

Kharia: A transparent language*

Sterre Leufkens

Amsterdam Centre for Language and Communication

The transparency of the South-Munda language Kharia is measured by checking whether it exhibits the non-transparent properties listed in the introduction of this issue. It turns out that Kharia is transparent to a high degree. The most important non-transparent properties that are attested are apposition, discontinuity, some fusional morphology and phonological adaptations. Transparent properties of Kharia are the consistent use of clitics for phrase-marking (instead of head-marking by affixes) and the absence of functionally specialized lexeme classes.

1 Introduction

In this issue of *Linguistics in Amsterdam*, two questions are raised: (i) is there any systematicity in how languages may lose transparency, i.e. acquire opacity; and (ii) how can this systematicity be explained? The answers can be found by means of measuring and comparing the transparency of several languages. The Munda language Kharia will be the subject of such a study here.

Kharia is spoken in India, mostly in the state of Jharkhand and adjoining areas (see Figure 1¹). According to *Ethnologue* (Lewis (ed.) 2009), it currently has approximately 292,000 speakers. Kharia belongs to the South Munda languages, which are part of the Austro-



Figure 1: The area where Kharia is spoken.

* I am indebted to Kees Hengeveld, Magaly Grandez, Norval Smith and Roland Pfau for their valuable comments on earlier versions of this paper. John Peterson kindly helped me by sharing all his writings and thoughts on Kharia.

Abbreviations used in glosses: 1, 2, 3 = 1st, 2nd, 3rd person, ACT = active voice, CLASS = classifier, CNTR = contrastive, COMP = complementizer, COP = copula, DU = dual, DUR = durative, FOC = focus, GEN = genitive, HON = honorific, HUM = human, INF = infinitive, INFER = inferential, INSTR = instrumental, IRR = irrealis, MID = middle voice, NEG = negation, NHUM = non-human, OBL = oblique, PASS = passive, PERF = perfect, PL = plural, PROGR = progressive, PRS = present tense, PST = past tense, POSS = possessive, RDP = reduplicated form, SEQ = sequential converb, SG = singular, VOC = vocative.

¹ Map of India retrieved from <http://www.world-geographics.com/maps/asia/map-of-states-in-india/>.

Asiatic language family. It is in regular contact with Sadri (also known as Sadani, an Indo-Aryan language), Hindi (Indo-Aryan as well), Mundari (a Munda language, like Kharia), and the Dravidian language Kurukh (Peterson 2011: 5).

Typologically speaking, Kharia is a strongly isolating² language. Word order is predominantly head-final. A remarkable feature of Kharia (and Munda languages in general) is that it appears not to have morphosyntactic lexeme classes; it is said to be precategorial. Lexemes in precategorial languages (also called flexible languages, see e.g. Hengeveld 1992; Hengeveld & Van Lier 2010) can be used in all propositional functions, that is, as both heads and modifiers of both referential and predicative phrases, without any formal adaptations. In contrast, in non-flexible languages, certain lexemes can only be used in specific functions. In English, for instance, only specific lexemes (viz. verbs) can be used as heads of predicative phrases. They have to be adapted to be used as heads of referential phrases, for instance by adding the suffix *-ing*. These formal requirements make it necessary to distinguish between verbs and nouns. In Kharia however, the functional flexibility of lexemes renders a distinction between nouns, verbs, adjectives and adverbials irrelevant. This is a rare phenomenon and also one that is heavily debated (cf. Evans & Osada 2005 and the commentary in the same issue for a discussion). I will not go into this debate any further here, but let me be clear that my basic assumptions in this article are that 1) flexibility is a valid concept and 2) Kharia is flexible.

In this article, I will measure the degree of transparency of Kharia, by going through the list of transparent features as created by Hengeveld (this issue). For each feature on this list, I will answer the question whether Kharia's realization is transparent or not. My main data source for this is Peterson (2011), a recent and very thorough description of the Kharia language.

Note that the notion 'transparent' has gotten different interpretations in the literature (cf. Leufkens, forthcoming), but is defined in this issue as a one-to-one relation between meaning and form. The list of features thus addresses mismatches at the different interfaces between levels and within levels. In Section 2, it will be assessed whether there are any mismatches between pragmatics and semantics in Kharia. Section 3 deals with potential mismatches between on the one hand pragmatics and semantics, and on the other hand morphosyntax. In Section 4, mismatches between the upper three levels and the Phonological Level will be discussed. Section 5 deals with phenomena that take place at the Morphosyntactic Level and in Section 6, features pertaining to the

² Peterson (2011: 1) states that Kharia is 'predominantly agglutinating', but that would mean that it mainly makes use of affixes. Later on, Peterson actually argues that Kharia makes abundant use of clitics, which is a property of isolating languages.

Phonological Level are described. Finally, in Section 7 the results will be summarized and I will draw conclusions.

2 Interpersonal – Representational

In this section, the appearance of mismatches between the Interpersonal Level (pragmatics) and the Representational Level (semantics) will be considered. This involves two features: appositional structures (nominal apposition and cross-reference) and the availability of units as heads of predicate phrases.

2.1 *No apposition, no cross-reference*

Apposition is the phenomenon that within one utterance, an entity is referred to twice or more; either by lexical means or by a combination of lexical and grammatical expressions (cross-reference). This is opaque as it involves a relation between two Referential Subacts at the Interpersonal Level and a single unit at the Representational Level³.

Apposition is not prohibited in Kharia, as is evident from example (1). The phrases *Rata's mother* and *Darhi's wife* are both used to make reference to the same individual.

Peterson (2011: 84)

- (1) u kayom onɖor=kon [raʈa=yaʔ ayo=ɖom],
 this talk hear=SEQ Rata=GEN mother=3.POSS
 [daʈhi=yaʔ saw-ʈay=dom] gam=te:
 Darhi=GEN spouse-woman=3.POSS say=ACT.PRS
 'Hearing this matter, Rata's mother, [i.e.] Darhi's wife, said: ...'

It could be argued here that the second phrase adds information to the first (viz. that Rata's mother is Darhi's wife), so that the two phrases do not express the same meaning – in that view, this construction is transparent. Only apposition of two completely interchangeable elements would be truly non-transparent. However, to prevent myself from being too lenient in analyzing Kharia as transparent, I will adopt the more strict approach here: the phrases at least have

³ One could argue that this does not involve a many-to-one relation between meaning and form, but rather between pragmatic and semantic units. As transparency is about meaning-to-form relations, apposition would not be a violation of transparency. However, apposition also violates a one-to-one relation between semantics and morphosyntax, as there are two morphosyntactic constituents corresponding to one semantic unit. Hence no matter whether the non-transparent relation is at the IL/RL or at the RL/ML interface, apposition is opaque.

an overlapping meaning aspect, which makes this construction opaque at least to some extent.

A special case of apposition is cross-reference, where one of the elements in apposition is a grammatical unit. Cross-reference can for instance appear in pro-drop languages where person marking on the verb can refer on its own and hence constitutes a Referential Subact. If an argument is expressed not only by this verbal marking but by an independent pronoun as well, there are two Referential Subacts corresponding to one Individual. This procedure is in FDG distinguished from agreement; the latter involves constructions where the verbal marking and the independent lexical expression are *both* obligatory (Hengeveld & Mackenzie, 350).

Kharia exhibits cross-reference. It is a pro-drop language (Peterson 2011: 168), which means that person markers on the predicate are referential. Hence, when an argument is expressed as an independent lexical unit as well, cross-reference appears, for instance in example (2). Both *in* and *=o[?]j* are Referential Subacts, used to refer to the first person participant.

Peterson (2011: 95)

- (2) *in daʔ biʰ=o[?]j*
 1.SG water pour.out=ACT.PST.1.SG
 ‘I poured water out.’

Kharia is opaque with respect to apposition, as it allows for a two-to-one relation between Referential Subacts and entities at the Representational Level.

2.2 No limitations on which semantic units can be chosen as predicates

Transparency is maintained when every pragmatic or semantic unit can fulfil all propositional functions, i.e. when it can be used both as a head and as a modifier of both a predicative phrase and a referential phrase. Opacity arises when the use of a unit in one of these functions triggers morphosyntactic adaptations.

As pointed out in the introduction of this article, Kharia allows every lexeme to be used in all propositional functions. A unit can be used as a predicate and as an argument, and as a modifier of one of these, without any additional morphosyntactic requirements. Examples (3) - (6) show that units of different semantic and morphosyntactic type can be the head of a predicate phrase.

Peterson (2011: 76)

- (3) *lebu dɛl=ki*
 man come=MID.PST
 ‘The man came.’

- (4) bhagwan lebu=ki ro qel=ki
 God man=MID.PST and come=MID.PST
 ‘God became man [=Jesus] and came [to earth].’
 ‘‘God manned and came.’’

Peterson (2011: 226)

- (5) beto[?]q=si[?]q=ip
 hunger=PERF=1.SG
 ‘I am hungry.’
 ‘‘Hungered-me.’’

Kerkeṭṭā (1990: 31 in Peterson, 2011: 116)

- (6) u buṭha=kiyar=te=ko bay ja[?]b=si[?]. idib tunbo?
 this old.man=DU=OBL=CNTR madness grab=PERF night daytime
 ‘‘kerson=e la! kerson=e la!’’ lo[?]=na=kiyar.
 marry=ACT.IRR VOC marry=ACT.IRR VOC DUR=MID.IRR=DU
 ‘My father [and his wife > DU] have gone mad. Day and night they’ll
 keep on ‘‘Marry! Marry!’’-ing.’

In each example, the predicate phrase is headed by a different kind of unit. In (3) the predicate consists of an event-denoting head (*qel*) and an enclitic. In (4), the head of the predicate phrase is the lexeme *lebu* ‘man’, denoting an Individual (which functions in (3) as the head of a referential phrase). In (5) a unit denoting a property is used predicatively, and in (6) we see that even a quoted element can serve as the head of a predicate phrase.

We thus have seen that all kinds of constituents can be used as the head of a predicative phrase. Examples **Error! Reference source not found.** and (8) illustrate that the same is true for referential phrases.

Peterson (2011: 81)

- (7) kosu ranga buṇ bhiren=ta[?]jd=ip
 sick cold INSTR flounder=MID.PROGR=1.SG
 ‘I am plagued by (litt. floundering with) sickness and illness (litt. coldness)’

Peterson (2011: 86)

- (8) ho=kaṭ=a[?] silo[?] yo=yo[?]j
 that=SG.HUM=GEN plow look=ACT.PST.1.SG
 ‘I looked at his plowing.’

In **Error! Reference source not found.**, two units denoting properties are used as heads of referential phrases. In (8), an event-denoting unit has that function, without requiring any formal adaptations.

Finally, examples (9) and (10) show that there are also no restrictions on which units can be used as modifiers.

Peterson (2011: 79)

- (9) [...] jahã ɲoʔ=na cij
 indef.NHUM eat=INF thing
 ‘some of the food’

Peterson (2011: 86)

- (10) Biha karay lebu goʔj goʔd=ki
 marriage do man die go=MID.PST
 ‘The man who married (“did marriage man”) died.’

In (9), an indefinite functions as a modifier in a referential phrase, whereas in (10) the complex phrase *marriage do* fulfils that function.

Taken together, the examples illustrate that Kharia is transparent with respect to flexibility of lexemes, as there are no limitations on the use of lexemes of different semantic and morphosyntactic nature in different propositional functions.

3 Interpersonal/Representational – Morphosyntactic

This section will deal with potential mismatches between pragmatic and semantic units on the one hand, and morphosyntactic ones on the other. This involves three opaque phenomena: syntactic functions (as opposed to semantic or pragmatic functions), discontinuity (violations of domain integrity) and sensitivity of function marking to the morphosyntactic nature of the input unit.

3.1 *No grammatical relations (but pragmatic or semantic alignment)*

Functional Discourse Grammar (henceforth FDG; Hengeveld & Mackenzie, 2008) distinguishes three types of alignment, corresponding to three of the levels of organization. The first is interpersonal alignment, where sentences are organized according to pragmatic functions (Topic, Focus) and reference (definiteness, specificity, etc.). The second type of alignment is representational alignment, which concerns the organization of clauses and phrases in terms of semantic functions (Actor, Undergoer, etc.) or designation (animacy, person, etc.). The third type of alignment is morphosyntactic alignment, where predicate-argument structure is organized in terms of syntactic functions

(Subject, Object) of constituents. Only in the latter type of alignment, FDG speaks of grammatical relations. In all other cases, the relations are not ‘grammatical’ but in fact semantic or pragmatic (Hengeveld & Mackenzie, 2008: 317). Interpersonal and representational alignment systems are transparent by definition, since in such systems linguistic organization is motivated by semantic or pragmatic information.

Kharia relies to a considerable extent on representational alignment, as the semantic roles of arguments determine the expression of the predicate. Predicates are obligatorily marked for either active or middle voice. The former is associated with volitionality, transitivity and non-durationality, whereas the latter indicates non-volitionality, intransitivity and durational aspect (Peterson, 2011: 276, 286). The notions evolve around two prototypical types of events: one where there is a clear agent that causes the event to happen, and one where a situation or stative event is described that overcomes participants, rather than being caused by them. Many predicates can be marked for both active and middle, depending on whether the Speaker wants to portray the described situation more like an action or more like a process.

One-place predicates are often in the middle voice, as it is associated with intransitivity. However, if the argument bears clear Actor-properties (if it is a human, volitional performer of an action), active voice is used despite the predicate being intransitive. An example of this is (11). The predicate *biʔthuŋ=oʔ* ‘to spit’ is an event with a typical agent, volitionally causing the event, and therefore takes an active voice.

Peterson (2011: 284)

- (11) *ho=kaʔ koŋ-koŋ=ga biʔthuŋ=oʔ*
 that=SG.HUM know-RDP=FOC spit=ACT.PST
 ‘S/he deliberately spit.’

In example (12) however, the argument is a prototypical Undergoer in that it non-volitionally undergoes a change of state.

Peterson (2011: 285)

- (12) *ho=kaʔ (*koŋ-koŋ=ga) urumdaʔ=ki*
 that=SG.HUM (*know-RDP=FOC) sweat=MID.PST
 ‘S/he (*deliberately) sweated.’

Alignment is hence sensitive to the semantics of the one-place predicate and its argument: active voice combines with an Actor argument, whereas a middle voice involves an Undergoer argument. This indicates that Kharia has a representational alignment system, hence no grammatical relations.

However, there is a passive construction in Kharia, where the semantic distinction between Actor and Undergoer is neutralized. This means that there is evidence for a *syntactic* function Subject.

Peterson (2011: 368)

- (13) ho=kaŋ am=te pothi=te ter=o?
 that=SG.HUM 2.SG=OBL book=OBL give=ACT.PST
 ‘S/he gave you the book.’
- (14) am=te pothi ter d̥om=ki
 2.SG=OBL book give PASS=MID.PST
 ‘A book was given to you.’

In (13), we see that an Undergoer argument has to be marked with =te. However in the passive sentence (14), the Undergoer argument *pothi* is unmarked. We know that this is a genuine passive construction, as the Actor-argument (s/he) could be added by means of a by-phrase (*buy*-phrase; Peterson, personal communication). The Undergoer is hence treated grammatically as though it is an Actor, in other words: the semantic role is irrelevant for expression of argument relations. Thus, even though the alignment system of Kharia is predominantly semantically conditioned, we here have a case of morphosyntactic alignment: there is a Subject in Kharia.

To see whether there is an Object function as well, let us have a look at more transitive and ditransitive predicates. Their expression in Kharia is mostly semantically conditioned as well. Whereas transitive clauses are strongly associated with the active voice, it is also possible to use a middle voice in two-place predicates. Again, this distinction is in fact a distinction between events typically involving Actors and events typically involving Undergoer arguments. Example (15) shows this difference.

Peterson (2011: 144)

- (15) [...] ho maŋa boʔ=te d̥am=ki d̥aʔ=te yo=yoʔ
 that cave place=OBL arrive=MID.PST water=OBL look=ACT.PST
 ‘He arrived at the cave [and] looked at the water.’

The predicate ‘to arrive’ is seen here as an event involving an Undergoer, as ‘arriving’ is not something the participant is actively doing. Therefore the middle voice is used. *yo* ‘to look, to see’, however, is treated as an action and gets active voice marking. The distinction is clearly a semantic one, which shows the dominance of semantic roles in Kharia’s alignment system again.

Different object-like semantic roles (Undergoer, Experiencer, Recipient and even Locative arguments) are in fact never neutralized in Kharia. In examples (13) and (15) we have already seen that two of these roles (Locative

and Undergoer) are marked with an oblique marker =te. Possibly, these functions are always marked identically, which would mean that the semantic distinction is neutralized. We would then speak of a syntactically relevant Object function. However, this is in my opinion not the correct analysis. Object-like arguments actually *can* be expressed differently and I will argue that their expression is determined semantically.

Undergoers can be distinguished from other semantic roles in two ways. Firstly, they are optionally marked with =te whereas Recipients, Experiencers and Locative arguments *obligatorily* receive that marker. Whether =te is used on an Undergoer depends on the speaker: some speakers prefer to use it only when a countable, definite noun is referred to (Peterson, 2011: 144). In example (15) above, the oblique marker on *ḍa?* ‘water’, for instance, is dropped by some speakers because a mass noun is involved. At least for this group of speakers, there is a difference between Undergoers on the one hand and other arguments on the other, which is a semantically based and hence transparent distinction.

The second distinction between Undergoers and Recipients in Kharia appears in passive sentences. From sentence (13), repeated as (16), two passive constructions are derived; one foregrounding the Undergoer in (14) (repeated here as (17)) and one foregrounding the Recipient in (18). In (17), the Undergoer *book* loses its oblique marker and behaves like an Actor. This is impossible for the Recipient in (18). The Undergoer and the Recipient are treated differently here.

Peterson (2011: 368)

- (16) ho=kaṭ am=te pothi=te ter=o?
 that=SG.HUM 2.SG=OBL book=OBL give=ACT.PST
 ‘S/he gave you the book.’
- (17) am=te pothi ter ḍom=ki
 2.SG=OBL book give PASS=MID.PST
 ‘A book was given to you.’
- (18) *am pothi=te ter ḍom=ki=m
 2.SG book=OBL give PASS=MID.PST=2.SG
 ‘You were given the book.’

The same is true for Locatives: these cannot be subject in a passive sentence, hence the ungrammaticality of example (20).

Peterson (2011: 368)

- (19) ho=kaṭ u ṭhãṛo=te ḍoko=ki
 that=SG.HUM this place=OBL sit.down=MID.PST
 ‘S/he sat down in this place.’

- (20) *u thãŋo ɖoko ɖom=ki
 this place sit.down PASS=MID.PST
 ‘This place was sat down in.’

In sum, Kharia’s alignment system relies mostly on semantic information. There is no syntactic function Object in Kharia, as Undergoers, Recipients and Locatives are treated differently in the grammar. However, since semantic information is neutralized in passive sentences, it is necessary to distinguish the grammatical relation Subject. There is hence some opaque organization in Kharia, even though in general, the formal expression of argument relations is strongly semantically based.

3.2 *No discontinuity*

In a fully transparent language, what belongs together at the Representational Level stays together at lower levels. This principle is violated by the phenomenon of discontinuity. At least two sources exist for discontinuity: extraposition and infixation.

Kharia has an infix <nV> (Peterson 2011: 101-104), which used to be a nominalizer. In the modern language it “derives contentive morphemes from an underlying morpheme”; the meaning of the derived lexeme cannot be predicted from the meaning of the source lexeme. For instance, *bunui* ‘pig’ is derived from *bui* ‘raise an animal’, and *jonoʔ* ‘broom’ from *joʔ* ‘sweep’. As this infix is no longer productive, it is doubtful whether we should see this as a real infix – perhaps we should analyse it as lexicalized.

Another infix <[(o)ʔ]> or <[(o)ʔb]> (Peterson 2011: 230) serves as a causative marker, for example in *botoŋ* ‘fear’ versus *bo<ʔ>toŋ* ‘scare’. Its status is not doubtful at all – this is a productive infix with a consistent meaning.

Discontinuity can also be caused by extraposition, i.e. dislocation of a heavy part of a phrase or clause, for instance, in the sentence *the guy has arrived who is going to fix my lock* where the complex Relative Clause is not adjacent to its head (Hengeveld, this issue). I have not found such discontinuous configurations in Kharia, but I also have no evidence that it is not allowed.

3.3 *Function marking not sensitive to the nature of input*

As discussed in section 2.2, in a fully transparent language all semantic units can be used in all propositional functions: as heads and modifiers of predicative and referential phrases, without any morphological restrictions or requirements. Additionally, units of all degrees of morphosyntactic *complexity* should be available for all functions, without any restrictions or morphological requirements. Transparent function marking is hence function marking that

applies identically to all hosts. If the host (the marked unit) consists of more than one word, the transparent option would look like [x y z]=f (I am grateful to dr. R. Pfau for this illuminating illustration). I will call this sort of function marker, that can take phrases as its host, a clitic. In the literature, we find a lot of different criteria and definitions of clitic-hood, but for me the ability to mark phrases will be decisive. A transparent language is expected to have clitics (phrase-marking) rather than affixes (head-marking).

In Kharia, pragmatic and semantic functions are almost always marked by means of clitics. Peterson (2011) is somewhat implicit about why he analyzes Kharia's bound morphemes as clitics and not as affixes. Firstly, he shows (p. 35ff) that there are units that are phonologically dependent on other units. He calls these units phonological clitics (p. 52), without considering the option that they might be affixes. He then goes on to state that these phonological clitics are also clitics in a morphosyntactic sense, as they mark phrases, not words. This can be hard to prove in a head-final language like Kharia – if a marker attaches to the end of the phrase, one cannot see whether the phrase or the head is marked. Throughout the grammar, Peterson gives convincing examples of phrase-marking, for instance (22) and (23), but he does not give conclusive evidence for every separate bound morpheme. Finding such evidence indeed appears undoable (even more so for me in this article), since Kharia has so many bound morphemes. I will therefore assume clitic-status for all bound morphemes in Kharia, unless I find clear counterevidence.

There are clitics expressing a range of different functions in Kharia, for example pragmatic status (=ga 'FOC'), case and number (=ya? 'GEN'), information on tense and voice (=o? 'ACT.PST'), and so on. These clitics can attach to any kind of constituent. A function marker is hence indifferent to the morphosyntactic nature and complexity of its host constituent. As said above, the phrase-marking status of a marker can be hard to prove in head-final languages. However, example (22) provides such proof.

Peterson (2011: 54)

- (21) u=je[ʔ]=ko ho=ki=ya?=ga heke
 this=SG.NHUM=CNTR that=PL=GEN=FOC COP.PRS
 'But THIS is THEIRS.'

Peterson (2011: 403)

- (22) ore^ʔj kontan bui=na=ko ho=ki=ya? dhatam aw=ki
 ox cow raise=INF=CNTR that=PL=GEN custom COP=MID.PST
 'But RAISING OXEN AND COWS was their custom.'

We see that the same clitic (contrastive focus marker =ko) is used on a simple Noun Phrase in (21) and on a complex Complement Clause in (22). In the latter,

we know that it is not just the head *raising* that is marked, because the sentence is used in a conversation about customs of people: ‘raising oxen and cows’ is the custom that is contrasted to other customs.

Another convincing example of phrase-marking is found in (23), where the genitive marker attaches to an entire complex unit.

Peterson (2011: 55)

- (23) la? u sembho ro ðakay rani=kiyar=a? nãw jan
 then this Sembho and Dakay queen=DU=GEN nine CLASS
 beʔt=ðom=kiyar aw=ki=kiyar
 son=3.POSS=HON COP=MID.PST=HON
 ‘Then this Sembho and queen Dakay had nine sons.’ (Lit.: ‘This Sembho and Queen Dakay’s nine sons were.’)

The use of clitics is a very pervasive and salient strategy in Kharia. It is also a salient feature in Quechua, another language that is argued to be transparent to a great extent (M. Grandez, this issue). This could mean that phrase-marking is a prominent feature of relatively transparent languages. I will come back to this in the conclusion section.

4 Interpersonal/Representational/Morphosyntactic – Phonological

This section discusses features at the interface between the Morphosyntactic Level (and all levels above it) and the Phonological Level. The features under consideration are parallel phrasing and influence of phonological weight on word order.

4.1 *Phonological phrasing and morphosyntactic phrasing run parallel*

Transparency holds when units at the Morphosyntactic Level are parallel to units at the Phonological Level. Hence in a fully transparent language, a Morphosyntactic Word corresponds to a Phonological Word, and the same is true for Morphosyntactic Phrases and Phonological Phrases.

Not much work has been done on prosody in Kharia. Therefore, the argument given here is somewhat tentative. Preliminary analyses show that all lexemes in Kharia have a low-high (LH) pitch contour. Clitics have no pitch contour of their own (which is the reason that they are seen as clitics in the first place: they are phonologically dependent on their hosts). Hence a Phonological Word in Kharia consists of a LH pitched lexeme plus one or more enclitics. Such a phonological unit has the same boundaries as the morphosyntactic unit it corresponds to: a morphosyntactic stem plus, optionally, function markers. This relation is fully transparent.

There is, however, one case in Kharia where Morphosyntactic Words and Phonological Words are not parallel, viz. the case of ditropic clitics (Cysouw 2005; Peterson calls them ‘anticipatory clitics’). Person-marking clitics usually attach to the right end of predicates, as in (24). But when the sentential negator *um* is present, the clitic takes that as a host, as we see in (24).

Peterson 2011: 335

- (24) a. *ter=e=ijn*
 give=ACT.IRR=1.SG
 ‘I will give.’
 b. *um=ijn ter=e*
 NEG=1.SG give=ACT.IRR
 ‘I will not give.’

The negator and the clitic form one Phonological Word, while they cannot be considered one unit at any higher level. Alignment at the Phonological Level and at the higher levels is not parallel.

Phonological Words in Kharia do not combine to form Phonological Phrases together – lexemes do not form larger prosodic groupings (Peterson, 2011: 118). Hence we have no evidence (yet) for the existence of a Phonological Phrase in Kharia.

There is work to be done here, mainly on the status of demonstratives. In particular, it is not clear whether these are independent words or proclitics (Peterson 2011: 44). Some demonstratives have the typical LH pitch contour when preceding a bisyllabic unit, which qualifies them as independent Phonological Words. However, in other cases, demonstratives seem to attach to hosts like clitics do, as they do not allow other words to intervene between them and the following units. Furthermore, when preceding a monosyllabic unit, demonstratives can form a Phonological Word together with that unit. The analysis of demonstratives as either clitics or independent words has consequences for the analysis of phrasing in Kharia. If we were to analyse demonstratives as independent words, there would be evidence for a Phonological Phrase, since we then see grouping of more than one Phonological Word. However, if we see them as clitics, we have no evidence of groupings.

In sum, Kharia is non-transparent with respect to parallelism between morphology and phonology, because of ditropic clitics. This is the only opacity with respect to this feature – the possible absence of groupings larger than Words is not opaque.

4.2 *Phonological weight does not influence morphosyntactic placement*

In many languages, there is a tendency to put clauses that are phonologically complex to the right of the sentence. Phonological weight overrides interpersonal, representational or morphosyntactic alignment in such cases. In contrast, in a fully transparent grammar, only pragmatic and semantic factors determine morphosyntactic placement.

In Kharia, constituent order is usually determined on pragmatic basis (Peterson 2011: 426). However, it is possible to move a heavy phrase to the right. A case in point is example (25).

Peterson (2011: 398)

- (25) ... u goʔjuŋ=te socay=ga col=ki=ki no “i jãut
 this path=OBL think=FOC go=MID.PST=PL COMP what animal
 heke hoy?”
 COP.PRS INFER
 ‘... they walked along this path, thinking “What animal could it be?”’

The underlined unit, the Undergoer argument, would usually immediately follow the predicate, *socay=ga*. However, the Complement Clause is dislocated because of its weight. The fact that weight can influence morphosyntactic placement is a non-transparent property of Kharia.

5 **The Morphosyntactic Level**

In this section, I will describe phenomena taking place at the Morphosyntactic Level, that have no correlate or trigger at the Interpersonal Level or Representational Level, i.e. expletives, tense copying, argument raising, non-semantic classification systems (grammatical gender, declension, conjugation) and agreement. A reviewer points out that these features all involve mismatches between the upper levels and the Morphosyntactic Level, and wonders why they are not discussed in the appropriate section (i.e. Section 3). The difference is that for all features in Section 3, there is in fact relevant pragmatic or semantic information, but that information is overridden by morphosyntactic considerations. The features to which the current section is devoted, all involve cases where there is no pragmatic or semantic information present at all – only a morphosyntactic rule or property.

5.1 No expletive elements

Expletive elements are elements that exist only at the Morphosyntactic Level and are triggered by a morphosyntactic rule or process. They are not explained by pragmatic or semantic considerations and hence non-transparent.

Travis (1984: 218) finds an implicational hierarchy of expletive elements (‘pleonastics’ in her terminology):

Argumentless passives and unaccusatives > expletives of displaced NPs >
expletives of displaced CPs > weather predicates⁴

This means that if a language uses expletives for argumentless passives and unaccusatives, it will also use expletives in the other situations. If expletives are used in one of these four functions in a language, they will be used for weather predicates. Hence, the absence of expletives used with weather predicates in Kharia would be evidence that Kharia has no expletives at all.

Kharia uses no expletives for weather predicates, as is illustrated in example (26).

Peterson (2011: 241)

- (26) *ɖaʔ gim=oʔ*
 Water rain=ACT.PST
 ‘It rained.’
 ‘Water rains.’

The argument of the weather predicate is a word with a clear semantic correlate. The use of an active voice indicates that the water is seen as volitionally performing the action of raining. It is treated as a semantic Actor-argument. As it is not a semantically empty unit, the one-form-one-meaning relation is maintained.

5.2 No tense copying

Tense copying (also called sequence of tenses or *consecutio temporum*) is seen in FDG as a copying mechanism taking place at the Morphosyntactic Level (but cf. Leufkens 2009 for a discussion of the phenomenon). The tense operator of the main clause is copied to an embedded clause; this is called operator agreement (Hengeveld & Mackenzie 2008: 351). As a result, the tense of the embedded clause gets the same (absolute) tense as the main clause. Thus, a new

⁴ Travis includes ‘> referential NPs’ at this end of her hierarchy, but I do not agree with her that Referential NPs are expletive elements. Therefore, I leave them out of the discussion here.

operator comes into being at the Morphosyntactic Level, without a counterpart at the Representational Level. Tense copying violates a one-to-one relation between levels and is therefore opaque.

Kharia rarely shows tense inflection in embedded clauses. But it is attested in example (27), where we see that the embedded predicate form *heke* has a present tense, whereas the main clause carries a past tense.

Peterson (2011: 405, underlining mine)

- (27) muda tam jou [lebu=ki umay kon=sikh=o?
 But now up.to person=PL NEG.3.PL know=PERF=ACT.PST
 [no u=je? i daru heke]
 COMP this=SG.NHUM what tree COP.PRS
 ‘But up to this day, people have not found out what tree it is.’

Kharia hence does not exhibit a tense copying rule and is transparent with respect to this feature.

5.3 *No raising*

In many languages, we observe so-called raising constructions (Hengeveld & Mackenzie, 2008: 368): an argument, semantically belonging to an embedded clause, appears as an argument of the main clause. For instance, the subject of a Complement Clause becomes the object of the main clause. This procedure creates non-transparency, as semantic groupings and morphosyntactic groupings are not parallel after argument-raising.

John Peterson has not found any examples of raising in Kharia (personal communication). It is hence very unlikely that raising is allowed in Kharia.

5.4 *No grammatical gender, declension, conjugation*

When we say that a language has a gender system, it means that the nouns of that language are in some way divided into classes. Gender can be either natural (classification according to biological sex, only for humans and sometimes animals) or grammatical (classification on the basis of an abstract feature of nouns). Natural gender is transparent, as the nominal classification has a semantic trigger. Grammatical gender is non-transparent: it is not pragmatically or semantically motivated.

Declension and conjugation are other types of opaque classification systems. Such classifications are based on the form of nouns and verbs, respectively. Latin, for example, has declensions: Latin nouns can be divided into five classes that behave the same, formally, when inflected for case. Since such classification has a morphological but not a semantic basis, it is opaque.

As stated above in Section 2.2, there is strictly speaking no morphosyntactic class of nouns in Kharia, but let us take a look at the heads of referential phrases. These heads cannot be divided into genders, since they do not trigger different agreement and are not obligatorily marked for some (semantic or abstract) feature. Biological sex can, of course, be expressed, for instance by means of the suffix *-ḍay* for females (e.g. *kulam* 'brother', *kulam-ḍay* 'sister', Peterson 2011: 139), but this only applies to humans and animals and is optional. There is hence no grammatical gender in Kharia. Kharia also shows no classification on the basis of form, so there are no declensional or conjugational classes either.

5.5 No agreement

Agreement is the marking of a property of a certain element on some other element. In FDG, agreement is distinguished from cross-reference (Hengeveld & Mackenzie, 2008: 350), as discussed in section 2.1. Two types of agreement are possible: operator agreement (e.g. tense copying, as discussed in section 5.2) and argument agreement. The latter is attested in, for instance, French, where an explicit subject is obligatory while the verb is also marked for person (*nous chant-ons* '1.PL sing-1.PL'). Since both subject marker and pronoun are Subacts of Reference, there is one meaning (1.PL) for two forms, which is obviously non-transparent.

As discussed above, Kharia is a pro-drop language and is therefore considered to exhibit cross-reference and no subject-verb agreement. There is also no agreement between heads and modifiers, as illustrated in example (28).

Peterson (2011: 81)

- (28) *ho rusuŋ oʔ*
 that red house
 'that red house'

The absence of agreement is a transparent property of Kharia.

5.6 No fusional morphology

When two units fuse into one at the Morphosyntactic Level, this single new form corresponds to two units at the Representational and Interpersonal Levels. This is called fusional morphology and it is opaque. According to Hengeveld (2007: 38), fusional morphology can be of two types. The first is cumulation: "the expression of more than one category in one morpheme". The second type is stem alternation, obtaining when the expression of a semantic category alters a lexical stem. Traditionally, the term 'fusional morphology' applies only to

cumulation, but in this paper, I will include stem alternation in this notion as well.

Stem alternation is shown to be more frequent when the functional rigidity of morphemes is larger in a language (Hengeveld 2007). In other words: the more flexible a language is, the less stem alternation we find. Since Kharia is a prototypical flexible language, it can be expected that if there is fusional morphology at all, it will not involve changes of lexical stems.

I found some instances of fusion in Kharia, notably in voice-tense markers on the predicate. These markers consist of one syllable which codes a combination of a voice (active or middle) and a tense (past or present), e.g. =*ki* ‘MID.PST’ (Peterson 2011: 240). The two parameters cannot be taken apart; they are ‘melted’ into one morpheme. We are thus dealing with a case of fusional morphology and hence an example of opaqueness in Kharia. It involves cumulation, as grammatical items are combined without altering a lexical stem. Person and number are also cumulated in predicate markers and pronouns, e.g. *iŋ* ‘1.SG’.

The question whether Kharia has stem alternation is a bit more difficult to answer. There are two cases in point. The first is a group of predicates that are marked for middle and active voice not by enclitics, but by alternations in the last syllable of the stem. An example is *bagɽe* / *bagɽay* ‘be(come) bad (MID)’ / ‘ruin (ACT)’ (Peterson 2011: 221). However, all words that show this stem alternation are loanwords from Sadri. Consequently, this may be stem alternation, but it is not characteristic for Kharia.

Another case of stem alternation occurs at the boundary of stems and clitics. For instance the final /l/ of the stems *col* ‘go’ and *ɖel* ‘come’ is deleted before enclitics starting with /n/ (Peterson 2011: 34). The stem is fused with the clitic, which violates domain integrity. The violation is not ‘severe’ in the sense that the lexical stem is still recognizable, but this is opacity in a strict sense. Kharia, then, is non-transparent with respect to this feature.

6 The Phonological Level

This section discusses phenomena at the Phonological Level that have no trigger or correlate at any higher level. This involves several assimilation strategies (degemination, nasalization, insertion, deletion) that occur between adjacent phonemes. Such assimilations are opaque, because they obscure boundaries and violate domain integrity.

6.1 No degemination

If two consonants stand next to each other (either within a word or across a word boundary), they tend to be pronounced as one consonant in many languages. An

example is the Dutch word *krabpaal* ‘cat tree’ (/krap/ + /pal/ = /krapaal/), where the double /p/ is reduced to a single /p/. The double consonant is called a geminate, and the process of reducing it to one consonant is called degemination.

In Kharia, geminates are maintained, as we see in example (29). Here, the geminate is phonemic.

Peterson (2011: 31)

(29)	u=naʔ	un=na=ʔ
	/unaʔ/	/un:aʔ/
	this=GEN	place=INF=GEN
	‘of this’	‘to place (GEN)’

However, there seem to be a few cases of opaque degemination. An example is the form *koleʔj* ‘to curse one another’, a combination of *kol* ‘reciprocal’ and *leʔj* ‘curse’ (Peterson 2011: 31). Even though this not a frequent phenomenon, it reflects an instance of opacity in Kharia.

6.2 No nasalization

Nasalization is the process where a vowel that is adjacent to a nasal also becomes nasal. Peterson (2011: 27) states that there is some non-phonemic nasalization in Kharia, but he gives no examples. If such cases indeed exist, we would be dealing with yet another example of opacity in Kharia’s phonology.

6.3 No insertion

In Kharia, when an enclitic starting with a vowel attaches to a lexeme ending in a vowel, a -w- or -y- is added in between, thus preventing diphthongization (Peterson 2011: 35). This means that a sound is inserted for articulatory reasons – not because of a semantic or pragmatic reason. This is non-transparent.

6.4 No deletion

In many languages, when vowels are adjacent to each other, one is deleted. In Kharia, clitics that end in a vowel often lose this vowel when the following clitic starts with a vowel. This is what happens in example (30).

Peterson (2011: 64)

(30)	am=pe	+	=yaʔ	>	am=p=aʔ
	2=2.PL		=GEN		2=2.PL=GEN

The deletion of the vowel reduces the forms of the clitics =*yaʔ* and =*pe*. These reductions are not pragmatically or semantically motivated and hence opaque.

Summing up, we see that Kharia shows instances of phonological assimilation: degemination, nasalization, phoneme insertion and deletion. Even though there are only few examples, the presence of phonological assimilation is a non-transparent property of Kharia. Note, however, that the phonological adaptations mostly involve clitics. Only very rarely do alternations occur in lexical stems.

7 Conclusion

In this paper, the behaviour of Kharia with respect to several transparent features was analysed. Kharia shows several opaque features. Non-transparent features attested in Kharia are apposition (including cross-reference) and a passive construction, resulting in some morphosyntactic argument alignment. Moreover, Kharia exhibits fusional morphology, non-parallel alignment because of ditropic clitics, discontinuity as a result of infixes (and, possibly, as a result of extraposition), some phonological adaptation processes, and influence of phonological weight on word order.

It might appear from this opsomming that Kharia is a highly opaque language. However, taking into account the unimportance of some of these features in the grammar, this leads to a different view. The evidence for a syntactic function Subject (the passive construction), as well as the evidence for non-parallel alignment (ditropic clitics) and phonological assimilation (few instances that mostly apply to clitics, not to lexical stems) is infrequent and not ‘severe’. The dominant traits of Kharia’s grammar are in fact semantically or pragmatically based.

Kharia maintains a one-to-one relation between units at the four levels of linguistic organization with respect to all other features. The flexibility of Kharia, that is, the fact that lexemes are precategorical, is a key factor here. This enables pragmatic and semantic units to be expressed irrespective of morphosyntactic information. Another important transparent feature in Kharia is the extensive use of phrase-marking clitics. Clitics typically attach to all units regardless of morphosyntactic category or complexity. Again, morphosyntactic information is disregarded in Kharia – the direct expression of pragmatic and semantic information is not overruled by formal processes. The dominance of semantics in Kharia is also visible in the expression of argument structure. Even though, as said, there is a passive construction, the alignment of Kharia is predominantly semantically grounded.

Kharia then, turns out to have a preference for semantically based structure over morphosyntactically based structures. Non-transparent features

often involve superficial, articulation-based processes. The larger architectural properties of Kharia favour semantics (function) over morphosyntax (form). Kharia, then, appears to be a highly transparent language. Whether this is true, will become clear from a comparison with other languages in the epilogue of this issue.

References

- Evans, Nicholas & Toshiki Osada (2005). Mundari: The myth of a language without word classes. *Linguistic Typology* 9 (3); 351-390.
- Hengeveld, Kees (1992). Non-verbal predication: Theory, typology, diachrony. Berlin: Mouton de Gruyter. (Functional Grammar Series 15).
- Hengeveld, Kees (2007). 'Parts-of-speech systems and morphological types'. *ACLCLC Working Papers* 2 (1); 31-48.
- Hengeveld, Kees & Lachlan Mackenzie (2008). *Functional Discourse Grammar. A typologically-based theory of language structure*. New York: Oxford University Press.
- Hengeveld, Kees & Eva van Lier (2010). An implicational map of parts of speech. *Linguistic Discovery* 8 (1); 129-56.
- Leufkens, Sterre (2009). *Tense copying in Functional Discourse Grammar*. Unpublished Research Master paper, University of Amsterdam.
- Leufkens, Sterre (forthcoming). The transparency of creoles. *Journal of Pidgin and Creole Languages*.
- Lewis, Paul M. (ed.) (2009). *Ethnologue: Languages of the World*, sixteenth edition. Dallas, Texas: SIL International. Online version: www.ethnologue.com.
- Peterson, John (2011). *A grammar of Kharia. A South Munda language*. Leiden/Boston: Brill.
- Travis, Lisa (1984). *Parameters and effects of word order variation*. Cambridge: MIT dissertation.