Encoding event structure in Urdu/Hindi VerbNet

Annette Hautli-Janisz
Department of Linguistics
University of Konstanz
`annette.hautli@`
`uni-konstanz.de`

Tracy Holloway King
Search Science
eBay Inc.
`tracyhollowayking@`
`gmail.com`

Gillian Ramchand
Institute for Linguistics
University of Tromsø
`gillian.ramchand@`
`uit.no`

Abstract

We propose a new kind of event structure representation for computational linguistics, based on the theoretical framework of First-Phase Syntax (Ramchand, 2008). We show that the approach not only gives a theoretically well-motivated set of subevents and related semantic roles, it also posits the levels of representation needed for analyzing a linguistic phenomenon that has repeatedly caused problems in computational systems, namely the treatment of complex predication. In particular, we look at V+V complex predicates in Urdu/Hindi and show that Ramchand's subevent decomposition implemented in a VerbNet-style resource allows for a consistent semantic analysis of these complex events. We also show how the proposed event representation can be added to existing resources in the language, in particular the Hindi-Urdu Treebank and Hindi PropBank.

1 Introduction

With the advent of large-scale computational semantic analyses, an issue that repeatedly crops up is how verbal event structure can be represented. In this paper we propose a new way of representing events and semantic roles in computational linguistics, based on the theoretical linguistic framework of First-Phase Syntax (Ramchand, 2008). The approach makes predictions as to how events should be encoded across languages; moreover it provides a computationally attractive way of encoding them in a lexical resource. We demonstrate this by tackling a notoriously difficult phenomenon, namely the analysis of complex predicates (CPs) in Urdu/Hindi, and show that First-Phase Syntax not only provides a well-motivated analysis for simplex verbs, but also posits the levels of representation needed for providing a consistent and computational analysis for CPs. We encode the representation in a VerbNet-style resource for Urdu/Hindi and show that it can also be incorporated into existing lexical resources, namely the Hindi-Urdu Treebank (Bhatt et al., 2009) and Hindi PropBank (Hwang et al., 2010; Vaidya et al., 2012).

The paper proceeds as follows: After providing a brief overview of related work in Section 2, we introduce First-Phase Syntax and its application in computational linguistics and also provide a linguistic background to Urdu/Hindi CPs (Section 3). We then show how these complex predicates are handled in First-Phase Syntax and how the information is incorporated in the VerbNet-style lexical resource for Urdu/Hindi (Section 4). This is followed by a discussion on how the information can be incorporated into other resources for the language (Section 5). Section 6 concludes the paper.

2 Related work

For English, one of the central resources for encoding the syntactic and semantic information on verbs is VerbNet (Kipper-Schuler, 2005). VerbNet uses the temporal ontology proposed by Moens and Steedman (1988), an approach that has proven highly useful in the past and is still employed in many computational applications. However, with the substantial progress of theoretical linguistic work in the area of formalizing event structure, the
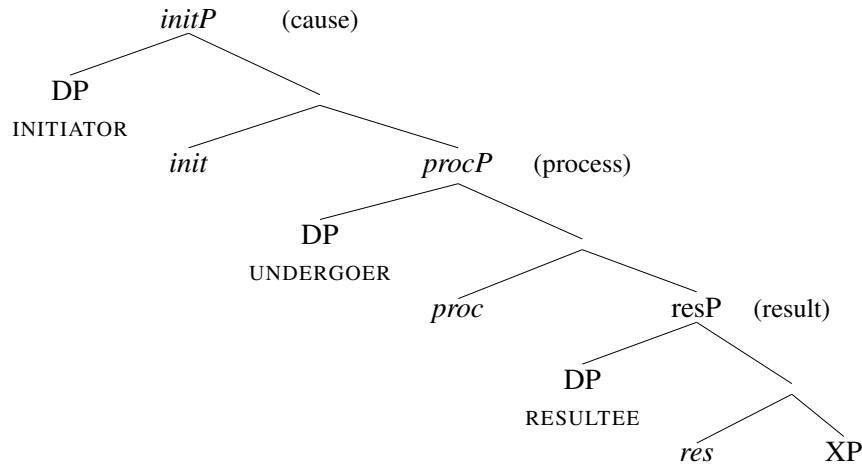
Figure 1: Event decomposition and projection by Ramchand (2008)

field has gained a deeper understanding of the underlying structure of events.

One key aspect of the automatic meaning representation of verbs is the assignment of semantic roles to the participants of the event. Here, Verb-Net uses thematic roles (e.g. Agent, Patient, Theme) based on work of Gruber (1965), Fillmore (1968) and Jackendoff (1972). In contrast, FrameNet (Baker et al., 1998) is based on a cognitive approach to event structure and defines frames which are characterized by frame-specific roles, yielding a large number of distinct roles. Finally, in Prop-Bank (Palmer et al., 2005), verb arguments are given numerical values: Arg0 has agentive or causer-like properties, whereas Arg1 is more patient-like.

For Urdu/Hindi, only a few lexical resources exist to date. In the spirit of English PropBank, Hindi-Urdu PropBank (Vaidya et al., 2011; Vaidya et al., 2012) uses PropBank-style thematic roles, accounting for CPs by merging the roles of main verbs and light verbs. In the Hindi/Urdu Treebank (Bhatt et al., 2009), these PropBank-style roles are combined with the kāraka roles assumed by Pāṇini (see Butt (2006) for a discussion of Pāṇini's system).

In general, the issue with thematic roles is that they are difficult to define and hard to consistently apply across verb classes (let alone across languages). As we will show in the following, the semantic roles assumed in First-Phase Syntax are language-independent and can be motivated by language-internal entailments based on event struc-

ture. Moreover, the complex predicates found in Urdu/Hindi call for an analysis that is theoretically well-motivated and can be consistently and productively applied across the verbal inventory.

In the following we introduce the framework of First-Phase Syntax (§3.1) and provide an overview of the phenomenon of complex predication in Urdu/Hindi (§3.2).

3 Background

3.1 First-Phase Syntax

First-Phase Syntax (Ramchand, 2008) is an approach which proposes hierarchical linguistic representations that directly encode structural semantic interpretational properties in the domain of event structure. In the framework, an event maximally decomposes into three subevents: an initiation subevent, a process subevent and a subevent denoting a result state. Each subevent licenses a semantic role and has its own projection in the tree. Figure 1 shows the general architecture: The [init] projection is responsible for introducing the external argument, i.e. the causer of the event ('subject' of cause = INITIATOR), the specifier of the process subevent undergoes the action denoted by the verb ('subject' of process = UNDERGOER) and the result state of the event is licensed by *resP* ('subject' of result = RESULTEE). The initiating as well as the resultative subevent are stative, whereas the process subevent has a dynamic interpretation. The "glue" between subevents is one of causation: The [init] subevent
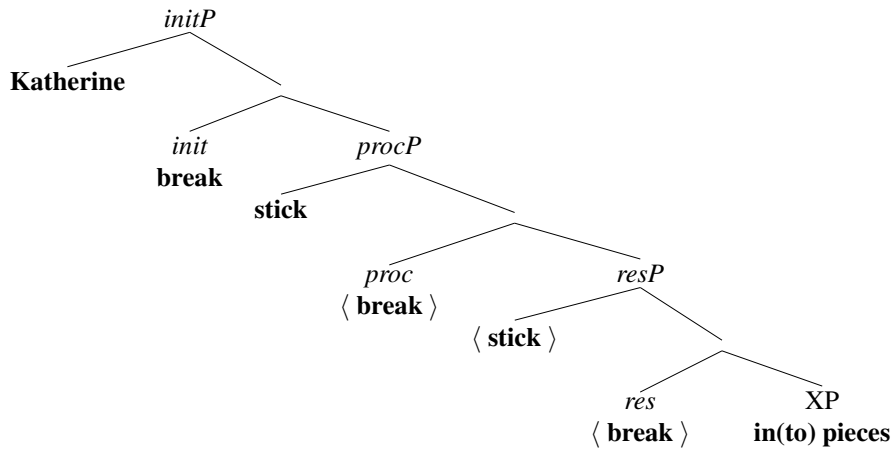
Figure 2: First-phase representation of (1)

causes the [proc] subevent to happen which brings about a change of state, which in turn leads to a result state under [res]. The reason for this decomposition is that across languages, the system allows for the identification of the general parts of verb meaning and therefore provides a set of principles that languages adhere to.

For example, the English verb 'to break', as shown in example (1) with the analysis in Figure 2[1], licenses three subevents, namely [init], [proc] and [res]. Here, Katherine is the INITIATOR of the event, with the stick being the UNDERGOER as well as the RESULTEE of the breaking event. The optional phrase 'in(to) pieces' is a RHEME, a semantic role which contributes the predicational/rhematic content to the state described by the result projection. The syntactic diagnostics for the subevent decomposition in English are the following: The [init] subevent is licensed by the ungrammaticality of the causative form of the verb, with [proc] being licensed by the grammaticality of a durative event modification like 'for hours'. The result subevent is licensed by the ungrammaticality of the latter.

(1) Katherine broke the stick in(to) pieces.

In order to make the approach compatible with computational lexical resources, we assume that each verb corresponds to a lexical entry which contains the subevental structure of the verb and the se-

mantic roles that it licenses.[2] Since one argument can carry more than one semantic role, e.g. the stick in Figure 2 is both the UNDERGOER and the RESULTEE of the breaking event, the subevental structure is indexed as shown in (2). The subscripts i and j indicate that the semantic role of the INITIATOR in [init$_i$] is filled by a different argument than the roles licensed by the [proc] and the [res] subevent ([proc$_j$, res$_j$]). As will be shown in Section 4, this information can be encoded in a VerbNet-style resource.

(2) break: [init$_i$, proc$_j$, res$_j$]

First-Phase Syntax has a number of properties that set it apart from other approaches to event structure and semantic role representation and make it attractive to use in computational linguistics. First of all, it preempts the problem that thematic roles are hard to delimit and to define. In First-Phase Syntax, each semantic role is licensed by a subevent which in turn is tied to a syntactic diagnostic that identifies it. These diagnostics can vary from language to language, but have to be consistent within a language. Defining these syntactic criteria does away with the recurring

---

[1]The terminals in angle brackets represent the Minimalist assumptions of insertion and movement of lexical items in the tree.

[2]Unlike some explicitly syntactic decompositional accounts of argument structure (e.g. Distributed Morphology), Ramchand assumes that the syntactically relevant part of the verb entries containa information on the event structures licensed as well as the relevant coindexation relations among subevents. Thus, her lexical entries are similar to what we assume here. Unlike lexical decompositional accounts, Ramchand does not assume that lexical entries contain richer argument role information independent of the relationship to event structure.

problem of thematic role assignment, which generally lacks an explicit demarcation. Moreover, in the light of lexical resource development, these criteria greatly facilitate the annotation process.

Another property of First-Phase Syntax is that composite semantic roles are explicitly allowed, i.e. an entity can be both the INITIATOR and the UNDERGOER of an event. This feature is not accounted for in other semantic role encodings, but it adds considerable expressive power to the system without extending the set of roles. Moreover, the roles proposed by Ramchand (2008) have the benefit that they are abstract enough to allow for a consistent semantic analysis and are valid across languages. The set of roles bears a striking resemblance to the topmost level of the role hierarchy introduced by Bonial et al. (2011), an attempt to find a more coarse-grained and language-independent set of semantic roles for the mapping between different resources. A direct comparison of the two sets is difficult, because the role set of Bonial et al. is motivated by a conceptual view of semantics. Nevertheless, Ramchand's roles are in the spirit of other approaches that aim at establishing a more general set of semantic roles in theoretical and computational linguistics, with the First-Phase roles having the additional benefit of being tied to concrete syntactic diagnostics.

3.2 Urdu/Hindi complex predicates

A central characteristic of the verbal system of Urdu/Hindi is the heavy usage of complex predicates (CPs) that can appear in V+V, N+V, A+V and P+V combinations (Hook (1974), Masica (1976), Mohanan (1994), Butt (1995), Raza (2011), inter alia). The formation of CPs is a highly productive process with around 20 light verbs participating. As a consequence of the expressive power of CPs, the number of simple verbs in Urdu/Hindi (∼700 verb roots (Raza, 2011)) is comparatively fewer than in many other languages.

In general, Urdu/Hindi CPs comprise two verbs: The first verb is the main verb and contributes the main propositional content of the clause. The second verb is finite and serves as the light verb of the CP, contributing a bleached-out version of its full verb meaning to the event denoted by the main verb of the CP. In Urdu/Hindi, different types of CPs exist: One type of CP are *aspectual complex predicates*

(Butt, 1995) where the light verbs contribute a sense of "completion, suddenness, directionality, benefaction, etc." (Masica, 1976, p. 143): The example in (3) (Butt, 1995, p. 91) shows a construction with *gɪr-na* 'to fall' as the main verb of the clause and the light verb *ja-na* 'to go', which adds completeness to the falling event. Example (4) (Butt, 1995, p. 34) shows a *permissive complex predicate* with the main verb *ja-na* 'to go' and the permissive light verb *de-na* 'to give'. Here, the light verb adds an argument to the clause which is not licensed by the main verb, namely the 'lettee', *Anjum*. A third type of complex predicate, the *complex predicate of motion* (Hautli-Janisz, 2013), is illustrated in example (5): Here, the main verb *kud-na* 'to jump' is complemented by the light verb *nɪkal-na* 'to emerge', which adds the source argument *makan=se* 'from the house' to the clause and adds a general telic path reading.

CPs in Urdu/Hindi are problematic for shallow as well as deep parsing approaches. Their frequency and productivity means that a static list is insufficient, but a dynamic resource has to account for the fact that the syntactic structures, semantic roles and event structures of two verbs need to be merged to form a single predicational head. This, we claim, can be done with the First-Phase Syntax approach presented above and we implement it using a class-based approach like VerbNet. The methodology is described in the remainder of the paper.

4 Encoding event structure

The two levels of representation that are generally assumed in VerbNet are the syntactic and the semantic/conceptual representation. Each verb is characterized by a set of syntactic frames or alternations that it participates in. From the viewpoint of syntax, a frame is characterized by the obligatory syntactic constituents and the semantic roles that these constituents play in the event.

One difference in the syntactic representation between English and Urdu/Hindi VerbNet (henceforth UHVN) is due to a structural difference between the two languages. English has a fixed word order and the order of constituents in the description and the order of elements in the syntactic frame indicate which constituent occupies which thematic role in the frame. This way of relating syntactic to seman-

(3) ɑnjʊm **gɪr gɑ-yi**
Anjum.F=Erg **fall go-Perf.F.Sg**
'Anjum fell (completely).'

(4) ɑnjʊm=ne sɑddɑf=ko **ja-ne dɪ-ya**
Anjum.F=Erg Saddaf.F=Dat **go-Inf.Obl give-Perf.M.Sg**
'Anjum let Saddaf go.'

(5) cor mɑkan=se bahɑr **kud nɪkl-a**
thief.M.Sg.Nom house.M.Sg=Source outside **jump emerge-Perf.M.Sg**
'The thief jumped out of the house.' (Hook 1974, p. 69)

tic information cannot be directly transferred to languages with a free word order such as Urdu/Hindi, which require the resource to be more explicit about the way the syntactic and semantic role information is connected. To account for this, case information is recorded in the syntactic frames to capture the mapping of semantic roles to syntactic constituents.

Another difference between English and Urdu/Hindi is the existence of several classes of light verbs, in addition to the standard VerbNet classes representing main verbs. Extending the VerbNet system, we posit a special class for light verbs in Urdu/Hindi and within this class are several subclasses. The syntactic and semantic structures of the light verbs and main verbs constrain the possible CPs in the language and their interpretation.

4.1 Simple verbs

The main verb component of a CP has an underlying semantics which includes the First-Phase subevents and the roles of its arguments. A basic intransitive motion verb like *gɪr-na* 'to fall' has a VerbNet entry as shown in Figure 3: The verb licenses a [proc] subevent, with the nominative argument in the clause occupying the UNDERGOER role. The semantic representation employs the 'motion' predicate also used in English VerbNet, showing that the UNDERGOER undergoes motion in the process subevent (motion(proc, UNDERGOER)).

Paths and locations of motion are also encoded as in English VerbNet, in particular following the proposal made by Hwang et al. (2013). For that, the rhematic position of the [res] subevent, interpreted as the LOCATION role in motion events, is split into INITIAL_LOCATION and DESTINATION.

Main verb *gɪr-na* 'to fall'	
Frame:	1.2.1
Description:	NP.UNDERGOER V
Syntax:	NP (nom) = UNDERGOER V
Semantics:	motion(proc, UNDERGOER)
Example:	*ɑnjʊm gɪri.* 'Anjum fell.'

Figure 3: Example of [proc] event structure in UHVN

The representation in UHVN is illustrated in Figure 4: The verb *nɪkal-na* 'to emerge' licenses two subevents, namely [proc, res], with the semantic roles of the UNDERGOER and the RESULTEE combined on the nominative argument (the subject). The INITIAL_LOCATION is characterized by an NP with locative case marking. The semantic representation of *nɪkal-na* 'to emerge' is similar to the one in Figure 3: The UNDERGOER performs a motion in the process subevent, but with the additional information on the path of motion that is defined by the INITIAL_LOCATION, with DESTINATION and TRAJECTORY left unspecified in the frame. In the result subevent, the RESULTEE is not at the INITIAL_LOCATION anymore (!at(INITIAL_LOCATION)).

4.2 Light verbs

In CPs, the light verb only contributes a bleached version of its full verb counterpart to the event. In CP formation, the VerbNet lexical entry of the main verb combines with that of the light verb. The same main verb may combine with different light verbs, and it is the VerbNet syntax of the light verb

Main verb *nɪkal-na* **'to emerge'**	
Frame:	1.2.4
Description:	NP.UNDERGOER+RESULTEE NP.INITIAL_LOCATION V
Syntax:	NP (nom) = UNDERGOER + RESULTEE
	NP (loc) = INITIAL_LOCATION
	V
Semantics:	motion(proc, UNDERGOER)
	path(proc, INITIAL_LOCATION, ?TRAJECTORY, ?DESTINATION)
	result_state(res, RESULTEE, !at(INITIAL_LOCATION))
Example:	*amra kamre=se nɪkli.*
	'Amra emerged from the room.'

Figure 4: Semantic representation of location and path in UHVN

Permissive light verb *de-na* **'to give'**	
Frame:	0.1
Description:	NPINITIATOR XP* V Vlight
Syntax:	NP (erg/nom) = INITIATOR
	XP*
	Vn
	Vlight
Semantics:	permission(init, INITIATOR, UNDERGOER)
Example:	*anjʊm=ne saddaf=ko gaɽi calane di.*
	'Anjum let Saddaf drive the car.'

Figure 5: Syntactic frame for the permissive light verb *de-na* 'to give' in UHVN

which governs the surface realization of the arguments. The light verb may introduce new semantic arguments and the structure of the First-Phrase roles (Figure 1) governs which light verbs can combine with which main verbs.

Figure 5 shows the lexical entry for the permissive light verb *de-na* 'to give' (as exemplified in (4)), which adds a permission-giving argument to the event denoted by the main verb. On the syntactic level, the light verb contributes an argument which can alternate between nominative and ergative case marking (Mohanan (1994), Butt and King (2005), inter alia). This is the INITIATOR of the event. Otherwise, no restriction on the number and role of the other arguments in the clause is assumed, represented by XP* for any kind of phrase appearing zero or more times. For the CP, the light verb *de-na* 'to give' expects the main verb in the nominal form (Vn). For the semantic representation, we introduce the predicate 'permission', showing that in the initiation subevent, the INITIATOR gives permission to

the UNDERGOER (permission(init, INITIATOR, UNDERGOER)). The remaining semantic information of the event, e.g. the aspect of motion as in (4), is contributed by the main verb *ja-na* 'to go'.

As shown in example (5), *nɪkal-na* 'to emerge' can, in addition to its full-verb counterpart, also serve as a light verb in CPs of motion. Figure 6 shows the light verb entry in UHVN: Similar to the permissive light verb, the syntactic frame comprises the arguments that are required by the light verb, here a nominative-marked argument which is the RESULTEE and a locative-marked argument which denotes the INITIAL_LOCATION. As the light verb can only combine with main motion verbs of the class 'iTHlAnA-1.1', a syntactic restriction needs to be encoded in the VerbNet entry: In UHVN, this information is attached to the main verb entry in the syntactic frame (V: synres=iTHlAnA-1.1), facilitating an automatic lookup and analysis of valid CP constructions. If no constraint is encoded, as is the case for the permissive light verb *de-na* 'to give' in

Light verb of motion *nɪkal-na* **'to emerge'**	
Frame:	1.4.2.1
Description:	NP (nom) NP (loc) XP* V Vlight
Syntax:	NP (nom) = RESULTEE
	NP (loc) = INITIAL_LOCATION
	XP*
	V: synres = iTHlAnA-1.1
	Vlight
Semantics:	path(proc, INITIAL_LOCATION, ?TRAJECTORY, ?DESTINATION)
	result_state(res, RESULTEE, !at(INITIAL_LOCATION))
Example:	*cor makan=sE kud nɪkla*
	'The thief jumped out of the house.'

Figure 6: Syntactic frames for light verbs of motion in UHVN

Figure 5, the light verb forms CPs with verbs from across the verbal inventory.

The semantic representation of the light verb does not contribute motion information, but solely contributes the 'path' and 'result_state' predicates, in parallel to the encoding of these notions for main verbs. This reduced information in the VerbNet entry of the light verb reflects the view in theoretical linguistics that light verbs only contribute a bleached version of their full verb meaning.

4.3 Representing CPs

As shown above, light verbs comprise a separate class in UHVN. Many light verbs in Urdu/Hindi have full verb counterparts and hence have multiple entries in UHVN: one for the light verb meaning and one (or more) for their full, main verb meaning. For a semantic representation of CPs, the VerbNet information of the main and the light verb is merged: The syntactic constraints of the light verb are checked against the information contained in the main verb entry; if they are fulfilled, the information regarding event structure, semantic roles and semantic predicates is combined.[3] In the following we illustrate the resulting CP analysis using example (5) with the CP of motion *kud nɪkal-na* 'to jump out of (lit. to jump emerge)'. The representation looks as in (6): The motion information in the first line is contributed by the main verb *kud-na* 'to jump',

which licenses a [proc] subevent in which an UNDERGOER undergoes motion. The path and the resultative information (second and third line) come from the light verb *nɪkal-na* 'to emerge', which licenses a [res] subevent and, in its light verb usage, only contributes the 'path' and 'result' information.

Similarly, the analysis of the permissive CP *ja-ne de-na* 'to let go (lit. to go give)' exemplified in (4) is a combined version of the representations of the main and light verbs: The 'permission' information and the INITIATOR in (7) is contributed by the light verb *de-na* 'to give', with the motion information coming from the main verb *ja-na* 'to go'.

This treatment of CPs in UHVN reflects the theoretical linguistic approach to analyzing these constructions: The light verb only contributes a bleached version of its full-verb-information to the CP and constrains the types of arguments and the combinatorial possibilities of the verbs. Both of these aspects are accounted for in the resource.

4.4 Interim summary

Overall, the UHVN approach to encoding event structure makes use of three subevents, namely initiation, process and result. In order to represent motion events and CPs, we use the semantic roles of INITIATOR, UNDERGOER and RESULTEE. For the VerbNet-encoding of the path, the First-Phase roles of LOCATION and RHEME are further split into INITIAL_LOCATION, TRAJECTORY and DESTINATION.

---

[3]Ramchand explicitly assumes the merging of information in First-Phase Syntax by way of the *underassociation* principle, whereby semantic roles that can be identified based on their encyclopedic content can be unified.

(6) motion(proc, UNDERGOER)
path(proc, INITIAL_LOCATION, ?TRAJECTORY, ?DESTINATION)
result_state(res, RESULTEE, !at(INITIAL_LOCATION))

(7) permission(init, INITIATOR)
motion(proc, UNDERGOER)

5 Implementation in other resources

The subevents and semantic roles of UHVN are compatible with information contained in other lexical resources for Urdu/Hindi, in particular Hindi/Urdu PropBank (Palmer et al., 2005; Hwang et al., 2010) and the Hindi/Urdu Treebank (Bhatt et al., 2009). In Hindi/Urdu PropBank, the semantic role information of each verb is stored in a frame, which can, with the semantic role encoding assumed in First-Phase Syntax, be extended to encode two different layers of semantic role assignment. Figure 7 shows the annotation for *nɪkal-na* 'to emerge': Whereas the PropBank entry assigns Arg0 (actor) and Arg2-sou (source attribute) to the arguments of the frame, the UHVN representation assigns the UNDERGOER as well as the RESULTEE role to the moving entity, with the source location complemented by the INITIAL_LOCATION role.

Main verb *nɪkal-na* 'to emerge'	
Hindi/Urdu PropBank	**Urdu/Hindi VerbNet**
Arg0	UNDERGOER
	RESULTEE
Arg2-sou	INITIAL_LOCATION

Figure 7: Hindi/Urdu PropBank and UHVN roles

The light verb entry of *nɪkal-na* 'to emerge' in Hindi/Urdu PropBank contains an Argm, an argument modifier role, which combines with nouns in N+V CPs (Vaidya et al., 2013). This entry, as shown for complex predicates of motion, can be extended, adding the semantic roles of RESULTEE and INITIAL_LOCATION shown in Figure 6.

The semantic roles can also be added to the Hindi/Urdu Treebank, where the dependencies between verbs and arguments are encoded using the karaka roles of Pāṇini. Here, *nɪkal-na* 'to emerge' receives the roles k1 'karta' (most independent participant in an event) and k2 'karma' (locus of the re-

sult implied by the verb root). Extending the layer of annotation with the semantic roles established here would provide an interesting comparison of different principles of annotating participants of an event.

6 Discussion and conclusion

Overall, for its use in computational linguistics, First-Phase Syntax has a number of attractive properties: First of all, participants in an event can have more than one semantic role, enhancing the expressiveness of the system without increasing the number of roles. Secondly, having a set of syntactic criteria that govern the assignment of the semantic roles facilitates the process of extending the resource, as annotators can more easily decide what the correct semantic role of an argument is. Moreover, the assumptions made in First-Phase Syntax provides a framework for analyzing a notoriously difficult construction in Urdu/Hindi, namely CP formation.

Another benefit of First-Phase Syntax is its crosslinguistic validity. As shown in Ramchand (2008), the framework can be applied across languages and linguistic phenomena. Instead of having different annotation schemes emerge to accommodate constructions in languages other than English, the framework can serve as a guiding principle to encode event structure consistently across languages.

Using the class-based approach of VerbNet to analyzing CPs of different types has been shown to be a clean and theoretically well-motivated way of dealing with CPs in this kind of resource. The bleached content of the light verbs is reflected at the syntactic as well as semantic level of the VerbNet entries. The syntactic constraints as to their combinatorial possibilities with main verbs allow for a consistent and efficient computational treatment. This, together with the event decomposition and semantic roles assumed in First-Phase Syntax, paves the way for a cross-linguistic, theoretically well-motivated computational analysis of event structure.

References

Collin F. Baker, Charles J. Filmore, and John B. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of the 17th International Conference on Computational Linguistics (Coling-ACL'98)*, pages 86–90.

Rajesh Bhatt, Bhuvana Narasimhan, Martha Palmer, Owen Rambow, Dipti Sharma, and Fei Xia. 2009. A Multi-Representational and Multi-Layered Treebank for Hindi/Urdu. In *Proceedings of the Third Linguistic Annotation Workshop, ACL-IJCNLP 2009*, pages 186–189.

Claire Bonial, William Corvey, Martha Palmer, Volha Petukhova, and Harry Bunt. 2011. A Hierarchical Unification of LIRICS and VerbNet Semantic Roles. In *Proceedings of the ICSC Workshop on Semantic Annotation for Computational Linguistic Resources (SACL-ICSC 2011)*, pages 483–489.

Miriam Butt and Tracy Holloway King. 2005. The status of case. In V. Dayal and A. Mahajan, editors, *Clause Structure in South Asian Languages*, pages 153–198. Berlin: Springer Verlag.

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

Miriam Butt. 2006. *Theories of Case*. Cambridge: Cambridge University Press.

Charles J. Fillmore. 1968. Lexical entries for verbs. *Foundations of Language*, 4(4):373–393.

Jeffrey Gruber. 1965. *Studies in Lexical Relations*. Ph.D. thesis, MIT, (reprinted in Lexical Strutures in Syntax and Semantics. Amsterdam, North-Holland, 1976).

Annette Hautli-Janisz. 2013. Moving Right Along: Motion verb sequences in Urdu. In M. Butt and T. Holloway King, editors, *In Proceedings of the LFG13 Conference*, pages 295–215.

Peter Edwin Hook. 1974. *The Compound Verb in Hindi*. The University of Michigan: Center for South and Southeast Asian Studies.

Jena D. Hwang, Archna Bhatia, Claire Bonial, Aous Mansouri, Ashwini Vaidya, Nianwen Xue, and Martha Palmer. 2010. Propbank annotation of multilingual light verb constructions. In *Proceedings of ACL 2010: The Fourth Linguistic Annotation Workshop*, pages 82–90.

Jena D. Hwang, Martha Palmer, and Annie Zaenen. 2013. Representing paths of motion in verbnet. In T. Holloway King and V. de Paiva, editors, *From Quirky Case to Representing Space: Papers in Honour of Annie Zaenen*, pages 155–166. Stanford: CSLI Publications online.

Ray Jackendoff. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.

Karin Kipper-Schuler. 2005. *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon*. Ph.D. thesis, University of Pennsylvania.

Colin P. Masica. 1976. *Defining a Linguistic Area*. Chicago and London: The University of Chicago Press.

Marc Moens and Mark Steedman. 1988. Temporal ontology and temporal reference. *Computational Linguistics*, 14:15–38.

Tara Mohanan. 1994. *Argument Structure in Hindi*. Dissertations in Linguistics. Stanford: CSLI Publications.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1):71–106.

Gillian Ramchand. 2008. *Verb Meaning and the Lexicon: A First-Phase Syntax*. Cambridge: Cambridge University Press.

Ghulam Raza. 2011. *Subcategorization Acquisition and Classes of Predication in Urdu*. Ph.D. thesis, Universität Konstanz.

Ashwini Vaidya, Jinho Choi, Martha Palmer, and Bhuvana Narasimhan. 2011. Analysis of the hindi proposition bank using dependency structure. In *Proceedings of the Fifth Linguistic Annotation Workshop (ACL'11)*, pages 21–29.

Ashwini Vaidya, Jinho Choi, Martha Palmer, and Bhuvana Narasimhan. 2012. Empty Argument Insertion in Hindi PropBank. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*, pages 1522–1526.

Ashwini Vaidya, Martha Palmer, and Bhuvana Narasimhan. 2013. Semantic roles for nominal predicates: Building a lexical resource. In *Proceedings of the 9th Workshop on Multiword Expressions (MWE 2013)*, pages 126–131.