Rightward scrambling is tough: CP complements and rightward scrambling in a Hindi-Urdu Treebank

“8.3 ‘Rightward’ scrambling is tough.”
(Bhatt, Farhudi, and Rambow 2011: 48)

1. Introduction
I want to be able to examine naturally occurring examples of constituents that appear to the right of the verb in verb-final Hindi-Urdu.

This includes finite complement CPs, which MUST appear to the right of the verb, as in (1), and other optionally “scrambled” constituents, as in (2).

(1) a. Sita-ne kahaa thaa ki Mohan aayaa thaa. [Subj V Aux CP]
   Sita-ERG said AUX that Mohan come AUX
   ‘Sita said that Mohan had come.’
   b. *Sita-ne [CP ki Mohan aayaa thaa] kahaa thaa *[Subj CP V Aux]

(2) Sita-ne Mohan-ko dikhaa-ii ek kitaab [Subj IO V DO]
   Sita-ERG Mohan-DAT show-PFV a book
   ‘Sita showed a book to Mohan’

Goals of the talk:
- Provide a breakdown of the range of postverbal material in Hindi-Urdu
- Examine the ways in which the Treebank currently handles postverbal material and displacement to the right edge
- Investigate how this interacts with theoretical proposals in the literature concerned with finite complement CPs and rightward scrambling
- Explore ways to fine-tune the Treebank’s representation of postverbal material to make it an optimal tool for investigating sentences like (1)-(2)

2. Finite Complement CPs
Observation: Finite complement CPs appear to the right of the finite verb in (head-final) Hindi-Urdu while all other non-clausal complements of the verb appear to the left.

(3) Sita-ne kahaa thaa ki Mohan aayaa thaa. [Subj V Aux CP]
   Sita-ERG said AUX that Mohan come AUX
   ‘Sita said that Mohan had come.’
(4) Sita-ne Mohan-ko ek kitaab dikhaa-ii [Subj IO DO V ]
   Sita-ERG Mohan-DAT a book show-PFV
   ‘Sita showed a book to Mohan’
Important questions:
- Are these clauses in fact complements of the selecting verb? Or something else?
- If they are complements, do they originate in the preverbal position?
- If so, how do they get to the postverbal position? By what kind of displacement? In what component of the grammar?
- How are complement clauses related to other finite clauses that may appear postverbally or other non-clausal constituents (“rightward scrambled” phrases)?

I will ultimately argue that because their movement motivations and the syntactic characteristics of movement are quite different, traces of postverbal verb-complement CPs and other postverbal material (rightward scrambled phrases) should be differentiated in the Treebank.

The current state of things in the Treebank:
At the level of Dependency Structure (DS)
Complement clauses are vakya karma – sentential objects. The arc is tagged k2, meaning it is being treated like the complement of the verb at this level.

(5) Relation-DS-k2: us-ne kah-aa ki Ram kal nahiN ay-egaa
he-ERG tell-PERF that ram tomorrow NEG come-FUT
‘He told that Ram will not come tomorrow.’
(Barati et al 2009 4.1.4.C)

What will happen at Phrase Structure (PS)?
The complement position (at which complement clauses are assumed to originate) will be marked with a trace, marking syntactic displacement. The CP complement will then be adjoined at some higher position on the right edge.

The trace, *EXTR* for “extraposition” is shared by all phrases undergoing rightward movement (so for both movements in (1)-(2)) (Bhatia et al, 2010)

(6)
3. Complexities
   • CP complement syntax
   • What about co-occurrence with a pronominal associate *yeh*?
   • What about extraposed clausal complements of nouns and relative clauses?
   • What about null complement anaphora (NCA)?

3.1 The syntax
   For all intents and purposes, finite complement CPs behave as though they were in their base position, though it doesn’t look it.

   Well-known binding and NPI licensing facts:
   (8) a. Har aadmii-ne, soc-aa ki us-ne, Siita-ko dekhaa.
       each man-**ERG** think-**PFV** that he-**ERG** Sita-**ACC** see-**PFV**
       Every man thought that he saw Sita.’
   b. Siita-ne *(nahiN) kah-aa ki koi bhii aay-aa
       Sita-**ERG** NEG say-**PFV** that someone even come-**PFV**
       ‘Sita did not say that anybody came.’

   Two ways to understand this: (1) the CP has never moved from its base position (either an exceptional or non-exceptional right-hand complement), or (2) the CP is reconstructed to its base position for interpretation.

   Does this displacement take place in the narrow syntax? Following the line of argumentation in McCloskey (2000) and Bennet, Dowd, Elfner, and McCloskey (2011) for Irish, the core problem with a syntactic movement account is that we are left with no way to explain:
   • the positioning facts of finite complement CPs and
   • the lack of interpretive effects arising from their (obligatory) displacement.
Positioning Facts:
First, while other constituents generated in direct object position are free to remain in situ or move both left and right, CPs are restricted to the clause-peripheral position.

(9) a. Sita-ne Arjun-se bolaa thaa ki Mohan aayaa thaa.  [DP DP V Aux CP]
   Sita-ERG Arjun-INST said AUX that Mohan come AUX
   ‘Sita said to Arjun that Mohan had come.’
   b. *Sita-ne Arjun-se [cp ki Mohan aayaa thaa] bolaa thaa *[DP DP CP V Aux]
   c. *Sita-ne [cp ki Mohan aayaa thaa] Arjun-se bolaa thaa *[DP CP DP V Aux]

Second, finite complement CPs must appear rightmost in their containing clause. They cannot intervene between the verb and a postverbal DP, nor can they intervene between the verb and a postverbal CP that is not a verbal complement:

(10) a. *V<sub>CP</sub> CP Aux
   b. *V<sub>CP</sub> CP DP

   Raam-ERG that man-DAT said [that Sita left] [REL came AUX]
   ‘Ram told that man who had come that Sita had left.’
   b. Raam-ne us aadmii-ko kahaa [jo aayaa thaa] [ki siitaa gayii] S-IO-V-RC-CP
   Raam-ERG that man-DAT said [REL came AUX] [that Sita left]
   ‘Ram told that man who had come that Sita had left.’
   (Mahajan 1997b: 115, judgments my informants)

For these and other reasons, I have proposed a PF alignment account for finite complement CPs (Manetta, forthcoming).

3.2 Co-occurrence with yeh
Finite CP complements can co-occur with the pronominal associate yeh

(12) us-ne yeh kah-aa ki Sita der-se ay-egii
   he-ERG EXPL said-PRF that Sita delay-with come-FUT
   ‘He said that Sita will come late.’

The DS treats this structure very differently from the way it treats regular clausal complements. This is a relation *samanadhikaran* “noun elaboration” (DS-rs). *Yeh* is the complement (k2) of the verb, and then *ki* has the relation rs to the pronoun.

(13) \[ kahaa \]
    \[ usne \]
    \[ yeh \]
    \[ rs \]
    \[ ki \]
    \[ ccof \]
    \[ ayegii… \]
This is of course also the way in which clausal complements to nouns look at DS.

(14) maiN yeh baat ki Siita der-se ay-egii jaan-ti huN
    I.NOM this fact that Sita delay-with come-FUT know-HAB AUX
    ‘I know this fact that Sita will be late.’

(15) jaanti huN
    k1
    main yeh baat
    rs
    ki
    ccof
    ayegii…

But of course, these are only optionally extraposed, unlike true verbal complement CPs, which MUST be extraposed **whether the pronominal associate is present or not**.

(16) a. us-ne yeh kah-aa [ki Siita der-se ay-egii]
    he-ERG EXPL said-HAB that Sita delay-with come-FUT
    ‘He said that Sita will come late.’
    b. * us-ne yeh [ki Siita der-se ay-egii ] kah-aa
    he-ERG EXPL that Sita delay-with come-FUT said-HAB
    ‘He said that Sita will come late.’

(17) a. maiN yeh baat [ki Siita der-se ay-egii ] jaan-ti huN
    I.NOM this fact that Sita delay-with come-FUT know-HAB AUX
    ‘I know this fact that Sita will be late.’
    b. maiN yeh baat jaan-ti huN [ki Siita der-se ay-egii ]
    I.NOM this fact know-HAB AUX that Sita delay-with come-FUT
    ‘I know this fact that Sita will be late.’

Do we want these to have the same underlying structure or should they be somehow differentiated? Is *yeh* really the complement of the verb in (16), or is the clause? If it is *yeh*, how can we understand the fact that (16) and (17) behave differently?

The assumption that some pronominal/null element is the true complement of the verb and that the clause is adjoined is essential to a major strand of analysis of Hindi-Urdu subordinate clauses (Dayal 1994, 1996, Lahiri 2002).

3.3 Relative clauses
Extraposition of postnominal relative clauses in Hindi-Urdu is also completely optional.

(18) [vo kitaab [jo sale-par he]] achchii he
    that book REL sale-on is good is
    ‘That book that is on sale is good.’
We might want to examine the same kinds of potential conditioning factors for extraposition as does Strunk 2010 for relative clauses in the German Treebank: indefiniteness of the head noun, length of the relative clause, embeddedness, and restrictiveness.

There are obviously parameters governing the potential for displacement of noun complement CPs and relative clause CPs to the right edge that are irrelevant for verbal complements (since they must obligatorily appear there).

3.4 Null complement anaphora

(20) A: Raam aaj der-se aa-egaa.
    ‘Ram today delay-with come-FUT
    ‘Ram will come late today.’

B: mujhe pataa he me.DAT known is
    ‘I know.’

Bhatt and Farudi (2010) suggest that NCA might be analogous to the pronominal associate cases – that is, it is a silent pronominal element, not a silent clausal one.

(21)

Of course, this analysis assumes that when yeh and the clausal complement are co-present, that it is yeh that is the complement of the verb, not the clause itself. If we move away from this analysis, then we might want the elided material in (20) to be a clausal proform.

3.5 A proposal

When yeh is serving as a pronominal associate, treat it as a clausal expletive. Here is how the Penn Treebank handles “it-extraposition” in English:
Extraposed sentences with `it'

- *It* is recognised as surface subject;
- Extraposed clause is attached at S level with *EXP* attach;
- Extraposed clause is interpreted as a subject of a pleasure (*it* ignored for argument interpretation).

(S (NP-SBJ (NP It))
 (S *EXP*-1))
 (VP is
 (NP a pleasure))
 (S-1 (NP-SBJ *)
 (VP to
 (VP teach
 (NP her)))))

pleasure(teach(*someone*, her))

This would give us a nice contrast between (16) and (17):

(22) DS

This would give us a nice contrast between (16) and (17):

(22) DS

This would give us a nice contrast between (16) and (17):

(22) DS

This would give us a nice contrast between (16) and (17):

(22) DS
4. Complement CPs and rightward scrambling in syntactic theory

Claims made about complement CPs:
1) Generated in complement position and extraposed in the syntax (Mahajan 1988)
2) Never generated in complement position, adjoined to CP/IP and coindexed with a DP (Dayal 1994, 1996)
3) In its low base position (SVO/antisymmetric approaches) (Mahajan 1997)
4) Generated as a complement to DP (empty or pronounced), then extraposed (Lahiri 2002)
5) Generated in complement position, part of a rightward moved remnant-VP (Bhatt and Dayal 2007)
6) Generated in complement position and aligned rightward at PF (Manetta forthcoming)

• Only the first is consistent with the Treebank representation.
• The second brings up the issue of how to understand semantic coindexation not tied to movement in the Treebank. The DS would need to be very different to accommodate this view.
• The fourth is much along the lines of the approach to NCA and the current DS approach (assume a DP complement to V whenever possible). Both the second and the fourth are seriously challenged by the contrast in (16)-(17).
• The fifth would mean a very elaborate change to the PS, and also has important implications for rightward scrambling, below.
• The sixth brings up the issue of PF movement for a syntactic Treebank – would this mean that the final representation in the Treebank would have the CP in situ (preverbal position), and would not reflect final surface word order? This is clearly problematic.

The diversity of approaches is indicative of how challenging these CPs are to account for. For Treebanking purposes, we want to focus on how to best extract information about CPs of all types.

5. Rightward scrambling
Other constituents besides CPs can optionally appear postverbally:

(24) Siitaa-ne Raam-ko dekh-aa
    Sita-ERG Ram-ACC see-PERF
    ‘Sita saw Ram.’
(25) Siitaa-ne dekhaa Raam-ko
(26) Raam-ko dekhaa Siitaa-ne
(27) dekhaa Siitaa-ne Raam-ko
(28) dekhaa Raam-ko Siitaa-ne

SVO
OVS
VSO
VOS

These seem to be discourse old things of a variety of types (e.g. recurring/resumed topic) (Karimi 2005), or have a contrastive function, or may even be in the partially ordered set (poset) relation (a la Prince 1998) with other entities in the discourse, but we don’t have nearly as complete a picture for these as we do for leftward scrambling (see Gambhir 1981, Butt and King 1996).

We might want to know more about the discourse significance of rightward scrambling, motivations, etc, that the Treebank could potentially tell us.
• What kinds of arguments usually scramble rightward?
• With what relative frequency?
• What is their role in the discourse?
• What other constituents scramble rightward?
• What is the interpretive significance of appearing postverbally?

5.1 Leftward scrambling in the Treebank
How might this be different from leftward scrambling?
(29) Atif-ne yeh film tiin baar dekh-ii
    Atif-ERG this film (f) three times see-PERF.F
    ‘Atif saw this film three times.’
Of course, at DS the arguments are represented in their usual relationship to the verb. At PS, the trace of a scrambled constituent is labeled *SCR* (Bhatt, Farudi, and Rambow 2011)

(30)

```
NP-P       VP-Pred
    NP   P    NP_1  VP-Pred
      Atif  ne   Dem  Num  NP  VP-Pred
            yeh  N  film  NP  V  *SCR* 1  dekhii
```

5.2 How would the Treebank currently handle rightward scrambling?

In the DS layer, the argument would be in its usual relation. In the PS layer, it would be represented by the trace for rightward movement, *EXTR* and would appear adjoined postverbally just like for CPs (Bhatia et al 2010)

(31)

```
NPVP
  NPVP-Pred  N
    NP  V  Ram-ko
      Sitane  *EXTR* 1  dekhaa
```

5.3 Is this the ideal representation?

Rightward-scrambled constituents are neither completely like postverbal CPs nor like leftward-scrambled constituents:
The facts about the syntax of rightward scrambled DPs: they behave for most purposes as though in their pre-rightward position EXCEPT wrt the interpretation of wh-XP. This has been understood to mean that they are obligatorily reconstructed (Bhatt and Dayal 2007, Manetta, forthcoming).

Problematic conflations:

• Both the Treebank, as well as some theoretical approaches, conflate postverbal CPs with rightward scrambled constituents.
• Searching the Treebank for *EXTR* will yield all types of postverbal material, even though they are syntactically quite distinct.
• On the other hand, it would obviously be inappropriate to group leftward and rightward scrambling, having rightward scrambling also leave *SCR*. These are also clearly different syntactic animals (for Korean: Lee 1993, Rambow and Lee 1994, Ko 2007, Ko and Choi 2008, 2009).

Data for reference

Condition C:

(32) a. *siitaa-­ne usei tumhaaraa raam-­1­ko likhaa hua khat dikhaayaa sita-­ERG him you.­GEN raam-­DAT written AUX letter showed S-­IO-­DO-­V
   b. *siitaa-­ne usei dikhaayaa tumhaaraa raam-­1­ko likhaa hua khat sita-­ERG him showed you.­GEN raam-­DAT written AUX letter S-­IO-­V­-­DO
   c. siitaa-­ne tumhaaraa raam-­1­ko likhaa hua khat usei dikhaayaa sita-­ERG you.­GEN raam-­DAT written AUX letter him showed S-­DO-­IO-­V

‘Sita showed a letter written by you to Ram, to him.’

(Mahajan 1997:195,198)

Reciprocal Binding:

(33) a. ??? [ek duusre-­kei baccô]-­ne [Anu aur Ramaa]-­koi dekhaa. SOV each.­other-­GEN.OBL kids-­ERG Anu and Ramaa-­ACC saw
   b. ??? [ek duusre-­kei baccô]-­ne dekhaa [Anu aur Ramaa]-­koi i each.­other-­GEN.OBL kids-­ERG saw Anu and Ramaa-­ACC SVO
   c. [Anu aur Ramaa]-­koi [ek duusre-­kei baccô]-­ne dekhaa. Anu and Ramaa-­ACC each.­other-­GEN.OBL kids-­ERG saw OSV

‘[Each other’s] kids saw [Anu and Ramaa].’

(Bhatt and Dayal 2007:289)
Wh-Echo reading:

(34) Sita-ne dhyaan-se dekh-aa thaa kis-ko? [S V Aux Wh DP]

Sita-ERG care-with see-PFV be.PST who-ACC
≠ ‘Who had Sita looked at carefully?’ (wh-question reading)
= ‘Sita looked carefully at WHO?’ (echo reading)

6. Another proposal

Postverbal non-CPs should not be represented as *EXTR*, which should be reserved for CPs, but instead a new type of *SCR* trace – perhaps *RSCR*

(35)

This proposal allows us to separately pick out just the class of:
- rightward scrambled non-CP-complement constituents,
- or just the class of postverbal complement CPs,
- or just the class of leftward scrambled constituents,

depending on our needs. It also better reflects very real syntactic differences between the types of postverbal material.

What may yet need a more “fine-grained” approach are the different types of rightward moved CPs, more on this below.

7. How do other treebanks represent rightward scrambling/movement?

This doesn’t turn out to be very informative. Many of those Treebanks of so-called “free word order” languages (Turkish, Persian, German) don’t make use of traces to represent displacement. Other Treebanks have a relatively impoverished set of traces.

- PennParsed Corpora of Historical English: *T* for wh-movement, *ICH* (interpret constituent here) for all other A-bar movement including rightward extraposition
- Penn Treebank (Switchboard corpus): right dislocation – no conindexation, dislocated element receives dash tag TPC; all others (appositives) pseudoattached with *ICH*
- TiGer: dependent constituents attached to governing node, then branches are crossed to indicate displacement
- TüBa-D/G: discontinuous element is labeled with its role in the main clause (i.e. what it modifies)
- Penn Arabic Treebank: Pro-drop subjects and passive traces , *T* WH-traces, NP-TPC trace to subject, *ICH* Rightward movement , *RNR*
• Penn Korean Treebank: leftward scrambling -- *T*
• Persian/Turkish treebanks lack traces (like TiGer)

The Hindi-Urdu Treebank is already relatively rich in empty category types – why not enrich this slightly further.

Table 2 (adapted from Bhatia et al 2010)

<table>
<thead>
<tr>
<th>Description</th>
<th>Type</th>
<th>Where Inserted</th>
<th>Label</th>
<th>Coindexed at PS?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trace of phrase undergoing extraposition</td>
<td>trace</td>
<td>PS</td>
<td><em>EXTR</em></td>
<td>movement</td>
</tr>
<tr>
<td>Trace of phrase moving to case position in verbal projection</td>
<td>trace</td>
<td>PS</td>
<td><em>CASE</em></td>
<td>movement</td>
</tr>
<tr>
<td>Trace of moved relative pronoun</td>
<td>trace</td>
<td>PS</td>
<td><em>RELTR</em></td>
<td>movement</td>
</tr>
<tr>
<td>Trace of scrambling (leftward movement)</td>
<td>trace</td>
<td>PS</td>
<td><em>SCR</em></td>
<td>movement</td>
</tr>
<tr>
<td>Trace of head moving to incorporate</td>
<td>trace</td>
<td>PS</td>
<td><em>HEAD</em></td>
<td>movement</td>
</tr>
<tr>
<td>Trace of scrambling (rightward movement)</td>
<td>trace</td>
<td>PS</td>
<td><em>RSCR</em></td>
<td>movement</td>
</tr>
</tbody>
</table>

8. The far right
We are left with this (at least) four-way typology of post-verbal elements:

Table 3

<table>
<thead>
<tr>
<th>Type</th>
<th>Displacement is...</th>
<th>Trace in Treebank</th>
<th>Proposed Trace in Treebank</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP - verb complement</td>
<td>obligatory</td>
<td><em>EXTR</em></td>
<td><em>EXTR</em></td>
</tr>
<tr>
<td>CP – noun complement</td>
<td>optional</td>
<td><em>EXTR</em></td>
<td><em>EXTR</em></td>
</tr>
<tr>
<td>CP – relative clause</td>
<td>optional</td>
<td><em>EXTR</em></td>
<td><em>EXTR</em></td>
</tr>
<tr>
<td>XP - scrambled</td>
<td>optional</td>
<td><em>EXTR</em></td>
<td><em>RSCR</em></td>
</tr>
</tbody>
</table>

• What remains to be better understood here is the contrast between the clauses that optionally extrapose and the clauses that obligatorily do so.
• It is certainly possible that the factors governing optional clause extraposition are very different from the factors governing the optional rightward scrambling of other, non-clausal constituents.
• Obviously, we might be interested in the range of factors governing optional rightward displacement, none of which apply in the case of CPs that are the complements of verbs.
References

Bennett, Ryan, Andrew Dowd, Emily Elfner, and Jim McCloskey. 2010. Lightest to the Right: An Anomalous Displacement in Irish. ms. University of California, Santa Cruz.


Ko, Heejeong, and Jae Young Choi. 2009. Rightward Movement and Output Economy.


