Word Order Variation in Urdu/Hindi Wh-Constituent Questions

Miriam Butt

University of Konstanz

Workshop on Information Structure: Form and Interpretation
Wien, July 2018

Universität
Konstanz

DFG Deutsche
Forschungsgemeinschaft

Context

- Part of a Research Unit (FOR 2111) *Questions at the Interfaces* at Konstanz
  - Looking at non-canonical uses of questions across languages
  - Joint work with María Biezma, Tina Bögel and Farhat Jabeen
- Generally trying to understand the interplay between **prosody, morphosyntax** and **semantics/pragmatics**.
- This talk:
  - focus on word order variation in Urdu/Hindi wh-constituent questions

Overall Goal

Understand word order variation in questions

- not via syntax-specific movement triggers
- but via an integration of i(nformation)-structural concerns (*information packaging*; Chafe (1976), Vallduví (1992), Krifka (2008))

Overall Goal

Information Structure and Questions:

- Word order in Hindi/Urdu is known to be correlated with i-structure (e.g., Gambhir 1981, Kidwai 2000, Butt and King 1996, 1997).
- Basic Idea: leverage this knowledge towards understanding pragmatic effects found with word order variation of wh-elements (or "k-elements") in Hindi/Urdu.
- Hypothesis: Word order variation indicates strategies for *Common Ground Management* in the sense of Krifka (2008).
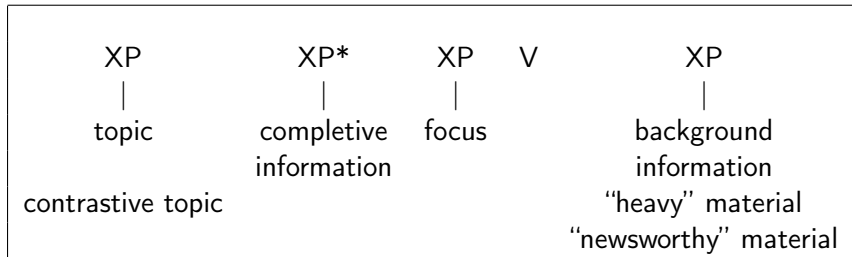
Word Order

- The default word order in Urdu/Hindi is **SOV**.
- Major constituents can scramble.
- Discontinuous NPs are made use of as a regular part of the language.
- Unlike in Japanese and Korean, for example, material can appear after the verb.
- **Topics** generally initial, **focus** immediately preverbal (Gambhir 1981, Kidwai 2000)

Word Order

Topological View of Urdu/Hindi Clause Structure (based on Butt and King 1996, 1997, Gambhir 1981)

XP	XP*	XP	V	XP
|	|	|		|
topic	completive information	focus		background information
contrastive topic				"heavy" material
				"newsworthy" material

Rampant Pro-Drop

- Any argument can be dropped
  (no licensing via agreement or other morphological or syntactic
  means).
- Pro-drop licensed by context: the center of discourse can be dropped
  (and objects can only be dropped if subjects are also dropped), see
  Prasad (2000, 2003) for details.

Polar Questions

- Polar questions have the same SOV default order as declaratives.
- Question vs. declarative status is signaled via intonation.

Declarative: Intonational phrase boundary is L-L%

(1)　ʃahina=ne　　norina=ko　　mara$_{\text{L-L\%}}$
　　　Shahina.F=Erg Norina.F=Acc hit-Perf.M.Sg
　　　'Shahina hit Norina.'　　　　　　　　　　　　　　　(Declarative)

Polar Question: Intonational phrase boundary is L/H-H%

(2)　ʃahina=ne　　norina=ko　　mara$_{\text{L/H-H\%}}$
　　　Shahina.F=Erg Norina.F=Acc hit-Perf.M.Sg
　　　'Did Shahina hit Norina?'　　　　　　　　　　　　(Polar Question)

Polar *kya* 'what'

- Polar questions can optionally be expressed with *kya* 'what'.

  (3) (**kya**) ʃahina=ne   norina=ko   mara?
        what Shahina.F=Erg Norina.F=Acc hit-Perf.M.Sg
        'Did Shahina hit Norina?'

- Grammars and previous literature report polar *kya* as appearing only clause initially in Urdu/Hindi.

Polar *kya* 'what'

- In contrast, Bhatt and Dayal (2014) point out that it can appear anywhere in the clause.

  (4) (kya) ɑnu=ne (kya) uma=ko (kya) kɪtab   (%kya) d-i
       what A.F=Erg what U.F=Dat what book.F.Sg.Nom what   give-Perf.F.Sg
       (kya)?
       what
       'Did Anu give a/the book to Uma?'

- Bhatt&Dayal establish that polar *kya* is NOT a question marker.
    - It is optional in matrix clauses.
    - Generally disallowed in embedded clauses (complements of "rogative" predicates like 'wonder' and 'ask' are an exception).

Polar *kya* 'what'

Current State of Our Art (Biezma et al. 2018)

- Polar *kya* is a focus sensitive item which serves to constrain the set of possible answers viable in the context of an utterance.
- It imposes restrictions on what the question is about.
- Polar *kya* questions convey some assumptions regarding the possible answers that plain information-seeking questions do not convey.

Wh-Questions in Urdu/Hindi

- Urdu/Hindi has traditionally been characterized as a **wh-in-situ** language (Bayer and Cheng 2015).

  (5) a. sita=ne   dʰyan=se   **ram=ko**   dekʰ-a   tʰ-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   dʰyan=se   **kıs=ko**   dekʰ-a   tʰ-a?
           Sita.F=Erg carefully   who.Obl=Acc see-Perf.M.Sg be.Past-M.Sg
           'Who had Sita looked at carefully?'

- But wh-words can in principle appear anywhere in the clause.
- They have the same kind of scrambling possibilities as normal NPs (Manetta 2012).

Scrambling of Wh-Constituents

- The scrambling possibilities have overwhelmingly been accounted for in terms of syntax-specific movement (triggered by features).
- For a comprehensive overview over the state-of-the-art around 2003, see Bhatt (2003).
- Notions such as topicalization, focus and echo questions are invoked, but seemingly as a by-product of an overall movement analysis.

Hypothesis: Word order variation is systematically leveraged to express pragmatic information.

Some Previous Analyses

- Bhatt and Dayal (2007) analyze the *wh*-constituent within the verbal complex as being due to "short distance verb topicalization".
- This topicalization does not appear to be independently motivated.
- **Immediately Postverbal Position:** Deriving [Subj V DO-wh Aux] (based on Bhatt and Dayal 2007:295)
    - Start with: [S [DO-wh V] Aux]
    - Leftward scrambling of DO-wh $\longrightarrow$
        [S [DO-wh$_i$ [t$_i$ V]] Aux]
    - Leftward scrambling (topicalization) of V $\longrightarrow$
        [S [[t$_i$ V]$_j$ [DO-wh$_i$ t$_j$]] Aux]
    - Covert wh-movement $\longrightarrow$
        [DO-wh$_i$ [S [[t$_i$ V]$_j$ [t$'_i$ t$_j$]] Aux]]

Some Previous Analyses

- Manetta (2012):
    - topic and focus is the result of leftward scrambling
    - backgrounded/old information is the result of rightward scrambling
    - Postverbal *wh*-words are interpreted as echo questions.
    - *wh*-words carry features which differ according to position.[1]

Regular *wh*-word	Echo *wh*-word
[*i*wh]	[*i*wh]
[*u*Q]	[*u*E]

---
[1]i=interpretable feature, u=uninterpretable feature

Interim Summary:

- Both Bhatt&Dayal and Manetta invoke information structural notions.
- However, these are not included formally as part of the analytical framework.
- Both Bhatt&Dayal and Manetta take a predominantly syntactic perspective.

Focus and Questions

- The **default** position for wh-words is actually not the in-situ position.
- It is the immediately **preverbal** position.
- Recall that this is the default focus position (Gambhir 1981, Butt and King 1996, 1997, Kidwai 2000).

   (6)   a.   **sita=ne**   ram=ko   dek^h-a   t^h-a
                Sita.F=Erg Ram.M=Acc see-Perf.M.Sg be.Past-M.Sg
                'Sita had seen Ram.'

           b.   ram=ko   **kɪs=ne**   dek^h-a   t^h-a?
                Ram.M=Acc who.Obl=Erg see-Perf.M.Sg be.Past-M.Sg
                'Who saw Ram?'

Focus and Questions

- Féry et al. (2016) conducted a comparative study of Hindi and Indian English.
- They asked questions like:
    - *In front of the well, who is pushing the car?* (Questioning the Subject)
    - *In front of the well, what is the man pushing?* (Questioning the Object)
- They found the following word orders in the responses.

	SOV	OSV
Subject Questioned (n=28)	6	22
Object Questioned (n=26)	26	–

$\implies$ Default information focus position is immediately preverbal.

Focus and Questions

From an Alternative Semantics (Rooth 2016) perspective this is expected:

- wh-constituent questions open up a set of alternatives to be considered (and answered), just like focus
- so the focus position is the natural position for wh-constituents

Other Word Orders?

- The literature has to date mostly concentrated on the following orders.
- (7) is generally analyzed as an echo-question (Mahajan 1997, Bhatt and Dayal 2007, Manetta 2012)

  (7) sita=ne   d^{h}yan=se   dek^h-a   t^h-a   **kıs=ko**?
       Sita.F=Erg carefully   see-Perf.M.Sg be.Past-M.Sg who.Obl=Acc
       'Sita had looked at carefully at who?'   (wh postverbal)

  (8) sita=ne   d^{h}yan=se   [dek^h-a   **kıs=ko**   t^h-a?]
       Sita.F=Erg carefully   see-Perf.M.Sg who.Obl=Acc be.Past-M.Sg
       'Who had Sita (really) looked at carefully?'   (wh in verbal complex)

- (8) has been analyzed as an instance of secondary focusing of the wh-word in the sense of Büring (2015) by Butt et al. (2016).
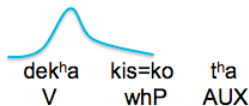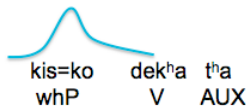
Verb focus and Secondary Focus

- The verb is the primary focus in (8), but the wh-constituent is also contextually prominent and placed within the domain of the primary focus.

  (9)   sita=ne   d^{h}yan=se [dek^h-a   **kɪs=ko**   t^h-a?]
              Sita.F=Erg carefully   see-Perf.M.Sg who.Obl=Acc be.Past-M.Sg
              'Who had Sita (really) looked at carefully?'   (wh in verbal complex)

- Prosodic phrases generally have a L*+H pattern.
- Focus is marked via a larger pitch excursion.
- wh-words generally have an H*
- In the postverbal position, this H* is not realized on the wh-word.
- The verb is clearly in focus with an H*.

kis=ko   dek^{h}a t^{h}a   dek^{h}a   kis=ko   t^{h}a
whP   V   AUX   V   whP   AUX

Other Word Orders?

- But what about all the other word orders that are in principle possible?
- Not easy to tease out what is going on (secondary focus analysis involved heavy lifting in terms of syntax+prosody+pragmatics).
- To get a further sense of word order variation and use — study of Bollywood scripts (with attendant movies).
- This talk: first look at overall data patterns.

Bollywood Scripts

We have machine readable data (and the movies) for the following scripts:

1. Ankhon Dekhi (2014)
2. Dedh Ishqiya (2014)
3. Dum Laga Ke Haisha (2015)
4. Jab We Met (2007)
5. Lootera (2013)
6. Masaan (2015)
7. NH10 (2015)
8. Queen (2014)
9. Socha Na Tha (2005)
10. Talvar (2015)
11. Titli (2014)
12. Udaan (2010)

Wh-Questions in Bollywood Scripts

- Wh-words (kwords) were extracted from these 12 scripts.
- This talk evaluates a subset of the results:
  - Core arguments (with the exception of *kya* 'what')
  - Adjuncts: 'where' and 'when'

Wh-Questions in Bollywood Scripts

Word Order Distribution in the 12 Bollywood Scripts
(subset of overall wh-words)

Distribution	Core Arguments (without *kya*)	Adjuncts ('where', 'when')	Total
Single Word	28	14	42
Initial	9	10	19
Medial	2	12	14
Preverbal	118	209	327
In Verbal Complex	0	5	5
Postverbal/Final	6	7	13
Embedded	12	17	29
No Verb	14	5	19
Total	189	279	468

Some Analysis

- The overwhelming number of cases is found in the default immediately preverbal focus position.

    (10) or ye joker **kon** hai   bike pe?
          and this joker who be.Pres.3.Sg bike on
          'And who is this joker on the bike?'   (Jab We Met)

- Wh-word within the verbal complex indeed instantiates verb focus (secondary focus of wh-word)

    (11) ZAMINDER: **rɑh** kahã̄ rahe   ho tʋm?
          LANDOWNER: stay where Prog.2 you
          'Where are you staying*foc*?   (Lootera)

Some Analysis

- None of the clause final examples of wh-words instantiates an echo question.

  (12) a. ZAMINDAR: manikpur=mē bʰi bɪjli   a   gɑ-yi
            LANDLORD: Manikpur=in also electricity.F come go-Perf.F.Sg
            'Electricity has come even to Manikpur.'
       b. PAKHI: or **ruk-e-g-i**   kɑb=tɑk?
            PAKHI: and stay-3.Sg-Fut-F.Sg when=til
            'And how long will it stay?'   (Lootera)

- Rather, they seem to be instances of verb focus with secondary focus on the wh-word as well.

Some Analysis

- The wh-words are found only rarely in medial position.
- The adjuncts are found in medial position more readily than the arguments.
- All of the examples with *kahã* 'where' in medial position involve a non-literal meaning.
- They do not express questions.

(13) a. na ɑbʰi **kahã** naʃta   hu-a
          no now where breakfast.M become-Perf.M.Sg
          'No breakfast hasn't been able to happen yet.'   (Dum Laga Ke Haisha)

     b. hʊm **kahã** ʊn=ke   standard=ke   hãĩ!
         we   where they.Obl=Gen.Obl standard=Gen.Obl be.Pres.3.Pl
         'We are not of their standard (class)!'   (Dum Laga Ke Haisha)

Some Analysis

- The corpus has yielded instances of what has been dubbed as *Repeat Questions* (Disselkamp 2017)
- Repeat questions involve situations where an interlocutor is trying to get an answer to a question and has to repeat it several times.
- Disselkamp observes for German that rather than repeating exactly the same question, interlocutors vary the surface form.

   (14)   a.   Wo   ist die Katze?
                    where is the cat
                    'Where is the cat?'

            b.   Wo   die Katze ist ./?
                    where the cat   is
                    'Where may the cat be?'

Some Analysis

- Repeat Questions can also be observed in Hindi/Urdu.
- The word order is varied.
- The following are uttered in a context where friends are getting into a car, have trouble starting it and are simultaneously trying to ensure everybody is there.

(15) a. rohan   kahã hai?
         Rohan.M where be.Pres.3.Sg
         'Where is Rohan?'   (Udaan)

     b. ...

     c. kahã hai   rohan?
         where be.Pres.3.Sg Rohan.M
         'Where is Rohan?'   (Udaan)

Summary and Outlook

- Hindi/Urdu indeed uses word order variation for information-structure packaging.
- More generally, uses word order variation for Common Ground (CG) Management in the sense of Krifka (2008)
  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.
- Explaining word order variation solely in terms of syntax-internal features or constraints is a dead end.

Summary and Outlook

- Need an architecture that seemlessly integrates morphosyntactic and prosodic information so that the semantics and pragmatics of (non-canonical) questions can be calculated compositionally:

  $\Longrightarrow$ LFG (Butt et al. 2016, 2017).

Acknowledgements

Thanks!

Very many thanks go to Rajesh Bhatt and Veneeta Dayal for ever interesting discussison on Hindi/Urdu and to Ghulam Raza for data help.

References I

Bayer, Josef and Lisa Lai-Shen Cheng. 2015. Wh-in-Situ. In M. Evaraert and H. van Riemsdijk, eds., *The Blackwell Companion to Syntax*. Oxford: Blackwell Publishing.

Bhatt, Rajesh. 2003. Topics in the syntax of the modern Indo-Aryan languages — *wh*-in-situ and *wh*-movement. Class Handout, UMASS Amherst.

Bhatt, Rajesh and Veneeta Dayal. 2007. Rightward scrambling as rightward movement. *Linguistic Inquiry* 38(2):287–301.

Bhatt, Rajesh and Veneeta Dayal. 2014. Polar *kyaa*. Talk presented at the *Workshop on Non-Canonical Questions and Interface Issues*, Konstanz, February 2014.

Biezma, María, Miriam Butt, and Farhat Jabeen. 2018. Polar Questions vs. *Kya*-Questions in Hindi/Urdu. Talk presented at the GLOW 41 Semantics Workshop *The grammar and pragmatics of interrogatives and their (special) uses*, Budapest, April.

Büring, Daniel. 2015. A theory of second occurrence focus. *Language, Cognition and Neuroscience* 30(1–2):73–87.

Butt, Miriam, Tina Bögel, and Farhat Jabeen. 2017. Polar *kya* and the Prosody-Syntax-Pragmatics Interface. In M. Butt and T. H. King, eds., *Proceedings of the LFG'17 Conference*, pages 125–145. CSLI Publications.

Butt, Miriam, Farhat Jabeen, and Tina Bögel. 2016. Verb Cluster Internal Wh-Phrases in Urdu: Prosody, Syntax and Semantics/Pragmatics. *Linguistic Analysis* 40(3–4).

Butt, Miriam and Tracy H. King. 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 Tracy Holloway King. 1997. Null elements in discourse structure. Written to be part of a volume that never materialized.

Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics and point of view. In C. Li, ed., *Subject and Topic*, pages 27–55. New York: Academic Press.

Disselkamp, Gisela. 2017. The Pragmatics of Speaker Repeat Questions. Poster presented at *ESSLLI* 2017.

References II

Féry, Caroline, Pramod Pandey, and Gerrit Kentner. 2016. The prosody of focus and givenness in Hindi and Indian English. *Studies in Language* 40(2):302–339.

Gambhir, Vijay. 1981. *Syntactic Restrictions and Discourse Functions of Word Order in Standard Hindi*. Ph.D. thesis, University of Pennsylvania, Philadelphia.

Kidwai, Ayesha. 2000. *XP-Adjunction in Universal Grammar: Scrambling and Binding in Hindi-Urdu*. Oxford: Oxford University Press.

Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica* 55(3–4):243–276.

Mahajan, Anoop. 1997. Rightward scrambling. In D. Beerman, D. LeBlanc, and H. van Riemsdijk, eds., *Righward Movement*, pages 185–213. Amsterdam: John Benjamins.

Manetta, Emily. 2012. Reconsidering rightward scrambling: Postverbal constituents in Hindi-Urdu. *Linguistic Inquiry* 43(1):43–74.

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.

Rooth, Mats. 2016. Alternative semantics. In C. Féry and S. Ishihara, eds., *Oxford Handbook of Information Structure*. Oxford: Oxford University Press.

Vallduví, Enric. 1992. *The Informational Component*. New York: Garland Press.