

**MOVING RIGHT ALONG:
MOTION VERB SEQUENCES IN URDU**

Annette Hautli-Janisz
University of Konstanz

Proceedings of the LFG13 Conference

Miriam Butt and Tracy Holloway King (Editors)

2013

CSLI Publications

<http://csli-publications.stanford.edu/>

Abstract

In this paper I survey the phenomenon of motion verb sequences (MVSS) in Urdu/Hindi, a combination of two motion verbs denoting a complex motion event. First noted by Hook (1974), the construction exhibits interesting syntactic and semantic properties and behaves unlike other complex verbal expressions found in the language. The paper shows that MVSSs should be treated as a special type of complex predicates, *complex predicates of motion*, complementing the various types of complex predicates already established in Urdu/Hindi (e.g., Mohanan (1994), Butt (1995)). This paper provides a first formal analysis of the construction and accounts for the types of combinations, word orders and argument structures that are possible in the language.

1 Introduction

Urdu/Hindi is known to exhibit various types of complex verbal constructions, including N+V, ADJ+V and V+V complex predicates (CPs) (e.g., Mohanan (1994); Butt (1995)). A lesser known construction, but one that occurs fairly frequently is that of motion verb sequences (MVSSs), where two motion verbs combine in a verbal phrase to express motion and direction in one complex event. (1) exemplifies the construction with the two motion verbs *kud-na* ‘to jump’ and *nikal-na* ‘to emerge’ which results in the interpretation of ‘jumping out’.

- (1) cor makan=se bahar **kud niki-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)

Hook (1974) discusses the construction in the context of complex predication, but is puzzled by several of its properties: In contrast to other complex predicates in Urdu/Hindi, the lexical semantic load seems to be equally distributed on both verbs and syntactic properties which are common among simple verbs but not complex predicates in the language are possible in these constructions, for instance causativization and the ability to swap the verbs. Moreover, the wealth of combinatorial possibilities between different motion verbs is surprising.

This paper investigates the phenomenon of motion verb sequences in Urdu/Hindi and shows that they should be treated as a new type of complex predicate, the *complex predicate of motion*. I propose an analysis that accounts for the types of combinations, word orders and argument structures that are possible in Urdu and I sketch a formal analysis of the construction. The paper proceeds as follows: section 2 provides a brief overview of complex predicates in Urdu, followed by a presentation of the motion verb data and their syntactic properties. The section is complemented by a quantitative investigation of the phenomenon, which provides

[†]I am greatly indebted to Asad Mustafa from KICS Lahore, Pakistan, for his support in data collection and Rajesh Bhatt, Miriam Butt, Gilian Ramchand and Melanie Seiss for helpful discussions.

hints as to its pervasiveness in the language. In section 3, the paper continues with a discussion of the syntactic status of MVSS, situating the construction with respect to similar constructions and their analyses in other languages. Based on these insights, section 4 provides an analysis of MVSS in Lexical Functional Grammar, followed by a general discussion and conclusion of the paper in section 5.

2 The phenomenon

2.1 In general: Verb sequences in Urdu

Complex predicates (CPS) are a special type of V+V sequence in Urdu, where a verb in the root form denoting the main action is modified by a so-called light verb, which merges with the argument structure of the main verb (Butt, 1995). CPS are a very common and in fact preferred way of expressing events in Urdu and they complement the rather small set of around 700 simple verbs. The light verbs, which also serve as main verbs in the language, contribute a bleached version of their full verb meaning. Table 2.1 presents an overview of light verbs, mainly those which also serve as full motion verbs.

	Light verbs	Contribution
Aspectual CPS	<i>par-na</i> ‘to fall’	Inception (no conscious control)
	<i>uṭh-na</i> ‘to rise’	Inception
	<i>ja-na</i> ‘to go’	Telicity
Permissive CP	<i>de-na</i> ‘to give’	Adds a causer to the event

Table 1: Common light verbs in Urdu/Hindi

The light verbs in (2) and (3) support the event denoted by the verb in the root form: in (2), the main verb *gir-na* ‘to fall’ is aspectually modified by *par-na* ‘to fall’, which contributes a sense of sudden inception without conscious control. In (3), the light verb *de-na* ‘to give’ adds an external argument to the event of *ja-na* ‘to go’ and constitutes the so-called permissive construction.

(2) am **gir par-a**
 mango.M.Sg.Nom **fall fall-Perf.M.Sg**
 ‘The mango fell (suddenly).’

(3) anjom=ne saddaf=ko **ja-ne di-ya**
 Anjum.F.Sg=Erg Saddaf.F.Sg=Acc **go-Inf.Obl give-Perf.M.Sg**
 ‘Anjum let Saddaf go.’

Butt (1995) shows that both CPS are monoclausal, but aspectual and permissive CPS differ in so far as in (2), the argument structure of the light verb *par-na* ‘to fall’ merges with the argument structure of the main verb *gir-na* ‘to fall’, whereas in (3), the light verb *de-na* ‘to give’ merges its argument structure with *ja-na* ‘to go’, but moreover adds the permission-giving entity *anjom* ‘Anjum’.

2.2 Motion verb sequences

Superficially, motion verb sequences in Urdu/Hindi are parallel to aspectual CPS in that at most two (motion) verbs are put in sequence, where the first motion verb is in the root form, complemented by a second, finite motion verb. The constructions in (4) to (7) exemplify the phenomenon.

(4) sand gayō=ki ṛevār **baṛ^h** **doṛ-a**
 ox.M.Sg.Nom cow.F.Pl.Obl=Gen.F herd.M.Sg **advance run-Perf.M.Sg**
 ‘The ox charged into a herd of cows.’

(5) sand hamare mākan=mē **g^hus cal-a**
 ox.M.Sg.Nom Pron.1.Pl.Obl.Gen house.M.Sg=Loc **enter move-Perf.M.Sg**
 ‘An ox got into our house.’

(6) ṁs=ki ank^hō=mē ansu **a b^har-e**
 Pron.3.Sg=Gen.F tear.F.Pl=Loc tear.M **come advance-Perf.M.Pl**
 ‘Tears welled up in her eyes.’

(7) g^hoṛa **doṛ b^hag-a**
 horse.M.Sg.Nom **run run-Perf.M.Sg**
 ‘The horse ran away.’

In all examples, the combination of two motion verbs yields a complex motion event, for instance combining *baṛ^h-na* ‘to advance’ and *doṛ-na* ‘to run’ as in (4) yields the bounded-path interpretation ‘running to’. The combinatorial possibilities are not restricted to directional motion verbs combining with manner of motion verbs as in (4) and (5), but constructions with both verbs denoting direction (*a-na* ‘to come’ and *b^har-na* ‘to advance’ in (6)) or manner (*doṛ-na* ‘to run’ and *b^hag-na* ‘to run’ in (7)) are also possible. Moreover, different valencies do not prevent motion verbs from combining, as shown in (5) with the transitive verb *g^hus-na* ‘to enter’ and the intransitive *cal-na* ‘to walk’ which combine to mean ‘walking/getting into’.

Oddity #1 An interesting property of MVSS is that some combinations facilitate the swapping of motion verbs. This means that the root verb becomes the finite verb and vice versa, while retaining the overall interpretation of the sentence. In (8), the verbs *cal-na* ‘to walk’ and *ṁṛ-na* ‘to fly’ render the interpretation of ‘flying up’, regardless of the particular order they occur in. The same behavior is found for *nikal-na* ‘to emerge’ and *b^hag-na* ‘to run’ in (9).

(8) hava=ke ek j^honke=ke sat^h
 wind.M.Sg=Gen.Obl one gust.M.Obl=Gen with
 patang **ṁṛ cal-i** / **cal ṁṛ-i**
 kite.F.Sg.Nom **fly move-Perf.F.Sg** / **move fly-Perf.F.Sg**
 ‘The kite flew up with a gust of wind.’ (Hook 1974, p. 57)

- (9) ek kala sap bal=se
 one black.M.Sg snake.M.Sg.Nom snake-pit.M.Sg=Instr
b^hag nīkī-a / nīkī b^hag-a
run emerge-Perf.M.Sg / emerge run-Perf.M.Sg
 ‘A black snake shot out of the snake pit.’

In both examples, the order of the motion verbs in the sequence is flexible and the MVSS denote the same event, despite the different syntactic configuration that the motion verbs are found in. This property is not found for aspectual and permissive CPs in Urdu/Hindi, probably due to the fact that the light verb is so light that it cannot provide the main propositional content of the clause. The contribution of the finite verb in MVSS is therefore “heavier” than in Urdu/Hindi CPs, an observation further confirmed by the following examples.

Oddity #2 Some constructions allow for the causativization of (at least one of) their motion verbs. The examples from (10) to (12) show the causativized versions of the constructions in (7), (8) and (9), respectively. While in (10), the finite verb *doṛ-a-na* ‘to run-Caus’ is in the causative¹, (11) shows that in other cases, the verb in the root form, *uṛ-a-na* ‘to fly-Caus’, causativizes. In (12), both verbs, *b^hag-a-na* ‘to run-Caus’ and *nīkī-na* ‘to emerge.Caus’ are in the causative.

V₁.base + V₂-Caus

- (10) malīk=ne g^hore=ko b^hag doṛ-a-ya
 owner.M.Sg=Erg horse.M.Sg.Obl=Acc run run-Caus-Perf.M.Sg
 ‘The owner made the horse run away.’

V₁-Caus + V₂.base

- (11) hava patang=ko uṛ-a cal-i
 gust.F.Sg.Nom kite.M.Sg=Acc fly-Caus move-Perf.M.Sg
 ‘The gust made the kite fly up.’

V₁-Caus + V₂-Caus

- (12) malīk=ne sap=ko bal=se
 owner.M.Sg=Erg snake.M.Sg=Acc snake-pit.M.Sg.Obl=Instr
b^hag-a nīkī-a
run-Caus emerge.Caus-Perf.M.Sg
 ‘The owner made the snake shoot out of the snake pit.’

¹In fact, *b^hag doṛ-a-na* (and its inverse *doṛ b^hag-a-na*) ‘to run away’ are the sole instances of Urdu/Hindi MVSS where causativization only applies to the finite verb. This might be due to the existence of the nominal compound *b^hag doṛ* ‘a lot of running around’, which might be used as a “simple” verbal predicate in this construction, taking the causative suffix *-a-*.

In all cases, an external argument, the causer, is added to the overall event structure. If, as in (12), both verbs are in the causative form, the external argument is shared between the two verbs. Verbs that do not causativize as simple verbs also do not allow for causativization in MVSS. In turn, verbs that can causativize as simple verbs do not necessarily allow for causativization in MVSS, where certain constraints seem to hold between the two verbs. Again, the nature of the finite verb is different from aspectual and permissive light verbs in Urdu/Hindi, as those cannot causativize.

Challenge Despite the seeming flexibility in MVSS, some combinations are clearly ungrammatical, as shown in (13) with *ring-na* ‘to crawl’ and *g^hus-na* ‘to enter’. Native speaker intuition says that in those cases, the lexical semantic entailment of *ring-na* ‘to crawl’ as a slow movement is not compatible with the determination and force contributed by *g^hus-na* ‘to enter’.

- (13) * bacca kamre=mẽ **ring g^hus-a**
 child.M.Sg.Nom room.M.Sg.Obl=Loc **crawl enter-Perf.M.Sg**
 ‘The child crawled into the room.’

What complicates the matter is that some MVSS exhibit a varying degree of speaker acceptance, as is the case for example (14), which is grammatical for the Urdu informants, but unacceptable for Hindi speakers.

- (14) √/* baccah kamre=se **ring nkl-a**
 child.M.Sg.Nom room.M.Sg.Obl=Instr **crawl emerge-Perf.M.Sg**
 ‘The child crawled out of the room.’

Based on the data in Hook (1974) and my own fieldwork data for Urdu/Hindi, motion verb sequences in the language exhibit a number of interesting properties that have so far not been attested for other verbal complexes in the language. In order to obtain a better grasp of the phenomenon, the following quantitative investigation based on different corpora sheds some light on the construction in present-day usage.

2.3 A quantitative investigation of MVSS

The preceding section shows that there are considerable idiosyncrasies in MVSS and the aim of the quantitative investigation is to provide insights into the pervasiveness of the phenomenon as well as the combinatorial patterns. Moreover, investigating large amounts of data might show tendencies for some motion verbs to appear in specific slots in the sequence.

The investigation makes use of three different Urdu corpora, namely a corpus crawled from the BBC Urdu website (BBC), the CLE corpus (Urooj et al., 2012) (CLE) and the Urdu section of the Hindi-Urdu Treebank (Bhatt et al., 2009)

(HUTB). In total, the corpora have around 16.1 million tokens. Due to the adjacency of the motion verbs and the fact that only two verbs can combine, an automatic bigram analysis suffices to extract all MVS instances. These were then aggregated over the whole corpus and the number of times the motion verbs appear as simple verbs was recorded. This serves as an approximation as to how common the verb is overall and how preferred it is in combination with other motion verbs. Table 2 gives an overview of the results, with the number of simple motion verbs found in the corpus, the number of MVSS and the number of unique MVSS.²

	BBC	CLE	HUTB
# of tokens	8,018,600	7,984,827	96,388
# of simple motion verbs	13,035	11,709	181
# of MVS	146	677	6
# of unique MVSS	33	81	3
% of MVSS	1.1%	5.8%	3.3%

Table 2: Statistics on motion verbs in the three corpora

The investigation shows that the percentage of MVSS compared to the overall usage of motion verbs is comparatively low and ranges from 1.1% in the BBC corpus to 5.8% in the CLE corpus. This might be due to the fact that MVSS preferably occur in literary text and the only corpus which accounts for this text genre is the CLE corpus, which in turn has the highest percentage of MVSS.

Moreover, the number of unique MVSS in comparison to the overall number of MVSS shows that some combinations are clearly preferred and in fact used across corpora, for instance *b^hag nīkal-na* ‘to run out of (lit. to run emerge)’, *baṛ^h caṛ^h-na* ‘to climb up (lit. to advance climb)’ and *ṡtar caṛ^h-na* ‘to climb down (lit. to descend climb)’. In addition, the verbs *b^hag-na* ‘to flee/run’, *doṛ-na* ‘to run’ and *cal-na* ‘to move/walk’ are often used as the finite verb in MVSS, with a range of different root verbs. The most flexible motion verb is *nīkal-na* ‘to emerge’ which can be used both as a root and a finite verb in a range of combinations. The investigation also shows that direct causative MVSS are less frequent than their base counterparts, whereas MVSS with indirect causatives do not exist.

2.4 Intermediate summary

What the empirical basis shows is that MVSS in Urdu/Hindi are complicated insofar as they exhibit a variety of patterns on different levels: the combinatorial possibilities regarding their lexical semantics as well as their valency, the ability to swap and the way causativization can apply. The properties imply that the construction is unlike other verbal complexes in the language, most notably those of aspectual and permissive complex predicates, which seem similar from their surface structure.

²MVSS with *ja-na* ‘to go’ as V_2 are not counted, as those combinations can be aspectual CPS denoting completion, following Butt (1995).

3 The status of the finite verb in MVSS

3.1 Mono- versus biclausality

An important question is whether the finite verb in MVSS functions as a light verb similar to aspectual and permissive light verbs or whether MVSS are modifying constructions, where the root verb modifies the finite verb. A prerequisite for complex predicatehood is monoclausality. Concerning monoclausality in Urdu/Hindi CPS, Butt (1995) proposes a number of tests, for instance the behavior of the CPS in anaphora and control constructions. However, the MVSS considered here are mostly intransitive and therefore Butt's tests for monoclausality cannot be applied reliably across constructions. Instead, I test the grammaticality of MVSS in passive constructions and their behavior with respect to negative polarity items.

Passivization Passivization in Urdu/Hindi is done via the passive auxiliary *ja-na* 'to go' which attaches to the verbal phrase. As shown in chapter 3 on passive alternations in Urdu/Hindi, if the passive auxiliary *ja-na* 'to go' combines with intransitive verbs, the interpretation of the passive construction is one of ability: the subject is able to perform an action (Butt and King, 2001). As an example, see (15): (15a) shows the active construction with the verb *g^hus-na* 'to enter' where the subject of the clause is nominative. In the ability passive alternant in (15), the subject receives the instrumental marker =*se* and is interpreted as having the ability to enter.

- (15) a. laṛki kamre=mē g^hus-i
 girl.F.Sg.Nom room.M.Sg.Obl=Loc enter-Perf.F.Sg
 'The girl entered the room.'
- b. laṛke=se kamre=mē g^hus-i ga-yi
 girl.F.Sg.Obl=Instr room.M.Sg.Obl=Loc enter-Perf.F.Sg go-Perf.F.Sg
 'The girl was able to enter the room.'

The example in (16a) shows a construction with the MVS *g^hus cal-na* 'to enter walking', where the finite verb *cal-na* 'to walk' licenses the subject *saṅḍ* 'ox' of the event. The verb in the root form, *g^hus-na* 'to enter' contributes the locational oblique *makan* 'house', which is not licensed by *cal-na* 'to walk'. If the root verb *g^hus-na* 'to enter' was embedded under the finite verb *cal-na* 'to walk', passivization should not be possible. However, the grammaticality of the passive alternant in (16b) shows that both motion verbs are in fact in the same clause: the nominative subject in (16a) becomes the instrumental-marked subject *saṅḍ=se* 'by the ox' in (16b), when the passive auxiliary *ja-na* 'to go' is attached to the verbal phrase. MVSS thus behave parallel to simple verbs.

- (16) a. saṅḍ makan=mē g^hus cal-a
 ox.M.Sg.Nom house.M.Sg=Loc enter move-Perf.M.Sg
 'An ox got into the house.'

- b. sand=se makan=mẽ g^hos cal-a ga-ya
 ox.M.Sg=Instr house.M.Sg=Loc **enter move-Perf.M.Sg** go-Perf.M.Sg
 ‘The ox was able to get into the house.’

This pattern holds for all Urdu/Hindi MVSS and it provides initial evidence with respect to the monoclausality of MVSS, a fact that will be further confirmed by the behavior of MVSS with negative polarity items.

Negative polarity items Using evidence from negative polarity items (NPIs) put forth by Bhatt (2005), Butt (to appear) shows that permissive complex predicates in Urdu/Hindi behave like other monoclausal constructions in the language, in that the NPI reading is obtained because the negation particle and the polarity item are in the same clause. Using the NPI test reliably distinguishes monoclausal constructions like permissive CPs from biclausal constructions like the instructive.

The construction in (17) shows that according to Bhatt (2005), motion verb sequences in Urdu/Hindi are indeed monoclausal: The polarity item *ek b^hi* ‘only one’ is attached to *bal=se* ‘from the snake pit’, an argument which is licensed by the finite verb in the clause, *nikal-na* ‘to emerge’, which is transitive. If the verb in the root form, *b^hag* ‘run’ were in an embedded clause, the negation particle *nahĩ* ‘not’ could not be placed in front of it and yield the NPI reading of the whole clause.

- (17) [sap] ek b^hi [bal=se]
 snake.M.Sg.Nom one Emph snake-pit.M.Sg=Instr
 nahĩ **b^hag nĩkl-a**
 not **run emerge-Perf.M.Sg**
 ‘The snake did not shoot out of even one snake pit.’

Similarly, this holds for intransitive motion verb sequences, as shown in (18).

- (18) ek b^hi patang nahĩ uṛ cal-i
 one Emph kite.F.Sg not fly move-Perf.F.Sg
 ‘Not even one kite flew up with a gust of wind.’

The evidence from both passivization and NPIs shows that MVSS in Urdu/Hindi are clearly monoclausal constructions and therefore parallel the behavior of aspectual and permissive complex predicates found in the language. This leads to the question as to how these constructions should be treated syntactically, in particular whether they belong to the class of serial verbs or complex predicates. This is elucidated in the following.

3.2 MVSS: serial verbs or complex predicates?

Using the concept of sequential motion verbs in order to express complex motion is a common phenomenon across languages, in particular in many West African,

Papua New Guinean and Australian languages. A cross-linguistic overview of the pattern is shown in the following examples, with (19) for Korean (Zubizarreta and Oh, 2007), (20) for Edo (Baker and Stewart, 1999; Ogie, 2003), (21) for Thai (Wechsler, 2003) and (22) for Dagaare, a West-African language spoken in North-Western Ghana.³ The exact concept that is conveyed by the construction can only be approximated by the English translation.

- (19) John-i kongwen-ey **kel-e ka-ss-ta**
 John-Nom park-Loc **walk-L go-Past-Decl**
 ‘John walked to the park.’ Korean (Zubizarreta and Oh, 2007, (7))
- (20) Òzó **rhùlé-rè làá òwá**
 Ozo **run-Past enter** house
 ‘Ozo ran into the house.’ Edo (Ogie, 2003, (19))
- (21) Piti **den khâw** rooŋrian
 Piti **walk enter** school
 ‘Piti entered the school walking.’ Thai (Wechsler, 2003, (2))
- (22) ó **varef kpéf waf lef** la a die poó
 Pron.3.Sg **jump.Perf enter.Perf come.Perf fall.Perf** Part Def room inside
 ‘S/he jumped (and) fell into the room.’ Dagaare

With respect to their syntactic treatment, the constructions in (19) to (22) are mostly analyzed as serial verbs, a syntactic class which has not yet been attested for Urdu. Instead, two kinds of complex predicates (aspectual and permissive CPS) have been established for Urdu/Hindi (Butt, 1995). Serial verbs and complex predicates share three crucial properties: First, the construction is characterized by a succession of verbs in a single clause with one subject. Secondly, the verbs behave as a single unit with respect to tense. Lastly, the verbs in the sequence share arguments. All three criteria match the properties of Urdu/Hindi MVSS.

Despite the fact that Urdu MVSS share a number of properties with serial verbs, they also exhibit major differences: In particular the verbs in the sequence do not contribute delimited subevents of the overall event, but the subevent of the root verb in Urdu MVSS merges with the subevent denoted by the finite verb. Moreover, Urdu MVSS do not meet the criterium of causativization in serial verbs set forth by Aikhenvald (2006), in that not only the first verb causativizes, but in fact either of the two verbs in the sequence can appear in the causative or even both. In addition, it has not been attested that Urdu MVSS share their objects, the only exception being the case of causativization, where the addition of the external argument renders the former subject the new object, which is shared by both verbs. I am fully aware that this discussion of MVSS in the light of serial verbhood only approximates of

³The example was provided by an anonymous reviewer.

what the literature provides as criteria for different languages. Nevertheless, all properties taken together, I claim that MVSSs are not prototypical serial verbs.

In contrast to the serial verb, the concept of the complex predicate in Urdu has a set of well-defined criteria and therefore facilitates a comparison with phenomena like MVSSs. Table 3 shows the comparison of MVS properties with the set of criteria proposed by Butt (1995), Butt and Geuder (2001) and Butt (2010), which set complex predicates apart from serial verbs.

Criteria	MVSSs
Light verbs do not have a systematic semantic contribution.	✓
CPs have a complex argument structure.	✓
Light verbs contribute a bleached version of their lexical semantics.	✓
Only a reduced set of verbs function as light verbs.	✓

Table 3: Comparison of CP properties with MVSSs

The properties of Urdu MVSSs mostly correspond to those exhibited by aspectual and permissive complex predicates in Urdu. Using the example in (23), I briefly exemplify the criteria that are characteristic for complex predicates and are also found for Urdu/Hindi MVS.

- (23) sap bal=se nikal b^hag-a
 snake.M.Sg.Nom snake-pit.M.Sg=Instr emerge run-Perf.M.Sg
 ‘The snake shot out of the snake pit.’

Most importantly, the way arguments are merged and contributed by the motion verbs in the MVS is complex. In (23), the main verb *nikal-na* ‘to emerge’ licenses a SUBJ and an OBL, whereas the finite verb *b^hag-na* ‘to run’ is intransitive and only licenses a SUBJ. Overall, the MVS is transitive. The complexity is increased in cases where one verb in the MVS causativizes and adds an external argument which is not licensed by the other verb. A detailed discussion on the way arguments are merged follows in section 4.

Moreover, example (23) shows that the lexical semantic contribution of the finite motion verb is a bleached version of its full-verb counterpart. The snake in (23) cannot actually perform a running motion in the way that the verb is used for a human agent, but what is conveyed by using *b^hag-na* ‘to run’ as the light verb is the notion of speed. The same holds for other light verbs of motion, e.g. the near-synonym of *b^hag-na* ‘to run’, *doṛ-na* ‘to run’, which contributes the same notion of speed in its light verb usage. In total, the data in Hook (1974) and the three corpora show that around ten light verbs of motion participate in the construction. A classification based on their lexical semantic and syntactic entailments follows in Section 4.

Despite the similarities, in order to anchor motion verb sequences in the set of complex predicates, the notion of the light verb in Urdu/Hindi has to be extended in a number of aspects: First, aspectual and permissive light verbs do not

allow for causativization, an alternation that is generally possible for Urdu/Hindi MVSSs. Moreover, it is not possible to swap the verbs of aspectual and permissive CPSs, a property that holds for a restricted number of MVSSs. Lastly, the number of “light” verbs in MVSSs is larger than the set of verbs found in the complex predicates established for Urdu (one permissive light verb, *de-na* ‘to give’, and around 15 aspectual light verbs (Hook, 1974)). However, the quantitative investigation in section 2.3 shows that there are clear tendencies for some verbs to appear more frequently and in combination with a range of different main verbs.

In sum, the above investigation has shown that Urdu MVSSs are closer to complex predicates than they are to serial verbs. Nevertheless, MVSSs are not prototypical Urdu CPSs as established by Butt (1995), and the characterization of complex predicates in the language has to be slightly extended in order to accommodate these motion constructions, in particular with respect to the characteristics of the light verbs available in the language. However, in principle, MVSSs work according to the criteria for complex predicatehood set forth by Butt (1995) and I suggest that MVSSs should be analyzed as *complex predicates of motion*. The following section provides a formal account of the construction and sheds more light on the lexical semantic concepts underlying the light verbs of motion.

4 An LFG account

The following analysis of complex predicates of motion in Lexical Functional Grammar (LFG) (Bresnan and Kaplan, 1982; Dalrymple, 2001) accommodates the different kinds of MVSSs. The motion verb in the root form as well as the light verb of motion contribute syntactic and lexical-semantic properties of the event.

4.1 Argument sharing

The data show that indeed two groups of light motion verbs exist in Urdu/Hindi: those that contribute arguments that are unified with the arguments of the main verb and those that merge arguments as well as add extra arguments to the event, while both light verbs additionally contribute lexical semantic content. In terms of Butt (1998), the first type of light verb triggers *event fusion* in that “the highest arguments of each a-structure are unified with one another” (p. 145). This mechanism accounts for aspectual complex in Urdu/Hindi predicates and is transferrable to CPSs of motion, illustrated with the example in (24). The main motion verb *ur-na* ‘to fly’ licenses a theme argument, while the light verb *cal-na* ‘to walk’ unifies its theme argument with it and moreover contributes information on the continuity of the motion event. Using lexical mapping theory as in Bresnan and Kanerva (1989), the theme *patang* ‘kite’ then maps onto the subject of the construction.

For PATH, I build on Jackendoff’s (1990) assumption that the notion of path is one of the “semantic parts of speech” and is specified by a set of attributes. For complex predicates of motion, I claim that the specific shape of the PATH is in fact instantiated by the light motion verbs in Urdu/Hindi, complementing the range of spatial postpositions. Table 4 shows the PATH attributes proposed by Jackendoff (1990) and the corresponding light verbs in Urdu.

Jackendoff’s PATH attributes	Light verbs of motion in Urdu
to	<i>g^hos-na</i> ‘to enter’
toward	<i>baṛ^h-na</i> ‘to advance’
away-from/from	<i>nikal-na</i> ‘to emerge’

Table 4: Light verbs of path in Urdu

Following the key notions of Talmy (1985) and Slobin (2004), motion events are also characterized by the configuration (or manner) with which they are carried out. As for PATH, I assume that this concept is conveyed by a number of light motion verbs which, based on the entailments of their full verb meaning, modify the manner with which the motion is carried out. So far, the literature does not provide a set of semantic attributes which are abstract enough to describe the contribution of the Urdu/Hindi motion light verbs, therefore I base the attributes on the results of the quantitative investigation in section 2.3 and the lexical semantic contribution of the respective light verbs. From this, a set of configurational attributes can be derived. Table 5 summarizes a number of light verbs which encode the configuration or manner in Urdu/Hindi motion CPS.

CONFIG attributes	Light verbs of motion in Urdu
continuity	<i>cal-na</i> ‘to walk’
speed	<i>b^hag-na</i> ‘to run’
	<i>doṛ-na</i> ‘to run’
	<i>ur-na</i> ‘to fly’

Table 5: Light verbs of motion configuration in Urdu

Nevertheless, the data show that these light verbs cannot be added arbitrarily to motion events, in particular the two near synonyms *b^hag-na* ‘to run’ and *doṛ-na* ‘to run’ cannot be used interchangeably. The lexical constraints that need to be fulfilled in order for the light verbs to be compatible still need to be worked out.

4.3 Constituent structure

The verbal phrase of complex predicates of motion is grouped under one constituent (VCmotion), where the main motion verb (Vmain) precedes the light motion verb (Vlight-motion). The c-structure for (25) is shown in Figure 1.

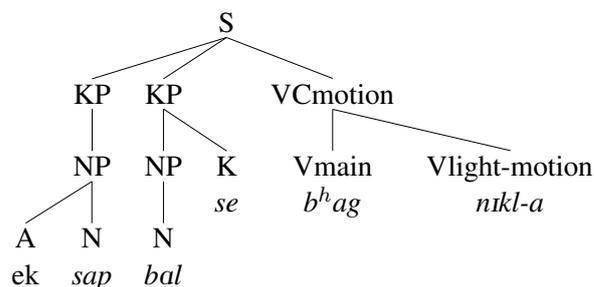


Figure 1: C-structure analysis of motion CPs

This analysis parallels the treatment of other complex predicates in Urdu, in particular their computational analysis in the Urdu ParGram grammar (Butt et al., 1999; Butt and King, 2007; Bögel et al., 2009).

4.4 Functional structure

The contribution of motion verbs in complex predicates of motion with respect to argument sharing and lexical semantic contribution is stored in the lexicon. For this, I adopt the basic idea in Butt (2010), who argues for one underlying underspecified entry which can play out as a light or full verb. The lexical-semantic features of PATH and CONFIG are kept under a [LEX-SEM MOTION] feature, as the [LEX-SEM] f-structure is already used for other lexical semantic information such as agentivity, which is syntactically represented by the ergative case marker in Urdu/Hindi. In the following, I present the analyses for the two kinds of light verbs.

Case #1: The finite motion verb is an event-fusional light verb of motion. The event-fusional light verbs of motion such as *cal-na* ‘to move/walk’ and the near-synonyms *doṛ-na* and *b^hag-na* ‘to run’ contribute lexical semantic information and unify their arguments with those of the main verb. The construction in (26) exemplifies the f-structure treatment of a CP with *b^hag-na* ‘to run’: Its lexical semantics is bleached or “light” in the sense that it loses its actual running interpretation, but it contributes a sense of speed to the main motion event.

- (26) *sap* *bal=se* ***nikal b^hag-a***
 snake.M.Sg.Nom snake-pit.M.Sg.Obl=Instr **emerge flee-Perf.M.Sg**
 ‘The snake shot out of the snake pit.’

The functional structure in Figure 2 shows that similar to aspectual CPs in Urdu/Hindi, the verb in the root form, *nikal* ‘emerge’ is the main verb of the sentence and licenses a SUBJ and an OBL. The light verb *b^hag-na* ‘to run’ does not have its own predicate value, but contributes its lexical semantic information under [LEX-SEM MOTION CONFIG].

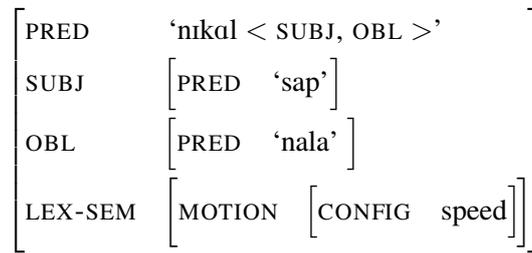


Figure 2: F-structure for (26)

This analysis of the light verb makes the correct predictions for the behavior of the construction in the causative in that in CPs with event-fusional light verbs, only the main verb in the root form can be causativized. This is exemplified in (27) for the causative alternant of the construction in (26): The external causer *malik* ‘owner’ is added to the event, licensed by the causative form of the main verb *nikal* ‘emerge.Caus’.

- (27) a. *malik=ne* *sap* *bal=se*
owner.M.Sg=Erg snake.M.Sg.Nom snake-pit.M.Sg.Obl=Instr
nikal ***b^hag-a***
emerge.Caus flee-Perf.M.Sg
‘The owner made the snake shot out of the snake pit.’

Figure 3 shows that the analysis of the causative CP construction is parallel to the construction with simple verbs proposed by Butt (1998) and Butt and King (2006), in that the main verb *nikal* ‘emerge’ is embedded under a causative predicate A-CAUSE, which licenses the subject of the sentence, *malik* ‘owner’. The subject in the inchoative variant in (26), *sap* ‘snake’, turns into the object in the causative variant. The event-fusional light verb *b^hag-na* ‘to run’ again contributes LEX-SEM information.

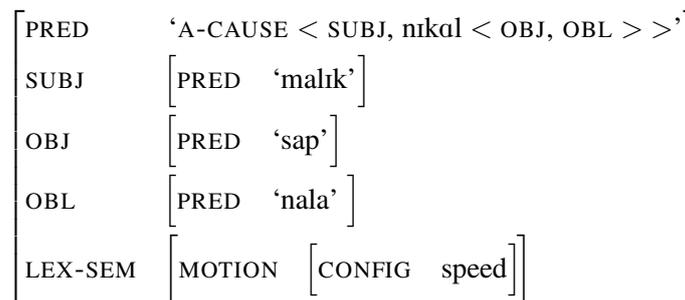


Figure 3: F-structure for (26)

In the following, I present the analysis for argument-fusional light verbs of motion, which in addition to their lexical semantics contribute extra arguments to the motion event.

Case #2: The finite motion verb is an event-fusional light verb A more complex case arises when the light verb in a motion CP adds arguments which are not licensed by the main verb. As an example, consider the construction in (28). The verb *nikal-na* ‘to emerge’ is a very common light verb in motion CPs and contributes lexical semantic information close to its full verb meaning in that it emphasizes the path out of a source location, but it also licenses a source oblique, here *makan* ‘house’, which is not licensed by the main verb of the sentence, *kud-na* ‘to jump’.

(28) cor makan=se bahar kud niki-a
 thief.M.Sg.Nom house.M.Sg=Instr outside jump emerge-Perf.M.Sg
 ‘The thief jumped out of the house.’

One solution would be to use the finite light verb as the main predicate of the sentence and treat the verb in the root form as a modifier of that verb, however this goes against the syntactic evidence for monoclausality presented in section 3. The solution I propose makes use of the restriction operator introduced by Butt et al. (2003) for complex predicate formation, in that the subject of the light verb is restricted out in order to allow the subject of the main verb take its place. This means that the root verb, here *kud-na* ‘to jump’, is treated as the main predicate of the construction, which also licenses an event filled by *nikal-na* ‘to emerge’, which in turn licenses an oblique source location. As a whole, the two predicates license two grammatical functions, SUBJ and OBL. The lexical semantic information on PATH additionally contributed by *nikal-na* ‘to emerge’ is encoded under [LEX-SEM MOTION]. Figure 4 illustrates the analysis.

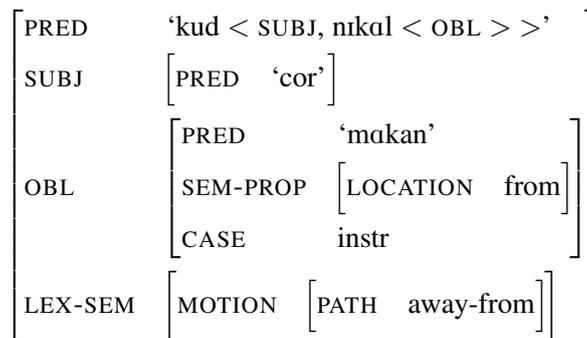


Figure 4: F-structure for (28)

An interesting case arises with the causative alternant of argument-fusional CP constructions: In these CPs, both predicates are required to be marked with the causative, otherwise the construction is ungrammatical. This is illustrated by example (29), the causative variant of (28), where the verbs *kud-a-na* ‘to jump-Caus’ as well as *nikal-na* ‘to emerge.Caus’ are in the causative. The requirement that the finite verb has to be in the causative confirms the “less light” status of

these finite verbs in that they actively contribute to the subcategorization frame of the construction.

- (29) malik=ne cor makan=se bahar
owner.M.Sg=Erg thief.M.Sg.Nom house.M.Sg=Instr outside
- kud-a nikal-a**
jump-Caus emerge.Caus-Perf.M.Sg
‘The owner made the thief jump out of the house.’

The analysis I propose here is shown in Figure 5: The causative subevent introduced by both motion verbs is recorded as the single outermost predicate A-CAUSE, which embeds the two motion verbs and their subcategorization frames. This analysis abstracts away from the fact that both verbs have a causative subcategorization frame on their own, but accounts for the fact that the motion verbs contribute the same causer to the event.

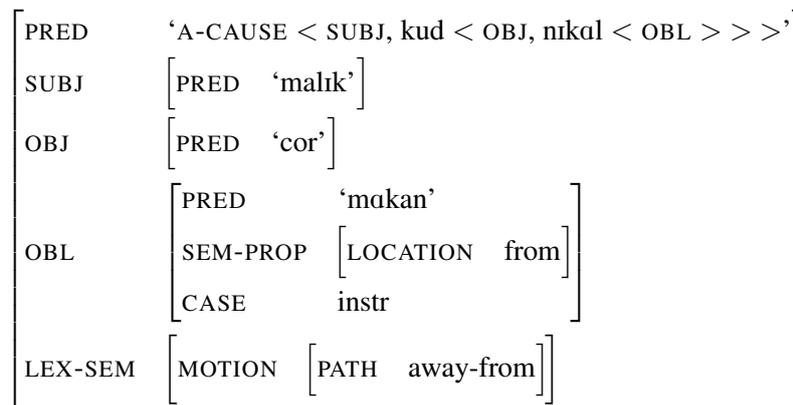


Figure 5: F-structure for (29)

In sum, the f-structure analyses show that light motion verbs are not uniform in the way they merge with main motion verbs in complex predicates of motion, in particular with respect to predicate-argument structure. As a consequence of the analysis, swapped motion verb constructions do not have the same f-structure, although their interpretation is the same.

5 Discussion and conclusion

The survey of motion verb sequences in Urdu/Hindi has shown that they should be treated as a new type of complex predicate in the language, the *complex predicates of motion*. This implies a new set of light verbs, namely *light verbs of motion*, which are shown to behave differently than aspectual and permissive light verbs established by Butt (1995). On the one hand, their influence on the overall event structure is more prominent than what is known from other complex predicates

in the language, in particular due to causativization and the ability to swap light verb and main verb. On the other hand, the constraints that are contributed by each motion verb and the exact mechanisms that prevent or license motion verb combinations are less transparent, not least because of dialectal differences.

However, light verbs of motion share important conceptual properties with other light verbs in the language, in particular, they merge their arguments and contribute lexical semantic information, a “bleached” version of their full verb meaning. However, the group of light verbs of motion is not uniform due to the fact that they share their arguments in different ways: The *event-fusional light verbs of motion* share their arguments and contribute lexical semantic features, whereas the *argument-fusional light verbs of motion* contribute additional arguments, which are not licensed by the main verb. This variation yields two different analyses on the level of f-structure: For the event-fusional CPs, the main verb is the sole predicate of the sentence, whereas in the argument-fusional case, the main verb merges its arguments with those of the embedded light verb. The causative alternants of motion CPs are parallel to simple verb causatives in that the causer is licensed by an A-CAUSE predicate, which embeds the argument structure of the motion CP.

The LFG analysis also records the lexical semantic contribution of the light motion verbs and groups it according to the two notions PATH and CONFIGURATION, two key notions of expressing motion events. The attributes of PATH are cross-linguistically well-established and have been formalized in Jackendoff (1990). The contribution of the CONFIG attributes, on the other hand, is harder to grasp and highly language-dependent, as the description of manner of motion is more difficult to formalize. In the case of Urdu/Hindi complex predicates of motion, the contribution of the light verbs regarding CONFIG is inferred based on the evidence coming from the quantitative investigation of the phenomenon and the way different motion verbs consistently modify motion events language-internally.

An interesting area for further research is to investigate the exact workings that determine the grammaticality and ungrammaticality of certain combinations. The restrictions do not seem to hold on a syntactic, but rather on a lexical semantic level that goes beyond the encoding of PATH and CONFIG. This relates to the question as to what aspects of meaning apart from path and configuration are exactly contributed by the verbs in the sequence. Resolving these issues will also pave the way for a more formal account of the argument structure which deals both with argument sharing as well as lexical semantic composition.

References

- Aikhenvald, Alexandra Y. 2006. Serial Verb Constructions in Typological Perspective. In A. Aikhenvald and R. Dixon (eds.), *Serial verb constructions: a cross-linguistic typology*, Oxford: Oxford University Press.
- Baker, Mark and Stewart, Osamuyimen T. 1999. On Double-Headedness and the Anatomy of the Clause, manuscript.

- Bhatt, Rajesh. 2005. Long distance agreement in Hindi-Urdu. *Natural Language and Linguistic Theory* 23(757-807).
- Bhatt, Rajesh, Narasimhan, Bhuvana, Palmer, Martha, Rambow, Owen, Sharma, Dipti and Xia, Fei. 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.
- Bögel, Tina, Butt, Miriam, Hautli, Annette and Sulger, Sebastian. 2009. Urdu and the Modular Architecture of ParGram. In *Proceedings of the Conference on Language and Technology (CLT09)*, CRULP, Lahore, Pakistan.
- Bresnan, Joan and Kanerva, Jonni M. 1989. Locative inversion in Chichewâ: A case study of factorization in grammar. *Linguistic Inquiry* 20(1), 1–50.
- Bresnan, Joan and Kaplan, Ronald M. 1982. Lexical-Functional Grammar: A Formal Theory for Grammatical Representation. In J. Bresnan (ed.), *The Mental Representation of Grammatical Relations*, Cambridge, MA: The MIT Press.
- Butt, Miriam. 1995. *The Structure of Complex Predicates in Urdu*. Stanford: CSLI Publications.
- Butt, Miriam. 1998. Constraining argument merger through aspect. In E. Hinrichs, A. Kathol and T. Nakazawa (eds.), *Complex predicates in nonderivational syntax*, pages 73–113, New York: Academic Press.
- Butt, Miriam. 2010. The light verb jungle: still hacking away. In M. Amberber, B. Baker and M. Harvey (eds.), *Complex Predicates*, pages 48–78, Trondheim: Cambridge: Cambridge University Press.
- Butt, Miriam. to appear. Control vs. Complex Predicates. *Natural Language and Linguistic Theory* .
- Butt, Miriam and Geuder, Wilhelm. 2001. On the (Semi)Lexical Status of Light Verbs. In N. Corver and H. van Riemsdijk (eds.), *Semi-lexical Categories: On the content of function words and the function of content words*, Berlin: Mouton de Gruyter.
- Butt, Miriam and King, Tracy Holloway. 2001. Non-Nominative Subjects in Urdu – A Computational Analysis, Paper to be presented at the International Symposium on ‘Non-nominative Subjects’ organized by ILCAA.
- Butt, Miriam and King, Tracy Holloway. 2006. Restriction for Morphological Valency Alternations: The Urdu Causative. In *Festschrift for Ronald Kaplan*, Stanford: CSLI Publications.
- Butt, Miriam and King, Tracy Holloway. 2007. Urdu in a Parallel Grammar Development Environment. *Language Resources and Evaluation* 41(1), 191–207.

- Butt, Miriam, King, Tracy Holloway and Maxwell, John. 2003. Complex Predicates via Restriction. In M. Butt and T. Holloway King (eds.), *Proceedings of the LFG03 Conference*, pages 92–104.
- Butt, Miriam, King, Tracy Holloway, Niño, María-Eugenia and Segond, Frédérique. 1999. *A Grammar Writer's Cookbook*. Stanford, CA: CSLI Publications.
- Dalrymple, Mary. 2001. Lexical Functional Grammar. In *Syntax and Semantics*, volume 34, Academic Press.
- Hook, Peter Edwin. 1974. *The Compound Verb in Hindi*. The University of Michigan: Center for South and Southeast Asian Studies.
- Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge, MA: The MIT Press.
- Mohanan, Tara. 1994. *Argument Structure in Hindi*. Dissertations in Linguistics, Stanford: CSLI Publications.
- Narasimhan, Bhuvana. 2003. Motion events and the lexicon: a case study of Hindi. *Lingua* 113, 123–160.
- Ogie, Ota. 2003. About Multi-Verb Constructions in Edo. In D. Beermann and L. Hellan (eds.), *In Proceedings of the Workshop on Multi-Verb Constructions, Trondheim Summer School 2003*.
- Slobin, Dan. I. 2004. The many ways to search for a frog: Linguistic Typology and the expression of motion events. In S. Strömquist and L. Verhoeven (eds.), *Relating events in narrative: Vol. 2. Typological and contextual perspectives*, pages 219–257, Mahwah, NJ: Lawrence Erlbaum Associates.
- Talmy, Leonard. 1985. Lexicalization patterns: semantic structure in lexical forms. In T. Shopen (ed.), *Language Typology and Syntactic Description: Grammatical Categories and the Lexicon*, volume 3, pages 57–149, Cambridge: Cambridge University Press.
- Urooj, Saba, Hussain, Sarmad, Adeeba, Farah, Jabeen, Farhat and Parveen, Rahila. 2012. CLE Urdu Digest Corpus. In *Proceedings of Conference on Language and Technology (CLT12)*.
- Wechsler, Stephen M. 2003. Serial Verbs and Serial Motion. In D. Beermann and L. Hellan (eds.), *Proceedings of the Workshop on Multi-Verb constructions*, pages 1–27.
- Zubizarreta, Maria Luisa and Oh, Eunjeong. 2007. *On the Syntactic Composition of Manner and Motion*. Cambridge, MA: The MIT Press.