

**QUESTIONS AND INFORMATION STRUCTURE IN
URDU/HINDI**

Miriam Butt

University of Konstanz

Proceedings of the LFG14 Conference

Miriam Butt and Tracy Holloway King (Editors)

2014

CSLI Publications

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

Abstract

The overall goal of this paper is to open up a new perspective on questions in Urdu/Hindi. The paper focuses particularly on word order variation involving *wh*-elements. An analysis is developed which seeks to understand these variations not in terms of syntax-specific movement triggers, but via an integration of i(nformation)-structure, more precisely in terms of strategies for information packaging and Common Ground Management.

1 Introduction

The overall goal of this paper is to open up a new perspective on questions in Urdu/Hindi. The empirical domain is word order variation found with *wh*-elements. The analysis proposed understands these variations not in terms of syntax-specific movement triggers, but via an integration of i(nformation)-structure, more precisely in terms of *information packaging* (Chafe 1976, Vallduví 1992, Krifka 2008).

The basic idea is as follows. Given that word order variation in Urdu/Hindi in declaratives correlates with information status (e.g., Gambhir 1981, Kidwai 2000, Butt and King 1996, 1997), then why not assume that word order variation plays a similar role in questions? In particular, the analysis involves an assumption of *Common Ground Management* (CG Management) in the sense of Krifka (2008) coupled with existing ideas on the relationship between i-structure and questions developed within LFG (Mycock 2006, 2013).

The paper takes a closer look at word order variation found in: 1) constituent questions; 2) polar questions involving the question marker *kya* ‘what’. The analysis sees word order variation in constituent questions as expressing pragmatic information with respect to CG Management. The polar *kya*, on the other hand, serves to partition the clause into a focus part and a background/given part, whereby the background part is not available for questioning.

Relevant background information on word order and information structure in Urdu/Hindi declarative clauses is presented in Section 2. Section 3 discusses constituent question formation. The main components for an alternative analysis are provided in section 4 and a sample analysis for immediately preverbal vs. immediately postverbal *wh*-constituents is presented in section 5. Positional variation with respect to the polar question marker *kya* ‘what’ is dealt with in section 6. The paper concludes with section 7.

2 Word Order and Information Structure in Urdu/Hindi

The default word order in Urdu/Hindi is SOV. All major constituents can scramble. Hindi/Urdu is not strictly verb-final, which means that material can appear after the verb. Topics are generally found clause initially, focus in the immediately preverbal positions (e.g., Gambhir 1981, Kidwai 2000).

However, one cannot assume a simple one-to-one mapping between position and information structure. Gambhir (1981) has shown that the clause-final position, for example, has several functions. One central function is de-emphasis (which mostly involves pronominals), dubbed BACKGROUND in Butt and King (1996, 1997). However, postverbal material can also signal added emphasis on new information that is presented, e.g., in a TV/radio announcement style (or for the creation of suspense). Processing may also play a role in that “heavy” items can be shifted to the end of the clause (cf. heavy NP-shift in English).

Similarly, Gambhir (1981) shows that the clause initial position is not always a topic, but can also be used for scene setting. Given her description, this can be thought of as “frame setting” in Krifka’s (2008) sense. An illustrative example from English is given in (1), where *healthwise* is not a topic, but a frame setter.

- (1) A. How is John?
 B. {Healthwise}, he is [FINE]_F.

Urdu/Hindi also contains some discourse particles. The discourse particle *to* generally marks preceding constituents as topics (cf. Kidwai 2000) or frame setters. The particle *hi* ‘only’ serves to emphasize and/or focus preceding constituents.

Based on the above patterns and on a small corpus study of Bollywood movie dialogs, Butt and King (1996, 1997) developed a four-way i-structure analysis in terms of [\pm prom(inent)] and [\pm new], as shown in Table 1.

Type	Definition	Position
Topic	[−New,+Prom(inent)]	Clause-Initial
Focus	[+New,+Prom(inent)]	Immediately Preverbal
Background	[−New,−Prom(inent)]	Postverbal
Completive Information	[+New,−Prom(inent)]	Between Topic and Focus

Table 1: Four major types of i-structure categories (Butt and King 1996, 1997)

Their proposal incorporated ideas by Vallduví (1992) on information packaging, Choi’s (1996) realization of information structure within LFG and Kiss’ (1995) notion of discourse configurationality. Butt and King proposed that a clause could be partitioned into four main information structural components: topic, focus, background and completive information. An example of a sentence containing all four types of i-structure categories is provided in (2), where “T” stands for topic, “F” for focus, “CI” for completive information and “B” for background.

- (2) [nadya]_T (to) [ab^hi] [tɔfi]_{CI} [bazar=se]_F xarid
 Nadya.F.Nom indeed just now toffee.F.Nom market.M=from buy
 rah-i t^h-i [mere=liye]_B
 stay-Perf.F.Sg be.Past-F.Sg I.Gen.Obl=for
 ‘Nadya was just buying toffee at the market for me.’

Butt and King’s system is fairly simple and in need of expansion. For example, the system as presented does not deal with embedded clauses, nor does it provide a complete analysis of when and how pro-drop is possible.¹ Some critical discussion and further development of the system can be found in Mycock (2006, 2013).

3 Constituent Questions

3.1 Basic Data

Urdu/Hindi has traditionally been characterized as a *wh*-in-situ language (but also see Bayer (2006)). This is illustrated in (3), where (3b) is the interrogative version of (3a). The *wh*-element *kis=ko* in (3b) occurs in exactly the same linear position in (3a) as *ram=ko* does in (3a).

- (3) a. sita=ne d^hyan=se **ram=ko** dek^h-a t^h-a
 Sita.F=Erg carefully Ram.M=Acc see-Perf.M.Sg be.Past-M.Sg
 ‘Sita had looked at Ram carefully.’
- b. sita=ne d^hyan=se **kis=ko** dek^h-a t^h-a?
 Sita.F=Erg carefully who.Obl=Acc see-Perf.M.Sg be.Past-M.Sg
 ‘Who had Sita looked at carefully?’

However, the traditional characterization is not quite correct as the default position for *wh*-elements is actually the immediately preverbal position, which has in turn independently been established as a focus position (Kidwai 2000). (4) shows the default word order for the questioning of an ergative subject. The *wh*-element is not in-situ, but immediately preverbal.

- (4) ram=ko d^hyan=se **kis=ne** dek^h-a t^h-a?
 Ram.M=Acc carefully who.Obl=Erg see-Perf.M.Sg be.Past-M.Sg
 ‘Who had looked at Ram carefully?’

Similarly, the default order for a how-question is illustrated in (5c). The *wh*-element is in the immediately preverbal position and not in the in-situ position. When it appears in-situ, as in (5b), the clause expresses an additional dimension of meaning, e.g., some degree of wonder.

- (5) a. sita=ne **d^hyan=se** ram=ko dek^h-a t^h-a
 Sita.F=Erg carefully Ram.M=Acc see-Perf.M.Sg be.Past-M.Sg
 ‘Sita had looked at Ram carefully’

¹Any argument can in principle be dropped in Urdu/Hindi. Pro-drop cannot be argued to be licensed via agreement or any other morphological or syntactic factors. Butt and King (1997) suggest that pro-drop is linked to old information status. An in-depth corpus study conducted by Prasad (2000, 2003) suggests that grammatical relations are a significantly conditioning factor so that object pro-drop is more likely if the subject of the clause has been dropped.

- b. sita=ne **kaise** ram=ko dek^h-a t^h-a?
 Sita.F=Erg how Ram.M=Acc see-Perf.M.Sg be.Past-M.Sg
 ‘How had Sita managed to see Ram?’ (expresses degree of wonder)
- c. sita=ne ram=ko **kaise** dek^h-a t^h-a?
 Sita.F=Erg Ram.M=Acc how see-Perf.M.Sg be.Past-M.Sg
 ‘How had Sita looked at Ram?’ (default order for a how-question)

Wh-constituents are not restricted to the default focus position or the in-situ position — they have the same kind of scrambling possibilities as non-interrogative NPs (Manetta 2012). However, when they appear in non-default positions, this correlates with an additional pragmatic effect (cf. (5b) vs. (5c)). To date, a comprehensive analysis of these pragmatic effects does not exist.

3.2 Scope and Scope Marking

The left clausal periphery has received the most overall attention to date. This is because the literature has focused on why, given a general assumption of covert *wh*-movement for Urdu/Hindi, *wh*-words in embedded clauses cannot take matrix scope (e.g., Mahajan 1990, Srivastav 1991, Dayal 1994, 1996, 2014, Lahiri 2002a, Bhatt and Dayal 2007, Manetta 2010, 2012).

This issue is the most obvious one to tackle from an LF-based perspective in which the *wh*-word is assumed to move to a position where it can act as a scope operator (usually SpecCP). Since the *wh*-word in Urdu/Hindi can stay in situ, LF-based approaches have to assume that the *wh*-word undergoes covert (invisible) movement to the appropriate operator position. But then, what prevents covert movement from applying in examples such as (6), where the *wh*-element cannot take matrix scope?

- (6) ravi jan-ta t^h-a
 Ravi.M.Nom know-Impf.M.Sg be.Past-M.Sg
 [ke sita=ne d^hyan=se **kis=ko** dek^h-a t^h-a]
 that Sita.F=Erg carefully who.Obl=Acc see-Perf.M.Sg be.Past-M.Sg
 ‘Ravi used to know [who Sita had looked at carefully].’
 *‘Who did Ravi use to know [Sita had looked at carefully]?’

I do not go into the various (movement or copy-theory) solutions proposed in the literature since *wh*-extraction is not the main focus of the paper (see Mycock (2006) for an LFG approach). Instead, I focus on some empirical issues.

Empirically, there are two ways of achieving matrix scope for embedded *wh*-elements. The *wh*-constituent can either appear in the matrix clause (“extraction”), as shown in (7).² Alternatively, the so-called *scope marking* construction can be employed, as illustrated in (8)–(9).

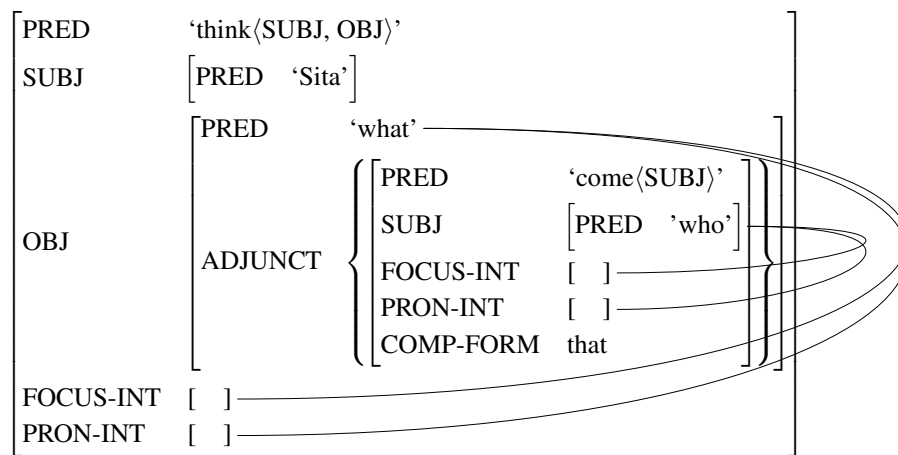
²The acceptability of long-distance extraction as in (7) has been disputed (e.g., Stepanov and Stateva 2006). The first reported instances are in Gurtu (1985). It has since been established that they are definitely acceptable under certain intonational contours (Dayal 2014).

(7) **kya**_i tom jan-te ho
 what you.Pl.Nom know-Impf.M.Pl be.Pres.2.Pl
 [ki us=ne ____i ki-ya]?
 [that 3.Sg.Obl=Erg do-Perf.M.Sg]
 ‘What did you know that he did?’ (Srivastav 1991:766) **wh-extraction**

(8) ravi **kya** jan-ta he
 Ravi.M.Nom what.Nom know-Impf.M.Sg be.Pres.3.Sg
 [ke sita **kis=ko** pasand kar-ti he]?
 that Sita.F who.Obl=Acc liking do-Impf.F.Sg be.Pres.3.Sg
 ‘Who does Ravi know Sita likes?’
 Lit.: ‘What does Ravi know, who does Sita like?’ **scope marking**

In the scope marking construction, *kya* ‘what’ is introduced in the matrix clause and the embedded *wh*-element remains in-situ. The matrix *kya* signals that the embedded *wh*-element must be interpreted as having matrix scope. Note that the *kya* in the scope marking construction is the same *kya* that acts as a standard *wh*-element and means ‘what’, cf. (7). Dayal has argued for an indirect dependency account by which the embedded *that*-clause is anaphorically related to the *kya* in the matrix clause, which in turn is not seen as an expletive, but as a “thematic” ‘what’, i.e., a full *wh*-element (Srivastav 1991, Dayal 1994, 1996, 2014).

(9) Scope Marking Construction, f-structure for (7)



I follow Dayal’s overall indirect dependency analysis and render the ‘that’-clause as an adjunct modifying the thematic *kya*, as shown in (9). This instantiates the indirect dependency approach, by which the *kya* ‘what’ is analyzed as a proper (not expletive) argument of the matrix verb and the embedded ‘that’-clause modi-

fies the *kya* argument of the matrix clause.³ The analysis also follows the overall approach to questions established within the ParGram effort.⁴

In contrast, Mycock (2006) adopts a direct dependency approach (cf. McDaniel 1989) for her LFG analysis of Hungarian scope marking. A direct dependency approach has also been proposed for Hindi by Manetta (2010), but Dayal (2014) shows that this does not make the right predictions.

3.3 Word Order Variation in Constituent Questions

In this section, I return to the issue of word order variation with respect to *wh*-elements. Recall that a *wh*-constituent can appear anywhere an NP can (Manetta 2012). In particular, the following examples have recently been discussed in some detail (Bhatt and Dayal 2007, Manetta 2012).

- (10) a. sita=ne d^hyan=se **kis=ko** dek^h-a t^h-a?
 Sita.F=Erg carefully who.Obl=Acc see-Perf.M.Sg be.Past-M.Sg
 'Who had Sita looked at carefully?' (*wh*-in-situ/preverbal focus)
- b. sita=ne d^hyan=se dek^h-a t^h-a **kis=ko**?
 Sita.F=Erg carefully see-Perf.M.Sg be.Past-M.Sg who.Obl=Acc
 'Sita had looked carefully at who?' (*wh* postverbal)
- c. sita=ne d^hyan=se dek^h-a **kis=ko** t^h-a?
 Sita.F=Erg carefully see-Perf.M.Sg who.Obl=Acc be.Past-M.Sg
 Reading 1: 'Who had Sita looked at carefully?' (*wh* in verbal complex)
 Reading 2: 'Who had Sita really looked at carefully?'
 (i.e., she had not looked at anybody carefully)

(10a) is a standard *wh*-question with the *wh*-element in the default preverbal focus position (it is also in-situ). (10b) is analyzed as an echo question by Bhatt and Dayal (2007). For (10c), Bhatt and Dayal (2007) propose a Rightward Remnant Movement analysis in which the verb moves for topicalization purposes.

Manetta (2012) investigates a larger range of data than Bhatt and Dayal and argues that their analysis does not cover enough empirical ground. She insteads advocates a scrambling account in which movement is triggered by probe-goal relationships involving features such as Q(uestion), *wh*, E(cho)) and the EPP (Extendend Projection Principle). While these features can be motivated theory-internally, several hold no explanatory power when viewed from an external perspective.

However, rather than delving into this issue, I would like to pursue an alternative analysis which invokes pragmatic, information structural concerns in order

³A reviewer expressed worry on how to restrict the appearance of such CP adjuncts. The adjunct version of a CP is constrained by the following factors: a) it must contain a *wh*-phrase (checked for via PRON-INT (interrogative pronoun)); b) the matrix clause must contain a *kya*; c) the matrix verb must be of the right type (cf. Lahiri 2002b); d) if a COMP analysis is possible, then that is preferred via OT-Marks (Frank et al. 2001).

⁴<http://typo.uni-konstanz.de/redmine/projects/pargram/wiki/Questions>

to explain the word order variation. Manetta and Bhatt and Dayal already each invoke information structural notions as part of their analysis. Bhatt and Dayal assume verb topicalization and Manetta explicitly refers to existing work on information structure as well as scrambling in Urdu/Hindi (Gambhir 1981, Butt and King 1996, Dayal 2003). She concludes that Topic/Focus is the result of leftward scrambling while Backgrounded/Old Information is the result of rightward scrambling. However, both Manetta and Bhatt and Dayal propose syntax-specific movement triggers, rather than assuming an independent information-structural component.

For one, recall that (10b) has been analyzed as an echo question by Bhatt and Dayal. Bhatt and Dayal posit that the postverbal *wh*-constituent cannot be interpreted as a standard information-seeking question because it is trapped in a remnant VP, which acts as an island. Manetta derives the echo reading from the old/background information analysis of the postverbal position. However, echo questions have been analyzed as involving a type of focus (Artstein 2002, Truckenbrodt 2012). Given this, the echo reading of *kis=ko* in (10b) cannot follow from the background/old information connection drawn by Manetta. It is also not immediately obvious how being trapped in a VP island would focus a constituent.

Now consider the occurrence of the *wh*-constituent within the verbal complex, as in (10c). When the *wh*-element is in this position, an extra pragmatic meaning dimension can be added to the question. One interpretation of (10c) is that the speaker is not actually expecting an answer to the question (giving rise to a type of rhetorical question). This does not follow from Manetta's account, nor does it follow from Bhatt and Dayal's short distance topicalization of the verb. Further examples of this type, along with an alternative analysis, are provided in section 5.

4 Information Packaging

The proposal put forward here is that word order variation of *wh*-constituents should not be understood primarily in terms syntactic considerations, but in terms of *information packaging* (Chafe 1976, Vallduví 1992).

4.1 Krifka's proposal

Krifka (2008) cuts through the existing plethora of proposals for topic and focus by providing a combinatory interaction between semantics and pragmatics. He proposes that information structure consists of two major parts:

1. Common Ground Content: truth conditionally relevant information
2. Common Ground Management: pragmatics, packaging of information to fulfill communicative needs/structure the discourse in a certain way.

Among other issues, Krifka argues that understanding information structure in terms of features such as [\pm new] (as done in Butt and King (1996, 1997) for

example) is not useful. He shows that these features does not yield the right semantic/pragmatic effects. Instead, he proposes to understand focus and topic in terms of CG Management. Krifka sees both topic and focus as being interpreted with respect to alternative sets, as per Rooth's (1985) Alternative Focus Semantics. This means that the relevant instance out of a larger set of possible alternatives is identified as part of the communicative effort. With respect to both topic and focus, the notion of *givenness* is centrally important — this is what is (or what is assumed to be) already in the Common Ground. The particulars of CG Management are realized via a type of “File Card Semantics” à la Heim (1982). The overall idea is that for each entity introduced as part of a discourse, a “file card” is opened up and information about that entity, including whether it is given or not, is recorded.

The idea of a file card semantics is reminiscent of what was more fully fleshed out within Discourse Representation Theory (Kamp and Reyle 1993). A variant is also assumed by Vallduví (1992) and it appears to be very similar to the “sorting key” notion adopted by Mycock (2006). For the purposes of this paper, I remain agnostic as to which particular technology to adopt.

Important for the analysis is Krifka's separation of information structure into CG Content vs. CG Management and the idea of understanding topic and focus in terms of Rooth's alternative semantics. In particular, the use of word order variation for information packaging falls under CG Management.

4.2 Mycock's LFG Perspective

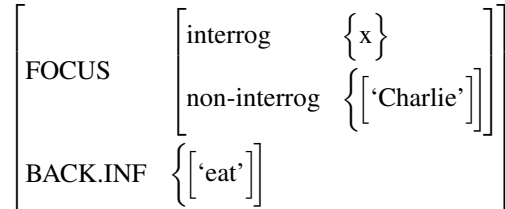
Mycock (2006) investigates *wh*-constituent questions from a typological perspective and works out an analysis within a parallel LFG architecture in which phonology/prosody and morphosyntax can contribute to the overall interpretation of an utterance on an equal footing. Mycock posits that all question phrases must have focus status. This status can be encoded at one or more levels of linguistic structure. Thus, focus status could be signaled via syntactic position and/or intonation or (an interaction of) other linguistic devices. Mycock explicitly integrates information structure into her overall analysis. Her proposal provides an immediate account of why the default position for *wh*-constituents in Urdu/Hindi is the immediately preverbal position: this position syntactically encodes focus status and is thus one very basic way to ensure that a *wh*-phrase has focus status.

With respect to truth-conditional semantics, Mycock adopts Ginzburg and Sag's (2000) propositional abstract semantics. Under this approach, *wh*-words introduce a parameter which is to be filled in. Interrogative vs. non-interrogative focus is distinguished explicitly via a feature, as shown (12) for the example in (11). Additionally, Butt and King's four-way i-structural distinction between topics, focus, background and completive information is adopted.

The calculation of the semantics of interrogative scope is effected via a meaning constructor [**interrog-scope**], which can be introduced via the syntax (annotation on c-structure rules) or via prosodic information.

(11) [What]_{FOCUS} did [CHARLIE]_{FOCUS} eat?

(12) **i-structure**



4.3 Combining Ideas

Mycock does not deal with non-canonical interpretational effects produced by word order variation. Krifka does not deal (explicitly) with questions or with the effects of information structure in SOV languages like Urdu/Hindi. Each of the proposals has elements that are important for an overall analysis of the word order variation found with *wh*-elements in Urdu/Hindi.

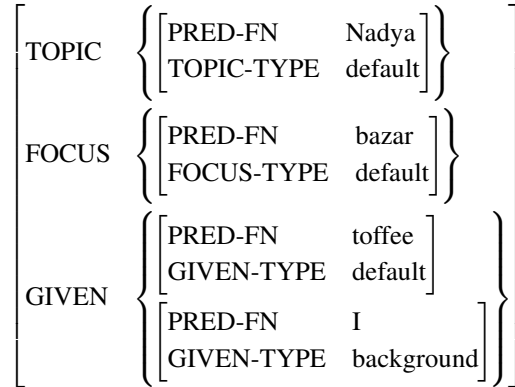
I assume an LFG architecture in which i-structure is represented as a separate projection (as per King 1997 and Mycock 2006) and in which information about i-structure flows together from morphosyntax and prosody. However, instead of feature-based notions of topic, focus, background and completive information, I adopt Krifka's basic notions of topic, focus and givenness. At the same time, I do allow for finer grained distinctions within those categories (encoded via an X-TYPE feature). This is also consistent with Krifka's approach.

The revised i-structure analysis of (2), repeated here as (13), is as in (14). The focus is on 'bazar', all other information is registered as given, with the postverbal material additionally being marked as backgrounded. The values for Topic, Focus and Given are sets because one can have multiple instances of these.

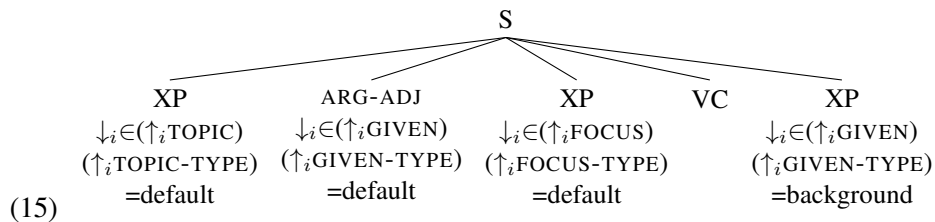
The default values for X-TYPE are provided via c-structure annotations in the syntactic positions associated with them. These default values can be "overridden" by information coming from morphology, prosody, lexical items such as the focus clitic *hi* (cf. Butt and King 1998) or the overall structure of the discourse. In practice, the annotations take the form of disjunctions. The default disjunct applies in the absence of other, more particular information.

(13) [nadya]_T (to) [ab^hi] [ʈʌfi]_{CI} [bazar=se]_F xarid
 Nadya.F.Nom indeed just now toffee.F.Nom market.M=from buy
 rah-i t^h-i [mere=liye]_B
 stay-Perf.F.Sg be.Past-F.Sg I.Gen.Obl=for
 'Nadya was just buying toffee at the market for me.'

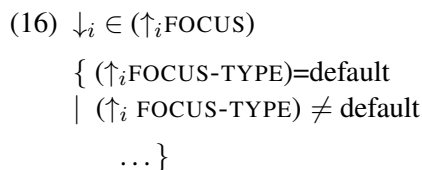
(14) i-structure for (13)



The tree in (15) shows the basic c-structure assumed for Urdu. There is little to no evidence for a matrix-level VP in Urdu and major constituents can scramble. This is reflected in the flat, exocentric structure of (15). The ARG-ADJ stands for a metacategory that expands into any or no number of arguments and adjuncts.



The *i* subscript on the annotations indicates that the information is projected to the i-structure. As mentioned, the annotations are more complex than indicated in (15), with a typical annotation featuring a disjunct as in (16) whereby the second half of the disjunct tests for relevant conditions when the focus type is not default.



The overall semantics and pragmatics of an utterance are assumed to be calculated via a Discourse Representation Theory (DRT) style of analysis, or, more precisely, a Segmented DRT (SDRT) style analysis (e.g., Asher and Lascarides 2003, Lascarides and Asher 2007), whereby the i-structure information flows into a CG Management system. A CG Management system in the sense of Krifka has so far not been incorporated into SDRT; however, pragmatic information management/update has been integrated. The exact mechanics of this go beyond the scope of this paper. As a theory of syntax, LFG is in principle compatible with several different types of semantic/pragmatic analyses and I am assuming that the spirit of

the Krifka approach to information packaging can be integrated into SDRT. SDRT itself is in principle compatible with LFG.

The approach taken here differs from Dalrymple and Nikolaeva (2011), who reject a Krifka type structured meaning approach by which a sentence is partitioned into given vs. not (Krifka 1992, 2008). They instead see information structure as partitioning sentence meaning into i-structure categories. In contrast, I follow Krifka in viewing information structure as providing instructions for CG Management. The overall ramifications of the differing proposals with respect to i-structure need to be investigated in detail, but again go beyond the scope of this paper.

5 Analysis: immediately postverbal *wh*-constituent

This section takes a closer look at *wh*-constituents in the immediately postverbal position within the verbal complex as in (10c). The verbal complex in Urdu/Hindi can consist of various light verbs, auxiliaries and modals (cf. Butt and Rizvi 2010).⁵ The examples in (17) and (18) are from a Bollywood movie and an actual conversation, respectively. They both feature *kya* ‘what’.

- (17) varna nic^he log soc^h-ē-g-e pata nahī
 otherwise underneath people think-3.PI-Fut-3.M know not
 ye log us kamre=mē kar **kya** rah-e hē
 this people that.Obl room.M.Sg.Obl=in do what stay-Perf.M.Pl be.Pres.3.Pl
 itn-i der=se
 this much-F.Sg time.F.Sg=Inst
 ‘Otherwise the downstairs people will think: What are these people doing in that room for such a long time?’
 (Lit. Otherwise the downstairs people will think: Don’t know WHAT these people are doing in that room for such a long time.)
 From Bollywood Movie *Socha Na Tha*

- (18) un=ke g^har=mē **saman** t^h-a
 Pron.3.Pl=Gen.M.Obl house.M.Sg=in luggage.M.Sg.Nom be.Past-M.Sg
kya pahle?
 what before
 ‘What possessions did they even have in their house before (then)?’
 (implies: they had no possessions before) Parveen Butt, July 2014

Tests show that the *wh*-constituent cannot appear anywhere else within the verbal complex. The *wh*-constituent is generally stressed. These two facts speak for a privileged immediately postverbal position within the verbal complex that is associated with focus. A further piece of evidence concerns negation. Manetta (2013)

⁵When the verb occurs on its own, the immediately postverbal position is ambiguous: it could be within the verbal complex, or it could be postverbal material outside of the verbal complex.

- (20) a. ye log kamre=mẽ **kar kya** rah-e hẽ
 this people room.M.Sg.Obl=in do what stay-Perf.M.Pl be.Pres.3.Pl
 ‘What are these people doing in the room?’
- b. vo **k^ha** rah-e hẽ
 they eat stay-Perf.M.Pl be.Pres.3.Pl
 ‘They are eating.’

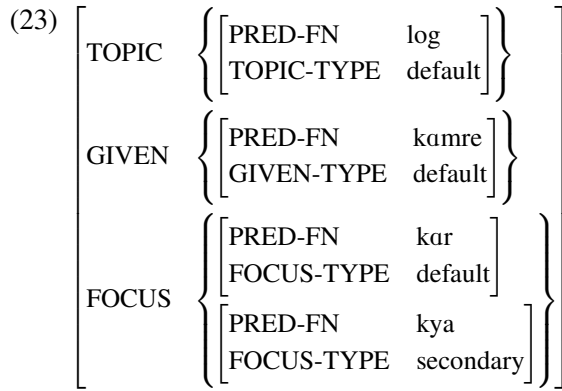
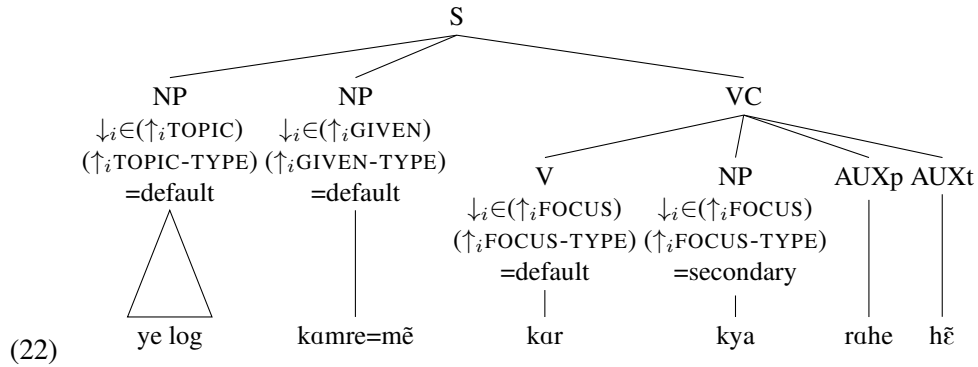
I propose that if the verb is to be focused in a question, the question word (or phrase) is placed in the immediately postverbal position within the verbal complex in order for the verb to receive the standard intonation associated with focus in Urdu/Hindi (cf. Patil et al. 2008 on standard focus patterns in Hindi).⁸ In terms of Krifka’s Alternative Semantics approach to information structure this could be understood as follows: the placement of a *wh*-element in this position signals that the speaker is not expecting an answer for the *wh*-phrase and that therefore no set of alternative answers should be opened up for the XP containing the question word.

This approach offers an immediate explanation of why (21b) is pragmatically illformed, but (21a) is fine. Abush (2010) proposes that several elements, including questions and focus, trigger defeasible (soft) presuppositions. However, the *wh*-element *konsi* ‘which’ carries with it a strong presupposition that there is a set of alternative answers to the question posed and that this set is non-empty, i.e., that the questioned entity exists. Placing *konsi* ‘which’ in the immediately postverbal position thus leads to an interpretational clash: a strong presupposition of the existence of an entity in an alternative set of answers vs. a presupposition that no set of alternatives should be considered since the answer is not at issue in the first place.

- (21) a. sita [konsi yāzal] ga-ti
 Sita.F.Nom which.M.Sg song.M.Sg.Nom sing-Impf.F.Sg
 rah-ti t^h-i?
 stay-Impf.F.Sg be.Past-F.Sg
 ‘Which song (ghazal) did Sita keep singing?’
- b. #sita ga-ti [konsi yāzal]
 Sita.F.Nom sing-Impf.F.Sg which.M.Sg song.M.Sg.Nom
 rah-ti t^h-i?
 stay-Impf.F.Sg be.Past-F.Sg
 ‘Which song (ghazal) did Sita keep singing?’

As a sample analysis, (22) and (23) provide c- and i-structure representations of (20a). The postverbal position functions as a syntactic secondary focus position. Additionally, when this position is filled, it goes hand in hand with the verb being in focus (one could think of this as a type of construction). The linear order of auxiliaries (and light verbs) within the verbal complex in Urdu is very strict, the analysis here follows the pattern described in Butt and Rizvi (2010).

⁸This analysis is originally due to an observation by Farhat Jabeen.



6 Polar Questions

This section shows that the overall information structural perspective on word order variation with respect to *wh*-elements also provides an analysis for the ordering possibilities of *kya* ‘what’ in polar questions.

As shown in (24), polar questions in Urdu/Hindi take the syntactic form of declaratives. Intonation must be used to distinguish between a declarative (24a) and an interrogative (24b). As shown in (25) and (26), polar questions can also be overtly marked morphosyntactically via *kya* ‘what’. This *kya* has been dubbed “Polar *kya*” by Bhatt and Dayal (2014).

- (24) a. anu=ne uma=ko kitab d-i
 Anu.F=Erg Uma.F=Dat book.F.Sg.Nom give-Perf.F.Sg
 ‘Anu give a/the book to Uma. declarative
- b. anu=ne uma=ko kitab d-i?
 Anu.F=Erg Uma.F=Dat book.F.Sg.Nom give-Perf.F.Sg
 ‘Did Anu give a/the book to Uma? polar question
- (25) **kya** anu=ne uma=ko kitab d-i?
 what Anu.F=Erg Uma.F=Dat book.F.Sg.Nom give-Perf.F.Sg
 ‘Did Anu give a/the book to Uma? polar *kya*

Traditional grammars only mention the clause initial position for polar *kya* (e.g., Glassman 1977). However, Bhatt and Dayal (2014) report a relatively free distribution of polar *kya*, cf. (26).

- (26) (**kya**) anu=ne (**kya**) uma=ko (**kya**) kitab (**kya**) d-i?
 what A.F=Erg what U.F=Dat what book.F.Sg.Nom what give-Perf.F.Sg
 ‘Did Anu give a/the book to Uma?’

Bhatt and Dayal suggest that the different possible positions result from topicalization. That is, in (26), ‘Anu’ is topicalized when it precedes the *kya* and when the *kya* appears immediately preverbally, then all three of the arguments have been topicalized. Bhatt and Dayal adduce evidence for topicalization from interactions with weak indefinites, idiomatic objects and gapping. Again, a clear connection is drawn between the position of the question element and information structure.

Bhatt and Dayal (2014) seek to understand polar *kya* as a speech act operator in the sense of Krifka (2014). I propose an alternative analysis, which is still consonant with Bhatt and Dayal’s finding that the items to the left of *kya* show evidence of topicalization. However, rather than assuming topicalization, I propose that the polar *kya* in clause medial position partitions a clause into given vs. not, as per Krifka’s Structured Meaning approach (Krifka 1992, 2008). Everything to the left of *kya* must be interpreted as given, everything to the right as not.

Evidence for this analysis comes from data⁹ as in (27), which involves alternative questions. As can be seen, anything to the right of *kya* is available for questioning. However, material to the left of *kya* is not. This is consonant with an analysis under which everything to the left of *kya* is part of what is presupposed/given and not available for focus and hence not for questioning.

- (27) a. ram=ne sita=ko **kya** kitab d-i ya āguṭʰi?
 Ram.M=Erg Sita.F=Dat what book.F.Nom give-Perf.F.Sg or ring.F.Nom
 ‘Did Ram give a book or a ring to Sita?’
- b. ram=ne **kya** sita=ko kitab d-i
 Ram.M=Erg what Sita.F=Dat book.F.Nom give-Perf.F.Sg
 ya amra=ko/*ravi=ne?
 or Amra.F=Dat/Ravi.M=Erg
 ‘Did Ram give a book to Sita or Amra?’
 *‘Did Ram or Ravi give a book to Sita?’

Under this approach polar *kya* is a type of focus-sensitive operator that determines which parts of a clause are backgrounded/presupposed and which are open for further discussion. When *kya* appears in clause initial position, there is no given part of the clause.

Further support for the analysis comes from the fact that, as shown in (28), polar *kya* cannot appear immediately postverbally. This restriction makes sense

⁹The data are due to Rajesh Bhatt.

if the immediately postverbal position is indeed a special focus position in which lexically contentful constituents can be interpreted. In this position focus-sensitive operators (which create their own focus domains) are misplaced.

- (28) *sita=ne d^hyan=se ram=ko dek^h-a **kya** t^h-a?
 Sita.F=Erg carefully see-Perf.M.Sg who.Obl=Acc what be.Past-M.Sg
 ‘Did Sita carefully look at Ram?’

Technically, the partitioning of a clause into a given part vs. a focus part is accomplished via annotations in the c-structure introduced by the lexical entry of polar *kya*.¹⁰ Via f-precedence everything that is to the left of polar *kya* can be determined to be GIVEN at i-structure. This is illustrated in (29) for (27a).

- (29) i-structure for (27a)

TOPIC	$\left\{ \begin{array}{l} \text{PRED-FN} \quad \text{Ram} \\ \text{TOPIC-TYPE} \quad \text{default} \end{array} \right\}$
FOCUS	$\left\{ \begin{array}{l} \text{PRED-FN} \quad \text{kitab} \\ \text{FOCUS-TYPE} \quad \text{default} \\ \text{PRED-FN} \quad \text{ãguṭhi} \\ \text{FOCUS-TYPE} \quad \text{default} \end{array} \right\}$
GIVEN	$\left\{ \begin{array}{l} \text{PRED-FN} \quad \text{Ram} \\ \text{GIVEN-TYPE} \quad \text{background} \\ \text{PRED-FN} \quad \text{Sita} \\ \text{GIVEN-TYPE} \quad \text{background} \end{array} \right\}$

7 Summary

This paper has examined word order variation in constituent and polar questions in Urdu/Hindi. Rather than understanding the word order variation in terms of syntactic triggers, the paper has laid out an approach which leverages the correlation between i-structure status and linear position established for declarative clauses. The paper further proposes to understand i-structure in terms of Krifka’s (2008) proposals for information structure, in particular, in terms of CG Management.

In constituent questions, word order variation was proposed to signal strategies for GC Management. The paper concerned itself particularly with the immediately postverbal position within the verbal complex. This is analyzed as a secondary structural focus position. Furthermore, the verb is in focus and the CG Management expectation is for there to be no answer for the *wh*-constituent.

Word order variation with respect to polar *kya* is also analyzed via i-structure status. However, it is seen as an operator which serves to partition a clause into a

¹⁰This is analogous to how the meaning constructor [**interrog-scope**] is introduced by Mycock.

focus part and a background (given) part, whereby the given part is presupposed and thus not available for questioning.

Acknowledgements

I would like to acknowledge Rajesh Bhatt and Veneeta Dayal for help with the data and for the many interesting discussions on the theoretical aspects of questions in Urdu/Hindi. I would also like to thank Farhat Jabeen, Ghulam Raza and Ashwini Vaidya for further help with the data and the audience of the TripleA 1 Workshop at the University of Tübingen, in particular, Manfred Krifka, for valuable feedback. My colleagues Maribel Romero and Maria Biezma have been invaluable with respect to trying to navigate me through the forest of the semantics and pragmatics of questions. Thanks go to Tracy King and an anonymous reviewer for incisive comments, not all of which I have been able to do justice to. Finally, I would like to thank my brother Yassin for putting on a CD at just the right moment.

References

- Abush, Dorit. 2010. Presupposition Triggers from Alternatives. *Journal of Semantics* 27(1), 37–80.
- Artstein, Ron. 2002. A focus semantics for echo questions. In Á. Bende-Farkas and A. Riester (eds.), *Workshop on Information Structure in Context*, pages 98–107, Stuttgart: IMS: University of Stuttgart.
- Asher, Nicholas and Lascarides, Alex. 2003. *Logics of Conversation*. Cambridge: Cambridge University Press.
- Bayer, Josef. 2006. Wh-in-situ. In Martin Evaraert and Henk van Riemsdijk (eds.), *The Blackwell Companion to Syntax*, Oxford: Blackwell.
- Bhatt, Rajesh and Dayal, Veneeta. 2007. Rightward Scrambling as Rightward Movement. *Linguistics Inquiry* 38(2), 287–301.
- Bhatt, Rajesh and Dayal, Veneeta. 2014. Polar *kyaa*, talk presented at the *Workshop on Non-Canonical Questions and Interface Issues*, Konstanz, February 2014.
- Bögel, Tina and Butt, Miriam. 2012. Possessive clitics and *ezafe* in Urdu. In Kersti Börjars, David Denison and Alan Scott (eds.), *Morphosyntactic categories and the expression of possession*, pages 291–322, Amsterdam: John Benjamins, (Linguistik Aktuell/Linguistics Today 199).
- Butt, Miriam and King, Tracy H. 1996. Structural Topic and Focus without Movement. In M. Butt and T. H. King (eds.), *Proceedings of the First LFG Conference*, CSLI Publications.

- Butt, Miriam and King, Tracy Holloway. 1997. Null Elements in Discourse Structure, written to be part of a volume that never materialized, <http://ling.uni-konstanz.de/pages/home/butt/main/papers/nulls97.pdf>.
- Butt, Miriam and King, Tracy Holloway. 1998. Interfacing Phonology with LFG. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG98 Conference*, CSLI On-line Publications.
- Butt, Miriam and Rizvi, Jafar. 2010. Tense and Aspect in Urdu. In Patricia Cabredo-Hofherr and Brenda Laca (eds.), *Layers of Aspect*, pages 43–66, Stanford: CSLI Publications.
- Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics and point of view. In Charles Li (ed.), *Subject and Topic*, pages 27–55, New York: Academic Press.
- Choi, Hye-Won. 1996. *Scrambling: Optimality-theoretic Interaction between Syntax and Discourse*. Ph. D. thesis, Stanford University.
- Dalrymple, Mary and Nikolaeva, Irina. 2011. *Objects and Information Structure*. Cambridge: Cambridge University Press.
- Dayal, Veneeta. 1994. Scope Marking as indirect wh-dependency. *Natural Language Semantics* 2(2), 137–170.
- Dayal, Veneeta. 1996. *Locality in WH Quantification*. Dordrecht: Kluwer Academic Publishers.
- Dayal, Veneeta. 2003. Bare nominals: Non-specific and contrastive readings under scrambling. In Simin Karimi (ed.), *Word Order and Scrambling*, pages 67–90, Oxford: Blackwell.
- Dayal, Veneeta. 2014. Questions and the vP Phase in Hindi-Urdu, ms. Rutgers University.
- Dik, Simon C. 1997. *The Theory of Functional Grammar, Part 1*. Berlin: Mouton de Gruyter.
- É. Kiss, Katalin. 1995. *Discourse Configurational Languages*. Oxford: Oxford University Press.
- Frank, Anette, King, Tracy Holloway, Kuhn, Jonas and Maxwell III, John. 2001. Optimality Theory style constraint ranking in large-scale LFG grammars. In Peter Sells (ed.), *Formal and Empirical Issues in Optimality Theory*, pages 367–397, Stanford: CSLI Publications.
- Gambhir, Vijay. 1981. *Syntactic Restrictions and Discourse Functions of Word Order in Standard Hindi*. Ph. D. thesis, University of Pennsylvania, Philadelphia.

- Ginzburg, Johnathan and Ivan A. Sag. 2000. *Interrogative Investigations: The Form, Meaning, and Use of English Interrogatives*. Stanford, CA: CSLI Publications.
- Glassman, Eugene H. 1977. *Spoken Urdu*. Lahore: Nirali Kitaben.
- Gurtu, Madhu. 1985. *Anaphoric Relations in Hindi and English*. Ph.D. thesis, CIEFL, Hyderabad.
- Heim, Irene. 1982. *The semantics of definite and indefinite noun phrases*. Ph.D. thesis, University of Massachusetts, Amherst.
- Kamp, Hans and Reyle, Uwe. 1993. *From Discourse to Logic*. Dordrecht: Kluwer Academic Publishers.
- Kidwai, Ayesha. 2000. *XP-Adjunction in Universal Grammar: Scrambling and Binding in Hindi-Urdu*. Oxford: Oxford University Press.
- King, Tracy Holloway. 1997. Focus Domains and Information Structure. In Miriam Butt and Tracy Holloway King (eds.), *On-Line Proceedings of the LFG97 Conference*, CSLI On-line Publications.
- Krifka, Manfred. 1992. A Compositional Semantics for Multiple Focus Constructions. In Joachim Jacobs (ed.), *Informationsstruktur und Grammatik*, pages 17–53, Opladen: Westdeutscher Verlag.
- Krifka, Manfred. 2008. Basic Notions of Information Structure. *Acta Linguistica Hungarica* 55(3–4), 243–276.
- Krifka, Manfred. 2014. Embedding Illocutionary Acts. In Thomas Roeper and Margaret Speas (eds.), *Recursion: Complexity in Cognition*, Berlin: Springer.
- Lahiri, Utpal. 2002a. On the proper treatment of expletive wh in Hindi. *Lingua* 112, 501–540.
- Lahiri, Utpal. 2002b. *Questions and Answers in Embedded Contexts*. Oxford: Oxford University Press.
- Lascarides, Alex and Asher, Nicholas. 2007. Segmented Discourse Representation Theory: Dynamic Semantics with Discourse Structure. In H. Bunt and R. Muskens (eds.), *Computing Meaning: Volume 3*, pages 87–124, Kluwer Academic Publishers.
- Mahajan, Anoop. 1990. *The A/A-Bar Distinction and Movement Theory*. Ph.D. thesis, MIT.
- Manetta, Emily. 2010. Wh-expletives in Hindi-Urdu: The vP phase. *Linguistic Inquiry* 41(1), 1–34.

- Manetta, Emily. 2012. Reconsidering Rightward Scrambling: Postverbal Constituents in Hindi-Urdu. *Linguistic Inquiry* 43(1), 43–74.
- Manetta, Emily. 2013. Copy theory in wh-in-situ languages: Sluicing in Hindi-Urdu. *Journal of South Asian Linguistics* 6, 3–22.
- McDaniel, Dana. 1989. Partial and Multiple Wh-Movement. *Natural Language and Linguistic Theory* 7(4), 565–604.
- Mycock, Louise. 2006. *The Typology of Constituent Questions: A Lexical-Functional Grammar Analysis of wh-questions*. Ph.D. thesis, University of Manchester.
- Mycock, Louise. 2013. Discourse Functions of Question Words. In M. Butt and T.H. King (eds.), *Proceedings of the LFG13 Conference*, Stanford: CSLI Publications.
- Patil, Umesh, Kentner, Gerrit, Gollrad, Anja, Kügler, Frank, Féry, Caroline and Vasishth, Shravan. 2008. Focus, Word Order and Intonation in Hindi. *Journal of South Asian Linguistics* 1(1), 55–72.
- Prasad, Rashmi. 2000. A Corpus Study of Zero Pronouns in Hindi: An Account Based on Centering Transition Preferences. In *Proceedings of DAARC2000*, pages 66–71, Lancaster University.
- Prasad, Rashmi. 2003. *Constraints on the Generation of Referring Expressions, with Special Reference to Hindi*. Ph.D. thesis, University of Pennsylvania, Philadelphia.
- Raza, Ghulam and Ahmed, Tafseer. 2011. Argument Scrambling within Urdu NPs. In *Proceedings of the LFG11 Conference*, Stanford, CA: CSLI Publications.
- Rooth, Mats. 1985. *Association With Focus*. Ph.D. thesis, University of Massachusetts, Amherst.
- Samek-Lodovici, Vieri. 2005. Prosody-syntax interaction in the expression of focus. *Natural Language and Linguistic Theory* 23, 687–755.
- Srivastav, Veneeta. 1991. Subjacency Effects at LF: The Case of Hindi wh. *Linguistic Inquiry* 22(4), 762–769.
- Stepanov, Arthur and Stateva, Penka. 2006. Successive cyclicity as residual wh-scope marking. *Lingua* 116, 2107–2153.
- Truckenbrodt, Hubert. 2012. An analysis of prosodic F-effects in interrogatives: Prosody, syntax and semantics. *Lingua* 124, 131–175.
- Vallduví, Enric. 1992. *The Informational Component*. New York: Garland Press.
- Zubizarreta, Maria Luisa. 1998. *Prosody, Focus and Word Order*. Cambridge, MA: The MIT Press.