) a. *nei* *maŋɖem* *ɖu-loŋ=nei*
1PL.EXCL why stay-MID.FUT=1PL.EXCL
'Why should we stay?'

b. *nei* *maŋɖem=nei* *ɖu-loŋ*
1PL.EXCL why=1PL.EXCL stay-MID.FUT
'Why should we stay?'

The targeted use of the index can also cause variation regarding its placement within a complex verb phrase (cf. Section 5.4.2). In the explicator verb construction in (43b), the index attaches to the second verb, the explicator verb conveying the aktionsart of the predicate; this is the most frequent placement of indexes within explicator verb constructions. In (43a), a few utterances in advance, however, the index attaches to the lexical verb, thus highlighting the semantics of the event to a stronger degree.

(43) a. *goj-gu=niŋ* *ui-loŋ* *ɖio?* *[...]*
die-MID.PST=1SG go-MID:FUT QUOT
'I will die [...]'

b. *oh eno?=niŋ goj-gu* *ui-loŋ=niŋ=be* *ɖio?*
oh here=1SG die-MID.PST go-MID:FUT=1SG=HON QUOT
'Oh, here is it I will die.'

An index does not attach to a particular host in a clause because it has to, casually speaking, go *somewhere*, but is placed in a manner best suitable for the discourse-oriented need on the side of the speaker, namely to ensure the desirable amount of attention. That verbs still make up the majority of hosts can be traced back to two facts: firstly, many clauses in Gutob consist of (simple or complex) verbs only, therefore leaving no possibility for an index to go anywhere else. Secondly, predicates are more often not part of a presupposition (i.e. the piece of information conveyed by the speaker which is not shared by the hearer) than other constituents in a clause (Lambrecht 1994: 296) and might thus be especially prone to receiving an index as a marker of noteworthiness.

5.6 Conclusion

Our corpus study of index placement in Gutob has revealed that non-verbal indexes in the language are neither a fringe phenomenon nor limited to particularly unusual conditions. The use of preverbal indexes is not exceptional, but is applied actively and frequently by speakers in order to structure their discourse with regard to the piece of information they consider especially noteworthy. Thus, although some elements are more prone to become the host of an index, no lexical item in Gutob is an *a priori* host of an S/A index.

In this respect, Gutob is different from the closely related Kharia, where either the verb or the negative particle preceding the predicate becomes the mandatory host of the person index. It also differs from North Munda languages like Santali, where the index can also attach to various hosts immediately preceding the verb, but is syntactically confined to this position if it does not attach to the verb itself.

Indexing in Gutob is not syntactically obligatory, nor is the placement of the index predetermined. With regard to the question of whether index placement in Gutob can be referred to as *rule governed*, we agree with Givón (2011: 189) who, reflecting on indexes in Ute, states, that "[i]f by 'rule governed' one means the traditional generative statement, with purely syntactic conditioning of the choice of options, the answer is surely no. If, on the other hand, one means that the choices are non-random, and motivated by communicative or cognitive factors, the answer is probably yes." The choices behind

the variablity in Gutob index placement and its contribution to information management are better understood now. The same would be desirable for many other cases of *optional* marking: conditions might be unknown, or the grammaticality of an utterance might not be at stake, but speakers' choices are surely not arbitrary.