Syntactic composition and selectional preferences in Hindi Light Verb Constructions

Ashwini Vaidya, Owen Rambow, Martha Palmer

October 6, 2017

1   Abstract

Previous work on light verb constructions (e.g. *chorii kar* 'theft do; steal') in Hindi describes their syntactic formation via co-predication (Ahmed et al., 2012, Butt, 2014). This implies that both noun and light verb contribute their arguments, and these overlapping argument structures must be composed in the syntax. In this paper, we present a co-predication analysis using Tree-Adjoining Grammar, which models syntactic composition and semantic selectional preferences without transformations (deletion or argument identification). The analysis has two key components (i) an underspecified category for the nominal and (ii) combinatorial constraints on the noun and light verb to specify selectional preferences. The former has the advantage of syntactic composition without argument identification and the latter prevents over-generalization, while recognizing the semantic contribution of both predicates. This work additionally accounts for the agreement facts for the Hindi LVC.

2   Introduction

Light verb constructions have been defined as "two (or more) predicational elements that each contribute to a joint predication" (Butt, 2010). A light verb construction e.g. *give a kiss* consists of a 'light' verb *give* and a second predicating element viz. the noun *kiss*. In comparison to the simple event verb *kiss*, the light verb construction has a more nuanced meaning, where there is a 'transfer' of the kiss from giver to receiver. The verb *give* in this light verb construction contributes this meaning, which is more abstract as compared to its ordinary lexical usage as *give* in *give a book*.

Light verb constructions (henceforth, LVCs) are found across languages such as Japanese, Korean, Persian and English as well as German. In languages like Hindi, Urdu and Persian, these constructions are particularly pervasive. LVCs in Hindi are often subsumed under the term 'complex predicates' because light verbs in Hindi can combine with another verb, an adjective, adverb or even a borrowed English verb or noun. Complex predication of this sort is highly

productive in Hindi, as well as other South Asian languages (Masica, 1976). In this paper, we continue to refer to predicating noun and light verb combinations as 'light verb constructions', though they are also sometimes referred to as 'support verb' or 'complex verb' constructions.

One of the primary syntactic challenges for LVCs is monoclausality with two predicating heads. For simple predicates, such as the examples seen in (1) and (2), a single set of semantic roles will map to their respective grammatical functions. E.g. subject and object to Agent and Patient respectively. In the case of LVCs as seen in (3), two predicating heads will result in a composite argument structure (see also Figure 1).

(1) Simple predicate

    mohan=ne   shyam=ko   ṭalaash-aa
    Mohan.M.Sg=Erg Shyam.M.Sg=Acc search-Perf.M.Sg

    'Mohan searched for Shyam'

(2) Simple predicate

    mohan=ne   apna homework   kiy-aa
    Mohan.M.Sg=Erg his.M homework.M.Sg do-Perf.M.Sg

    'Mohan did his homework'

(3) Light verb construction

    mohan=ne   shyaam=kii   ṭalaash   k-ii
    Mohan.M.Sg=Erg Shyam.M.Sg=Gen talaash.F.Sg do-Perf.F.Sg

    'Mohan made a search for Shyam'

This 'division of labour' problem, where the correct roles mapped to correct grammatical functions must eventually surface in a monoclausal structure is precisely the reason for the plethora of operations required for LVC analysis. Previous work e.g. *Argument Transfer* (Grimshaw and Mester, 1988) states that the number of arguments 'transferred' to the light verb can only take place in a particular order (Agents before Themes, for example).

Some analyses e.g. Hook and Pardeshi (2006) have considered the light verb to be a form of auxiliary. This allows the light verb construction to remain monoclausal without having to posit two predicating elements. Butt and Lahiri (2013) have argued that in fact light verbs in South Asian languages have been historically stable, and not subject to phonological attrition or semantic bleaching, which are characteristics of auxiliaries. Further arguments for monoclausality have shown that light verb constructions have a single subject in the clause (Butt, 1995). This implies that the syntactic representation must allow the representation of the argument structures of both the light verb and the predicating noun.
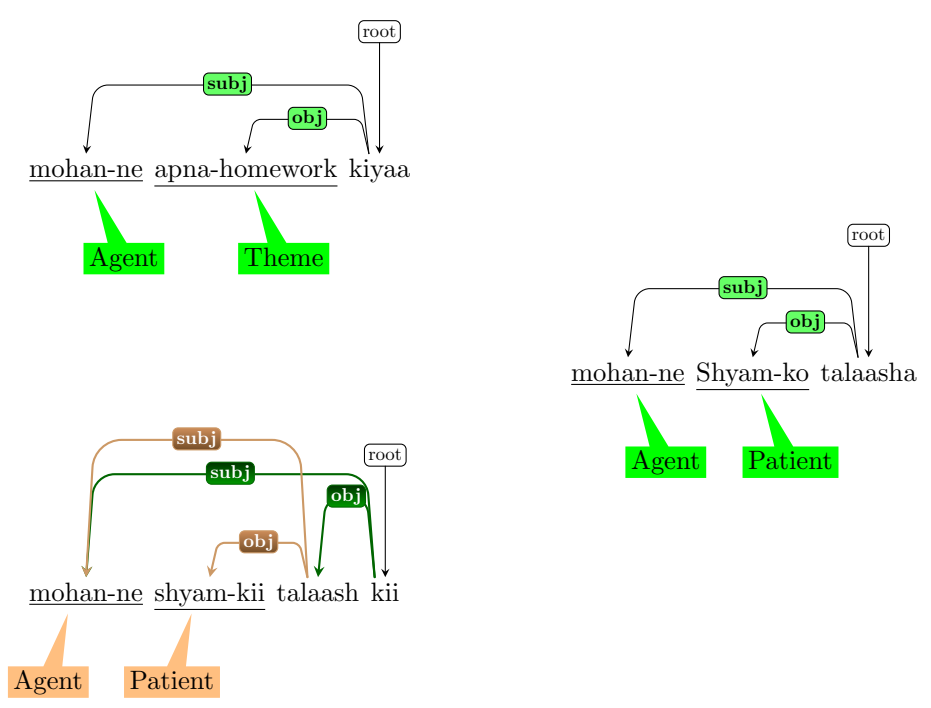
Figure 1: Argument mapping in simple predicates *kiyaa* 'do' (example 2), *talaasha* 'search' (example 1) vs. light verb construction *talaash kiyaa* 'search do; ', (example 3) with overlapping argument structures.

3

Another fallout of the monoclausality problem is the treatment of the light verb as simply a theta marker of arguments, which contributes nothing at all to the event description. Kearns (1988) and others describe predicate 'bleaching', where the light verb is stripped of any semantic contribution. The term 'bleaching' is somewhat misleading, as it does seem that the light verb contributes a core meaning e.g. volitionality, forcefulness, surprise or transfer (Hook, 1974). Going further, this meaning constrains the range of possible light verb constructions as well. For example, *make a mistake* is acceptable but not **take a mistake.* North (2005) has shown that nouns that are semantically similar tend to co-ocur with the same light verb. Such a semantic constraint explains the patterns of acceptable combinations in Hindi as well and this needs to be represented in the analysis.

Previous work on light verb constructions has increasingly turned to lexicalized grammars to address these questions. Lexicalized grammars have been used for linguistically precise representations and are particularly useful with respect to phenomena that interact with valence (Müller and Wechsler, 2014). Before we turn to our tree-adjoining grammar analysis, we examine an existing analysis of LVCs in Lexical-Functional Grammar.

3 Previous work: LFG

Mohanan (1994), Alsina et al. (1997) and Butt (1995) have advocated the joint predication of the noun and light verb in an LVC, where both parts of the predication contribute their meaning. Their treatment takes advantage of the parallel architecture of mutually constraining levels of analysis, particularly the functional-structure and argument-structure. At one level (f-structure), the LVC is monoclausal, and at another (a-structure), the argument frames of the two elements are composed in a merger operation.

When light verbs occur, they trigger a process of *argument merger* because they appear with an incomplete argument structure frame. Below, we show the a-structures for the noun and light verb as seen in example (4). The % notation stands for a variable whose value will be supplied by the noun's argument structure. The resulting *merger* of the two argument structures results in the composite argument structure of the LVC.

($\uparrow$ PRED) = 'kar (do) <AGENT, %PRED >'

($\uparrow$ PRED) = 'bahas (quarrel) <AGENT, THEME >'

(4) meNdak=ne   bicchu=se   bahas   kii
     frog.M.Sg=Erg scorpion.M.Sg=Inst quarrel.F.Sg do.Perf.F.S
     'The frog quarrelled with the scorpion. (Lit) The frog did (a) quarrel with the scorpion'

In order to compose the argument structure from the two co-predicators, a Restriction Operator is used to manipulate the f-structure (Butt et al., 2003).

The ability to restrict certain features allows the manipulation of the f-structure (valency property) of the light verb. The restriction operator identifies the noun as the %PRED value in the light verb's a-structure.

The lowest matrix argument for DO as seen in Figure 2 is identified with the highest embedded argument, which is also an agent (Butt et al., 2008). Now, there are three arguments to be linked at the level of f-structure using the linking relations that connect the grammatical functions SUBJ, OBJ with the thematic roles indicated by the features $[\pm o]$ and $[\pm r]$. The highest (and only) [o] argument is linked to SUBJ by the mapping principles (Bresnan, 2001). The instrumental argument gets $[+r]$ and is linked to OBL.

$$
\begin{array}{cccc}
\text{DO} < \overbrace{agent} & \overbrace{\text{QUARREL} < agent} & \text{theme>>} \\
[-o] & [-r] & [+r] \\
\downarrow & \downarrow & \downarrow \\
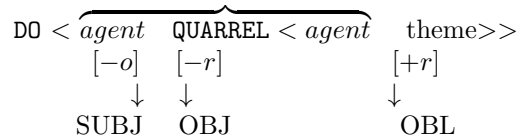\text{SUBJ} & \text{OBJ} & \text{OBL}
\end{array}
$$

Figure 2: Merged a-structures for DO and QUARREL

The object is $[-r]$ and is linked to OBJ. Note that in this example, the light verb shows agreement with the predicating noun, hence it acts both as the OBJ of the LVC, as well as predicate. In light verb constructions, the light verb will agree with the nominal, provided no other argument in the sentence has nominative (null) case (example (4)).

For simple predicates, the verb will always agree with the highest nominative (null) marked argument, which is not necessarily the syntactic subject. In simple verbs, in examples (5) and (6), the nominative argument is the subject and the object respectively. In example (7), there is no nominative argument available, thus the verb shows 'default' agreement with third person, masculine, singular.

Nouns such as *bahas* 'quarrel' in the example 4 also have another property– the light verb *agrees* with the predicating nominal *bahas* in number and gender. Therefore, the nominal is simultaneously an argument and a part of the LVC.

(5) laṛkii   laṛke=ko   dekʰ-egii
     girl.F.Sg.Nom boy.M.Sg=Acc see-fut.F.Sg

     '(The) girl will see (the) boy'

(6) laṛke=ne   capaaṭi   kʰaa-ii
     boy.M.Sg=Erg bread.F.Sg.Nom eat-Perf.F.Sg

     '(The) boy ate (the) bread'

(7) laṛkii=ne   auraṯ=ko   dekʰ-aa
     girl.F.Sg=Erg woman.F.Sg=Acc see-Perf.M.Sg

     '(The) girl saw (the) woman'

Mohanan (1997) points out that a small class of nominals such as *yaad* 'memory' or *istemaal* 'use' do not show this pattern of agreement. This appears to correlate with their ability to have non-subject arguments with nominative

or accusative case (see 8). In such cases, although the nominal is the only unmarked 'argument' the verb shows default agreement i.e. masculine singular. Nouns such as *yaad* also differ from nominals (such as *bahas* 'quarrel') with respect to sentential negation, gapping and passivization (Mohanan, 1997). (We refer the reader to the tests shown in Mohanan (1997), and do not repeat them here).

(8) siṭaa=ne   mohan=ko   yaad   ki-yaa
     Sitaa.F.Sg=Erg Mohan.M.Sg=Acc memory.F do-Perf.M.Sg

     'Sita remembered Mohan'

$$
\begin{bmatrix}
\text{PRED} & \text{'kar'}\left\langle\left[2{:}\text{meNdak}\right], \text{'bahas}\left\langle63{:}\text{biccHU}\right\rangle\text{'}\right\rangle \\
\text{SUBJ} & 2\begin{bmatrix}\text{PRED} & \text{'meNdak; frog'}\end{bmatrix} \\
\text{OBJ} & \begin{bmatrix}\text{PRED} & \text{'bahas; argument'}\end{bmatrix} \\
\text{OBL} & 63\begin{bmatrix}\text{PRED} & \text{'biccHU; 'scorpion'}\end{bmatrix}
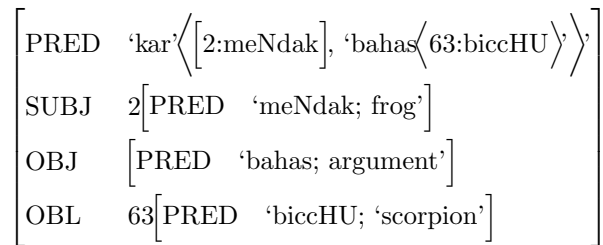\end{bmatrix}
$$

Figure 3: Final (abbreviated) F-structure for the LVC *bahas kar* 'debate do'. Note that *bahas* acts simultaneously as a co-predicator and argument of the light verb

The f-structure in Figure 3 shows the composed argument structure with the predicating noun. The hallmark of this analysis is that the syntactic composition takes place at a different layer of representation than the f-structure (which is monoclausal, containing only one SUBJ and OBJ and the OBL). The operation of *restriction* becomes necessary to compose the argument structures, followed by the process of *argument identification* to have the correct number of arguments at f-structure. In order to allow the argument structure specification of noun and light verb to exist side-by-side, this analysis requires that both a deletion as well as an identification step are maintained.

However, when the identification requires that an AGENT and a GOAL be identified as the same, there is a need for the light verb to effectively re-write the preverbal element's arguments. Lowe (2015) points out that in the LFG Linking account, the exact mechanism for argument identification is not made explicit. This particular problem with the linking of a-structures and f-structures has prompted revisions in the analysis, e.g. within LFG+glue. In this approach, the subcategorization specifications of the LVC are de-linked from its f-structure. Instead, the argument structure itself is represented using LFG+glue at a separate level of semantic analysis and the f-structure is relatively simpler, with only one real PRED i.e. the predicating noun, with the light verb contributing the feature AGENTIVE=+

In this paper, we propose that LVCs are formed via syntactic composition, but our analysis separates the monoclausal requirement of the LVC and the

semantic contribution of the light verb. In the LFG account, maintaining mono-clausality without a deletion rule and argument identification becomes a difficult task. Instead of combining two sets of argument structures, we design a single initial tree which contains all the arguments, but an underspecified category (i.e. it is not the elementary tree for the 'predicating noun'). We show that we can represent the LVC's argument structure using such a representation. Further, we describe our analysis of the acceptability of noun and light verb combinations based on their semantic properties. We incorporate this into our analysis using features, which will permit the appropriate combination of two predicating elements to take place.

In the following sections, we briefly introduce the TAG formalism that we will use and then motivate our TAG proposal. We then describe the design for the elementary trees in TAG and conclude with a summary and discussion of our work.

4   Introduction to lexicalized Tree-Adjoining Grammar

We briefly introduce the tree-adjoining grammar formalism that we use in this section. Tree-Adjoining Grammar (TAG) is a formal tree-rewriting system that is used to describe the syntax of natural languages (Joshi and Schabes, 1997). The primitive of a TAG grammar is an elementary tree, which is a fragment of a phrase structure tree labelled with both terminal and non-terminal nodes. The elementary trees are combined by the operations of **substitution** (where a terminal node is replaced with a new tree) or **adjunction** (where an internal node is split to add a new tree, see also Fig 4).

The elementary trees in TAG can be enriched with feature structures (Vijay-Shanker and Joshi, 1988). These can capture linguistic descriptions in a more precise manner and also capture adjunction constraints. TAG with feature structures is also known as FTAG (Feature-structure based TAG). A TAG can also be lexicalized i.e. each elementary tree has at least one lexical item as one of its terminal nodes. Lexicalized TAG enhanced with feature structures is known as Lexicalized Feature-based Tree-Adjoining Grammar (LF-TAG). This has been used for developing computational grammars for English (XTAG-Group, 2001), French (Abeillé and Candito, 2000) and Korean (Han et al., 2000). In our analysis, we will also use LF-TAG, but we will refer to it as LTAG for convenience.

Figure 4 shows the basic steps for composing elementary trees containing feature structures. Unless it is a substitution node, each node has a top and a bottom feature structure. Features can be shared among nodes in an elementary tree. In the tree for the verb *running*, the variable $\boxed{1}$ is used to show that the verb must share the same features as the subject NP.

The tree for *running* is an **initial tree** with a single terminal for its argument noun phrase (NP). The tree for *is*, on the other hand, is a special type of
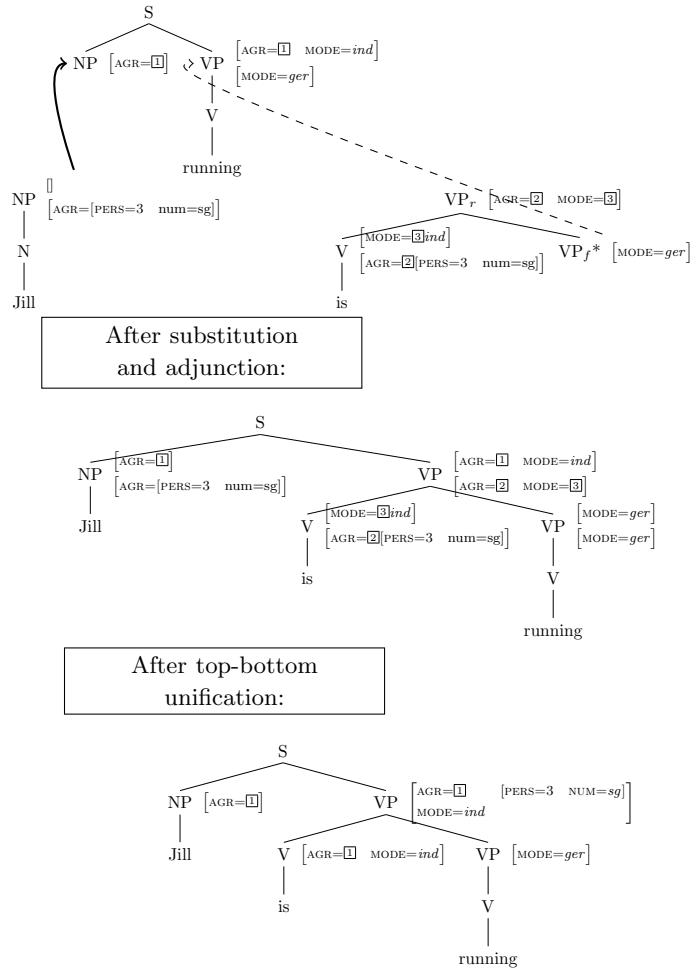
Figure 4: LTAG showing feature structures and constraints on adjunction (Example adapted from Kallmeyer and Osswald (2013)). The topmost trees show the operations of substitution (solid line) and adjunction (dashed line). Following these operations, we get a complete sentence 'Jill is running'. After both top and bottom nodes unify, the derivation is complete.

elementary tree called the **auxiliary** tree. It has a foot node (marked with an asterisk), whose labels are identical to its root node. The auxiliary tree will adjoin into the tree for *running* at the VP node only. The top and bottom feature structures for MODE at the VP node for *running*, have different values (*ind*icative and *ger*undive), and they cannot unify. This captures an adjunction constraint for *obligatory adjunction* and requires adjunction to take place at this node only.

During adjunction, the top feature structure at VP_r in the auxiliary tree (for *is*) will unify with the top of the adjunction site (VP). The bottom feature structure at VP_r in the auxiliary tree will unify with the bottom of the adjunction site. During substitution, the top node in the tree for *Jill* unifies with the node at NP in the initial tree for *running*. This results in the second tree in Figure 4, post the operations of substitution and adjunction. In a final derivation step, top and bottom feature structures at each node will unify, to give the final derived tree with a single feature structure at each node. The resulting tree is called a *derived* tree, but another by-product of the TAG analysis is also the *derivation* tree. This tree has numbered node labels that record the history of composition of the elementary trees. For example, the tree for *Jill is running* can be seen in Figure 5. The root of this tree is labelled with *running*, which is an *initial* tree of type S.
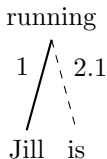

Figure 5: Derivation tree for 'Jill is running'. The dashed node indicates adjunction and the solid node indicates substitution. Node numbers indicate their position in the derived tree.

5   The TAG proposal

The TAG analysis that follows assumes that the LVC formation takes place at the syntactic rather than the lexical level. This is based on evidence that the LVC is syntactically flexible although it forms a single meaningful unit. In this respect, LVCs are similar to verb-particle constructions, or decomposable idioms (Sag et al., 2002). In general, tests for light verb constructions cannot include standard surface constituency diagnostic tests e.g. adjacency, scrambling, negation or adverbial modification (Butt, 2010). Rather, the real diagnostic test is whether the two predicates result in one event, which may not be modified separately (Butt, 2014).

On the other hand, the fact that light verb constructions express a single event has also resulted in proposals to treat the LVC as a single lexical entry.
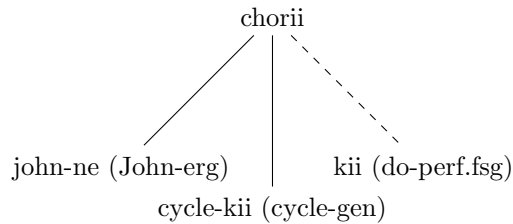
Samvelian and Faghiri (2013) analyze the Persian LVC as a 'construction' in the Construction Grammar framework. In their analysis, all LVCs are understood as being non-compositional, as both noun and light verb do not contribute a consistent meaning for every LVC combination. In effect, every LVC forms a separate lexical entry.

In terms of TAG, this gives us two options, first an analysis where the nominal projects all the arguments of the LVC, and is the initial tree for the construction. The second option is to represent both components as anchors of a single elementary tree– a single multi-word expression. Previous TAG analyses for English LVCs have both nominal and verb as anchors in the same elementary tree (XTAG-Group, 2001) and the arguments simply substitute into the this tree.

Figure 6 shows the derivation trees (cf. Figure 5) for the analysis options as described above for example (9). The LVC in question is *chorii kar* 'theft do'. The dashed line indicates adjunction into the elementary tree, whereas the solid line indicates substitution. In the noun-centric analysis, the light verb adjoins into the nominal's elementary tree. For the multi-word expression tree, *theft* and *do* are both anchors of the elementary tree.

(9) john=ne   cycle=kii   chorii   k-ii
     John.M-Erg   cycle.F-Gen   theft.F   do-perf.F.sg

     'John stole the bicycle'

Noun as anchor →   chorii

                          john-ne (John-erg)   kii (do-perf.fsg)
                                        cycle-kii (cycle-gen)

Multi-word anchor →   chorii-kii (theft-do.perf.fsg)

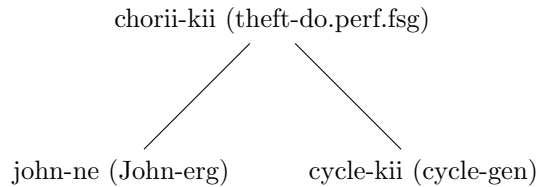   john-ne (John-erg)   cycle-kii (cycle-gen)

Figure 6: Derivation graphs showing the options for the analysis of *Ram ne cycle-kii chorii kii* 'Ram stole the cycle'. The LVC is *chorii kii*.

For the Hindi LVC, the multi-word option is more suitable for those LVCs

that are idiomatic or are formed via incorporation. An example of an idiomatic LVC in Hindi is *goḍ lenaa* 'lap take; adopt', which consists of a noun like *goḍ* 'lap' and a light verb *lenaa* 'take'. It expresses a meaning that is entirely different from the meaning of the individual elements. Davison (2005) also describes combinations such as *golii maarnaa* 'bullet hit; shoot', where a noun like *golii* 'bullet' is an instrument in the act of shooting. She suggests that such nouns are bare indefinites and are incorporated with the verb. It is plausible to treat cases of non-compositional as well as incorporated LVCs as single lexical entries that are not formed by a process of syntactic composition. Such cases are unlike those productive LVC cases that are composed in the syntax, and are better represented with noun and verb as the anchor of the tree. Such a division is in keeping with other analyses e.g Müller (2010) for Persian which maintains separate analyses for LVCs that are idiomatic vs. those that are compositional. This view maintains that both forms exist side-by-side and are formed differently in the syntax.

For the purposes of this paper (and the TAG analyses that follow), we focus on LVCs that do not have idiomatic meanings or incorporation. Such LVCs contain event nouns that subcategorize for their own arguments and combine with the light verb to form a single 'compositional' meaning. It seems reasonable that for such nouns, there is syntactic composition of the argument structures of both noun and light verb. Recent psycholinguistic evidence, both from behavioural data as well as ERP studies show that LVCs are likely to be constructed 'online', rather than being retrieved as single lexical entries like ordinary predicates. Processing light verb constructions incurs an additional cost, measured in reaction time differences and sustained neural activity after the onset of the verb (Wittenberg and Piñango, 2011, Wittenberg et al., 2014). This evidence strengthens the idea that non-idiomatic LVCs are likely to be composed in the syntax.

(10) cycle=kii   chorii john=ne   k-ii
     cycle.F-Gen theft.F John.M-Erg do-perf.F.sg
     'John stole the bicycle'

We also note that Hindi allows scrambling (movement) of NPs before the verb, and that scrambling can also affect the predicating noun. For example, it's possible to scramble (9) as (10). For such cases, it would be necessary to extend the LTAG analysis, similar to previous work done for German (for example Rambow (1994), Lichte (2007)). In the work on German, elementary trees are replaced by sets or trees with underspecified dominance. In order to account for scrambling, we will need to assume an analysis using sets of trees (e.g. multi-component TAG or MC-TAG). In the work on German, the source of predication does not move, since in German this is the verb.[1] However, we can simply add an initial tree with a trace which is co-indexed with the predicating noun. The predicating noun and its arguments are all treated as separate auxiliary trees in

---

[1]The verb in German can actually occur in different places, but it cannot scramble among its arguments within the *Mittelfeld*, the domain of scrambling in German.

the same treeset. The use of feature structures can be carried over to the multi component approach, since feature values can be co-indexed among different trees of the same set. This approach forces the syntactic realization of the predicate-argument structure, but allows for complete freedom of order among the arguments and the predicating noun (also across clause boundaries). We omit further details here.

In the following sections, we motivate the design of the elementary trees for the nominal and the light verb. We look at the underspecified category of the nominal and the feature structures that specify selectional preferences and agreement facts.

5.1   Category underspecification for the nominal

In this LTAG analysis the argument structures of noun and light verb are combined via the adjunction and substitution operations. The elementary tree of the nominal is an initial tree in our analysis and it also chooses a syntactic structure that will realize all its arguments. The light verb on the other hand is represented as an auxiliary tree, therefore it is an adjunct to the nominal's basic structure and contributes only features. However, as it is a predicate, it is also a special type of auxiliary tree viz., a predicative auxiliary tree (Abeillé and Rambow, 2000).

The initial tree of the nominal is lacking a category specification (Han and Rambow, 2000). We use the label X, which projects to an XP (Figure 7). We also assume that each node is specified with the feature CAT which has values like V or N. The CAT=V feature-value is shown on the initial tree because the [CAT=N] feature-value is not realized unless the light verb composes with the elementary tree of the nominal. Figure 7 shows this underspecification for the nominal *chorii*, which has the category X. The feature clash at XP_1 with the feature-value for TENSE ensures that adjunction takes place at this node. The TENSE feature also captures the fact that the tree for the nominal is neither completely verbal, nor completely nominal.

The underspecified category for the nominal's elementary tree has some precedent e.g. there are proposals for a *mixed category* analysis for the predicating nominal (Manning, 1993, Choi and Wechsler, 2001). The TAG underspecification can be understood as a disjunction in the category of the predicate nominal, allowing it to be resolved as a nominal when the light verb adjoins into it (11). When no adjunction takes place, the nominal can be understood as forming a phrase like (12). Therefore, the number of substitution nodes in the nominal's tree can also be motivated independently of the adjunction of the light verb.

(11)   john=ne   mary=se   baat   kii
       John.M-Erg Mary.F-Inst talk.F do.perf.F.sg
       'John talked with Mary'

(12)   john=kii   mary=se   baat
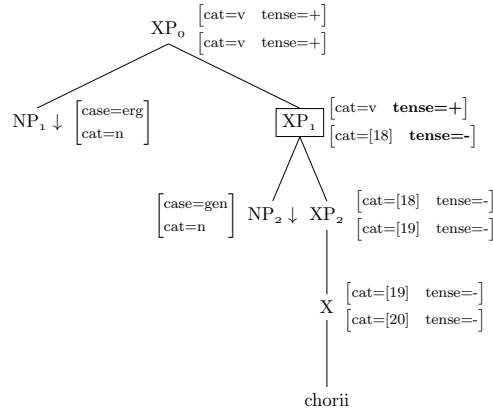   John.M=Gen Mary.F=Inst talk.F.Sg

XP$_o$ $\begin{bmatrix} \text{cat=v} & \text{tense=+} \\ \text{cat=v} & \text{tense=+} \end{bmatrix}$

NP$_1$ ↓ $\begin{bmatrix} \text{case=erg} \\ \text{cat=n} \end{bmatrix}$ XP$_1$ $\begin{bmatrix} \text{cat=v} & \textbf{tense=+} \\ \text{cat=[18]} & \textbf{tense=-} \end{bmatrix}$

$\begin{bmatrix} \text{case=gen} \\ \text{cat=n} \end{bmatrix}$ NP$_2$ ↓ XP$_2$ $\begin{bmatrix} \text{cat=[18]} & \text{tense=-} \\ \text{cat=[19]} & \text{tense=-} \end{bmatrix}$

X $\begin{bmatrix} \text{cat=[19]} & \text{tense=-} \\ \text{cat=[20]} & \text{tense=-} \end{bmatrix}$

chorii

Figure 7: Tree for underspecified nominal category, *chorii* as seen in example (9)

'John's talk with Mary'

Nouns that appear as part of LVCs in Hindi behave somewhat differently from English. For example, in English, LVCs such as *make an offer*, *give a groan* often combine a nominalized form of an English verb e.g. *offer* or *groan* with a light verb. Consequently, a light verb construction may be paraphrased by the verbal form of the noun in English e.g. *gave a lecture* may be paraphrased by *lectured*. In contrast, nouns that occur as part of Hindi LVCs are very rarely nominalizations of main verbs. Butt (2010) notes that light verbs in Hindi LVCs act as verbalizers in order to create new predicates and incorporate borrowed items into the language e.g. *email kar* 'email do; email'. Therefore, LVCs are sometimes described as "a preferred way of augmenting the creative potential of the language" (Kachru, 2006)[93]. The presence of borrowed preverbal predicates is another reason to motivate category underspecification.

5.2   Feature structures for selectional preferences

The previous section showed that the elementary tree of the nominal specifies the number of arguments for the LVC. This seems to imply that the light verb does not contribute to the event predication. This is untrue because the elementary tree of the nominal is in fact, underspecified and requires the light verb to adjoin. The underspecification analysis separates the subcategorization information from the syntactic composition process, which avoids the additional steps of argument identification and deletion. Moreover, feature structures on both noun and light verb ensure that combinatorial constraints are also specified.

Ahmed and Butt (2011) have suggested that the combinatory possibilities of N-V combinations in Hindi and Urdu are in part governed by the lexical semantic compatibility of the noun with the verb. Similar observations have been made

for English (Barrett and Davis, 2003, North, 2005). Sulger and Vaidya (2014) found that nouns in Hindi preferentially occur with a particular light verb based on their ontological properties (extracted from Hindi WordNet(Bhattacharyya, 2010)). For example, the noun *varsha* 'rain' has the ontological node description in Hindi WordNet as 'Natural_State,State,Noun' and has a high likelihood of occurring with a light verb that is also marked for stativity e.g. the light verb *hu-* 'become'. If the ontological properties of noun and light verb are compatible, LVC formation is possible.

These combinatorial constraints also suggest that it is not useful to think of the light verb merely as a licensor of predication, as it also contributes semantic information to the LVC. Sulger and Vaidya (2014) were able to identify four ontological properties that governed combinatorial constraints on noun and light verbs: *agentivity*, *stativity*, *transfer*, *separability* and *punctuality*. Each of these were associated with the light verbs *kar* 'do', *hu-* 'become', *de* 'give', *le* 'take' and *lag* 'get' respectively.

Further, there was also a strong tendency for a group of nouns to show alternations among the light verbs, particularly *kar* 'do' and *hu-* 'become'. Hindi has other such light verb pairs such as *aa* 'come' and *dilvaa* 'cause to give', which result in alternations with the same noun (Ahmed and Faraz, 2015). However, the *kar* and *hu-* alternation is particularly productive. Claridge (2000) has claimed that such alternations are lexical alternatives to a syntactic transformation and may occasionally be found in English e.g. *come to light* vs. *bring to light*.

In order to constrain combinations on noun and light verb, we use the feature `AGT=+` (Agentive) feature for combinations with light verbs that also contain this property e.g. *kar* 'do'. In the case of an alternation with *hu-*, this feature is marked negatively. This feature distinguishes the elementary tree of an agentive light verb like *kar* from a non-agentive one such as *hu-* and allows it to adjoin into an elementary tree with the right number of arguments.

It also implies that a noun such as *bahas* 'quarrel', which occurs with *kar* as well as *hu-* in examples (13) and (14) will have two elementary trees associated with it, one as `AGT=+` and the other `AGT=-`. From the LTAG point of view, this is not surprising as such an alternation is similar to the passive-active alternation, which has two elementary trees in the English XTAG grammar (XTAG-Group, 2001). However, Ahmed and Butt (2011) state that examples such as (14) are 'resultative state meanings', which probably implies that such cases are not considered LVCs in the LFG analysis.

We maintain that the alternation with *hu-* 'be' provides a useful lexical alternative to an alternative syntactic structure (such as a passive). The alternation of the light verb *ho* 'be' and *kar* 'do' is moreover a characteristic of a certain group of nominals only (not all can show this alternation e.g., *intizar* 'waiting' will not alternate with *hu-*). The event of `QUARREL-BE` is compositional in a manner similar to `QUARREL-DO` except that the agent argument is not present and the light verb has an intransitivizing effect.

(13) meNdak=ne   bicchu=se   bahas   kii
      frog.M.Sg=Erg scorpion.M.Sg=Inst quarrel.F.Sg do.Perf.F.S

'The frog quarrelled with the scorpion; Intended: The frog did (a) quarrel with the scorpion'

(14) bicchu-se   bahas   hu-ii
      scorpion.M-Inst debate.F.Sg be-Perf.F.S

'There was a debate with the scorpion; Intended: A debate with a scorpion happened'

Even for LFG, we can speculate that the argument merger analysis could apply to a-structure definitions (shown below). This time, while Restriction would be needed to merge the argument structures of BE and QUARREL, the noun QUARREL would need to have a second a-structure entry, without an agent argument. There will be no need for argument identification in the case of merging BE and QUARREL. Butt (2010) state that argument *suppression* of the sort required in QUARREL-BE is not predicted by the linking theory between a-structure and f-structure. The possibility of two a-structures for the same lexical entry have been proposed for other languages as well (Butt, 2014).

($\uparrow$ PRED) = 'hu- $<$ %PRED $>$'

($\uparrow$ PRED) = 'bahas $<$ THEME $>$'

In TAG, this analysis will result in more elementary structures associated with a nominal anchor. This is seen in English as well, e.g. for Wall Street Journal TAG grammar, the average elementary tree ambiguity for a given word was 47 trees (Bangalore and Joshi, 2010). This increased lexical ambiguity is compensated for by the fact that there are complex combinatorial constraints on the local elementary trees. Therefore, once lexical ambiguity is resolved, the resulting derivation gives us a complete parse. This can be advantageous for LVCs as both noun and light verb can be used independently as ordinary nouns or 'full' verbs. In an LVC context, once noun and light verb select the correct predicative elementary trees, the ambiguity is resolved. This also implies that the design of elementary trees for both noun and light verb needs to be specified in detail.

5.3 Feature structures for agreement

In addition to combinatorial constraints, the Hindi LVC also has particular properties relating to verbal person and gender agreement, which were introduced in section 0.3. Most predicate nominals will trigger number and person agreement with the light verb when they are the only unmarked noun phrase in the sentence. This property correlates with their ability to have non-subject arguments with either either genitive, instrumental or locative case (15–17). Exactly when the internal argument appears with nominative or accusative case, the agreement facts change: the light verb will not agree with the noun (see 18). The internal argument's case is dependent upon the lexical semantic property of the predicating noun (Mohanan, 1997).

(15) sitaa=ne   mohan=kii   tareef   k-ii.
     Sitaa.F.Sg-Erg Mohan.M.Sg=Gen praise.F do-Perf.F.Sg
     'Sita praised Mohan '

(16) leena=ne   abhay=se   nafrat   k-ii
     Leena.F.Sg=Erg Abhay.M.Sg=Instr hatred.F do-Perf.F.Sg
     'Leena hated Abhay'

(17) logon=ne   us aadmi=par   hamlaa   ki-yaa
     people.M.Pl=Erg that man.M.Sg=Loc attack.M.Sg do-Perf.M.Sg
     'The crowd attacked the man'

(18) sitaa=ne   mohan=ko   yaad   ki-yaa.
     Sitaa.F.Sg=Erg Mohan.M.Sg=Acc memory.F do-Perf.M.Sg
     'Sita remembered Mohan '

The feature PERF is required in order to constrain the aspectual features of the light verb when the subject has ergative case. This has to be made explicit in the feature structure to account for the correct case realization. The feature AGR has the values person (first, second or third) and gender (masculine or feminine). This allows the right morphological form of the inflected verb to adjoin into the nominal's tree. As seen from the examples above, this value also depends on the case realization on the arguments and hence needs to be specified on the nominal (and the light verb's) elementary tree.

The second agreement feature NAGR captures the case of the light verb that agrees with the predicating noun. Most light verbs will agree with the predicating noun (if no other nominative argument is present) At the same time, there is a small class of predicating nouns, where the NAGR value is negative because the light verb will show default (masculine singular) agreement despite the fact that no other nominative argument is present (Mohanan, 1997) (example (18)). The NAGR value helps to distinguish between these two cases.

Table 1 shows the complete list of feature structures used in the analysis. CASE is used to specify the values (either nominative, ergative,instrumental, dative or locative) as shown in Table (1). This feature can only be found on the nouns. The predicating nominal itself has no case specified (unless it is the non-agreeing type) and gets nominative case from the light verb post-adjunction.

6 An example

In this section, we will show a worked-out example of the syntactic composition of noun and light verb. In this example, elementary trees for nominal *tareef* 'praise' and light verb 'do' are composed. In (19) the LVC's composite argument

Feature	Value	Notes
CAT	(n)oun, (v)erb, (a)djective, (ad)verb	Syntactic category
CASE	Erg,Nom,Acc, Dat,Instr, Loc	Case clitic
TENSE	±	Presence of tensed verb
PERF	±	Perfective aspect
AGR	person, gender (fsg, msg etc.)	Verbal agreement (with nominal argument)
NAGR	±	Whether agreement with a predicating noun is possible
AGT	±	For ho(-) and kar(+) and Agentive class nouns

Table 1: A list of feature-values used for the TAG analysis

structure has the agent subject *logon* and the theme *pustak*. While the former has verbal case, the latter has genitive case and is an argument introduced by the predicating noun *tareef*.

(19) logon=ne   pusṭak=kii ṭareef   k-ii
      people.M=Erg book.F=Gen praise.F do-Perf.F
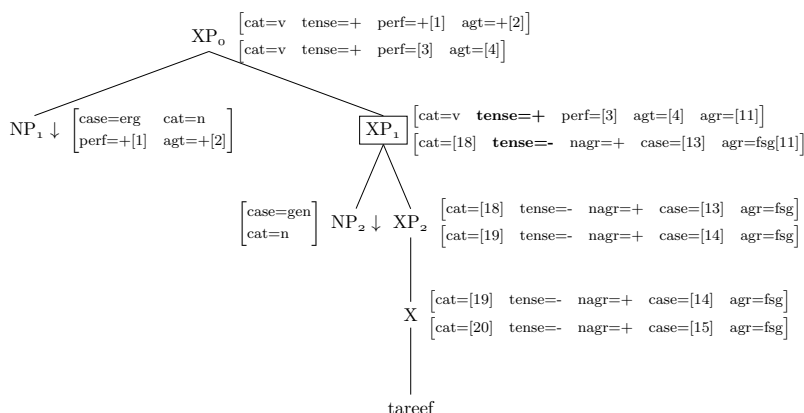   'People praised the book (Lit) People did praise of the book '

Figure 8: Tree for nominal *tareef* 'praise' (agentive), as seen in *logon ne pustak kii tareef kii* "People praised the book". The feature clash at XP_1 is marked with a box.

The nominal in Figure 8 has the composite argument structure for *tareef* 'praise', with an `AGT=+` (Agentive) light verb. It is an **initial** tree anchored by the lexical item *tareef* and the non terminals at NP_1 and NP_2 are marked with a ↓ for substitution with the actual lexical items.

NP_1 requires the ergative-marked agent argument. Agentive arguments would be found in combination with the `AGT=+` light verb *kar*. NP_1 in Figure 8 has the features for [`PERF=+`] and [`AGT=+`] as a consequence of having [`CASE=ERG`]. The agentive argument shares the values for `PERF` and `AGT` with the XP_o node.

This ensures that the light verb that adjoins into this tree will match the `PERF` and `AGT` values in NP_1.

The value for `NAGR` is positive as the light verb shows agreement with the nominal *tareef* for this example. Other arguments at NP_1 and NP_2 will not have an `AGR` feature unless they have nominative (null) case. When `NAGR` is positive, `AGR` gets its features from the nominal and these are passed up to the place of adjunction. When the light verb's agreement with the noun is not possible, e.g. if a higher nominative argument is available, the `AGR` feature will be populated by that argument. The value of `NAGR` can also be negative if the light verb does not agree with the nominal (or agrees only optionally). We explore these cases in more detail in the following sections.

The elementary tree for the nominal is not complete, because of the feature clash at XP_2 between [`TENSE=+`] vs. [`TENSE=-`]. This will force adjunction of the light verb into the nominal's tree.
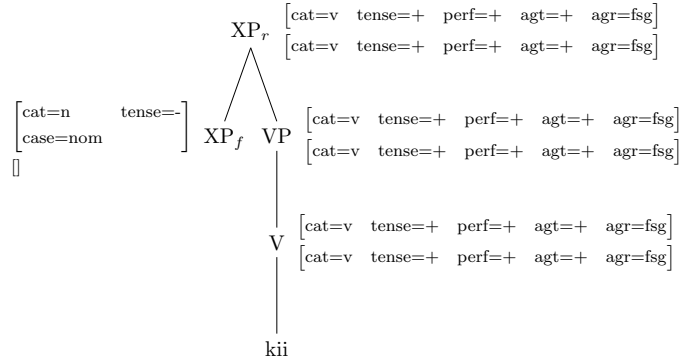

Figure 9: Elementary tree for light verb *kar* 'do' inflected as *kii* 'do.fem.sing.perf'

The light verb's tree will adjoin into the tree of the nominal. In order to model the light verb *kar* 'do' in Example 20, we will construct an auxiliary tree, anchored at *kar* 'do' as shown in Figure 9. The light verb *kar* is inflected for person, number, and gender as well as tense and aspect. In this particular example, it is tensed, feminine, singular and has perfective aspect; therefore it appears as *kii*, and its `AGR` value also has the correct feature `fsg`.

In Figure 9, the XP_r (root) node and its right-branching daughters are [`CAT=V`] with linguistic information about gender, number, tense and aspect. The feature `AGT=+` (agentive) at the top node implies that this auxiliary tree needs to unify with an initial tree that is also [`AGT=+`]. The XP_f (foot) node has [`TENSE=-`] and [`CAT=N`], which will enable it to adjoin into the elementary tree of a nominal. The `CASE` value is specified as `NOM` (nominative) as the light verb will assign nominative case to the noun. Figure 10 shows the tree post adjunction, and Figure 11 shows the composed argument structure post substitution and unification of the paired feature structures at each node.
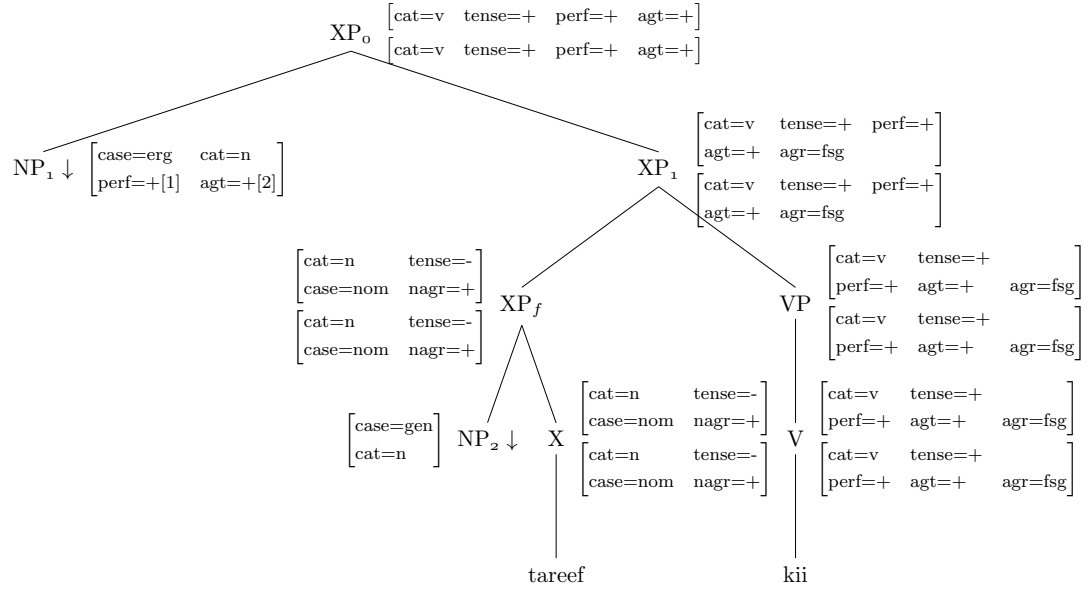
18

XP₀ $\begin{bmatrix} \text{cat=v} & \text{tense=+} & \text{perf=+} & \text{agt=+} \\ \text{cat=v} & \text{tense=+} & \text{perf=+} & \text{agt=+} \end{bmatrix}$

NP₁ ↓ $\begin{bmatrix} \text{case=erg} & \text{cat=n} \\ \text{perf=+[1]} & \text{agt=+[2]} \end{bmatrix}$

XP₁ $\begin{bmatrix} \text{cat=v} & \text{tense=+} & \text{perf=+} \\ \text{agt=+} & \text{agr=fsg} \\ \text{cat=v} & \text{tense=+} & \text{perf=+} \\ \text{agt=+} & \text{agr=fsg} \end{bmatrix}$

$\begin{bmatrix} \text{cat=n} & \text{tense=-} \\ \text{case=nom} & \text{nagr=+} \\ \text{cat=n} & \text{tense=-} \\ \text{case=nom} & \text{nagr=+} \end{bmatrix}$ XP$_f$

VP $\begin{bmatrix} \text{cat=v} & \text{tense=+} \\ \text{perf=+} & \text{agt=+} & \text{agr=fsg} \\ \text{cat=v} & \text{tense=+} \\ \text{perf=+} & \text{agt=+} & \text{agr=fsg} \end{bmatrix}$

$\begin{bmatrix} \text{case=gen} \\ \text{cat=n} \end{bmatrix}$ NP₂ ↓ X $\begin{bmatrix} \text{cat=n} & \text{tense=-} \\ \text{case=nom} & \text{nagr=+} \\ \text{cat=n} & \text{tense=-} \\ \text{case=nom} & \text{nagr=+} \end{bmatrix}$

V $\begin{bmatrix} \text{cat=v} & \text{tense=+} \\ \text{perf=+} & \text{agt=+} & \text{agr=fsg} \\ \text{cat=v} & \text{tense=+} \\ \text{perf=+} & \text{agt=+} & \text{agr=fsg} \end{bmatrix}$

tareef

kii

Figure 10: Post adjunction of the light verb's auxiliary tree into the initial tree *tareef* 'praise' at XP₂.

XP₀ $\begin{bmatrix} \text{cat=v} & \text{tense=+} & \text{perf=+} & \text{agt=+} \end{bmatrix}$

NP₁ $\begin{bmatrix} \text{case=erg} & \text{cat=n} \\ \text{perf=+} & \text{agt=+} \end{bmatrix}$

XP₁ $\begin{bmatrix} \text{cat=v} & \text{tense=+} & \text{perf=+} \\ \text{agt=+} & \text{agr=fsg} \end{bmatrix}$

logon-ne

$\begin{bmatrix} \text{cat=n} & \text{tense=-} \\ \text{case=nom} & \text{nagr=+} \end{bmatrix}$ XP$_f$

VP $\begin{bmatrix} \text{cat=v} & \text{tense=+} \\ \text{perf=+} & \text{agt=+} & \text{agr=fsg} \end{bmatrix}$

$\begin{bmatrix} \text{case=gen} & \text{cat=n} \end{bmatrix}$ NP₂ X $\begin{bmatrix} \text{cat=n} & \text{tense=-} \\ \text{case=nom} & \text{nagr=+} \end{bmatrix}$ V $\begin{bmatrix} \text{cat=v} & \text{tense=+} \\ \text{perf=+} & \text{agt=+} & \text{agr=fsg} \end{bmatrix}$
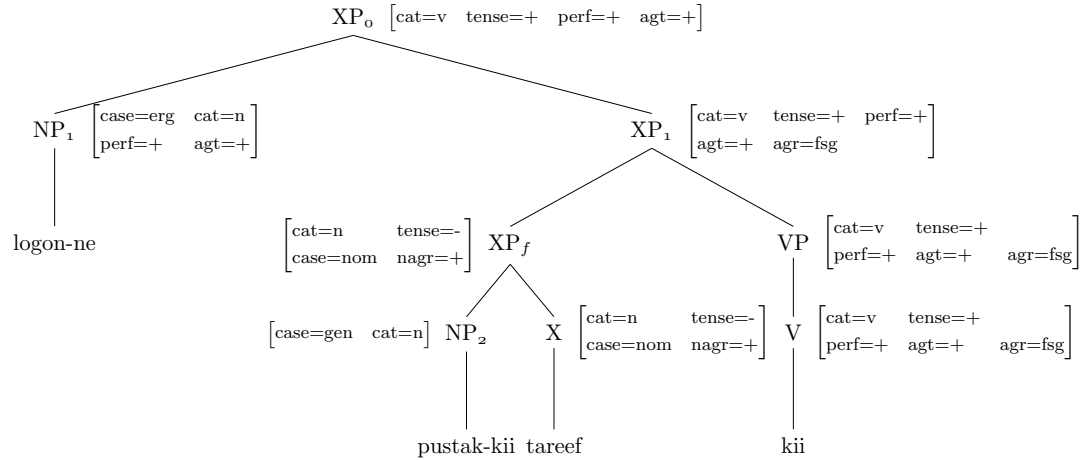
pustak-kii tareef

kii

Figure 11: After adjunction, substitution and feature structure unification, we get the complete argument structure. Substitution at the nodes NP₁ and NP₂ gives us *logon-ne pustak-kii tareef kii* 'People praised the book'

6.1 An example with alternations

In the previous section, *tareef* combined with light verb *kar* 'do'. The same noun *tareef* 'praise' may combine with the light verb *ho* (example 21). The noun *tareef* belongs to an 'alternating' class of nouns, which combine with *kar* 'do' as well as *hu-* 'become' in (21). (See the discussion about these nouns in section 0.5.2).

The example in (21) is a case where combination with *hu-* has an intransitivizing effect. It is not possible to add an experiencer subject to the example with light verb *ho* (22), though an agentive subject is implied and can be added using an adjunct phrase (Ahmed and Butt, 2011).

(20) logon=ne   pusṭak=kii   ṭareef   k-ii
      people.M=Erg book.F=Gen praise.F do-Perf.F

      'People praised the book (Lit) People did praise of the book '

(21) pusṭak=kii   ṭareef   hu-ii
      book.F.Sg=Gen praise.F be.Part-Perf.F.Sg/be.Pres

      'The book got praised; Lit: The praise of the book happened'

(22) *logon=ko   pusṭak=kii   ṭareef
      people.M=Dat book.F.Sg=Gen praise.F
      hu-ii
      be.Part-Perf.F.Sg/be.Pres.3.Sg

      '*People experienced the book's praise '

(23) pusṭak=kii   ṭareef   hu-ii
      book.F.Sg=Gen praise.F be.Part-Perf.F.Sg/be.Pres

      'The book got praised; Lit: The praise of the book happened'

The tree for non-agentive *tareef* will always combine with a light verb that is `AGT=-`, in this case *ho* 'be'. The agent is implicit in the event described by this verb, and could have been shown in this tree as an empty category, but the presence of the feature `AGT=-` itself is illustrative of this fact. In contrast with *kar* 'do', the auxiliary tree of the light verb *ho* 'be' will have [`AGT=-`]. In this way, we also ensure that agentive *tareef*'s elementary tree can combine only with a verb marked `AGT=+`, typically *kar* 'do'.

In that case, non-agentive *tareef* will anchor an elementary tree such as Figure 12. This elementary tree appears without an agentive argument, hence NP_1 will have the feature `AGT=-`. Figure 12 shows that the site of adjunction into *tareef* 'praise' (non-agentive) is at XP_1. As the nominal is the only nominative argument , the light verb will agree with *tareef* (and `NAGR=+`).

We note that light verb *ho* 'be' can also combine with stative nouns such as *nidhan* 'death', hence *ho* will have two elementary trees, one that will combine with stative nouns and the other with agentive nouns.
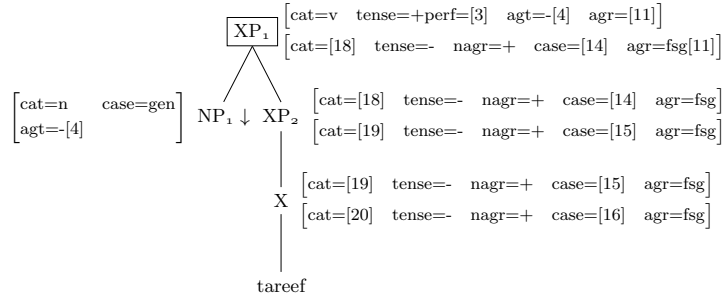
$$\text{XP}_1 \quad \begin{bmatrix} \text{cat=v} & \text{tense=+perf=[3]} & \text{agt=-[4]} & \text{agr=[11]} \\ \text{cat=[18]} & \text{tense=-} & \text{nagr=+} & \text{case=[14]} & \text{agr=fsg[11]} \end{bmatrix}$$

$$\begin{bmatrix} \text{cat=n} & \text{case=gen} \\ \text{agt=-[4]} \end{bmatrix} \quad \text{NP}_1 \downarrow \ \text{XP}_2 \quad \begin{bmatrix} \text{cat=[18]} & \text{tense=-} & \text{nagr=+} & \text{case=[14]} & \text{agr=fsg} \\ \text{cat=[19]} & \text{tense=-} & \text{nagr=+} & \text{case=[15]} & \text{agr=fsg} \end{bmatrix}$$

$$\text{X} \quad \begin{bmatrix} \text{cat=[19]} & \text{tense=-} & \text{nagr=+} & \text{case=[15]} & \text{agr=fsg} \\ \text{cat=[20]} & \text{tense=-} & \text{nagr=+} & \text{case=[16]} & \text{agr=fsg} \end{bmatrix}$$

tareef

Figure 12: Tree for nominal *tareef* (non agentive) as seen in *pustak-kii tareef huii* '(the) praise of the book happened'. The feature clash is at XP$_1$ and is marked with a box.

6.2 An example with an exceptional nominal

Nominals such as *yaad* 'memory' form part of an exceptional class, where the light verb is not required to agree with the nominal (See 8 in section 0.3). The light verb does not assign nominative case to the predicating noun *yaad* (Davison, 2005). Also, the light verb does not agree with the nominal and instead has the default masculine, singular, third person agreement features. The internal argument has accusative case.

(24) Light verb does not agree with nominal

sitaa=ne   mohan=ko   **yaad**   ki-yaa.
Sitaa.F.Sg=Erg Mohan.M.Sg=Acc memory.F do-Perf.M.Sg

'Sita remembered Mohan '

In order to represent cases such as *yaad*, we make use of the NAGR feature. For *yaad*, the value of NAGR is negative and adjunction into the tree of the nominal takes place lower, at X (Figure 14), and the light verb that will adjoin into it is found in Figure 13.

Table (2) shows a summary of the nouns and light verbs that have been captured using the analysis described above. The nominal and light verb can be classified with respect to the features that are needed for unification.

	AGENTIVE=+	AGENTIVE=−
NAGR=+	*tareef kar* 'praise do'	*tareef ho* 'praise be'
NAGR=−	*yaad kar* 'memory do'	

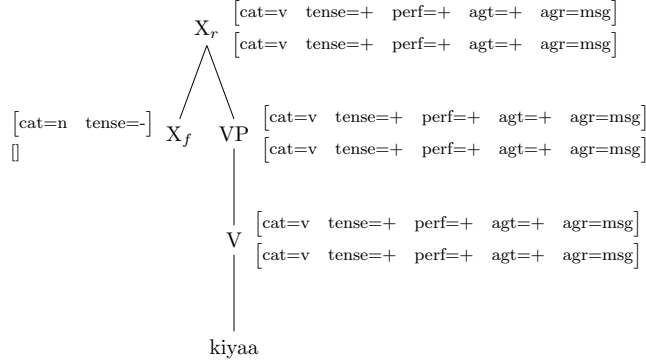Table 2: TAG features that capture syntactic composition in the light verb construction

Figure 13: Tree for light verb *kar* 'do' inflected as *kiyaa* 'do.masc.sing.perf' when it adjoins into the tree in Figure 14.
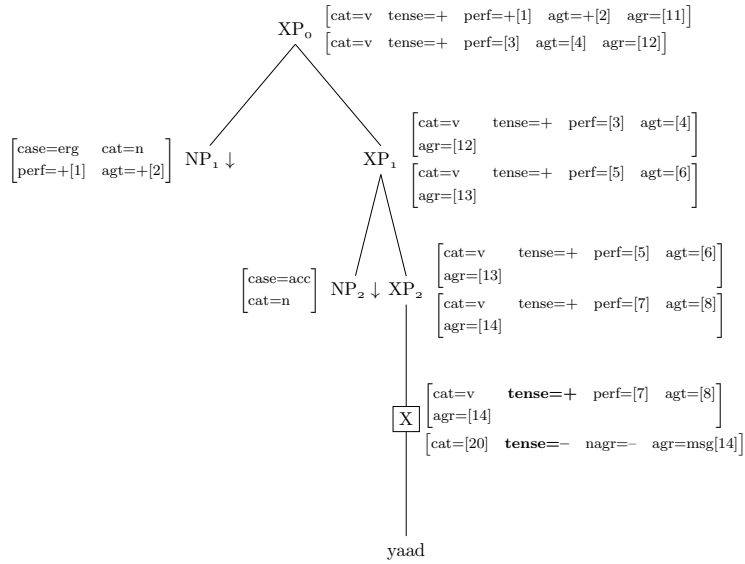

Figure 14: Tree for nominal *yaad* 'memory' (agentive), as seen in *Ram=ne Mohan=ko yaad kiyaa* "Ram remembered Mohan". The feature clash at X is marked with a box.

7   Summary

In our paper, we outlined two challenges for the formal description of LVC formation: a process of syntactic composition that required the merging of two predicate heads in a single clause, and second a process of semantic combination that checked the compatibility of the noun and verb's semantic features.

Using an underspecified cateogory and feature structures within TAG, we are able to capture the argument structure properties, alternations and agreement facts for the LVC. The process of syntactic composition as described in this analysis specifies each alternation on both the predicating nominal and the light verb. This is possible using the two operations of adjunction and substitution with complex elementary tree structures and is in keeping with the TAG maxim of "complicate locally, simplify globally" (Bangalore and Joshi, 2010).

From the point of view of TAG grammar extraction, the particular design of the elementary tree for the nominal maps directly to the subcategorization frames in Hindi PropBank (Vaidya et al., 2013). The nominal is also specified for all the arguments of the complex predicate in these frames, and this will directly help to determine the number of substitution nodes for the nominal's elementary tree. Elementary trees extracted from existing noun frames will eliminate an extra rule-writing step to extract predicate heads and their arguments from the Hindi Treebank. Instead, it would be possible to utilize hand-corrected frame files for elementary tree generation. These, in conjunction with phrase-structure conversion rules could be used to extract automatically TAG grammars from the Hindi Treebank (Bhatt and Xia, 2012).

8   Discussion

Syntactic composition is a more complex operation, hence it is worth revisiting why a simpler syntactic (or semantic) analysis might not be the right answer. Folli et al. (2005), Grimshaw and Mester (1988), Kearns (1988) treat the noun is the real predicate and the light verb a mere licenser of predication. But these analyses are not compatible with alternation facts where a change in the light verb results in a change in the number of arguments (e.g. the *kar-ho* 'do-be' alternation). Matsumoto (1996) has critiqued the Argument Transfer analysis in Grimshaw and Mester (1988) as it disallows intransitive LVCs. In the TAG analysis on the other hand, there is no such restriction.

A simpler semantic analysis, where the LVC is treated as a single lexical item also seems inappropriate as the compositional LVC cases contrast with their more idiomatic counterparts. For the idiomatic cases, a lexical analysis is more intuitive. The LVC is also syntactically flexible (see section 0.5), which makes it a phrasal category rather than a lexical one. Therefore, we are left with the syntactic composition choice, which also has the support of recent psycholinguistic processing studies that show a real-time cost in processing LVCs as compared to non-LVC examples (Wittenberg et al., 2014). This seems to indicate that a merger of syntactic and semantic information does take place

during comprehension.

Two LVC analyses that use formalisms other than lexicalized grammar are worth mentioning here. Pantcheva (2009) examines the Persian LVC using First-Phase syntax, where the light verb is not seen as essentially different from a main verb, except with respect to semantic bleaching. The light verb projects sub-event heads as *init*, *proc* and *res*, which each license their own argument. In the case of alternations (such as the *kar/ho* alternation), the intransitive light verb will lack an *init* subevent head which licenses the agentive argument. For those cases where the nominal predicate changes the argument structure e.g. *give a kiss* (2 arguments) vs. *give a sigh* (1 argument), a movement and merge operation captures the change in the sub-event structure of the light verb *give*. For the TAG analysis, the light verb anchors a different elementary tree compared to its 'full' counterpart and for cases like *give a sigh*, the elementary tree of *sigh* will simply reflect the right number of arguments. Hindi examples equivalent to *give a sigh* are *snaan kar* 'bath do; bathe', which has only one agent argument.

The second analysis within a construction grammar framework uses an inheritance hierarchy to analyze LVCs (Goldberg, 2003). Here, the analysis focuses on the fact that in Persian light verb constructions can have lexical properties (e.g. feed derivational processes) as well as phrasal properties because they are separable by auxiliaries and negation. In a sense, this is parallel to the Hindi LVC, noun and verb are syntactically separable, as seen earlier (although derivational processes are not possible). The nominal itself can be independently modified by an adjective. Such discontinuity among LVCs can be modelled via the adjunction operation in TAG. Although we do not model this property explicitly in this analysis, it has been demonstrated for other discontinuous multi- word expressions such as idioms (Abeillé and Schabes, 1989).

The Hindi LVC has sometimes been called 'multidimensional' (Mohanan, 1997), which is another way of describing its non-canonical syntactic and semantic mapping. Although such a process is grammatically complex and psycholinguistically costly, the LVC in Hindi is a highly productive phenomenon. It can be seen as a convenient alternative to a syntactic transformation such as a passive (the speaker need only swap a transitive light verb for an intransitive one). The LVC is also a way of introducing new predicates into the language (Kachru, 2006). These facts also raise the question of whether a non-canonical mapping is confined to phenomena at the 'periphery' of language, or whether it is another equally productive strategy that exists alongwith canonical cases. It may well be that the boundary between what is peripheral and central is continuous rather than discrete and LVCs, alongwith other types of multi-word predicates are examples of highly productive phenomena that prefer a non-canonical strategy for predication.

References

Abeillé, Anne and Marie-Hélène Candito. 2000. FTAG: A Lexicalized Tree-Adjoining Grammar for French. In *Tree Adjoining Grammars: Formalisms, Linguistic Analysis and Processing*. CSLI Publications.

Abeillé, Anne and Owen Rambow. 2000. Tree Adjoining Grammar: An Overview. In A. Abeillé and O. Rambow, eds., *Tree Adjoining Grammars: Formalisms, Linguistic Analysis and Processing*. CSLI Publications.

Abeillé, Anne and Yves Schabes. 1989. Parsing Idioms in Lexicalized TAGs. In *Fourth Conference of the European Chapter of the Association for Computational Linguistics*. http://acl-arc.comp.nus.edu.sg/archives/acl-arc-090501d3/data/pdf/anthology-PDF/E/E89/E89-1001.pdf.

Ahmed, Tafseer and Miriam Butt. 2011. Discovering Semantic Classes for Urdu N-V Complex Predicates. In *Proceedings of the International Conference on Computational Semantics (IWCS 2011), Oxford*. http://www.aclweb.org/anthology/W11-0132.

Ahmed, Tafseer, Miriam Butt, Annette Hautli, and Sebastian Sulger. 2012. A reference dependency bank for analyzing complex predicates. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*. http://www.lrec-conf.org/proceedings/lrec2012/pdf/474$_P$aper.pdf.

Ahmed, Tafseer and S.M Faraz. 2015. Corpus Analysis of the Light Verbs used in Urdu Psych Noun+Verb Predicates. In *2nd Kashmir International Conference on Linguistics, Muzaffarabad*.

Alsina, Alex, Joan Bresnan, and Peter Sells. 1997. Complex Predicates: Structure and Theory. In *Complex Predicates*. CSLI Publications, Stanford.

Bangalore, Srinivas and Aravind Joshi. 2010. Introduction. In S. Bangalore and A. Joshi, eds., *Supertagging: Using Complex Lexical Descriptions in Natural Language Processing*, pages 1–31. https://mitpress.mit.edu/sites/default/files/titles/content/9780262013871$_s$ch$_0$001.pdf: MIT Press, Cambridge.

Barrett, Leslie and Anthony R Davis. 2003. Diagnostics for determining compatibility in English support-verb-nominalization pairs. In *Proceedings of the 4th international conference on Computational Linguistics and Intelligent text processing (CICLing '03)*. http://link.springer.com/chapter/10.1007/3-540-36456-0$_9$.

Bhatt, Rajesh and Fei Xia. 2012. Challenges in Converting between Treebanks: a Case Study from the HUTB. In *Proceedings of META-RESEARCH Workshop on Advanced Treebanking held in conjunction with LREC-12*. https://pdfs.semanticscholar.org/1642/f480bc1a2a45740df5594f5b59713c3478ec.pdf.

Bhattacharyya, Pushpak. 2010. IndoWordNet. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, pages 3785–3792. https://www.cse.iitb.ac.in/ pb/papers/lrec2010-indowordnet.pdf.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell.

Butt, Miriam. 1995. *The Structure of Complex Predicates in Urdu*. CSLI Publications, Stanford.

Butt, Miriam. 2010. The Light Verb Jungle: Still Hacking Away. In M. Amberber, M. Harvey, and B. Baker, eds., *Complex Predicates in Cross-Linguistic Perspective*, pages 48–78. http://ling.uni-konstanz.de/pages/home/butt/main/papers/cp-volume.pdf: Cambridge University Press.

Butt, Miriam. 2014. Control vs. complex predication: Identifying non-finite complements. *Natural Language and Linguistic Theory* 32:165–190.

Butt, Miriam, Tracy Holloway King, and John T. Maxwell III. 2003. Complex Predication via Restriction. In *Proceedings of the LFG03 Conference*. https://web.stanford.edu/group/cslipublications/cslipublications/LFG/8/pdfs/lfg03buttetal.ps.

Butt, Miriam, Tracy Holloway King, and Gillian Ramchand. 2008. Complex Predication:How Did the Child Pinch the Elephant? In L. Uyechi and L. Wee, eds., *Reality Exploration and Discovery: Pattern Interaction in Language and Life*. http://ling.uni-konstanz.de/pages/home/butt/main/papers/mo-final.pdf: CSLI Publications, Stanford.

Butt, Miriam and Aditi Lahiri. 2013. Diachronic pertinacity of light verbs. *Lingua* 135:7–29. http://www.sciencedirect.com/science/article/pii/S0024384112002549.

Choi, Incheon and Stephen Wechsler. 2001. Mixed Categories and Argument Transfer in the Korean Light Verb Construction. In F. van Eynde, L. Hellan, and D. Beermann, eds., *Proceedings of the 8th International HPSG Conference*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.489.7004&rep=rep1&type=pdf.

Claridge, Claudia. 2000. *Multi-word Verbs in Early Modern English: A Corpus-based Study*. Editions Rodopi B. V., Amsterdam-Atlanta edn.

Davison, Alice. 2005. Phrasal predicates: How N combines with V in Hindi/Urdu. In T. Bhattacharya, ed., *Yearbook of South Asian Languages and Linguistics*, pages 83–116. Mouton de Gruyter.

Folli, Raffaella, Heidi Harley, and Simin Karimi. 2005. Determinants of event type in Persian complex predicates. *Lingua* 115.

Goldberg, Adele. 2003. Words by default: The Persian complex predicate construction. In E. Francis and L. Michaelis, eds., *Mismatch: Form Function Incongruity and the Architecture of Grammar*, pages 83–112. CSLI Publications, Stanford.

Grimshaw, Jane and Armin Mester. 1988. Light verbs and theta-marking. *Linguistic Inquiry* 9(2):205–232.

Han, Chung-hye and Owen Rambow. 2000. The Sino-Korean light verb construction and lexical argument structure. In *Proceedings of the Fifth International Workshop on Tree-Adjoining Grammars and Related Formalisms, TAG+5*. http://www.sfu.ca/ chunghye/papers/light-verb.pdf.

Han, Chung-hye, Juntae Yoon, Nari Kim, and Martha Palmer. 2000. A Feature based Lexicalized Tree Adjoining Grammar for Korean. Tech. rep., Institute for Research in Cognitive Science, University of Pennsylvania, http://www.cis.upenn.edu/~xtag/koreantag.

Hook, Peter. 1974. *The Compound Verb in Hindi*. University of Michigan, Ann Arbor.

Hook, Peter and Prashant Pardeshi. 2006. Are vector verbs eternal? In *11th Biennial Rice University Linguistics Symposium: Intertheoretical approaches to Complex Verb Constructions*. http://www.ruf.rice.edu/ lingsymp/Pardeshi$_p$aper.pdf.

Joshi, Aravind and Y. Schabes. 1997. Tree-adjoining grammars. In G. Rozenburg and A. Salomaa, eds., *Handbook of Formal Languages*, vol. 3, pages 69–124. Springer.

Kachru, Yamuna. 2006. *Hindi*. John Benjamins.

Kallmeyer, Laura and Rainer Osswald. 2013. Syntax-driven semantic frame composition in Lexicalized Tree Adjoining Grammars. *Journal of Language Modelling* 1(2):267–330.

Kearns, Kate. 1988. Light verbs in English. Manuscript, MIT (revised 2002).

Lichte, Timm. 2007. An MCTAG with tuples for coherent constructions in german. In L. Kallmeyer, P. Monachesi, G. Penn, and G. Satta, eds., *Proceedings of the 12th Conference on Formal Grammar*.

Lowe, John J. 2015. Complex predicates: an lfg+glue analysis. *Journal of Language Modelling* 3(2):413–462.

Manning, Christopher D. 1993. Analyzing the Verbal Noun: Internal and External Constraints. *Japanese/Korean Linguistics* 3:236–253.

Masica, Colin. 1976. *Defining a linguistic area: South Asia*. University of Chicago Press.

Matsumoto, Yo. 1996. A syntactic account of light verb phenomena in Japanese. *Journal of East Asian Linguistics* 5:107–149.

Mohanan, Tara. 1994. *Argument Structure in Hindi*. CSLI Publications, Stanford.

Mohanan, Tara. 1997. Multidimensionality of representation- NV complex predicates in Hindi. In A. Alsina, J. Bresnan, and P. Sells, eds., *Complex Predicates*. CSLI Publications, Stanford.

Müller, Stefan. 2010. Persian Complex Predicates and the Limits of Inheritance-Based Analyses. *Journal of Linguistics* 46(3):601–655.

Müller, Stefan and Stephen Wechsler. 2014. Two Sides of the Same Slim Boojum: Further Arguments for a Lexical Approach to Argument Strucuture. *Theoretical Linguistics* 40((1-2)):187–224.

North, Ryan. 2005. *Computational Measures of the Acceptability of Light Verb Constructions*. Ph.D. thesis, University of Toronto.

Pantcheva, Marina. 2009. First Phase Syntax of Persian Complex Predicates: Argument Structure and Telicity. *Journal of South Asian Linguistics* 2(1).

Rambow, Owen. 1994. *Formal and computational aspects of natural language syntax*. Ph.D. thesis, University of Pennsylvania.

Sag, Ivan A., Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger. 2002. Multiword Expressions: A Pain in the neck for NLP. In *Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics (CICLing'02)*, pages 1–15. http://link.springer.com/chapter/10.1007/3-540-45715-1$_1$.

Samvelian, Pollet and Pegah Faghiri. 2013. Re-thinking Compositionality in Persian Complex predicates. In *Proceedings of Berkeley Linguistics Society 39th Annual Meeting*. http://journals.linguisticsociety.org/proceedings/index.php/BLS/article/viewFile/3882/3578.

Sulger, Sebastian and Ashwini Vaidya. 2014. Towards Identifying Hindi/Urdu Noun Templates in Support of a Large-Scale lfg Grammar. In *Proceedings of the Fifth Workshop on South and Southeast Asian Natural Language Processing at COLING 2014*. http://www.aclweb.org/anthology/W14-5501.

Vaidya, Ashwini, Martha Palmer, and Bhuvana Narasimhan. 2013. Semantic roles for nominal predicates: Building a lexical resource. In *Proceedings of the 9th Workshop on Multi-word Expressions, NAACL-13*. http://www.aclweb.org/anthology/W13-1018.

Vijay-Shanker, K. and Aravind Joshi. 1988. Feature structure based Tree Adjoining Grammars. In *Proceedings of COLING 1988*.

Wittenberg, Eva, Martin Paczynski, Heike Wiese, Ray Jackendoff, and Gina Kuperberg. 2014. The difference between "giving a rose" and "giving a kiss": Sustained neural activity to the light verb construction. *Journal of Memory and Language* 73:31–42.

Wittenberg, Eva and Maria Mercedes Piñango. 2011. Processing light verb constructions. *The Mental Lexicon* 6(3):393–413.

XTAG-Group, The. 2001. A Lexicalized Tree Adjoining Grammar for English. Tech. rep., IRCS, University of Pennsylvania.