Light Verb Constructions in Universal Dependencies for South Asian Languages

Abishek Stephen, Daniel Zeman
Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics
Malostranské náměstí 25, CZ-11800 Praha, Czechia
{stephen, zeman}@ufal.mff.cuni.cz

Abstract
We conduct a morphosyntactic investigation into the light verb constructions (LVCs) or the verbo-nominal predicates in South Asian languages. This work spans the Indo-Aryan and Dravidian language families in treebanks based on Universal Dependencies (UD). For the selected languages we show how well the existing annotation guidelines fare for the LVCs. We also reiterate the importance of the core and oblique distinction in UD and its usefulness for making accurate morphosyntactic annotation judgments for such predicates.

Keywords: light verbs, universal dependencies, multiword expressions

1. Introduction
Universal Dependencies (UD) (de Marneffe et al., 2021) presents a morphosyntactically oriented approach to perform linguistic annotations anchored on binary dependency relations between intra-sentential units. These dependency relations hold primarily between content words, while function words are seen as carriers of morphosyntactic features, which typically “belong” to a content word. Such a mechanism is followed in UD to increase the typological parallelism between languages. The selection of the dependency head gets a little complicated in the case of a multiword expression (MWE) where two or more words combine into a single lexical unit with or without morphosyntactic implications (Masini, 2019). One of the MWE classes where this can be witnessed is the light verb construction (LVC).

LVCs (Section 3) have a peculiar semantic composition that may provoke specific approaches to their syntactic analysis; however, in the case of South Asian languages, profound morphosyntactic clues are available and should be taken into account. The current annotations in the treebanks of these languages in UD treat the LVCs as combinations of lexemes that morphosyntactically behave as single words and mark them using the dependency relation compound, or its subtype compound:lvc. In the case of South Asian languages this is problematic given the surface-identical noun incorporations and object-verb sequences. We illustrate it on two examples from the treebanks of Hindi (Figures 1 and 2) and Telugu (Figures 3 and 4). In each pair, the first example has an LVC annotated as compound while the second example with a similar construction treats the noun as an object (obj) of the verb. Our main research question is whether these distinctions are well-motivated and clearly defined based on morphosyntax. It implies some broader questions about argument selection criteria and core vs. oblique distinction in South Asian languages.

---

1[https://universaldependencies.org/u/overview/syntax.html](https://universaldependencies.org/u/overview/syntax.html)
2For our study we consider all the noun-verb sequences marked as compound or compound:lvc in the treebanks as LVCs or verbo-nominal predicates.
3[https://universaldependencies.org/u/dep/compound.html](https://universaldependencies.org/u/dep/compound.html)
Figure 3: A verbo-nominal construction in Telugu (MTG) annotated as compound.

Figure 4: A verbo-nominal construction in Telugu (MTG) annotated as object.

Hence, using the treebanks of Indo-Aryan and Dravidian languages (Table 1) from UD 2.13 (Zeman et al., 2023), we intend to bring to light the fundamental issues around the treatment of various noun-verb sequences. We illustrate that not all noun-verb sequences qualify to be marked as compound or compound:lvc. We will focus on how the morphosyntactic implications have been overlooked by illustrating supporting examples for the same. Furthermore, we also emphasize the essential distinction between core and oblique arguments in UD (Zeman, 2017) that encompass a crucial role in the morphosyntactic treatment of the noun-verb sequences.

The paper is organized into 6 sections. Discussion of related works happens in Section 2. In Section 3, we present a portrait of LVCs in the selected UD treebanks, organized by language families. In Section 4, we discuss the structural composition of the LVCs by differentiating between incorporation and compounding. In Section 5, the morphosyntax of LVCs finds adequate theoretical treatment, confronted with treebank practice in Section 6.

2. Related Work

Kahane et al. (2018) discusses how to analyze multiword expressions in treebanks based on UD. They mainly focus on distinguishing syntactically irregular MWEs from semantically non-compositional ones and highlight issues related to intra-treebank annotation inconsistencies created because of the MWEs. The analysis concerns the English and French treebanks in UD 2.1 and they note inter-corpus variation in the usage of the dependency relation compound. But the LVCs did not receive any attention.

Nivre and Vincze (2015) portrays how LVCs pose interesting challenges for linguistic annotation, especially from a cross-linguistic perspective. They present a survey of the different ways in which LVCs are analyzed in UD 1.1. They group the languages into 3 groups and compare how the LVCs consisting of a transitive verb and a direct object are handled. For example, they report that in the English phrase take a photo, photo is attached to the verb take as a direct object (dobj) because the English treebanks in version 1.1 did not distinguish LVCs whereas the treebanks of Swedish, German, and Irish distinguish LVCs through their syntactic structure.

Since our study takes into consideration the constructions labeled as compound or compound:lvc it is worthwhile to mention that in the Persian treebank (Seraji et al., 2016) the non-canonical subjects are analyzed with respect to LVCs and such constructions are labelled as compound:lvc. In the case of the Hungarian treebank (Vincze et al., 2017), the label dobj:lvc can be found between the nominal and verbal component of the LVCs, where the dobj part of the label marks that syntactically it is a verb–object relation but semantically, it is an LVC, marked by the lvc subtype.

Among the South Asian languages, Hindi has received a considerable spotlight for LVCs. Palmer et al. (2009) talks about the LVCs as support-verb

Table 1: Treebank sizes in UD 2.13.

<table>
<thead>
<tr>
<th>Language</th>
<th>Treebank</th>
<th>Sentences</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sanskrit</td>
<td>Vedic</td>
<td>3,997</td>
<td>27,117</td>
</tr>
<tr>
<td>Sanskrit</td>
<td>UFAL</td>
<td>230</td>
<td>1,843</td>
</tr>
<tr>
<td>Hindi</td>
<td>HDB</td>
<td>16,649</td>
<td>351,704</td>
</tr>
<tr>
<td>Hindi</td>
<td>PUD</td>
<td>1,000</td>
<td>23,829</td>
</tr>
<tr>
<td>Urdu</td>
<td>UDTB</td>
<td>5,130</td>
<td>138,077</td>
</tr>
<tr>
<td>Kangri</td>
<td>KDB</td>
<td>288</td>
<td>2,514</td>
</tr>
<tr>
<td>Bhojpuri</td>
<td>BHBT</td>
<td>357</td>
<td>6,665</td>
</tr>
<tr>
<td>Bengali</td>
<td>BRU</td>
<td>56</td>
<td>320</td>
</tr>
<tr>
<td>Marathi</td>
<td>UFAL</td>
<td>466</td>
<td>3,847</td>
</tr>
<tr>
<td>Sinhala</td>
<td>STB</td>
<td>100</td>
<td>880</td>
</tr>
<tr>
<td>Telugu</td>
<td>MTG</td>
<td>1,328</td>
<td>6,465</td>
</tr>
<tr>
<td>Tamil</td>
<td>TTB</td>
<td>600</td>
<td>9,581</td>
</tr>
<tr>
<td>Tamil</td>
<td>MWTT</td>
<td>534</td>
<td>2,584</td>
</tr>
<tr>
<td>Malayalam</td>
<td>UFAL</td>
<td>218</td>
<td>2,403</td>
</tr>
</tbody>
</table>

4Our analysis will largely be centered around the languages with larger treebanks.
constructions in Hindi-Urdu where eventive noun phrases combine with several verbs and are analyzed based on case marking. The analysis relies on the Proposition Bank (Palmer et al., 2005) scheme. Begum et al. (2011) focus on the identification of the noun-verb combinations based on the Hindi Dependency Treebank (HDTB). Müller (2019) shows an HPSG analysis and Vaidya et al. (2014) present a TAG (Joshi, 2005) analysis for predicates with the light verbs karāṇa ‘to do’ and honā ‘to be’ in Hindi, demonstrating that LVCs are a highly productive predicational strategy, challenging for computational grammars.

The PARSEME (Savary et al., 2023) multilingual annotated corpus of verbal multiword expressions also includes Hindi. The underlying hypothesis for the annotations is that verbal MWEs have some degree of semantic non-compositionality and the verb is considered to be the syntactic head.

Within the UD framework, typological studies around LVCs have not involved any of the South Asian languages so far.

### 3. Light Verb Constructions in UD

The LVCs belong to the class of complex predicates with a wide range of combinatorial potential where a verb (VERB) can combine with adjectives (ADJ), adverbs (ADV) or nouns (NOUN). Out of these, we focus on the verbo-nominal predicates comprising words with the part-of-speech tags NOUN and VERB. This subgroup is most similar to (and confusable with) object-verb sequences; it also has interesting morphosyntactic properties.

#### 3.1. Indo-Aryan Languages

The Indo-Aryan languages are characterized by split ergativity, subject-object agreement, canonical SOV word order, and the presence of post-nominal case marking. UD annotation guidelines capture these morphosyntactic nuances aptly although certain inconsistencies remain especially in the case of LVCs. Currently, in UD 2.13, treebanks of Bengali, Bhojpuri, Hindi, Kangri, Marathi, Sanskrit, Sinhala, and Urdu are valid and publicly available. Most of these treebanks use the dependency label compound to mark the verbo-nominal compounds or LVCs but the Bengali, Marathi, and Sinhala treebanks use the language-specific dependency sub-type label compound:lvc. Figure 5 illustrates a verbo-nominal compound in Hindi bātacīta karāṇa ‘to talk’ where the verb karāṇa ‘to do’ selects the noun bātacīta ‘chit-chat’ as the dependent. Other verbs constituting such constructions in the Hindi HDTB and Hindi PUD treebanks include honā ‘to be’, which is the second most frequent verb constituting verbo-nominal predicates after karāṇa ‘to do’, followed by lagāṇā ‘to put’.

In Urdu, denā ‘to give’ and lenā ‘to take’ also head verbo-nominal compounds along with kmā and honā. In Marathi, verbo-nominal compounds function as semantic verbs with varying degrees of lexicalization (Ravishankar, 2017). Here, too, the verbs karāṇe ‘to do’ and honē ‘to be’ are the most frequently selected verbal heads in LVCs. Bengali (Figure 6), Bhojpuri and Kangri also present a similar picture where the verbs ‘to do’ and ‘to be’ persistently head such constructions. There are two verbs that function as light verbs in Sinhala, viz. kāra ‘to do’, the volitive indicator, and ve ‘to be’, the involitive indicator (Liyanage et al., 2023). The current version of the Sinhala treebank (STB) contains 39 instances of noun-verb combinations marked as compound:lvc. Sinhala happens to be the only Indo-Aryan language in UD to select the noun as a head for LVCs (Figure 7).

In the Vedic Sanskrit treebank, complex syntactic structures are expressed through compounds, hence compounds are annotated as if their elements occurred in a non-composed form (Hellwig et al., 2020). Recombination of certain compounds into single words is reported in the Sanskrit UFAL treebank (Dwivedi and Zeman, 2018); the
Dravidian treebanks is that the distinction between LVCs and regular structures has largely relied on semantic cues or direct influence of the strategy used in the English UD treebanks. Intra-language morphosyntactic clues do not seem to have been considered.

4. Structural Composition of LVCs

According to Butt (2003), the “light” in LVCs indicates that although these constructions respect the standard verb complement schema, the verb cannot be said to be predicating fully but seems to be more of a verbal licensor for nouns. Moreover, the light verbs tend to have a “funny” syntax which distinguishes them from auxiliaries and main verbs. Additionally, Butt (2003) claims that such structures are monoclausal in nature where the predicational elements “co-predicate”. Such a view does not align well with saying that they form one lexical (and syntactic) unit, but using the compound relation in UD can be understood as saying exactly that. There seems to be a perturbing dichotomy around the lexicality of such sequences as shown in Figure 9, where two instances are analyzed as compounds and one is not. In order to establish a principled position on the structural composition of LVCs, we will now delve into the process of compounding and incorporation and discuss their entanglement with the predicate structure.

4.1. Compounding

We adopt the definition of compounds based on Haspelmath (2023b) as a construction consisting of two strictly adjacent slots for roots\(^8\) that cannot be expanded by full nominal, adjectival, or degree modifiers. Finkbeiner and Schlücker (2019) illustrate the non-expandability on a German example, where the adverb *sehr* ‘very’ cannot modify the first element in *Alt-bau* ‘old building’, i.e., *sehr Alt-bau* ‘very old building’ is not plausible.

On applying Haspelmath’s definition to Figure 9, we observe that the noun part of the compound *surū kara* ‘to start’ is a root morph whereas the other nouns *goli* ‘bullet’ and *cunautī* ‘challenge’ are derived nominal forms of their respective root morphs. If we assume this inference to be accurate, then *cunautī denā* ‘to challenge’ and *goli caḷānā* ‘to shoot’ should not be marked as compound. Hence if a noun-verb sequence shall be considered a compound, the nominal part should be a root without suffixes.

\(^8\)A root is a contentful morph (i.e., a morph denoting an action, an object, or a property) that can occur as part of a free form without another contentful morph (Haspelmath, 2023b).
Iftar Sangam was inaugurated by KKM President Ibrahim Kunnil.

When challenged by the army, the terrorists started firing.

The UD taxonomy has a more relaxed definition of compounds: it states that the compound relation should be used for combinations of lexemes that morphosyntactically behave as single words, and lexicalization or semantic idiomaticity should not be a criterion for identifying compounds. This entails that a lexicalized expression like make a decision in English does not qualify as an MWE or a compound in UD. Expressions that would qualify should have a single argument structure or in other words, the syntactic head of an LVC should select all the required arguments and the dependent noun should neither be modified nor have an argument structure of its own. But in the case of the Indo-Aryan languages, this does not seem to be the case.

In Marathi (Figure 10) the LVC prayatna karata ‘trying’ is tagged as compound:lvc where the noun prayatna ‘try’ heads the nsubj and xcomp dependency relations which is not consistent with the UD guidelines. For once we could assume it to be a language-specific decision but there are also examples like Figure 11 which say otherwise. In both the examples (Figure 10 and 11) the compound:lvc relation is headed by the verb karane ‘to do’ but the dependent nouns are different. This leads a UD user to the conclusion that in such predicates the nouns have arbitrarily chosen argument structure as no morphosyntactic motivations can be seen in the surface syntactic structure. Similar inconsistencies can also be found in other Indo-Aryan languages. This inconsistent behavior suggests that the annotation choices made for the LVCs are not strongly based on a concrete morphosyntactic mechanism.

Among Dravidian languages, Tamil and Malayalam have taken a left-headed approach considering the noun as the head whereas Telugu treats the verb as the syntactic head making the compound:lvc relation right-headed. The annotation of the LVCs is comparatively more consistent than in the Indo-Aryan languages but it seems to be heavily influenced by semantics or by the treatment of LVCs in the English treebanks. For example, the current version of the Malayalam UFAL treebank uses the compound:lvc relation for noun-verb and verb-verb sequences where the do-verb ceyyuka appears. No morphosyntactic motivation can be found in the respective documentation pages of the Dravidian languages.

We conclude that if a noun-verb construction is marked as compound:lvc, the syntactic head is eligible for modifications but not the dependent. If we need to annotate a child of the dependent node in the noun-verb sequence, then the sequence should be treated as verb with object.
4.2. Noun Incorporation

It is also worthwhile to mention the broader typological definition of incorporation by Haspelmath (2023a) according to which an incorporation is an event-denoting noun-verb compound construction in which the noun occupies an argument slot of the verb and occurs in a position where nominal patient arguments cannot occur. In most Indo-Aryan languages, verbo-nominal predicates must be analyzed as a lexical category but paradoxically enough, the noun is on par with a syntactically independent argument (Mohanan, 1995). Therefore, even though noun incorporation is a type of compounding of a syntactic object with the verb, both the object and the verb can have their own argument structures. It may thus be hard to find incorporation that satisfies Haspelmath’s definition in South Asian languages. Currently, the UD taxonomy has no special provisions to define incorporation and they are treated as compounds. As a result, there are no distinct annotations for an object-verb pair and a ‘conjunct verb’.10 The Hindi HDTB treebank in UD is converted from the Paninian Dependencies and in that scheme, conjunct verbs have a special tag `po(f (Tandon et al., 2016). It does not denote a dependency but rather represents the fact that the noun-verb sequence is an MWE. The logic behind the usage of the `po(f tag is based on the semantic coherence of the noun-verb sequence being a single predicative element although some morphosyntactic cues do come in handy (discussed in Section 5). Tandon et al. (2016) also acknowledges that the identification of conjunct verbs is problematic as it appears to be an issue for the syntax-semantics interface and the decision was left to the annotators at the cost of inconsistencies in the data. On conversion from the Paninian dependencies to UD all the `po(f relations were automatically changed to compound and the inconsistencies persist. This brings us to a juncture where distinguishing object-verb sequences from noun incorporation becomes necessary. For Dravidian languages, Sudharsan (1998) states that if the noun in a noun-verb sequence cannot be inflected for case or number and even cannot be modified by an adjective then it is the case of a noun incorporated into the verb. Since incorporated nouns do not take case or plural markers and external modifiers, they are morphosyntactically different from the regular object nouns. Similarly for Indo-Aryan languages or more specifically for Hindi-Urdu, Mohanan (2017) has also recommended very similar criteria for distinguishing objects and incorporated nouns. These criteria treat noun incorporation as a type of compounding but there are also cases where such syntactic tests are inadequate, for example in cases of independent syntactic argument structures. The nominal part can be a noun or a root morph. Usually, the root morphs do not have an argument structure of their own but a noun on the other hand has the potential to have its own argument structure in such noun-verb constructions (Mohanan, 1995). To qualify for a compound:lv relation the noun-verb sequence should have a single argument structure but that is not always true in case of noun incorporations. This indicates a need for a distinction between compounding and noun incorporation. In the following section, we find taxonomical differences between them but it will be also worthwhile to test how similar their morphosyntax is and how we can distinguish them from object-verb sequences.

5. Morphosyntax of LVCs

Subjects and objects in UD must satisfy the condition of being core arguments, which means that they should receive the language-specific coding and treatment associated with the grammatical functions `S, `A, and `P (Zeman, 2017; Andrews, 2007). This coding derives from primary transitive predicates and may include various strategies.
including case marking on nouns and agreement morphology on verbs. Nominals whose grammatical function is A or S are called subjects and their dependency relation to the verb is nsubj whereas the nominals whose grammatical function is P are called (direct) objects and their dependency relation to the verb is obj (Zeman, 2017). Turning back to Haspelmath’s definition of noun incorporation in Section 4, the incorporated noun cannot occupy the patient position and cannot have the function P. Hence, we illustrate the behavior of LVCs through morphosyntactic processes like verbal agreement, case marking, and nominal modification. This analysis will bring out the distinctions between compounds and object-verb sequences.

5.1. Case Marking

Hindi, Urdu, and some other Indo-Aryan languages follow a split-ergative pattern. Perfective clauses have the ergative alignment, imperfective clauses have a nominative-accusative alignment. In the latter, the subject is in the bare nominative form (without adpositions), while animate direct objects use the postposition ko. Inanimate direct objects may omit the postposition ko; if they use it, the object is understood as definite. The accusative (oblique) case is used with the postposition, but without it, the object stays in nominative. Indirect objects always use the postposition ko. In transitive perfective clauses, the subject takes the ergative postposition ne.

Nominal parts of LVC candidates are inanimate and thus harder to distinguish from direct objects. However, the ability to take the optional ko signals that the noun is an object.

A few true LVCs, such as śurū karana ‘to start’, can be transitive as a whole. Here, śurū is not an object and the whole compound may take a real object (which follows the above criteria for objects) or a complement clause. In most cases, however, the nominal part of the LVC is a direct object, and if the whole LVC is semantically transitive, then the external “object” is coded as a nominal modifier (with the genitive postposition kā) of the noun in the LVC. It should then be annotated as nmod in UD (pula kā nirmāna ‘construction of bridge’ in Figure 1). Even with śurū karana the genitive strategy is a possible alternative and occurred twice in HDTB. The predicating nominals in Hindi may also select arguments with other postpositions, such as par ‘on’, se ‘from’, or ko ‘to’ (Vaidya et al., 2016).

Eastern Indo-Aryan languages such as Bhojpuri do not have the ergative alignment in perfective clauses. Similarly to Hindi, animacy and definiteness play a role in marking of the direct object (Thakur, 2021). However, Bhojpuri uses the same postposition (ke) (Figure 14) for accusative, dative, and genitive, making it less obvious when it is selected by the nominal and not the verb.

In Dravidian languages too the arguments are postpositionally case-marked but in an agglutinative manner. In Tamil MWTT, we find examples like kumār muṉṉukku vantāṉ ‘Kumar progressed (in his career/ life)’ where the nominal component muṉṉukku ‘to the front’ of the compound: lvc is assigned the dative case and the subject proper noun Kumar takes the nominative case. Since muṉṉukku is treated as the root the analysis gets blurry but muṉṉukku vā ‘to progress’ might not qualify to be considered as a compound due to the dative case marking.

The presence of an adpositional phrase selected by the nominal differentiates compounding
from noun incorporation but this does not provide a suitable distinction between object-verb sequences and noun incorporations at least for the Indo-Aryan languages. In this light, we observe that currently most of the compound:lvc or compound relations describing noun-verb sequences are not true compounds as the nominal participant does show case marking.

5.2. Agreement

The split-ergative pattern in some Indo-Aryan languages allows for testing of object-verb agreement. In imperative clauses, the gender and number of the subject are cross-referenced by the verb’s morphology. In transitive perfective clauses, the ergative postposition ne blocks agreement with the subject; but unless the direct object is marked with ko, verbal morphology cross-references the gender and number of the object (rather than subject). If the postposition ko is present, the verb takes the default masculine singular form.\(^\text{11}\)

Agreement with the verb in transitive-perfective clauses is another signal that the nominal of an LVC candidate is an object rather than part of a compound. And it can also attest to the opposite: In mere pitā ne pujā surū kar di hai ‘my father has started the prayer’, the verb has a feminine form, agreeing with pujā, while both pitā ‘father’ and surū ‘start’ are masculine.

Eastern Indo-Aryan languages (e.g., Bhojpuri and Bengali), as well as Dravidian languages, follow the nominative- accusative pattern with subject-predicate agreement and no ergativity (Krishnamurti, 2003). In Telugu, the verb agrees with the subject when it is in the nominative case, whereas when there is a dative “subject”, the verb agrees with the incorporated noun (Nadimpalli and Lakshmi, 2022). Similar observations can be made for other Dravidian languages except for Malayalam where subject-verb agreement is absent.

To conclude this section, in many instances of noun-verb sequences agreement between the noun and the verb is observed and represents a deviation from typical compound behavior.

5.3. Modification

One of the signs of compounds is that their parts (and especially the dependent part) cannot be modified individually. We have seen that the patient in Hindi LVC candidates is often encoded as a modifier of the predicative nominal, which speaks against a noun-verb compound analysis. Similarly,\(^\text{11}\)

\(^{11}\)While in general postpositions block agreement in Indo-Aryan languages, Gujarati is an exception where verb agreement works despite postpositions (Subbarao, 2012, p. 97).

5.4. Word Order

Real compounds would not allow intervening words between the noun and the verb (at least not by Hasselmath’s definition of compounds). An intervention seems to be always possible at least by the negative particle: unhorine batāyā ki abhi pahale baica kā praśikṣaṇa surū nahīṁ huā hai. ‘He told that the training of the first batch has not started yet.’

5.5. Transitivity

The grammars of Indo-Aryan languages feature a systematic opposition of transitive (causative) and intransitive verbs. The intransitive counterpart of karanā in Hindi is honā ‘to be, become, happen’; as shown in Section 3, its cognates do the same job in the other languages. Whenever it is inappropriate to analyze X karanā as a compound, the same can be said about X honā. However, as honā is intransitive, X can hardly act as its object. In Hindi-Urdu this verb is also used as the copula, hence a copular analysis may be an alternative. Where the light verb cannot be a copula, we should probably go with secondary predication (xcomp).

6. LVCs in UD Revisited

Noun-verb compounds are very frequent in the current UD treebanks of South Asian languages. In Hindi HDTB, there are 6187 such compounds with the 5 most common verbs alone (out of which 4159 occurrences belong just to karanā ‘to do’). A similar pattern is found in the smaller Urdu treebank:
3542 occurrences with the top 5 verbs, including 2346 with \( kmā \) ‘to do’. The remaining treebanks are an order of magnitude smaller, yet we find 58 different compounds in Bhojpur and 31 in Hindi PUD occurring twice or more. Nevertheless, the treebanks are not always consistent and it is not uncommon to see the same noun-verb combination annotated sometimes as a compound and sometimes as an object.

For example, Hindi \( bātā karaṇā \) ‘to talk’ is a relatively frequent expression and it is usually annotated as \texttt{compound} (118 instances), though occasionally it is annotated as \texttt{obj} (25 instances). The noun \( bātā \) can occur with the postposition \( ko \) and then it is always annotated as the object (13 instances). It can occur in the plural (11 instances without \( ko \) and 2 instances with \( ko \)) and there can occasionally be other constituents between it and the verb. In transitive perfective clauses, the verb agrees with its feminine gender: \( Naṭavara Siṁha \) (Masc) \textit{ne} \textit{Nirupama Sena se bātā (Fem) \( ki \) (Fem) hai} ‘Natwar Singh had spoken to Nirupam Sen’. The noun \( bātā \) can be also modified by a nominal denoting the matter that is being talked about. All this is evidence that \( bātā \) should be syntactically analyzed as the object of \( karaṇā \). For more statistics across the treebanks, see the Appendix.

Furthermore, based on the arguments present in Section 5, we can conclude that in the present versions of the treebanks of South Asian languages, the treatment of noun-verb sequences or LVCs as compounds is not consistent because the interplay of surface level similarities between real noun-verb compounds and noun incorporations somehow weigh down the morphosyntactic cues. There should not be a problem if noun-verb compounds satisfying the UD guidelines are marked as \texttt{compound:lvc} just to differentiate it from other type of compounds. This would also handle most of the noun incorporations, but once the nominal participant is case marked, modified or triggering verbal agreement, the sequence should be analyzed differently. One of the solutions could be to label the relation \texttt{obj:lvc}, modifying \textit{Vincze et al. (2017)}’s proposal to fit the current UD version. By doing so, there will be a three-way distinction between noun-verb compounds and noun incorporations (with a single argument structure) marked as \texttt{compound:lvc}, object-verb sequences marked as \texttt{obj} and noun-incorporations with individual noun and verb argument structures as \texttt{obj:lvc}.

7. Conclusion

We have presented morphosyntactic clues for identifying light verb constructions in South Asian languages, which could prove instrumental in achieving consistent annotations of \texttt{compound} and \texttt{compound:lvc} dependency relations. While LVCs as semantically idiosyncratic constructions are widespread in these languages, we have shown that in many cases their syntactic behavior is transparent or very close to standard object-verb constructions. Their compound analysis should be reconsidered and the annotation could be changed to \texttt{obj} or \texttt{obj:lvc} based on the type of argument sharing.

We also touched upon the core vs oblique distinctions and highlighted the phenomenon of noun incorporations, which can be beneficial for tackling similar inconsistencies beyond the languages handled in this study.

8. Acknowledgements

This work was supported by the Grant 20-16819X (LUSyD) of the Czech Science Foundation (GACR); and LM2023062 (LINDAT/CLARIAH-CZ) of the Ministry of Education, Youth, and Sports of the Czech Republic; and the Charles University project GA UK No. 101924; and partially supported by SVV project number 260 698.

9. Bibliographical References


10. **Language Resource References**

Zeman, Daniel and Nivre, Joakim and Abrams, Mitchell and Ackermann, Elia and Aeppli, Noémi and Aghaei, Hamid and Agić, Željko and Ahmadi, Amir and Ahrenberg, Lars and Ajede, Chika and Akkurt, Salih and Furkan and Aleksandravičiūtė, Gabriélė and Alfina, Ika and Algom, Avner and Alnajjar, Khalid and Alzetta, Chiara and Andersen, Erik and Antonsen, Lene and Aoyama, Tatsuya and Aplonova, Katya and Aquino, Angelina and Aragon, Carolina and Aranes, Glyn and Aranzaibe, María Jesus and Arican, Bilge and Arnardóttir, Pórunn and Arutile, Gashaw and Arwidarasti, Jessica and Arsalan, Deniz and Asmazoğlu, Cengiz and Ateyah, Luma and Atmaca, Furkan and Attia, Mohammed and Atutxa, Alitziber and Augustin, Liesbeth and Avelãs, Mariana and Badmaeva, Elena and Balasubramani, Keerthana and Ballesteros, Miguel and Banerjee, Esha and Bank, Sebastian and Barbu Mititelu, Verginica and Barkarson, Starkaður and Basile, Rodolfo and Bas-mov, Victoria and Batchelor, Colin and Bauer, John and Bedir, Seyyit Talha and Behzad, Shabbnam and Belenli, Juan and Bengoechea, Kepa and Benili, Ibrahim and Ben Moshe, Yifat and Berk, Gözde and Bhat, Riyaz Ahmad and Biagetti, Érica and Bick, Eckhard and Bielinskiene, Agnė and Bjarnadóttir, Kristín and Blokland, Rogier and Bobicev, Victoria and Boizou, Loïc and Borges Völker, Emanuel and Börstell, Carl and Bosco, Cristina and Bouma, Gosse and Bowman, Sam and Boyd, Adriane and Bragaar, Anouch and Branco, António and Brokalté, Kristina and Burchardt, Aljoscha and Campos, Marisa and Candito, Marie and Caron, Bernard and Caron, Gauthier and Carvalheiro, Catarina and Carvalho, Rita and Cassidy, Lauren and Castro, Maria Clara and Castro, Sérgio and Cavalcanti, Tatiana and Cebiroğlu Eryiğit, Gülşen and Cecchini, Flavio Massimiliano and Celano, Giuseppe G. A. and Čeploň, Slavomír and Cesur, Nesilhan and Cetin, Savas and Çetinoğlu, Özlem and Chalub, Fabricio and Chamila, Liyanage and Chauhan, Shweta and Chi, Ethan and Chika, Tsai-Shi and Cho, Yongseok and Choi, Jinho and Chun, Jayeol and Chung, Jong-Min and Cignarella, Alessandra T. and Cinková, Silvie and Collomb, Aurélie and Çöltekin, Çağrı and Connor, Miriam and Corbetta, Claudia and Corbetta, Daniela and Costa, Francisco and Courtin, Marine and Crabbé, Benoît and Cristescu, Mihaela and Cvetkoski, Vladimir and Dale, Ingerid Løyning and Daniel, Philemon and Davidson, Elizabeth and de Alencar, Leonel and Dehouck, Mathieu and de Laurentiis, Martina and de Marneffe, Marie-Catherine and de Paiva, Valeria and Derin, Mehmet Oguz and de Souza, Elvis and Diaz de Ilarraza, Antanza and Dickerson, Cary and Dinakaramani, Arawinda and Di Nuovo, Elisa and Dione, Bamba and Dirix, Peter and Dobrovoljč, Kaja and Doyle, Adrian and
Dozat, Timothy and Droganová, Kira and杜兰, Magali Sanches and Dwivedi, Puneet and Ebert, Christian and Eckhoff, Hanne and Eguchi, Masaki and Eiche, Sandra and Eli, Marhaba and Elkahky, Ali and Ephrem, Binyam and Erina, Olga and Erjavec, Tomaz and Essaidi, Farah and Etienne, Aline and Evelyn, Wograin and FACundes, Sidney and Farkas, Richard and Favero, Federica and Ferdaousi, Jannatul and Fernanda, Marilia and Fernandez Alcalde, Hector and Fethi, Amal and Foster, Jennifer and Fransen, Theodorus and Freitas, Cláudia and Fujita, Kazunori and Gajošová, Katarína and Galbrath, Daniel and Gamba and Garcia, Marcos and Gårdenfors, Moa and Gerhard, Fabrício Ferraz and Gerdes and Kim and Gessler, Luke and Ginter, Filip and Godoy, Gustavo and Goenaga, lakes and Gojenola, Koldo and Gökkırmak, Memduh and Goldberg, Yoav and Gómez Guinovart, Xavier and González Saavedra, Berta and Griciūtė, Bernadeta and Groni, Matias and Grobol, Loïc and Grzüllis, Norman and Guillaume, Bruno and Guiller, Kirian and Guillot-Barbance, Céline and Günög, Tunga and Habash, Nizar and Hafsteinsson, Hinrik and Hajič, Jan and Haifajč jr., Jan and Hämäläinen, Mika and Hå Mj, Linh and Han, Na-Rae and Hanifmuti, Muhammad Yudistira and Harada, Takahiro and Hardwick, Sam and Harris, Kim and Haug, Dag and Heinecke, Johannes and Hellwig, Oliver and Hennig, Felix and Hladká, Barbora and Hlaváčová, Jaroslava and Hociung, Florinel and Höhle, Petter and Huang, Yidi and Huerta Mendez, Marivel and Hwang, Jena and Ikeda, Taka and Hanifmuti, Muhammad Yudistira and Harada, Takahiro and Hardwick, Sam and Harris, Kim and Haug, Dag and Heinecke, Johannes and Hellwig, Oliver and Hennig, Felix and Hladká, Barbora and Hlaváčová, Jaroslava and Hociung, Florinel and Höhle, Petter and Huang, Yidi and Huerta Mendez, Marivel and Hwang, Jena and Ikeda, Takumi and Ingason, Anton Karl and Ion, Radu and Iríma, Elena and Isha, Olajide and Islamaj, Artan and Ito, Kaoru and Jagodźińska, Sandra and Jannatul and Siratun and Jelínek, Tomáš and Jha, Apoorva and Jiang, Katharine and Johannsen, Anders and Jónsdóttir, Hildur and Jørgensen, Fredrik and Juntinen, Markus and Kašikara, Húner and Kabaeva, Nadezhda and Kahane, Sylvain and Kanayama, Hiroshi and Kanerva, Jenna and Kara, Neslihan and Karahöga, Ritván and Käsen, Andre and Kaydelen, Tolga and Kengatharaiyar, Sarveswaran and Kettnerová, Václava and Kharatyan, Lilit and Kirchner, Jesse and Klementieva, Elena and Klyachko, Elena and Kocharov, Petr and Kohn, Arne and Koksal, Abdullahit and Kopacewicz, Kamil and Korkiakangas, Timo and Köse, Mehmet and Koşevoy, Alexey and Kotsyba, Natalia and Kovalevskaitė, Jolanta and Krek, Simon and Krishnamurthy, Parameswaran and Kübler, Sandra and Kuqi, Adrian and Kuyprikçu, Özgüzhan and Kuzgun, Asli and Kwak, Sookyoung and Kyle, Kris and Laan, Kábi and Lalipala, Veronika and Lambertino, Lorenzo and Landó, Tatiana and Larasati, Septina Dian and Lavrentiev, Alexei and Lee, John and Lê Hồng, Phương and Lenci, Alessandro and Lertpradit, Saran and Leung, Herman and Levina, Maria and Levine, Lauren and Li, Cheuk Ying and Li, Josie and Li, Keying and Li, Yixuan and Li, Yuan and Lim, KyungTae and Lima Padovani, Bruna and Lin, Yi-Ju Jessica and Lindén, Kristina and Liu, Yang and Ljubešić, Nikola and Lobzhanidze, Irina and Loginova, Olga and Lopes, Lucelene and Lusito, Stefano and Luthfi, Andry and Luukko, Mikko and Lyshevskaya, Olga and Lynn, Teresa and Macketzan, Vivien and Mahamdi, Menel and Maillard, Jean and Makarchuk, Ilya and Makazhanov, Alibek and Mandl, Michael and Manning, Christopher and Manurung, Ruli and Maršan, Büşra and Márincer, Cătălina and Marecek, David and Marheinecke, Katrin and Markantonatou, Stella and Martinez Alonso, Héctor and Martín Rodríguez, Lorena and Martins, André and Martins, Cláudson and Mašek, Jan and Matsuda, Hiroshi and Matsumoto, Yuji and Mazzei, Alessandro and McDonald, Ryan and McGuinness, Sarah and Mendonça, Gustavo and Merzhevich, Tatiana and Miekkä, Niko and Miller, Aaron and Mischchenko, Karina and Missilä, Anna and Mittelut, Cătălina and Mitrofan, Maria and Miyao, Yusuke and Mojiri Foroushani, Amir and Molnar, Judit and Moloodi, Amir and Monemagni, Simo and More, Amir and Moreno Romero, Laura and Moretti, Giovanni and Moris, Shinsuke and Morki, Tomohiko and Moro, Shigeki and Mortensen, Bjartur and Moskalëvskiy, Bohdan and Muischen, Kadri and Munro, Robert and Murawaki, Yugo and Műürisepp, Kaili and Nawang, Pinkey and Nakhle, Mariam and Navarro Herniacek, Juan Ignacio and Nedolužhko, Anna and Nešporová, Zuzana and Nély, Györgyi and Nély, Rezsö and Nishioka, Naoko and Nishida, Mayuko and Nishida, Masato and Nishida, Minoru and Ogasawara, Hiroshi and Ortega, Yutaka and Otsuki, Eri and Ouyang, Shaoxun and Park, Hyunji and Patrício, Alice and Patrício, André and Passarotti, Marco and Patejuk, Agnieszka and Paulino-Passos, Guilherme and Pedone, Giulia and Peljak-Lapidfska, Angelika and Peng,
Siya and Peng, Siyao Logan and Pereira, Rita and Pereira, Silvia and Perez, Cenel-Augusto and Perkova, Natalia and Perrier, Guy and Petrov, Slav and Petrova, Daria and Peverelli, Andrea and Phelan, Jason and Pierre-Louis, Claudel and Pflitlulainen, Jussi and Pinter, Yuval and Pinto, Clara and Pintucci, Rodrigo and Pirinen, Tommi A and Pitler, Emily and Plamadja, Magdalena and Plank, Barbara and Poibeau, Thierry and Ponomareva, Larisa and Popel, Martin and Pretkalnina, Lauma and Prévost, Sophie and Prokopidis, Prokopis and Przepiórkowski, Adam and Pugh, Robert and Puolakainen, Tiina and Pyysalo, Samo and Qi, Peng and Querido, Andreia and Rääbis, Andriela and Rademaker, Alexandre and Rahoman, Mizanur and Rama, Taraka and Ramasamy, Loganathan and Ramisch, Carlos and Ramos, Joana and Rashef, Fatm and Rasooli, Mohammad Sadegh and Razemshankar, Vinit and Real, Livy and Rebeja, Petru and Reddy, Siva and Regnault, Mathilde and Rehm, Georg and Riabi, Arij and Riabov, Ivan and Riččler, Michael and Rimkutė, Erika and Rinaldi, Larissa and Rituma, Laura and Risjöyih, Putri and Rocha, Luisa and Rögnvaldsson, Eirikur and Roksanidc, Ivan and Romanenko, Mykhailo and Rosa, Rudolf and Roşca, Valentin and Rotavi, Davide and Rozonoyer, Ben and Rudina, Olga and Rueter, Jack and Rünnarsson, Kristján and Sadde, Shoval and Safari, Pegah and Sahala, Aleksis and Saleh, Shadi and Salomoni, Alessio and Samardžić, Tanja and Samson, Stephanie and Sanguinetti, Manuela and Sanjyur, Ezgi and Särg, Dage and Sartor, Marta and Sasaki, Mitsuya and Saurèle, Bäiba and Savary, Agata and Sawanakunanan, Yanin and Saxena, Shefali and Scannell, Kevin and Scarlata, Salvatore and Schang, Emmanuel and Schneider, Nathan and Schuster, Sebastian and Schwartz, Lane and Seddah, Djarné and Seeker, Wolfgang and Serafi, Mojgan and Shahzadi, Syeda and Shen, Mo and Shimada, Atsuko and Shirasu, Hiroyu and Shishkina, Yana and Shohibussirirr, Muh and Shvedova, Maria and Siwiet, Janine and Sigurdsson, Einar Freyr and Silva, Joao and Silveira, Aline and Silveira, Natalia and Silveira, Sara and Simi, Maria and Simionescu, Radu and Simkó, Katalin and Šimková, Mária and Simonarson, Haukur Barri and Simov, Kiril and Sitchinava, Dmitri and Sitter, Ted and Skacheduoba, Maria and Smith, Aaron and Soares-Bastos, Isabelba and Solberg, Per Erik and Sonnenhauser, Barbara and Sourov, Shafii and Sprugnoli, Rachele and Stamou, Vivian and Steingrimsson, Steinþórr and Stella, Antonino and Stephen, Abishek and Straka, Milan and Strickland, Emmett and Strnadová, Jana and Suhr, Alane and Sulestio, Yogi Lesmana and Sulubacak, Umut and Suzuki, Shingo and Swanston, Daniel and Szántó, Zsolt and Taguchi, Chihiro and Taji, Dima and Tamburini, Fabio and Tan, Mary Ann C. and Tanaka, Takaaki and Tanaya, Dipta and Taverni, Mirko and Tellia, Samson and Tellier, Isabelle and Testori, Marinella and Thomas, Guillaume and Tonelli, Sara and Torga, Liisi and Toska, Marsida and Trosterud, Trond and Turkhina, Anna and Tsarfaty, Reut and Türk, Utku and Tyers, Francis and Pórðarson, Sveinbjörn and Porsteinsson, Vilhjálmur and Uematsu, Sumire and Untilov, Roman and Urešová, Zdeňka and Uria, Larraitz and Uszkoreit, Hans and Utko, Andrius and Vagnoni, Elena and Vajjala, Sowmya and Vak, Socrates and van der Goot, Rob and Vanhove, Martine and van Niekerk, Daniel and van Noord, Gertjan and Varga, Viktó and Vedeniina, Ulia and Venturi, Giulia and Villemonte de la Clergerie, Eric and Vincze, Veronika and Vlasova, Natalia and Wakasa, Aya and Wallenberg, Joel C. and Wallin, Lars and Walsh, Abigail and Washington, Jonathan North and Wendt, Maximilian and Widmer, Paul and Wigderson, Shira and Wijono, Sri Hartati and Wille, Vanessa Berwanger and Williams, Seyl and Wirth, Mats and Wittner, Christian and Woldemariam, Tsegay and Wong, Tak-sum and Wróblewska, Alina and Wu, Qishen and Yako, Mary and Yamashita, Kayo and Yamazaki, Naoki and Yan, Chunxiao and Yasuoka, Koichi and Yavrumyan, Marat M. and Yenice, Arife Betül and Yildiz, Olcay and Yu, Zhuan and Yuliiawati, Arlisa and Žabokrtský, Zdeněk and Zahra, Shoroq and Zeldes, Amir and Zhou, He and Zhu, Hanzhi and Zhu, Yilun and Zhuravleva, Anna and Ziane, Rayan. 2023. Universal Dependencies 2.13. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. PID http://hdl.handle.net/11234/1-5287.

A. Appendix

Table 2 shows the most important relations going from a verb to a noun; in addition, it also shows compound relations going from a noun to a verb. It demonstrates that some treebanks favor the compound analysis much more than others, and three treebanks do not use the compound relation at all.

Table 3 shows some of the most frequent light verbs across the South Asian treebanks. Cognates are clearly observable in the Indo-Aryan languages but their preference in the individual languages varies (there are substantial differences even between Hindi and Urdu).
Table 2: Selected relations between verbs and nouns in UD 2.13 treebanks (only main relation types are shown, subtypes are merged with their main types). The relations go from the verb to the noun except for the “reversed compound” columns, where the noun is the parent node. **NV** means that the noun immediately precedes the verb; **NXV** means that the noun precedes the verb but there are one or more words between them; analogously, **VXN** means that the verb comes first, with at least one word between it and the noun. Frequencies are shown per 10K words; an empty cell means that the relation did not occur at all while zero means that it did occur but the normalized frequency is rounded down to 0.
<table>
<thead>
<tr>
<th>English</th>
<th>Hindi</th>
<th>HDTB</th>
<th>PUD</th>
<th>Urdu</th>
<th>Kangri</th>
<th>Bhojpuri</th>
<th>Bengali</th>
<th>Marathi</th>
<th>Sinhala</th>
<th>Malayalam</th>
</tr>
</thead>
<tbody>
<tr>
<td>do / make</td>
<td>karanā</td>
<td>118</td>
<td>56</td>
<td>knā</td>
<td>170</td>
<td>karaṇā</td>
<td>48</td>
<td>kara</td>
<td>38</td>
<td>63</td>
</tr>
<tr>
<td>make</td>
<td>karānā</td>
<td>3</td>
<td>1</td>
<td>krānā</td>
<td>1</td>
<td>krūnā</td>
<td>1</td>
<td>kara</td>
<td>63</td>
<td>1</td>
</tr>
<tr>
<td>do / make</td>
<td>karanā</td>
<td>3</td>
<td>1</td>
<td>krānā</td>
<td>1</td>
<td>krūnā</td>
<td>1</td>
<td>kara</td>
<td>63</td>
<td>1</td>
</tr>
<tr>
<td>happen / be</td>
<td>honā</td>
<td>22</td>
<td>12</td>
<td>hūnā</td>
<td>37</td>
<td>hoṇā</td>
<td>76</td>
<td>hō/bā</td>
<td>72</td>
<td>17</td>
</tr>
<tr>
<td>happen / be</td>
<td>honā</td>
<td>22</td>
<td>12</td>
<td>hūnā</td>
<td>37</td>
<td>hoṇā</td>
<td>76</td>
<td>hō/bā</td>
<td>72</td>
<td>17</td>
</tr>
<tr>
<td>give</td>
<td>denā</td>
<td>24</td>
<td>6</td>
<td>denā</td>
<td>36</td>
<td>denā</td>
<td>20</td>
<td>de</td>
<td>17</td>
<td>de</td>
</tr>
<tr>
<td>take</td>
<td>lenā</td>
<td>6</td>
<td>3</td>
<td>lenā</td>
<td>7</td>
<td>laiṇā</td>
<td>4</td>
<td>la</td>
<td>3</td>
<td>la</td>
</tr>
<tr>
<td>apply / put</td>
<td>lagānā</td>
<td>7</td>
<td>1</td>
<td>lgānā</td>
<td>2</td>
<td>laga</td>
<td>3</td>
<td>laga</td>
<td>3</td>
<td>laga</td>
</tr>
<tr>
<td>seem</td>
<td>laganā</td>
<td>1</td>
<td>1</td>
<td>lgnā</td>
<td>1</td>
<td>laganā</td>
<td>4</td>
<td>laganā</td>
<td>4</td>
<td>laganā</td>
</tr>
<tr>
<td>keep / put</td>
<td>rakhanā</td>
<td>2</td>
<td>1</td>
<td>rkhnā</td>
<td>7</td>
<td>rakhanā</td>
<td>4</td>
<td>rakhanā</td>
<td>4</td>
<td>rakhanā</td>
</tr>
<tr>
<td>stay</td>
<td>rahanā</td>
<td>0</td>
<td>0</td>
<td>rhnā</td>
<td>2</td>
<td>rhnā</td>
<td>2</td>
<td>rhnā</td>
<td>2</td>
<td>rhnā</td>
</tr>
<tr>
<td>create / make</td>
<td>banānā</td>
<td>2</td>
<td>0</td>
<td>bnānā</td>
<td>3</td>
<td>bnānā</td>
<td>3</td>
<td>bnānā</td>
<td>3</td>
<td>bnānā</td>
</tr>
<tr>
<td>be / become</td>
<td>banānā</td>
<td>2</td>
<td>0</td>
<td>bnānā</td>
<td>3</td>
<td>bnānā</td>
<td>3</td>
<td>bnānā</td>
<td>3</td>
<td>bnānā</td>
</tr>
<tr>
<td>come</td>
<td>ānā</td>
<td>2</td>
<td>3</td>
<td>ānā</td>
<td>3</td>
<td>ānā</td>
<td>8</td>
<td>ā</td>
<td>3</td>
<td>ā</td>
</tr>
<tr>
<td>drive</td>
<td>calānā</td>
<td>0</td>
<td>0</td>
<td>člānā</td>
<td>1</td>
<td>člānā</td>
<td>1</td>
<td>člānā</td>
<td>1</td>
<td>člānā</td>
</tr>
<tr>
<td>go / walk</td>
<td>calanā</td>
<td>1</td>
<td>2</td>
<td>čnā</td>
<td>1</td>
<td>čnā</td>
<td>1</td>
<td>čnā</td>
<td>1</td>
<td>čnā</td>
</tr>
<tr>
<td>meet</td>
<td>milanā</td>
<td>2</td>
<td>1</td>
<td>mlnā</td>
<td>1</td>
<td>mlnā</td>
<td>1</td>
<td>mlnā</td>
<td>1</td>
<td>mlnā</td>
</tr>
<tr>
<td>show / express</td>
<td>jatānā</td>
<td>3</td>
<td>3</td>
<td>jhtānā</td>
<td>1</td>
<td>jhtānā</td>
<td>1</td>
<td>udānā</td>
<td>4</td>
<td>udānā</td>
</tr>
<tr>
<td>pick</td>
<td>uṭhānā</td>
<td>2</td>
<td>1</td>
<td>āṭhānā</td>
<td>1</td>
<td>āṭhānā</td>
<td>1</td>
<td>āṭhānā</td>
<td>1</td>
<td>āṭhānā</td>
</tr>
<tr>
<td>cause</td>
<td>dilānā</td>
<td>1</td>
<td>1</td>
<td>dilānā</td>
<td>1</td>
<td>dilānā</td>
<td>1</td>
<td>dilānā</td>
<td>1</td>
<td>dilānā</td>
</tr>
<tr>
<td>put</td>
<td>dālanā</td>
<td>1</td>
<td>1</td>
<td>dālānā</td>
<td>1</td>
<td>dālānā</td>
<td>1</td>
<td>dālānā</td>
<td>1</td>
<td>dālānā</td>
</tr>
<tr>
<td>get / find</td>
<td>pānā</td>
<td>0</td>
<td>0</td>
<td>pānā</td>
<td>2</td>
<td>pānā</td>
<td>4</td>
<td>pānā</td>
<td>4</td>
<td>pānā</td>
</tr>
<tr>
<td>kill</td>
<td>māranā</td>
<td>1</td>
<td>1</td>
<td>mārnā</td>
<td>1</td>
<td>mārnā</td>
<td>1</td>
<td>mārnā</td>
<td>1</td>
<td>mārnā</td>
</tr>
<tr>
<td>fall</td>
<td>paḍanā</td>
<td>0</td>
<td>0</td>
<td>prnā</td>
<td>1</td>
<td>paṇā</td>
<td>8</td>
<td>paṇā</td>
<td>8</td>
<td>paṇā</td>
</tr>
</tbody>
</table>

Table 3: Selected lemmas of verbs that are connected with a noun via the compound relation (or its subtype), with the verb as the parent, in UD 2.13 treebanks. Frequencies are shown per 10K words; an empty cell means that the verb did not occur at all while zero means that it did occur but the normalized frequency is rounded down to 0.