Verbalization as Re-categorization of Lexical Categories in Santali

BISWANATH DASH, *Indian Institute of Technology Delhi*

ABSTRACT

A well-known lexicon-syntax debate in the generative tradition concerns whether word formation occurs in the lexicon or in syntax (Bruening, 2018; Embick & Noyer, 2007). This paper builds on the idea of word formation/ categorization as a syntactic process, focusing on verbalization. In the literature that takes categorization as a syntactic process, verbalization is considered either idiosyncratic or compositional (Arad, 2003). Typological literature (Rijkhoff & van Lier, 2013; Peterson, 2011, 2010; Rau, 2013) indicates that Austro-Asiatic (AA) languages such as Santali and Kharia possess flexible verbal categorization, where a root x can behave like both a noun (N) and a verb (V), defying the N-V distinction that is found in most languages. However, I show empirical support from Santali, an AA language spoken in the Indian states of Odisha and Jharkhand, to argue that verbal categorization is a compositional syntactic process in Santali, where any root must go through a categorization process forming an N or adjective (A) before getting verbalized.

Santali displays high semantic transparency in verbalization, where the verbalized items have a predictable meaning of an N or A. This paper analyzes Santali fluid verbalization and compares it with the kinds of verbalization seen in English. It also questions how re-categorization (verbalization of lexical categories, not roots) incurs a predictable meaning in the verbalized structures and which head of the structure takes care of the semantic transparency or compositionality in Santali.

1 Introduction

Categorization is the most rudimentary trait of human cognition (Harnad, 2017). The paper focuses on lexical categories, which are, according to Baker (2003), the fundamental concepts humans learn, providing special emphasis on verbalization [1] in Santali.

The distinction among lexical categories (LC), more specifically, between N and V, is considered to be the most robust categorial distinction. If any language has a distinction among its lexical categories at all, it is between N and V (Sapir et al., 1944; Whaley, 1996, p. 32; Evans, 2000, p. 103; Croft, 2002, p. 183). The universality of the distinction between N and V is attested in both typological and generative literature. Also, Baker (2003) considers A to be one of the universally present LCs.

A conflicting claim that has long been reported in the literature states that some languages do not have dedicated categories for basic communicative functions like reference,

[1] Verbalization refers to changing any non-verbal category, like N or A, to a V.

predication, and modification. These languages, instead, have fluid word classes, the members of which can carry out more than one of these communicative functions (Rijkhoff & van Lier, 2013). Table 1 presents a list of languages with fluid categorial distribution. These are few of the languages that are reported as the ones without a clear distinction among the LCs.

Language Family	**Languages**	**Reference**
Malayo-Polynesian	Tongan	Churchward, 1953; Broschart, 1997
	Samoan	Churchward, 1951
	Tagalog	Foley, 1998
	Riau Indonesian	Gil, 2013
	Sri Lanka Malay	Nordhoff, 2012
Wakashan family	Nootka	Hockett, 1958; Mithun, 1999
Turkic languages Australian languages Salishan languages		Van Lier & Rijkhoff, 2013
Austroasiatic languages	Santali	McPhail, 1953; Rau, 2013
	Kharia	Peterson, 2011, 2013

Table 1: Languages with a categorial fluidity

It is important to note that not having the distinction among the categories does not mean lacking distinction among the functions like reference, predication, and modification (Gil, 2005). Rather, when we assume that there is no N-V distinction, it means that there is no distinction in the mental lexicon. The subsequent step is to investigate the extent of fluidity among LCs in these flexible languages by using empirical data to understand the absence of categorial distinction.

The paper is organized as follows: section 2 presents empirical support from Santali, an AA language that is reported to have a flexible word class (Table 1), to show that even if there is no N-V distinction in the lexicon, syntax can differentiate between the two. It also uses some diagnostic tests to establish that the construction types under study are instances of verbalization. Section 3 compares Santali verbalization with similar constructions in English-type languages to differentiate between compositional and idiosyncratic verbalization. Subsequently, in section 4, I differentiate Santali verbalization as re-categorization

of already categorized elements from root-derived categorization. This section also discusses the functional head in the derivation responsible for the extreme fluid verbalization in Santali. Section 5 concludes the paper.

2 Fluid categorization in Santali

2.1 Santali

Santali is one of the widely spoken languages from the North Munda sub-group spoken dominantly by the Santal community, resulting in it being the third most spoken AA language. It is majorly spoken in central and eastern parts of India, including the states of Odisha, Jharkhand, Chhattisgarh, Madhya Pradesh, Uttar Pradesh, West Bengal, and Maharastra, etc. (Anderson, 2015) and also in eastern Nepal and western Bangladesh (Peterson, 2015). The specific empirical support for this study is provided by the data from Santali spoken in Odisha and the border regions of Odisha and Jharkhand (Figure 1).

Figure 1: Santali spoken regions in India

The highlighted regions in Figure 1 include districts like Saraikela, Purbi Singhbhum,

Paschimi Singhbhum, Gumla, and Simdega in the state of Jharkhand and Mayurbhunj, Sundargharh, and Jharsuguda in Odisha where Santali is predominantly spoken.

2.2 Verbalization as Fluid Categorization

As mentioned earlier, Santali, along with many other AA languages, is well known in the literature for its weak distinction between N and V. The broader classification substituting the rudimentary N-V classification is the contrast between an argument and a predicate. Any morpheme can attach with a case marker, number marker, and sometimes a definite marker and behave like an argument/N(P). Similarly, a concept can merge with tense, aspect, mood (TAM) markers, phi, and voice markers to behave as a predicate/V(P). This entails that there is no N-V distinction in the lexicon, since the same concept can behave as an N or a V in syntax, depending on the grammatical markers it attaches to. I use the terms N and V to denote 'argument' and 'predicate', respectively. In this paper, since we focus majorly on the verbalization, we will see how the TAM, phi and the voice marker verbalize any morpheme in Santali (1). In (1), the tense/voice marker *-en* is attached to the verb *sen* 'go' (1a). The same marker *-en* is also attached to the noun *raajaa* 'king' (1b) stamping out the N-V distinction in the lexicon and motivating a syntactic categorization assumption.

(1) Santali

a. uni sen-en-a-e
He go-PST.MID-FIN-3SG
'He left.'

b. uni raajaa-en-a-e
He king-PST.MID-FIN-3SG
'He king-ed.' (He became a king)

Peterson (2003, 2010, 2011, 2015) also shows similar evidence from Kharia that supports a syntactic account of categorization (2).

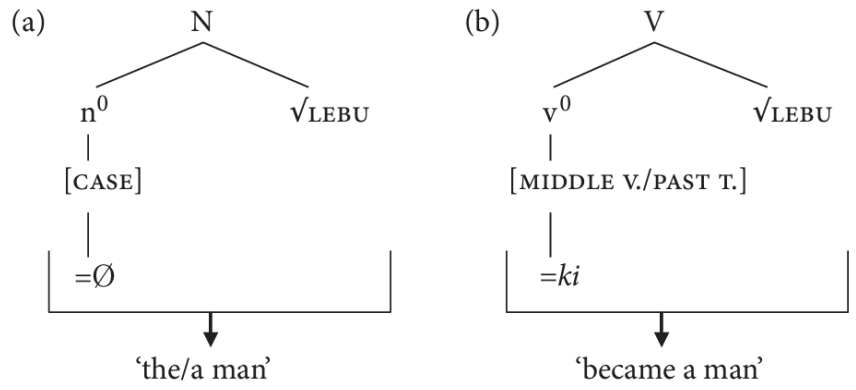
(2) Kharia

a. lebu Del-ki
man come-PST.MID
'The man came.'

b. bhagwan lebu-ki
God man-PST.MID
'God mann-ed.' (God became a man)

According to Peterson, the categorial status of a morpheme is decided depending on whether a root merges with a nominal (n) or verbal (v) categorizer (3).

(3) Kharia



In this paper, however, while analysing such a fluid categorial phenomenon in Santali, I claim that even if categorization in the AA languages like Santali and Kharia take place in the syntax, there is more to the process of categorization than what has been established in the literature (Peterson, 2003, 2011, 2015) (see Section 4 for more).

Before giving an analysis of how verbalization of non verbal entities happens in Santali and comparing it with similar looking phenomenon in other languages, we need to establish if the constructions (1) are really instances of verbalization. Looking at the constructions in (1), we can have two possibilities on the surface. First is that the non-verbal element like *raajaa* 'king' (1b) merges with the verbal markers and behaves as a verb. Secondly, there is a possibility of the presence of a null copula [2] that takes the verbal clitics which results in a verbalized kind of predicate on the surface. We now use some diagnostics to check if the verbalized looking structure is really verbalization or attachment of verbal markers on a null copula.

2.3 The Verbalization Tests

We perform three tests to see if the verbal clitics attach to the non-verbal entities like N or A, resulting in denominal/deadjectival verbalization or the verbal markers attach to the null copulas presenting the verbalization-like illusion in a structure like (1b) repeated as (4).

(4) uni raajaa-en-a-e
He king-PST.MID-FIN-3SG
'He king-ed.' (He became a king)

[2]The motivation for the null copula assumption comes from one of the comments during the FASAL (14) presentation. The idea was that since Santali verbalization structures give a regular become meaning, there could be a null copula present and the TAM PF forms, that look like cliticizing to the N/A, are actually markers of the null (become) copula. The tests in section 2.3 proves the verbalization claim, contrasting with the null copula assumption.

2.3.1 NP Scrambling

The first test concerns sentences with NP scrambling.
Test 1: NP1 NP2 -TAM.phi = NP2 NP1 -TAM.phi
Let us assume the possibility of presence of a null copula that takes the verbal markers in the predicate in (4). Therefore, we assume that in (4), *uni* 'he' is NP1 and *raajaa* 'king' is NP2, not a denominal verbal. There is a null copula after NP2, which hosts the verbal markers. If this assumption is true, Santali should behave like any free word order language where constructions like (4) are possible using a become copula resulting change of state semantics. Also, since Santali is a free word order language, we should be able to scramble both the NPs of a structure like (4) and get the same meaning. We can see that in Hindi, a free word order language that shows change of state semantics using a become copula, changing the order of NP1 and NP2 (5b) doesn't result ungrammaticality. However, when we apply the NP scrambling test to Santali, changing the order of the NPs in (6b) doesn't give the identical semantics as (6a).

(5) Hindi
   a. vah raajaa ban-a
      He king become-3SG.PST
      'He became a king.'
   b. raajaa vah ban-a
      King he become-3SG.PST
      'He became a king.'

(6) Santali
   a. uni raajaa-en-a-e
      He king-PST.MID-FIN-3SG
      'He became a king.'
   b. *raajaa uni-en-a-e
      King he-PST.MID-FIN-3SG
      'He became a king.'[3]

The null copula assumption is invalid according to the NP scrambling test.

2.3.2 Displacing verbal clitics

Verbal clitics are displaced in the second test.
Test 2: -TAM.phi NP1 NP2 = -TAM.phi NP1 NP2
Again, considering the free word order nature of Santali, if the null copula assumption is

[3]The sentence itself is not ungrammatical. When we scramble the NPs leaving the verbal clitics *in situ*, assuming there is a copula hosting it, the meaning of the sentence changes. The meaning of the sentence changes to 'The king became him', which is different from the expected meaning, 'He became a/the king'.

true, we should be able to displace the verbal clitics assuming they are hosted by the null copula, not the NP2. If the structure after displacement is grammatical, it will prove that constructions like (4) are not instances of verbalization. In Hindi, such a displacement is possible (7), in contrast to Santali (8).

(7) Hindi
  a. vah raajaa ban-a
     He king become-3SG.PST
     'He became a king.'
  b. ban-a vah raajaa
     become-3SG.PST he king
     'He became a king.'

(8) Santali
  a. uni raajaa-en-a-e
     He king-PST.MID-FIN-3SG
     'He became a king.'
  b. *-en-a-e uni raajaa
     -PST.MID-FIN-3SG he king
     'He became a king.'

(8b) shows that fronting a verbal clitic (with the assumed null copula) would yield an ungrammatical structure. This implies that there is no null copula and only *raajaa* can host the verbal clitic.

2.3.3 Inserting an intervener

In the third and the final test for verbalization we can check if the TAM and phi markers are attached to the NP or a null copula by inserting an interviner between the NP2 and the verbal clitics in (10a) and the resulting construction (10b) will be acceptable. Again, comparison with Hindi in (9), that has a become copula, shows that insertion of any grammatical marker after the NP2 doesn't incur ungrammaticality.

(9) Hindi
  a. vah raajaa ban-a
     He king become-3SG.PST
     'He became a king.'
  b. vah raajaa hi ban-a
     He king FOC become-3SG.PST
     'He became only a king.'

(10) Santali

a. uni raajaa-en-a-e
He king-PST.MID-FIN-3SG
'He became a king.'

b. *uni raajaa da-en-a-e
he king FOC-PST.MID-FIN-3SG
'He became only a king.'

If a null copula is the host to the verbal clitics, inserting any marker after the NP2 shouldn't create any issue, considering we are not inserting anything between the hypothetical null copula and the verbal clitics. Still, the ungrammaticality in (10b) shows that when we insert the focus marker *da*, it breaks the verbal structure of the denominal verbal and results in ungrammaticality.

These three tests show that the cliticization on the NPs in Santali take place because of verbalization where any category like N or A turn into verbs. Although a phenomenon like verbalization is not so unique across languages, as languages like English also has instances of verbalization (Clark & Clark, 1979), Santali verbalization structure are quite different from English-type languages and show much more productivity and regularity than other languages.

3 Comparing Santali Verbalization with English

Constructions like (1b) repeated as (4) are certainly not unique to Santali, as we see such instances of verbalization even in English (11). The sentences in (11) seem similar to to the Santali verbalized structure in (1b) and (4), since 'hammer', 'chair', 'water', 'tape', 'chain', etc. are prototypically used as nouns in English.

(11) English

a. He hammered the metal.
b. The professor chaired the meeting.
c. Peter taped the box.
d. The policeman chained the criminal.

While it may look like English and Santali are doing the same thing, we shall soon see that it is not the case. One of the major differences between English-type and Santali-type verbalization is that English verbalization is both idiosyncratic and compositional. In (11) two distinct types of verbalization can be noticed. Arad (2003) differentiates them as idiosyncratic and compositional[4] verbalization. The differences between idiosyncratic and compositional verbalization, on the surface, arises from the kinds of meaning each type

[4]Panagiotidis (2015) use the terms hammer-type and tape-type verbalization for idiosyncratic and compositional, respectively.

carries. 'Hammer' and 'chair' in (11a) and (11b) do not have any compositional correspondence with their nominal counterparts. On the other hand, 'tape' and 'box' in (11c) and (11d) have direct semantic correspondence with their nominal counterparts. 'hammer(v)' doesn't mean hit with a hammer, but 'tape(v)' mean seal with a tape. Arad shows example like (12) and (13) to argue that verbs in (12) are idiosyncratic and the meanings of the verbs are not dependent on a corresponding noun. The ungrammaticality in (13), on the other hand, shows that the verbs are compositional, and verbs like 'tape', 'chain', or 'button' cannot exist without the exact physical objects 'a tape', 'a chain', or 'a button'.

(12) English (Arad, 2003)
   a. He hammered the nail with a rock. (Kiparsky, 1982)
   b. String him up with a rope!
   c. She anchored the ship with a rock.

(13) English (Arad, 2003)
   a. *She taped the picture to the wall with pushpins.
   b. *They chained the prisoner with a rope.
   c. *Jim buttoned up his pants with a zipper.

Thus, English has both compositional as well as idiosyncratic verbalisation. Santali, however, displays only one type of verbalization, i.e., the compositional verbalization (14) [5]. The past tense middle voice marker *-en* in (14a) and (14b) produce intransitive structures and the past tense active voice clitic *-kidi* in (14c) and (14d) give a transitive structure of the intransitive counterparts.

(14) Santali
   a. uni dhiri-en-a-e
      He stone-PST.MID-FIN-3SG
      'He stoned.' (He became a stone)
   b. merhed martul-en-a-e
      metal hammer-PST.MID-FIN-3SG
      'The metal hammered.' (The metal became a hammer)
   c. jon uni-ke dhiri-kidi-a-e
      John he-ACC stone-PST.ACT-FIN-3SG
      'John stoned him.' (John changed him to become a stone)
   d. uni merhed-ke martul-kidi-a-e
      He metal-ACC hammer-PST.ACT-FIN-3SG
      'He hammered the metal.' (He changed the metal to become a hammer)

[5]Even if it shows tape-type compositional verbalization, Santali verbalization is much more regular and predictable than English, since, it shows only change of state semantics.

In Santali verbalization, there is no idiosyncrasy like some verbalized structures in English (11a) and (11b). The verbs in (14) show extreme predictability as they have direct semantic correspondence with their nominal counterparts. The Santali verbalized elements also display change of state (become) semantics (see Section 4.3 for more). Santali seems to have an extreme case of categorial fluidity in the case of verbalization; not only referent words like nouns (14), but any word can inflect for a verbal marker like tense and voice and get verbalized in syntax (15). Adjectives like *maaraang* 'big' (15a), demonstratives like *noa* 'this' (15b), kinship DPs like *ini bohya* 'my sister' (15c), animal names like *seta* 'dog' (15d), and even the most restricted kind of noun, i.e., proper names like *binit* 'Vineet' (15e) also get verbalized by taking the tense and voice clitics.

(15) Santali

a. daare maaraang-en-a
Tree big-PST.MID-FIN
'The tree bigged.' (The tree became big)

b. hana noa-en-a
That this-PST.MID-FIN
'that this-ed.' (That became this)

c. uni ini bohya-en-a-e
She my sister-PST.MID-FIN-3SG
'She my sister-ed.'[6] (She became my sister)

d. uni seta-en-a-e
He dog-PST.MID-FIN-3SG
'He dogg-ed.' (He became a dog)

e. uni binit-en-a-e
He Vineet-PST.MID-FIN-3SG
'He Vineet-ed.' (He became Vineet)

In comparison to the fluidity of verbalization in (15), English-type languages have some restrictions on verbalization.

The extreme compositionality seen in Santali verbalization (15) shows that the verbalized elements do not display any idiosyncrasy in meaning and carry the semantics of categories like, N, A, etc. There is always a change of state meaning in the formation of verbs from some category. This entails that verbalization in Santali is, in fact, re-categorization of an already categorized element. Now since, this is established from (15) that verbalization is re-categorization in Santali due to the lack of idiosyncrasy, the next step would require a detailed analysis of the idiosyncratic and compositional verbalization to show how the Santali verbalization takes place in Syntax.

[6]Context: My father married someone who had a daughter.

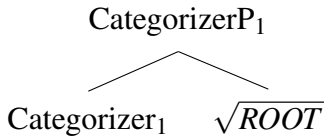
4 Verbalization is Re-categorization

It is evident from the extreme fluidity of the verbalization that there is a significant amount of overlap among categories in Santali. Words are underspecified with the categorial value in the lexicon. We employ a syntactic approach of word formation to explain Santali categorization. The fundamental assumption is that roots enter into the syntactic space without any categorial value. It is in syntax by the categorizer heads like *n* or *v*, the roots get their nominal or verbal categorial value. Since, we noticed both idiosyncratic and compositional verbalization in English (11) and extremely compositional verbalization in Santali, the subsequent step is to understand how the both types are generated in the syntax, considering syntax is the single universal derivational engine for word formation (Marantz, 1997, 2000).

4.1 Idiosyncratic and Compositional Verbalization on the Structure

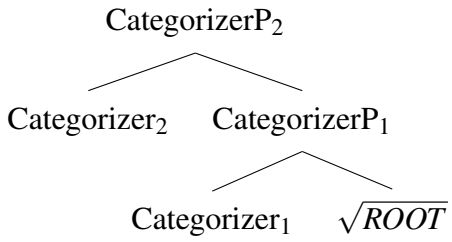
Arad (2003) discusses the differences between root-derived and word-derived verbalization to distinguish between idiosyncratic and compositional verbalization, respectively. According to the locality constraint on interpretation of roots (LCIR) (Arad, 2003), the interpretation of the root is restricted to the first categorizer[7] position in the derivational domain. In (16), $categorizer_1$ is the first categorizer, and the semantically underspecified root gets the interpretation when it merges with $categorizer_1$. The domain of the first categorizerP forms a closed interpretation domain (CID), and anything outside the CID does not have access to any atomic unit inside the domain, such as the root. The CID is the idiosyncratic/ non-productive domain.

(16)



Once the root merges with a categorizer, the meaning of the root is fixed in the first categorizerP. Any categorial derivation outside the CID, doesn't categorize the root, but recategorizes the first categorizer. In (17), the $categorizer_2$, that merges with the $categorizerP_1$, can only access the fixed interpretation of the root on the $categorizer_1$. It creates the compositional/ regular domain since it carries the semantics of the first categorizer head.

(17)



[7]The first categorizer position is described as the first phase position by Panagiotidis (2015).

From (14, 15), it is evident that Santali verbalization is extremely regular, displaying compositional semantics of an already categorized element, and never carries any idiosyncratic meaning. Therefore, the position of the Santali verbalizer head in the structure is the $Categorizer_2$ position in (17).

4.2 Against the Root-derived Approach

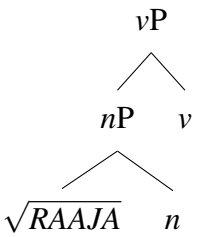
Since, according to Arad (2003), the idiosyncratic meaning of the root is restricted to the first merge position, the denominal and deadjectival kinds of verbalization doesn't take place by merely merging a verbalizer to a root. Based on this argument, we contradict the structure for the formation of N and V (3) in AA languages given for Kharia (Peterson, 2003, 2011, 2015) following the skeletal structure in (17).

In (1b) or (4), repeated here as (18), the V entails the interpretation of an N *raajaa* 'king'. When the V in the process of verbalization carries the meaning of an N, there must be an intervening nominal projection in the structure (Arad, 2003, p.759).

(18) uni raajaa-en-a-e
He king-PST.MID-FIN-3SG
'He king-ed.' (He became a king)

The *n* in (18) works as an intervener between the *v* and $\sqrt{RAAJAA}$, forming an N in the CID. According to the LCIR, $\sqrt{RAAJAA}$ gets nominalized in the CID and the meaning of it is fixed as a noun for the further derivations. When the *v* merges in the derivation, it accesses the nominal semantics of the $\sqrt{RAAJAA}$ from the *n* and verbalizes the noun *raajaa* 'king', resulting in the re-categorization of the N 19.

(19)



The intervening nominalizer brings nominal compositional meaning to the verbalized structure. The following section further elaborates on the compositionality dealing with the productive 'become' semantics in a verbalized structure.

4.3 Compositionality on the Structure

Santali carries a uniform *become* (change of state) semantics in the verbalized structure (15). At this stage, a pertinent question to ask is which head takes care of the *become* semantics in the structure. We adopt the analysis of Embick (2004) for resultative secondary

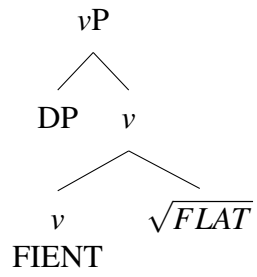
predicates (RSP) in English to see the position of become semantics on a verbalized structure.

Embick (2004) implements the analysis of Hale & Keyser (1993) for deadjectival verbals (20) where the *v*, that merges with the root to verbalize it, carries the *become* semantics as a feature [FIENT].

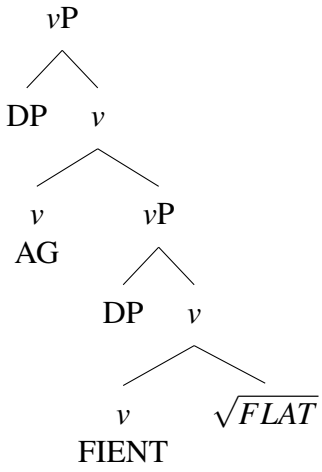
(20) a. The metal flatt-en-ed.
b. The smith flatt-en-ed the metal.

According to Embick, the [FIENT] feature is the become operator that denotes change of state or the transition event. In the process of verbalization, [FIENT] on the verbalizer (*v*), assigns the 'became flat' meaning to 'flat' (21). To explain the transitive structure in (20b), Embick puts the [FIENT] feature on the lower *v* in (22) that provides the 'changed something (the metal) to flat' meaning. The [AG] feature on the upper *v* in (22) is an agentive feature that licenses the agent on the external argument DP.

(21) Intransitive



(22) Transitive



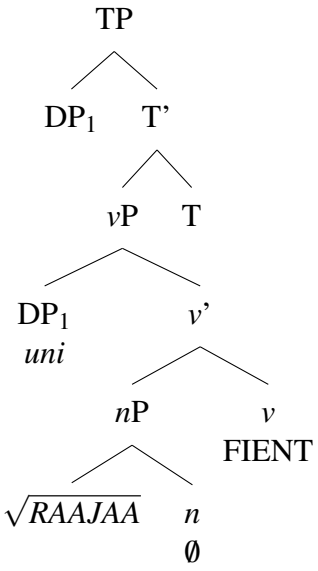
Since a productive *become* semantics is seen in Santali, unlike a restricted set of deadjectival verbalization in English, we adopt Embick's analysis to propose a [FIENT] feature on the verbalizer head in any word-derived verbalization (15) in Santali. Structures like (24) and (25) can explain the intransitive and transitive verbalized constructions like (23a) and (23b), respectively in Santali.

(23) Santali

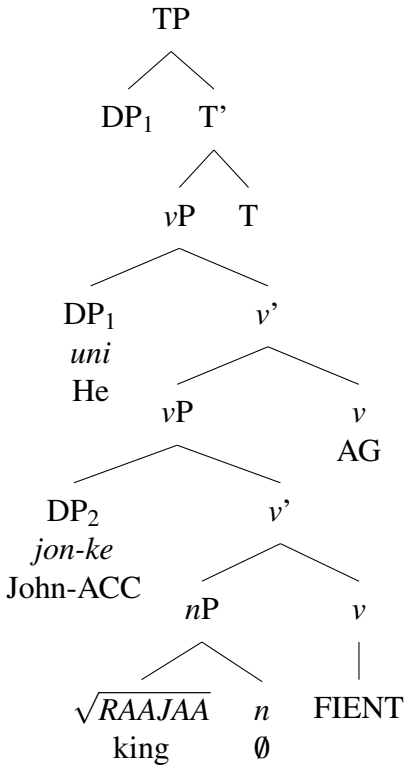
a. jon raajaa-en-a-e
John king-PST.MID-FIN-3SG
'John king-ed.' (John became a king)

b. uni jon-ke raajaa-kidi-a-e
He John-ACC king-PST.ACT-FIN-3SG
'He king-ed John.' (He changed John into a king)

(24) Intransitive



(25) Transitive



One of the major differences between Embick's analysis for deadjectival verbalization in English and our current analysis for word-derived (denominal, deadjectival, etc.) verbalization in Santali is that we emphasize on the re-categorization of already categorized elements by showing an intermediate categorizer.

5 Conclusion

One of the major findings of the current paper is that Santali verbalization is word-derived (re-categorized from a categorized element), not root-derived. Categorization in Santali-type languages is much more fluid than English-type languages in terms of productivity and compositionality. Secondly, unlike English-type languages, any concept or lexical category can be verbalized in syntax in Santali. The presence of the 'become' operator as a FIENT feature on the verbalizer head is the locus of the extreme verbal productivity in Santali. Even if syntax is the single derivational engine for word formation, there are language-specific variations of categorization depending on the productivity of word formation.

References

Anderson, Gregory DS. 2015. *The Munda languages*. Routledge.

Arad, Maya. 2003. Locality constraints on the interpretation of roots: The case of Hebrew denominal verbs. *Natural Language & Linguistic Theory* 21(4). 737–778.

Baker, Mark C. 2003. *Lexical categories: Verbs, nouns and adjectives*, vol. 102. Cambridge University Press.

Bruening, Benjamin. 2018. The lexicalist hypothesis: Both wrong and superfluous. *Language* 1–42.

Clark, Eve V & Herbert H Clark. 1979. When nouns surface as verbs. *Language* 767–811.

Croft, William. 2002. *Typology and universals*. Cambridge University Press.

Embick, David. 2004. On the structure of resultative participles in English. *Linguistic Inquiry* 35(3). 355–392.

Embick, David & Rolf Noyer. 2007. Distributed morphology and the syntax—morphology interface .

Evans, Nicholas. 2000. Kinship verbs. *Empirical Approaches to Language Typology* 103–172.

Gil, David. 2005. From repetition to reduplication in Riau Indonesian. In Bernhard Hurch (ed.), *Studies on reduplication*, 31–64. Berlin, Boston: De Gruyter Mouton.

Hale, Ken & Samuel Keyser. 1993. On argument structure and the lexical expression of syntactic relations. *An Annotated Syntax Reader* 312–327.

Harnad, Stevan. 2017. To cognize is to categorize: cognition is categorization. In *Handbook of categorization in cognitive science*, 21–54. Elsevier.

Kiparsky, Paul. 1982. Word-formation and the lexicon, Mid-America Linguistics Conference.

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In *Proceedings of the 21st Annual Penn Linguistics Colloquium*, vol. 4 2, 201–225.

Marantz, Alec. 2000. Roots: the universality of root and pattern morphology. In *Conference on Afro-Asiatic languages, University of Paris vii*, vol. 3, 14.

Panagiotidis, Phoevos. 2015. *Categorial features* 145. Cambridge University Press.

Peterson, John. 2003. Languages without nouns and verbs? an alternative to lexical classes in Kharia. In *Old and new perspectives on South Asian languages. Grammar and semantics. Papers growing out of the Fifth International Conference on South Asian linguistics (ICOSAL-5), held at Moscow, Russia in July*, .

Peterson, John. 2010. Language contact in Jharkhand: Linguistic convergence between Munda and Indo-Aryan in eastern-central India. *Himalayan Linguistics* 9(2).

Peterson, John. 2011. A grammar of Kharia: A south Munda language. In *A grammar of Kharia*, Brill.

Peterson, John. 2015. Kharia. In *The Munda languages*, 434–507. Routledge.

Rau, Felix. 2013. Proper names, predicates, and the parts-of-speech system of Santali. *Rijkhoff & van Lier (eds.)* 169–184.

Rijkhoff, Jan & Eva van Lier. 2013. *Flexible word classes: Typological studies of underspecified parts of speech*. OUP Oxford.

Sapir, Edward et al. 1944. Grading, a study in semantics. *Philosophy of Science* 11(2). 93–116.

Whaley, Lindsay J. 1996. *Introduction to typology: The unity and diversity of language*. SAGE publications.