Moving Right Along:
Motion verb sequences in Urdu

Annette Hautli
Universität Konstanz
LFG'13, Debrecen, Hungary,
July 19th, 2013

1 Introduction

- Various types of complex verbal constructions in Urdu/Hindi (e.g., Mohanan (1994), Butt (1995), Raza (2011)) with N+V, ADJ+V, P+V and V+V complex predicates.[1]
- **This talk** revolves around the phenomenon of motion verb sequences (MVSs) in Urdu (first noted by Hook (1973)), where two motion verbs are put in sequence.

(1) 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 1973, p. 69)

- Several properties are puzzling:
  - From a surface point of view, MVSs are similar to aspectual complex predicates (Butt, 1995): verb in the root form followed by a finite verb.
  - However, no aspectual contribution by the finite verb, but rather motional information.
  - Interpreting both verbs together results in a complex motion event.
  - Not restricted to a specific set of motion verbs.
  - **Problem:** How should Urdu MVSs be treated?

2 In general: V+V sequences in Urdu

Complex predicates are a common, frequently used and in fact preferred way of expressing events in Urdu (only about 700 simple verbs). Different kinds of **light verbs** in Urdu (as established by Butt (1995)):

	Light verbs	**Contribution**
Aspectual CPs	*paṛ-na* 'to fall'	Inception (no conscious control)
	uṭʰ-na 'to rise'	Inception
	ja-na 'to go'	Telicity
Permissive CP	*de-na* 'to give'	Adds a causer to the event

(2) am **gɪr pɑṛ-a**
mango.M.Sg.Nom **fall fall-Perf.M.Sg**
'The mango fell (suddenly).'

[1]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.

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

→ Aspectual and permissive light verbs contribute different information than the finite verb in an MVS.

- Up to now, no account of how MVSs should be treated and what an adequate syntactic representation should look like.

Roadmap:

- Show the various types of motion verb sequences (MVSs) and their syntactic properties.
- Draw conclusions on their constitution/distribution using a quantitative investigation.
- Situate MVSs with respect to similar constructions in other languages and related constructions in Urdu/Hindi.
- Provide an LFG account for the phenomenon.

3 How "light" is the finite verb?

3.1 Data

- At most two consecutive motion verbs in a verbal phrase.
- First motion verb in the root form, the second motion verb is finite and responsible for agreement and inflection.

(4) a. sanḍ gayõ=ki ṛevɑṛ **bɑṛ^h** **doṛ-a**
ox.M.Sg.Nom cow.F.Pl.Obl=Gen.Fem herd.M.Sg **advance** **run-Perf.M.Sg**
'The ox charged into a herd of cows.'

b. sanḍ hɑmare mɑkan=mẽ **g^{h}ʊs** **cɑl-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.'

c. g^{h}oṛa **doṛ** **b^{h}ag-a**
horse.M.Sg.Nom **run** **run-Perf.M.Sg**
'The horse ran away.'

Oddity #1: Some MVSs can swap their motion verbs.

- Root verb becomes the finite verb and vice versa.
- Overall interpretation of the sentence is retained.

(5) a. hɑva=ke ek j^{h}onke=ke sat^h pɑtɑng **uṛ** **cɑl-i**
wind.M.Sg=Gen one gust.M.Obl=Gen with kite.F.Sg.Nom **fly** **move-Perf.F.Sg**
'The kite flew up with a gust of wind.' (Hook 1973, p. 57)

b. hɑva=ke ek j^{h}onke=ke sat^h pɑtɑng **cɑl uṛ-i**
wind.M.Sg=Gen one gust.M.Obl=Gen with kite.F.Sg.Nom **move fly-Perf.F.Sg**
'The kite flew up with a gust of wind.'

(6) a. ek kala sap nale=se **b^{h}ag nɪkl-a**
one black.M.Sg snake.M.Sg.Nom pipe.M.Sg.Obl=Instr **run emerge-Perf.M.Sg**
'A black snake shot out of the pipe.'

b. ek kala sap nale=se **nɪkɑl b^{h}ag-a**
one black.M.Sg snake.M.Sg.Nom pipe.M.Sg.Obl=Instr **emerge flee-Perf.M.Sg**
'A black snake shot out of the pipe.'

Oddity #2: Some constructions allow for the causativization of (at least one of) their motion verbs.

1. V_1.base + V_2-Caus
2. V_1-Caus + V_2.base
3. V_1-Caus + V_2-Caus

→ Causativization not dependent on the position in the MVS, but

- on the individual motion verb (e.g. *ja-na* 'to go' cannot causativize).
- on the combination of motion verbs.

Challenge: Some combinations are clearly ungrammatical ((7) and (8)) or exhibit a varying degree of speaker acceptance (9).

(7) * ʃɑrabi kɑmre=se bahɑr **ḍɑgmɑga nɪkl-a**
drunkard.M.Sg room.M.Sg.Obl=Instr outside **stagger emerge-Perf.M.Sg**
'The drunkard staggered out of the room.'

(8) * bɑccɑh kɑmre=mẽ **ring g^{h}ʊs-a**
child.M.Sg.Nom room.M.Sg.Obl=Loc **crawl enter-Perf.M.Sg**
'The child crawled into the room.'

(9) √/* bɑccɑh kɑmre=mẽ **ring nɪkl-a**
child.M.Sg.Nom room.M.Sg.Obl=Loc **crawl emerge-Perf.M.Sg**
'The child crawled out of the room.'

→ Regional variation w.r.t. to the acceptance of MVSs, also differences between Urdu and Hindi.

3.2 A quantitative investigation of MVSs

Investigation of MVSs in three different Urdu corpora:

- Corpus crawled from the BBC Urdu website (BBC)
- The CLE corpus (Urooj et al., 2012) (CLE)

- The Hindi-Urdu Treebank (Bhatt et al., 2009) (HUTB)

→ In total, around 16.1 million tokens.

MVS **extraction:** Bigram collection of two motion verbs following each other.[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 different MVSs	33	81	3
% of MVSs	1.1%	5.8%	3.3%

Table 1: Statistics on motion verbs in the three corpora

General patterns:

- Some MVSs found across corpora:
  - *baṛʰ caṛʰ-na* 'to climb up (lit. to advance climb)'
  - *bʰag nɪkal-na* 'to run out of (lit. to run emerge)'
  - *ʊtar caṛʰ-na* 'to climb down (lit. to descend climb)'
- Causative MVSs are less frequent than their base counterparts..
- *bʰag-na* 'to flee/run' and *doṛ-na* 'to run' often in V_2 position.
- *cal-na* 'to move/walk' often V_2s in an MVS, with a range of different V_1s.
- Most flexible motion verb: *nɪkal-na* 'to emerge' used as V_1 and V_2 in a range of combinations.

3.3 Some conclusions

- Narasimhan (2003) claims that Urdu/Hindi is a *verb-framed language* (Talmy, 1991): Manner of motion is expressed by a participle construction, the path of motion is expressed by the main verb ('enter the room hobbling').
  - → No mentioning of MVSs and the way they express complex motion.
  - → "Stacking" of motion verbs as in the MVSs might be a way of avoiding elaborate adjunct constructions.
- Considerable amount of idiosyncrasy on a number of levels:
  - → No consistent explanation of combinatorial possibilities (e.g. manner of motion verbs versus directional motion verbs).
  - → Opaque rules as to the availability and interpretation of swapped MVSs.
  - → No consistent causativization pattern.
- Varying degree of lexical semantic content that is contributed by each motion verb.

[2] MVSs with *ja-na* 'to go' as V_2 are neglected as in this case I assume that *ja* 'go' is an aspectual complex predicate denoting completion, following Butt (1995).

4 The status of the finite verb in MVSs

Core question: Is the finite verb in the MVS a light verb?
Related question: Are MVSs monoclausal or biclausal?

Mostly intransitive motion verb sequences → Butt's (1995) anaphora and control tests for monoclausality cannot be applied.

Negation

- Negation (or any other modifier) can be either put in front of the MVS or between the motion verbs → scoping effects.

(10) a. sap snake.M.Sg.Nom bɑl=se snake pit.M.Sg=Instr nɑhĩ not **b^{h}ag flee/run** **nɪkl-a emerge-Perf.M.Sg**
'The snake didn't shoot out of the snake pit.'

b. sap snake.M.Sg.Nom bɑl=se snake pit.M.Sg=Instr **b^{h}ag flee/run** nɑhĩ not **nɪkl-a emerge-Perf.M.Sg**
'The snake fled but didn't emerge from the snake pit.'

- The negative polarity item can come before the MVS or in between.

(11) a. ek one b^{h}i Emph pɑtɑng kite.F.Sg nɑhĩ not **uṛ fly** **cɑl-i move-Perf.F.Sg**
'Not even one kite flew up with a gust of wind.'

b. ek b^{h}i pɑtɑng **uṛ** nɑhĩ **cɑl-i**

→ No clear indication whether MVSs are mono- or biclausal.

Complex predicate formation:

- MVSs are ungrammatical when an aspectual light verb is added.
  - "Light" motion verb in the same syntactic slot as the aspectual light verb?
- MVSs can be used in permissive complex predicates as established by Butt (1995).

(12) malɪk=ne owner.M.Sg=Erg g^{h}ore=ko horse.M.Sg=Acc [[**b^{h}ag flee/run** **doṛ-ne run-Inf.Obl**] **dɪ-ya give-Perf.M.Sg**]
'The owner let the horse run away.'

→ Implication: **MVSs are monoclausal.**

To conclude:

- In MVSs, both motion verbs are in the same clause.
- Empirical investigation shows tendencies for some verbs to appear in specific slots.
- Nevertheless, combinatorial possibilities seem to be vast.

5 A cross-linguistic perspective

Using MVSs is cross-linguistically not surprising; common method to express complex motion events, in particular in many West African, Papua New Guinean and Australian languages.

- Those constructions are mostly analyzed as serial verbs.
- In Urdu, two kinds of complex predicates (aspectual and permissive) have been established.

→ Are Urdu MVSs serial verbs or a (new) kind of complex predicate in the language?

MVSs and serial verbs

Serial verbs: Problem of defining a set of features of serial verbs that hold cross-linguistically (cf. Seiss (2009)) — Bowern (2008) offers the lowest common denominator:

Criterium	MVS**s**
Succession of verbs in a single clause with one subject.	√
The verbs behave as a single unit with respect to tense etc.	√
The verbs share arguments.	√
The verbs contribute whole subevents.	∼
The verbs share their objects (Baker, 1989).	—
In causative serial verbs, the causative comes first (Aikhenvald, 2006).	—

→ I do not consider MVSs to be prototypical serial verbs.

Are MVSs complex predicates (CPs)?

Clear-cut criteria for complex predicate-hood and the light verb features in Urdu (Butt, 1995, Butt and Geuder, 2001, Butt, 2010):

Criterium	MVS**s**
Succession of verbs in a single clause with one subject.	√
The verbs behave as a single unit with respect to tense etc.	√
CPs have a complex argument structure.	√
Light verbs do not have a systematic semantic contribution.	√
Light verbs contribute a bleached version of their lexical semantics.	√/∼
Only a reduced set of verbs function as light verbs.	∼

→ MVS are not prototypical Urdu CPs.

→ The notion of the light verb has to be extended.

→ **Claim: MVSs are instances of complex predication** in Urdu, with features that account for the interpretation of the light verb.

6 Why are MVSs important?

- Development of a lexical resource for Urdu verbs, in particular motion verbs.
- Lack of a solid theoretical basis of the syntactic representation of motion events in Urdu.

- Key information in a lexical resource with respect to motion verbs: encoding the path and configuration of motion (Hwang, Palmer and Zaenen (2013) on English VerbNet).
  - Requirement 1: Information on the nature of MVSs.
  - Requirement 2: Determination of the lexical semantic contribution of individual motion verbs in MVSs.
- Encode the phenomenon of MVSs in the lexical resource to allow for broad usability in computational applications.

7 An LFG account

Difficulty of establishing a unified account of information merging due to the (seeming) idiosyncrasy.

In general: MVSs have a complex argument structure where both verbs contribute arguments.

- Two groups of light motion verbs:
  1. Light verbs that contribute merging arguments and specific features of motion (the "real" light verbs)
  2. Light verbs that contribute merging arguments, specific features of motion and *additional* arguments (the "not so light" verbs)

Assumption #1: Build on Jackendoff's (1990) understanding that PATH is one of the "semantic parts of speech". **The specific form of the path is in fact represented by light motion verbs in Urdu.**

Jackendoff's PATH attributes	Light verbs of motion in Urdu
to	*gʰʊs-na* 'to enter'
toward	*baṛʰ-na* 'to advance'
away-from/from	*nɪkal-na* 'to emerge'
via	*gʊzar-na* 'to cross'

→ Specific motion verbs are used to express path in complex motion events, apart from the usage of directional postpositions.

Assumption #2: Motion events are also characterized by the configuration with which they are carried out. Encoding the configuration can account for light verbs that do not contribute path features.

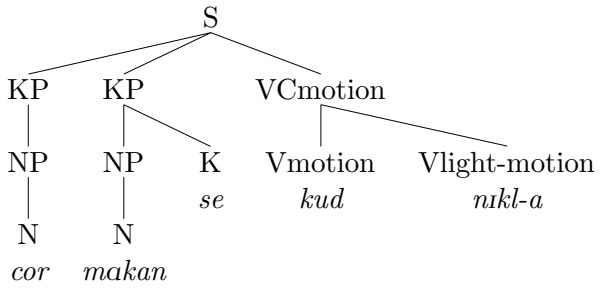
→ CONFIG attributes based on the corpus investigation:

CONFIG attributes	Light verbs of motion in Urdu
continuity	*cal-na* 'to move'
speed	*bʰag-na* 'to run'
	doṛ-na 'to run'
	uṛ-na 'to fly'

Parallel usage of motion verbs in Greek, e.g. *running* also encodes speed in motion events (S. Markantonatou, p.c.).

Constituent structure

- MVSs are grouped under one constituent (VCmotion), where **the main motion verb precedes the light motion verb**.



Functional structure

Lexicon-wise, basic idea as in Butt (2010):

- One lexical entry per motion verb, but two conjuncts:
  - A light verb usage (Vlight-motion) with lexical semantic features (and partial subcategorization information with case constraints)
  - A full verb usage with a fully specified subcategorization frame but underspecified lexical semantic features
- Lexical-semantic features under [LEX-SEM MOTION], the [LEX-SEM] f-structure is already used for information on agentivity that is manifested syntactically in Urdu.

Case #1: The second motion verb acts as a "real" light verb

1. *cal-na* **'to move/walk' as a light verb:**

- *cal-na* 'to move/walk' as the finite verb in an MVS contributes a sense of continuous movement.

  → "Light" in the sense that it looses its sense of walking.

  (13) sanḍ hɑmar-e mɑkan=mẽ **g^hʊs cɑl-a**
  ox.M.Sg.Nom Pron.1.Pl.Obl house.M.Sg=Loc **enter walk-Perf.M.Sg**
  'An ox got into our house.'

- Lexical entry of *cal-na* 'to move/walk':

  cɑl Vmain * (↑ PRED) = 'cɑl⟨(↑ SUBJ)⟩' ;
  Vlight-motion * (↑ LEX-SEM MOTION CONFIG) = continuity.

2. *doṛ-na* **'to run' and** *b^hag-na* **'to run' as light verbs:**

- *doṛ-na* 'to run' and *b^hag-na* 'to flee/run' are near synonyms and contribute a sense of speed and determination to the overall event.

→ "Light" in the sense that the actual running sense is suppressed.

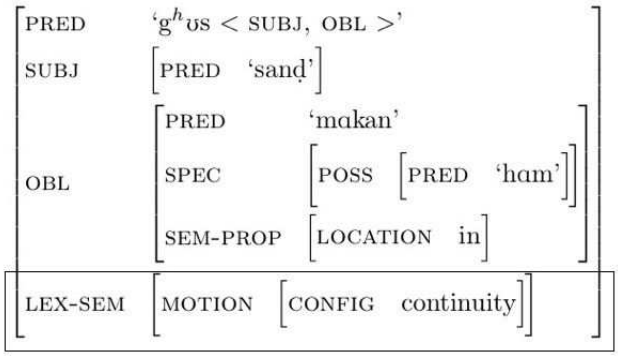


Figure 1: F-structure for (13)

(14) sap nale=se **nɪkʌl bʰag-a**
snake.M.Sg.Nom pipe.M.Sg.Obl=Instr **emerge flee-Perf.M.Sg**
'The snake shot out of the pipe.'

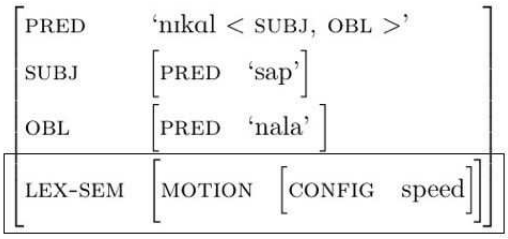


Figure 2: F-structure for (14)

Case #2: The second motion verb is "not so light"

1. *nɪkʌl-na* **'to emerge':**

- *nɪkʌl-na* 'to emerge' emphasizes the path out of a source location
- The source is required to be marked with instrumental *=se* and analyzed as an oblique.

  (15) cor mʌkan=se bahʌr **kud nɪkl-a**
  thief.M.Sg.Nom house.M.Sg=Source outside **jump emerge-Perf.M.Sg**
  'The thief jumped out of the house.'

- Lexical entry of *nɪkʌl-na* 'to emerge':

nɪkʌl	Vmain *	(↑ PRED) = 'nɪkʌl⟨(↑ SUBJ)(↑ OBL)⟩' ;
	Vlight-motion *	(↑ PRED) = 'nɪkʌl⟨(%ARGS (↑ OBL)⟩'
		(↑ OBL CASE) =c instr
		(↑ LEX-SEM MOTION PATH) = away-from.

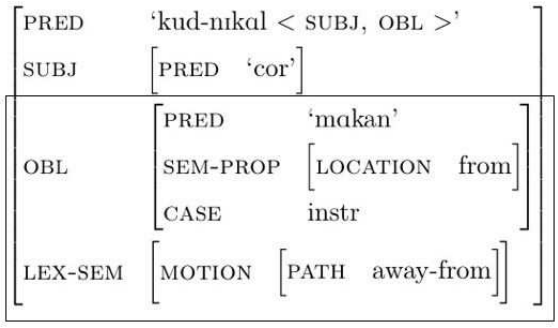


Figure 3: F-structure for (15)

Consequences:

- Swappable motion verbs do not have the same f-structure (consequent analysis of the finite verb as the light verb).

8 Discussion and conclusion

- Challenge of explaining the ungrammaticality of certain combinations remains → not solely found on the syntactic level, rather semantic constraints.
- I do not assume that information of PATH and CONFIG attributes between the main verb and the light verb regulates the combinatorial possibilities → semantic constraints, but on a different level than PATH and CONFIG.
- Answer might lie in what Levin and Rappaport Hovav (2011) term the *manner/result complementarity*: "Manner and result meaning components are in complementary distribution: a verb lexicalizes only one."

  → It seems that Urdu MVSs combine those two notions in a very flexible way.

  (16) g^{h}oṛa **doṛ** **b^{h}ag-a**
  horse.M.Sg.Nom **run** **run-Perf.M.Sg**
  'The horse ran away.'

  → The result of the running is the disappearance of the horse.

  → **Question:** Which motion verb contributes which aspect?

To conclude:

- MVSs in Urdu are in fact **complex predicates of motion**
- New group of light verbs in Urdu, namely **light verbs of motion**, which behave differently than aspectual and permissive light verbs established by Butt (1995):
  - In principle, all motion verbs can be light verbs.
  - Light verbs of motion exhibit a varying degree of "lightness".
    - Light verbs that solely contribute lexical semantic features, e.g. *cal-na* 'to move'.

    - Light verbs that contribute lexical semantic features and arguments, e.g. *nɪkal-na* 'to emerge'.

- Assumption of an elaboratted [LEX-SEM MOTION] f-structure with PATH and CONFIG information.

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, Brett and Harvey, Mark. 2010. Complex predicate formation. In M. Amberber, B. Baker and M. Harvey (eds.), *Complex Predicates*, pages 13–47, Cambridge: Cambridge University Press.

Baker, Mark. 1989. Object Sharing in Serial Verb Constructions. *Linguistic Inquiry* 20, 513–553.

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.

Bowern, Claire. 2008. The diachrony of complex predicates. *Diachronica* 25(2), 161–185.

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

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 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.

Hook, Peter Edwin. 1973. *The Compound Verb in Hindi*. The University of Michigan: Center for South and Southeast Asian Studies.

Hwang, Jena D., Palmer, Martha and Zaenen, Annie. 2013. Representing Paths of Motion in VerbNet. In T. Holloway King and V. de Paiva (eds.), *From Quirky Case to Representing Space: Papers in Honour of Annie Zaenen*, Stanford: CSLI Publications online.

Levin, Beth and Rappaport Hovav, Malka. 2011. Lexicalized Meaning and Manner/Result Complementarity. In B. Arsenijević, B. Gehrke, and R. Marín (eds.), *Subatomic Semantics of Event Predicates*, Dordrecht: Springer.

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.

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

Seiss, Melanie. 2009. On the difference between auxiliaries, serial verbs and light verbs. In M. Butt and T. Holloway King (eds.), *Proceedings of LFG09*, pages 501–519, Stanford: CSLI Publications.

Talmy, Leonard. 1991. Path to realization – via aspect and result. In *Berkeley Linguistics Society (BLS)*, volume 17, pages 480–519.

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)*.