

The interaction of light verbs and verb classes of Urdu

Tafseer Ahmed

Universität Konstanz

Germany

tafseer.khan@uni-konstanz.de

Abstract

The paper describes an attempt of identifying Urdu verb classes on the basis of the distribution of light verbs with different main verbs. We started with a frequency analysis of main + light verb sequences. The analysis of that data lead us to a thorough manual analysis of main + light verb sequences by using native speaker judgments. We focused on the three most frequent light verbs *dE* 'give', *IE* 'take' and *JA* 'go'. The verb classes were identified by considering acceptability/unacceptability of these light verbs with the main verbs. We identified some new classes. For example, we found that mental gain verbs e.g. *samajH* 'understand' are different from mental state verbs e.g. *Dar* 'fear'. The verb classes can also be used to disambiguate different senses of polysemous main verbs and different syntactical usages of the light verb forms.

1 Introduction

Urdu is an Indo-Aryan language. It is closely related to Hindi with a similar grammatical structure, but differences in script and vocabulary.

There is no comprehensive work on the verb classes of Urdu. Some authors have identified interesting classes and syntactic patterns for Urdu verbs. One of these classes consists of bodily expression verbs that allow an optional ergative marker (Butt 1995, Davison 1999). Another interesting class is of ingestive verbs (Saksena 1982, Butt 2006, Ramchand 2008). Khan (2009) identified six classes on the basis of non-canonical second argument of the verb. Ahmed (2010) clustered 184 Urdu verbs on the basis of related light verbs and aspectual auxiliaries and found four major classes. However much work remains to be done and the findings to date need to be verified and integrated with one another.

We follow Levin's (1993) classic assumption that the verb classes can be identified by their syntactic properties. She presented classes of English verbs using alternations related to English verbs. According to her, verbs that have similar syntactic properties also share semantic properties.

All of the alternations presented by Levin are not present in Urdu. An example is the beneficiary alternation that distinguishes some English verb classes. On the other hand, there are some other syntactic patterns that have not been discussed for Urdu. These can be used to classify Urdu verbs. Each light verb is acceptable with some main verbs. There exists a set of main verbs that are not acceptable with that light verbs. Hence the acceptability of certain light verbs with main verbs can be a criterion for identifying verb classes of Urdu.

Section 2 introduces Urdu light verbs. Section 3 explains how we identify the proposed verb classes for Urdu. Section 4 discusses the semantic properties of these classes and their relation to the semantic properties of the light verbs that are allowed/not allowed with these classes.

2 Light Verbs in Urdu/Hindi

In Urdu, we find sequence of verbs in which the main verb is followed by another verb (Schmidt 1999). The second verb of the sequence (that follows the main verb) can be an aspectual auxiliary, a modal or a light verb. A light verb is used to show completeness, suddenness or similar properties. In (1b), the light verb *paR* 'fall' is used that shows suddenness.

- (1) a. gARI cal-I
vehicle move-Perf.F.Sg
'The vehicle moved.'
- b. gARI cal paR-I
vehicle move fall-Perf.F.Sg
'The vehicle suddenly moved.'

Siddiqui (1971), McGregor (1972) and Hook (1974) provide lists of such verbs. Some of these verbs are: *dE* 'give', *lE* 'take', *A* 'come', *jA* 'go', *DAl* 'insert', *paR* 'fall', *beTH* 'sit', *uTH* 'rise', *dE* 'give', *rakH* 'put', *ban* 'get make', *lag* 'touch/hit', *nikal* 'come out', *Tahar* 'stop' and *cal* 'move'.

Butt and Geuder (2001) used the term light verbs for these verbs.¹ They argue that the light verbs are different from aspectual markers.

Most of the light verbs are not acceptable with all of the main verbs (Hook 1974, Butt and Geuder 2001). Every light verb is acceptable with a set of compatible verbs. Consider the example of the light verb *dE* 'give'. It is not acceptable with the verb *ruk* 'stop' as shown in (2b), however it is acceptable with the verb *cal* 'move'.

- (2) a. gARI cal dI
 vehicle.F.Sg move give.Perf.F.Sg
 'The vehicle moved.'
- b. *gARI ruk dI
 vehicle.F.Sg stop give.Perf.F.Sg
 'The vehicle stopped.'

However, the same verb *ruk* 'stop' is acceptable with the light verb *jA/ga* 'go'.

- (3) gARI ruk ga-yI²
 vehicle.F.Sg move give-Perf.F.Sg
 'The vehicle moved.'

In the previous literature, the semantic reasons for the use of these light verbs are mentioned, but there is no mention of their relation to verb classes.

An important issue with the light verbs is that these can be polysemous. McGregor (1972) pointed out that *jA* 'go' has a light verb usage to depict completion. However, it can also occur as a main verb (in conjunction) after another main verb.³ Similarly, we find main verb + main verb sequences for *dE* 'give' and *A* 'come'. Consider the following examples.

¹ In Urdu, noun + light verb and adjective + light verb are also used as complex predicate. However, we focus only on main verb + light verb sequences in this paper.

² The verb *jA* has the irregular form *ga* when used in perfective form. So the forms containing *ga* are the examples of the verb *jA*. Similarly, *dE/di/dI* 'give' and *lE/li/I* 'take' are variants of the same form in other examples.

³ There are other usages/senses of *jA* 'go' after the main verb. For example when *jA* comes after the perfective form of the verb, it is considered as a passive marker. However, the light verb *jA* that is used to represent completion is always used after the root form of the main verb. Hence it is ambiguous with the conjunction sequence only.

Similarly, all the light verbs are used only with specific form of the main verb preceding them.

- (4) a. cAnd nikal ga-yA
 moon emerge go-Perf.M.Sg
 'The moon emerged.'
- b. vuh [draxt kAT (kar)] ga-yA
 3SG tree cut having go-Perf
 'Having cut the trees, he went.'

While in (4a) *jA/ga* is used as a light verb, in (4b) it is a main verb. As *kar* 'having' can be dropped from the conjunctive clause, both sequences (verb + verb and verb + light verb) become form identical.

We find a similar ambiguity problems related to the light verb *dE* 'give'. Beside the light and main verb usages similar to (4a-b), the verb *dE* has another syntactic pattern. It introduces an additional dative or benefactive marked argument when it is used with certain verbs. The verb *xarId* 'buy' does not occur with the light verb *dE*. However *xarId* + *dE* has a dative marked beneficiary in the following example.

- (5) a. *us=nE mujHE kitAb xarId-I
 3SG=Erg 1SG.Da book win-Perf.F.Sg
 'He bought me a book.'
- b. us=nE mujHE kitAb
 3SG=Erg 1SG.Dat book
 xarId dI
 buy give-Perf.F.Sg
 'He bought me a book.'

Polysemy and identification of the correct/preferred sense is also concerned with the main verb. There are Urdu verbs that have more than one sense. Many of these senses are compatible with different light verbs. For example, the form *paRH* is used for both 'read/study' and 'read out' senses.

When *paRH* is followed by *dE*, it is used in 'read out' sense, as in (6b).

- (6) a. us=nE xat paRH li-yA
 3SG=Erg letter read take-Perf.M,Sg
 'He read the letter.'
- b. us=nE xat paRH di-yA
 3SG=Erg letter read give-Perf.M.Sg
 'He read out the letter.'

Moreover, most of the Urdu verbs have morphological causative counterparts. For example, *gir* 'fall' and *paRH* 'study' have causatives *gir-A* 'make fall' and *paRH-A* 'teach' respectively. However, there are some verbs like *badal* 'change' where the same form is used for both root and causative usages. The monovalent *badal* '(get) change' allows *jA* and rejects *dE*. On the other hand, divalent *badal* '(make) change' allows *dE* and rejects *jA*.

In summary, we know that different light verbs are used with different kinds of verb, hence these can be used to identify classes of Urdu verbs.

3 Verb classes based on light verbs

Our analysis as to the interaction of main verbs and light verbs started with the shallow processing of a corpus. We collected data related to main verb + light verb combinations by processing a (raw) corpus consisting of seven thousand documents containing 14 million tokens. The documents were obtained from CRULP's (www.crupl.org) Urdu corpus and websites www.urduweb.org and www.kitaabghar.com.

The manual inspection of this data suggests patterns and verb classes related to different light verbs. However, the data has some noise/unwanted results because of polysemous verbs and light verbs, as explained above. Other reasons were homophonous/homographic words and data sparseness for some verbs.

The frequency data and the polysemy problems were the motivation for the manual identification of verb classes. The frequency analysis helped in finding the major patterns, but the final decisions were made on the basis of native speaker's judgments. These judgments are crosschecked by Google search.

The frequency analysis shows that the light verbs *jA* 'go', *lE* 'take' and *dE* 'give' occurred with 127, 97 and 95 main verbs respectively. (There were 184 high frequency main verbs in our analysis.) The fourth most frequent light verb was *A* 'come' that occurred with 48 (out of 184) main verbs. Hence, we used acceptability of frequently used *jA*, *lE* and *dE* in the analysis of verb classes.

In Table 1, we display verb classes on the basis of their acceptability/preference with the light verbs *jA*, *lE* and *dE*. An acceptable sequence is marked as '+', an unacceptable sequence is marked with '-' and a semantically odd combination is marked as '?'.

The classes listed in Table 1 do not cover all the verbs of Urdu. We find that most of the divalent/transitives do not show special syntactic patterns with respect to the light verbs *dE* and *lE*. They accept both *dE* and *lE*.

4 Discussion

Table 1 shows that we find different groups of verb classes on the basis of acceptability of light verbs *dE*, *lE* and *jA*. In the following discussion,

we describe the semantic reasons of compatibility of the verb classes with these light verbs. We also discuss interesting verb classes found in this analysis.

Verb Class	Valency	<i>jA</i> 'go'	<i>lE</i> 'take'	<i>dE</i> 'give'
Change of state <i>gir</i> 'fall', <i>kaT</i> '(get)'	1	+	-	-
Ingestive <i>kHA</i> 'eat', <i>nigal</i> 'swallow'	2	+	+	-
Mental Gain <i>mAn</i> 'accept', <i>jAn</i> 'know'	2	+	+	-
Mental State <i>Dar</i> 'fear'	2	+	-	-
Perception <i>dEkH</i> 'see', <i>jHAnk</i> 'peep'	2,1	-	+	-
Grab <i>pakaR</i> 'grab', <i>tHAM</i> 'hold'	2	-	+	-
Send Away <i>pHEnK</i> 'throw', <i>bHEj</i> 'send'	2	-	-	+
Bodily expressions <i>hans</i> 'laugh', <i>cIx</i> 'scream'	1	-	?	+
Manner of Motion <i>ter</i> 'swim'	1	-	?	-
Manner of Displacement <i>uR</i> 'fly', <i>bHAg</i> 'run'	1	+	?	-
Sparkle <i>camak</i> 'shine', <i>mahak</i> 'smell (fragrantly)'	1	+	-	-

Table 1: Acceptability of some verb class and light verb sequences

In this analysis, we borrow the terms undergoer, resultee and rheme used by Ramchand (2008). However our analysis is not exactly similar to her analysis, therefore we use the terms 'undergoer', 'resultee' and 'rheme'. We consider the re-

ipients to be a type of resultee' and the received entity as a type of rheme'.

4.1 Verb classes related to *JA* 'go'

The verbs compatible with *JA* 'go' are those whose subject is an undergoer' i.e. it undergoes a change. One example of *JA* compatible verbs is the monovalent change of state verbs like *kaT* '(get) cut'. The state of the subject of these verbs is changed.

The other classes of *JA* compatible verbs are more interesting. The ingestive and mental gain/state verbs are divalent. The peculiar behavior of ingestive verbs in causativization and their unusual event structure have already discussed in Saksena (1982), Butt (2006) and Ramchand (2008). According to Ramchand (2008), the subject of ingestives is an undergoer. For this reason, the ingestives allow the light verb *JA*.

The subject of mental gain and mental state verbs undergoes a change. Hence, these verbs also allow the light verb *JA*. The verbs that do not accept *JA* are the ones whose subject is not an undergoer'.

There are two other interesting verb classes that accept *JA*. For *sparkle* verbs, the meaning/sense of the verb is changed when these are used with *JA* 'go' light verb. If the verb *camak* 'shine' is used in a sentence without any light verb, it means that the subject shines. However, when it is used with *JA* then it means that the subject becomes shiny.

Table 1 has a class *manner of displacement* that is different from the *manner of motion* class. Traditionally, the verbs *uR* 'fly' and *bHAg* 'run' are considered as manner of motion verbs. However these verbs allow *JA* 'go' which is related to change of state. For this reason, we introduce a special class 'manner of displacement' for these verbs.

The identification of *JA* accepting verb classes enables us to disambiguate (or find preferred reading) for the ambiguous verb + *JA* sequences as discussed in section 2. For example, since the verb *kaT* 'cut' does not belong to a *JA* accepting class, the sequence *kaT gayA* in (4b) must be a conjunctive clause.

4.2 Verb classes related to *IE* 'take'

Almost all of the verbs allowing *IE* 'take' are divalent. This light verb comes with the verbs whose subject can be a receiver/endpoint of an action. In other words, the subject can be a resultee' having a rheme'.

Table 1 shows that there are three kinds of syntactic patterns with respect to *IE* 'take' and *dE* 'give' light verbs.

There are verbs that allow *IE* and disallow *dE*. These are the verbs whose subject gets something and acts as resultee' having rheme'. Table 1 shows that *grab*, perception, mental gain and ingestive verbs belong to this kind of verbs. These verbs are semantically similar.

The ingestive and mental gain verb classes allow both *IE* and *JA* 'go' light verbs. The subject of these verbs is/can be a resultee' as well as an undergoer'. Beside these, there are some other verbs that show the same pattern because of the same semantic reasons. The verb *JI* 'win' behaves like ingestive and mental gain verbs (allow *IE* and *JA*). Similarly, *pahan* 'put on' behaves like grab verbs (allow *IE* only).

There are many verbs that allow both *IE* and *dE*. These verbs do not have any special requirement about receiving/giving of the subject. The subject can be a resultee' but it is not mandatory. These verbs can presumably be classified into finer classes on some other basis.

There are other verbs that do not allow *IE*. The subject of these verbs cannot act as a receiver or endpoint of a theme/result i.e. the subject cannot be a resultee'. The *send away* verbs are the example of these verbs that do not allow *IE*. As the subject of *bHEj* 'send' sends the object to some other place, it cannot be considered as the recipient or end point of the object that has been sent. Hence, the light verb *IE* that shows the reception/end point at the subject cannot be used with this verb.

The light verb *IE* distinguishes two different classes of mental or psych verbs. As shown in table 1, mental gain verbs e.g., *samajH* 'understand' allow both *IE* 'take' and *JA* 'go'. When someone understands some fact, he/she goes through a change of mental state (hence *JA* is allowed) by gaining the fact (hence *IE* is allowed). However the verb *Dar* 'fear' behaves differently.

- (7) a. vuh sANp=sE Dar ga-yA
3SG snake=Abl fear go-Perf.M.Sg
'He feared a/the snake.'
- b. *us=nE sANp=sE Dar li-yA
3SG=Erg snake=Abl fear take-Perf.M.Sg
'He feared a/the snake.'

As the stimulus in (7a-b) i.e. *sANp* 'snake' is not gained by the subject, the verb *Dar* 'fear' cannot be classified as a mental gain verb. We classify it in a class that is different from the class of

samajH 'understand' and *jAn* 'know' verbs. In Urdu, we have an independent evidence for this classification. The stimulus of *Dar* 'fear' is marked by the ablative marker. It shows that stimulus is a potential source and not the rheme of result. See Khan (2009) for more details.

4.3 Verb classes related to *dE* 'give'

The light verb *dE* 'give' is acceptable with the verbs that can have a receiver/endpoint that is different from the subject. It means the subject of these verbs cannot be a resultee' with rheme'.

The syntactic patterns of *dE* are similar to the patterns of *IE* 'take'. There are many divalent verbs that allow both *dE* and *IE*. These are the verbs whose subject is not necessarily a sender or receiver.

Beside these, there are verbs that allow *dE*, but does not allow *IE*. The *send away* verbs e.g. *bHEj* 'send' and *pHENk* 'throw' have a subject that cannot receive the theme i.e. it cannot be a resultee'. Hence, these verbs disallow *IE* and allow *dE*.

For a similar reason, *grab* verbs are not allowed with *dE*. The subject of these verbs is the receiver/endpoint and hence it is in conflict with the semantics of the light verb *dE*.

A similar observation for English light verb *give* was made in Newman (1996). He noted that *give*, in its extended meaning, is related to the emission. One can say *give a throw*, but *give a catch* is not acceptable. The reason is that the act of catching does not involve emission.

The monovalent verbs whose subject is not an undergoer' allow *dE* e.g., bodily expressions. However, the incompatibility of *dE* with manner of motion and displacement verbs e.g. *ter* 'swim' and *uR* 'fly' needs explanation in future work.

The verb classes related to *dE* help us in the disambiguation of some polysemous verbs. As described in section 2 and examples (6a-b), the verb *paRH* has two different senses. When *paRH* is used in 'read/study' sense, it acts as an ingestive verb and disallows *dE*. The other sense of *paRH* i.e. 'read out' is somewhat similar to bodily expression verbs, and hence it allows *dE*. Therefore, if we find a sequence of *paRH* and *dE*, we will consider it as an instance of 'read out' sense.

4.4 Verb classes related to *A* 'come'

Although we did not consider the light verb *A* 'come' as part of our analysis in Table 1, a consideration of its patterns of use bring out another interesting point. Rather than being sensitive to

event structure components such as resultee', rheme' and undergoer', it provides a sense of directionality of the action.

The verbs which accepts *A* 'come' turn out to be a subset of the ones which accept *JA* 'go'. However, only those verbs which have inherent potential directionality in their lexical semantics can be used with *A*. For example, the verbs *nikal* 'emerge', *ug* 'grow' and *baRH* 'increase' are related to direction.

(8) cAnd nikal ga-yA/A-yA
 moon.M.Sg emerge go-Perf/come-Perf
 'The moon emerged.'

The other direction-less change of state verbs e.g. *kaT* '(get) cut' does not allow *A* 'come' light verb.

(9) daraxT kaT gayA/*A-yA
 tree.M.Sg cut go-Perf/come-Perf
 'The tree got cut.'

5 Conclusion:

The study presented some classes of Urdu verbs on the basis of allowing combinations with the light verbs *JA* 'go', *IE* 'take' and *dE* 'give'. The identified classes show that light verbs are related to specific semantic classes. This work is an important step towards identifying Urdu specific (Levin-style) alternations that can give a comprehensive list of Urdu verb classes.

In this study we found that ingestive, mental gain and perception verbs behave similarly. Moreover, we found that mental/pysch verbs can be classified into (at least) two classes on the basis of light verb acceptability. The verbs of these two classes (mental gain and mental state) are semantically different from each other. Hence, the syntactic difference correctly determined the difference in semantics. Similarly, we identified two classes of manner of motion verbs.

The study needs further refinement especially in terms of semantic constructs explaining verb classes. However, the classes presented in Table 1 and the rough sketch of the semantic model that enable us to understand the problem and future directions for a complete solution.

Acknowledgment:

This research was carried out at Urdu ParGram project supported by the DFG (Deutsche Forschungsgemeinschaft).

References:

- Abul Lais Siddiqui. 1971. *Jamaul Qawaid (Comprehensive Grammar)*. Karachi.
- Alice Davison. 1999. Ergativity: Functional and Formal Issues. In Darnell et al. (eds.) *Functionalism and Formalism in Linguistics, Volume I: General Papers*, 177-208. John Benjamins.
- Anuradaha Saksena. 1982. *Topics in the Analysis of Causatives with an Account of Hindi Paradigms*. Berkeley: University of California Press.
- Beth Levin. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. , Chicago: University of Chicago Press.
- Gillian Ramchand. 2008. *Verb Meaning and the Lexicon: A First Phase Syntax*, Cambridge: Cambridge University Press.
- John Newman. 1996. *Give: A Cognitive Linguistic Study*. Berlin: Mouton de Gruyter.
- Miriam Butt. 1995. *The Structure of Complex Predicates in Urdu*. Stanford: CSLI Publications.
- Miriam Butt. 2006. *Theories of Case*. Cambridge: Cambridge University Press.
- Miriam Butt and Wilhelm Geuder. 2001. On the (semi)lexical status of light verbs. In *Semi-lexical categories: the function of content words and the content of function words*. Berlin: Mouton de Gruyter.
- Peter Edwin Hook. 1974. *The compound verb in Hindi*. Ann Arbor: University of Michigan.
- R. S. McGregor. 1972. *Outline of Hindi grammar*. Oxford: Oxford University Press.
- Ruth Laila Schimdt. 1999. *Urdu: An Essential Grammar*, London: Routledge.
- Tafseer Ahmed. 2010. Clustering Urdu verbs on the basis of related aspectual auxiliaries and light verbs. In *Proceedings of KONVENS (Conference on Natural Language Processing) 2010*. Saarbrücken.
- Tafseer Ahmed Khan. 2009. *Spatial Expressions and Case in South Asian languages*. PhD dissertation. Konstanz.