Discontinuous Genitives in Hindi/Urdu

Sebastian Sulger

Department of Linguistics
University of Konstanz

Workshop on Discontinuous Structures in Natural Language Processing
NAACL HLT 2016
June 17th, 2016
Some challenges for parsing posed by discontinuities

Practical challenges:
- Adapt the parser to be able to process discontinuous structures
- Reconstruct dependencies in the analysis, i.e., attach discontinuous parts to their syntactic heads

Theoretical challenge:
- Derive generalizations about what kinds of discontinuities are possible, and what kinds do not appear
- Find out what drives discontinuity; i.e., why it can or cannot take place across languages
Goal of this paper

This paper aims at contributing to all three of the above challenges.

- Discuss empirical properties of a specific case of discontinuity, observed with the Hindi/Urdu genitive
- Investigate data gathered from native speakers as well as treebank data [Bhatt et al., 2009]
- Formulate linguistic generalizations and implement using computational grammar:
  - Hindi/Urdu Pagram grammar [Butt and King, 2007, Bögel et al., 2009]
  - Use LFG as theoretical backdrop [Dalrymple, 2001, Bresnan, 2001]
  - Use XLE as parser [Crouch et al., 2015]
Outline

1. Introduction
2. Data
3. Constraints on genitive scrambling
4. LFG/XLE implementation
5. Conclusion
General description

- Hindi/Urdu: free-word order (non-configurational) language, canonical order SOV
- Genitive arguments: most common type of nominal argument (other: locative, instrumental)
- All nominal types may be modified by genitives (exception: pronouns)
- Morphosyntactic behavior:
  - Case clitic heading a case phrase (KP) [Butt and King, 2004]
  - Agreement with head noun (number, gender, morphological inflection)
  - Genitive KP can occur in various positions inside NPs
  - Multiple instantiations per NP are possible [Raza, 2010]
- Syncretism:
  - For the feminine: a single form $ki$ is used throughout the inflectional pattern
  - For the masculine: a single form $ke$ between the singular oblique and plural nominative and oblique
Some simple examples I

Masculine pattern:

1. a. ram=ka makan
   Ram.M.SG=GEN.M.SG house.M.SG
   ‘Ram’s house’

2. b. * ram=ki makan
   Ram.M.SG=GEN.F.SG/PL house.M.SG

3. c. * ram=ke makan
   Ram.M.SG=GEN.M.PL house.M.SG
Some simple examples II

Feminine pattern:

(2) a. nina=ki  beți
   Nina.F.SG=GEN.F.SG daughter.F.SG
   ‘Ram’s daughter’

b. * nina=ka  beți
   Nina.F.SG=GEN.M.SG daughter.F.SG

c. * nina=ke  beți
   Nina.F.SG=GEN.M.PL daughter.F.SG
Some simple examples III

Plural masculine pattern:

(3)  a.  nadya=ke  beṭe
    Nadya.F.SG=GEN.M.PL  son.M.PL
    ‘Nadya’s sons’

b.  * nadya=ka  beṭe
    Nadya.F.SG=GEN.M.SG  son.M.PL

c.  * nadya=ki  beṭe
    Nadya.F.SG=GEN.F.SG/PL  son.M.PL
Some simple examples IV

Plural feminine pattern:

(4) a. nadya=ki               bili
    Nadya.F.PL=GEN.F.PL  cat.F.PL
    ‘Nadya’s cats’

b. * nadya=ka               bili
    Nadya.F.SG=GEN.M.SG  cat.F.PL

c. * nadya=ke               bili
    Nadya.F.SG=GEN.F.SG/PL  cat.F.PL
Word order within NP

The word order within the Hindi/Urdu NP is relatively free.

- The canonical word order is: possessor first, possessum last
- The position of the genitive phrase varies with respect to other NP modifiers, such as adjectives or quantifiers
- NP modifiers occurring after the NP head are judged as ungrammatical by the informants
Illustrating word order I

(5)  

a. ram=ki  nili  gaṛi  
Ram.M.SG=GEN.F.SG  blue.F.SG  car.F.SG  
‘Ram’s blue car’

b. nili  ram=ki  gaṛi  
blue.F.SG  Ram.M.SG=GEN.F.SG  car.F.SG  
‘Ram’s blue car’

c. * nili  gaṛi  ram=ki  
blue.F.SG  car.F.SG  Ram.M.SG=GEN.F.SG
Illustrating word order II

\[(6)\]  

(a) \(\text{USTAD}=\text{KA} \quad \text{kuch hofyar talib-ilm}\)  
\(\text{teacher.M.SG=GEN.M.SG some smart student.M.PL}\)  
‘some smart students of the teacher’

(b) \(\text{USTAD}=\text{KA} \quad \text{hofyar kuch talib-ilm}\)

(c) \(\text{kuch USTAD}=\text{KA} \quad \text{hofyar talib-ilm}\)

(d) \(\text{kuch hofyar USTAD}=\text{KA} \quad \text{talib-ilm}\)

(e) \(\text{hofyar kuch USTAD}=\text{KA} \quad \text{talib-ilm}\)

(f) \(\text{hofyar USTAD}=\text{KA} \quad \text{kuch talib-ilm}\)
Functional ambiguity

It is assumed that Hindi/Urdu genitive modifiers may occur on a range of grammatical functions: SUBJ, OBJ, ADJUNCT

Evidence:

- Binding of reflexive pronoun
- Iterativity/optionality

(7) a. ram=ki tippani
   Ram.M.SG=GEN.F.SG comment.F.SG
   ‘Ram’s comment/criticism’  
   SUBJ

b. gaṛi=ki tabahi
   car.F.SG=GEN.F.SG destruction.F.SG
   ‘the car’s destruction’  
   OBJ

c. surx rang=ki mez
   red color.M.SG=GEN.F.SG table.F.SG
   ‘the table of red color’  
   ADJUNCT
Genitive scrambling I

- Genitive phrases may be dislocated outside of the NPs they modify (*genitive scrambling*)
- May occur to the left or to the right of head NP
- Possessor-first order inside NP is not necessarily preserved during scrambling
- [Fanselow and Féry, 2006]: leftwards dislocation *non-inverted scrambling*, rightwards dislocation *inverted scrambling*

(8) a. ram=ka
dost
ay-a
Ram.M.SG=GEN.M.SG friend.M.SG.NOM come-PERF.M.SG
‘Ram’s friend came.’ [Butt and Zinsmeister, 2009]

b. dost
ay-a
ram=ka
friend.M.SG.NOM come-PERF.M.SG Ram.M.SG=GEN.M.SG
‘Ram’s friend came.’ [Butt and Zinsmeister, 2009]
Genitive scrambling II

(9) a. ram=ne  us=ki  gaʀɿ
Ram.M.SG=ERG  PRON.3.SG.OBL=GEN.F.SG  car.F.SG.NOM
bazar=mẽ  dekʰ-i
market.M.SG=LOC.IN  see-PERF.F.SG

‘Ram saw her/his car in the market.’ (adapted from [Bögel and Butt, 2013], p. 301)

b. us=ki  ram=ne  gaɿ
PRON.3.SG.OBL=GEN.F.SG  Ram.M.SG=ERG  car.F.SG.NOM
bazar=mẽ  dekʰ-i
market.M.SG=LOC.IN  see-PERF.F.SG

‘His/her car, Ram saw in the market.’ (adapted from [Bögel and Butt, 2013], p. 301)
Genitive scrambling III

(10) a. \textit{us=ki} \textit{gaɾi} \textit{ram=ne}\[
\text{PRON.3.SG.OBL=GEN.F.SG} \text{ car.F.SG.NOM Ram.M.SG=ERG} \\
\text{bazar=mē} \quad \text{dek}^h\text{-i} \\
\text{market.M.SG=LOC.IN} \text{ see-PERF.F.SG}
\]

‘His/her car, Ram saw in the market.’ (adapted from [Bögel and Butt, 2013], p. 301)

b. \textit{gaɾi} \textit{ram=ne} \textit{us=ki}\[
\text{car.F.SG.NOM Ram.M.SG=ERG} \text{ PRON.3.SG.OBL=GEN.F.SG} \\
\text{bazar=mē} \quad \text{dek}^h\text{-i} \\
\text{market.M.SG=LOC.IN} \text{ see-PERF.F.SG}
\]

‘His/her car, Ram saw in the market.’ (adapted from [Bögel and Butt, 2013], p. 301)
Genitive scrambling — why?

- [Fanselow and Féry, 2006]: rich morphological agreement between constituents as a main factor influencing availability of discontinuous NPs across languages
- Situation as depicted here is a case in point: We are able to reconstruct dependencies by looking at phi-features of head & genitive
- Counter-example: Turkish
  - Discontinuous NPs
  - No agreement inside nominal projections
- More research needed to show why speakers choose to dislocate constituents (hunch: focus/topic configurations)
Outline

1. Introduction
2. Data
3. Constraints on genitive scrambling
4. LFG/XLE implementation
5. Conclusion
Local attachments are preferred if

- Where agreement morphology permits both scrambled as well as locally-attached genitives, local attachments highly preferred.
Local attachments are preferred II

\[(11)\]

(a) \(\text{us}=\text{ki} \quad \text{ga}\text{r}i \quad \text{nadya}=\text{ne}\)

\[
\begin{align*}
\text{PRON.3.SG.OBL}=\text{GEN.F.SG} & \quad \text{car.F.SG.NOM} \quad \text{Nadya.F.SG}=\text{ERG} \\
\text{bag}=\text{m}\text{ê} & \quad \text{dek}^h-\text{i} \\
\text{park.F.SG}=\text{LOC.IN} & \quad \text{see-PERF.F.SG}
\end{align*}
\]

‘Her/his car, Nadya saw in the park.’

(b) \(\text{gar}i \quad \text{nadya}=\text{ne} \quad \text{us}=\text{ki}\)

\[
\begin{align*}
\text{car.F.SG.NOM} & \quad \text{Nadya.F.SG}=\text{ERG} \quad \text{PRON.3.SG.OBL}=\text{GEN.F.SG} \\
\text{bag}=\text{m}\text{ê} & \quad \text{dek}^h-\text{i} \\
\text{park.F.SG}=\text{LOC.IN} & \quad \text{see-PERF.F.SG}
\end{align*}
\]

‘The car, Nadya saw in her park.’

preferred over

‘His/her car, Nadya saw in the park.’
Scrambling and case I

- Genitives may also be scrambled out of NPs that are overtly case-marked
- When they are, inverted scrambled genitives are ungrammatical
- I.e., genitives have to precede their head NPs in the clause
Scrambling and case II

(12)  a. bacco=ne  ram=ke  kal
child.M.PL.OBL=ERG  Ram.M.SG=GEN.M.SG.OBL  yesterday
kuṭṭe=ko  dek^h^-a
dog.M.SG.OBL=ACC  see-PERF.M.SG
‘The children saw Ram’s dog yesterday.’

b. * bacco=ne  kuṭṭe=ko  kal
child.M.PL.OBL=ERG  dog.M.SG.OBL=ACC  yesterday
ram=ke  dek^h^-a
Ram.M.SG=GEN.M.SG.OBL  see-PERF.M.SG
Scrambling from complement clauses I

- Genitive phrases OK to be scrambled from within non-finite complement clauses
- Ungrammatical to scramble them from within finite complement clauses
- In line with findings by [Mahajan, 1990], [Kidwai, 1999] as well as [Kidwai, 2000] (stating that scrambling of arguments from within finite complement clauses is generally not accepted, whereas scrambling from infinite complement clauses is)
Scrambling from complement clauses II

(13)  * us=ki ram=ne kah-a
      PRON.3.SG.OBL=GEN.F.SG Ram.M.SG=ERG say-PERF.M.SG
      kih [nina=ne garî dek^h-i]
      that Nina.F.SG=ERG car.F.SG.NOM see-PERF.F.SG

(14)  us=ki ram garî dek^h
      PRON.3.SG.OBL=GEN.F.SG Ram.M.SG.NOM car.F.SG.NOM see
      sak-a
      can-PERF.M.SG
      ‘His/her car, Ram could see.’
Genitives may not be scrambled out of adjuncts

Island behavior, i.e., the unavailability of constituents for movement/scrambling, is symptomatic for clausal adjuncts and is well-known throughout the literature, first discussed by [Ross, 1967]

Well-known diagnostic for distinguishing arguments from adjuncts, as discussed by, e.g., [Needham and Toivonen, 2011] in an LFG setting
No scrambling out of adjuncts II

(15)  

a. ram=ne  us=ki  
Ram.M.SG=ERG PRON.3.SG.OBL=GEN.F.SG  
bag=mē  haṭʰi  dekʰ-a  
park.F.SG=LOC.IN elephant.M.SG.NOM see-PERF.F.SG  
‘Ram saw an elephant in my park.’

b.  * us=ki ram=ne bag=mē haṭʰi dekʰ-a  
c.  * ram=ne bag=mē haṭʰi us=ki dekʰ-a
No scrambling from deep within I

- Not possible to scramble genitive phrases that are selected by nominals further down a path of grammatical functions
  - SUBJ SUBJ ADJUNCT
  - SUBJ OBJ ADJUNCT
  - OBJ SUBJ SUBJ
  - ...

- As soon as there is more than one GF along a path of GFs from the main clause to the genitive, the genitive may not be scrambled
(16)  

a. ram=ne orat=ke
   Ram. M.SG=ERG woman. F.SG=GEN. M.SG. OBL
   ḟōhār=ki gaṟi ḍekʰ-i
   husband. M.SG=GEN. F.SG car. F.SG. NOM see-PERF. F.SG
   ‘Ram saw the woman’s husband’s car.’

b. ram=ne [[[arat=ke]SUBJ ḟōhār=ki]SUBJ gaṟi]OBJ ḍekʰ-i

c. * orat=ke ram=ne ḟōhār=ki gaṟi ḍekʰ-i

d. * ram=ne ḟōhār=ki gaṟi orat=ke ḍekʰ-i
No scrambling from deep within III

(17)  a. nina=ne  sûrûx rûng=ke
      Nina.F.SG=ERG red  color.M.SG=GEN.M.SG
      mûkâna=kâ  dûrvâza  dék^h-a
      house.M.SG=GEN.M.SG  door.M.SG  see-PERF.M.SG
‘Nina saw the red house’s door.’

b. nina=ne  [[[sûrûx rûng=ke]_ADJUNCT  mûkâna=kâ]_SUBJ  dûrvâza]_OBJ
   dék^h-a

c.  * sûrûx rûng=ke nina=ne mûkâna=kâ dûrvâza dék^h-a

d.  * nina=ne mûkâna=kâ dûrvâza sûrûx rûng=ke dék^h-a
Outline

1. Introduction
2. Data
3. Constraints on genitive scrambling
4. LFG/XLE implementation
5. Conclusion
Genitive case marker

- Lexical entry for genitive case marker includes constraints (XLE notation: \( =c \)) on gender, number, morphological form
- Constraints are in the form of inside-out constraints, since genitive phrase may be embedded in either a subject, object, or adjunct
- Also makes sure that case marker only occurs on nominal f-structures (i.e., not as a verbal case) via \( NTYPE \) existential constraint

\[
(18) \quad kI \quad K * (\wedge \text{CASE}) =c \text{ gen} \\
\text{((\{SUBJ|OBJ|ADJUNCT\} \wedge) GEND)} =c \text{ fem} \\
\text{((\{SUBJ|OBJ|ADJUNCT\} \wedge) NTYPE).}
\]
Grammar rules

- Shuffle operator (,) in \textit{Nadj} template makes sure that NP modifiers occur in any order
- Annotation $! < h \hat{\phantom{h}}$ (making use of head precedence operator $< h$): currently annotated c-structure node (here: KP or AP) has to precede the c-structure node of the higher-level f-structure

\begin{equation}
\begin{align*}
a. \text{KP} & \rightarrow \text{NP} \\
& (K). \\
b. \text{NP} & \rightarrow \{\text{PRON} \\
& | \text{Nadj}\}. \\
c. \text{Nadj} = \text{KP}*: (\! \text{CASE} = \text{gen} \\
& \! < h \hat{\phantom{h}} \\
& \{\text{@SUBJ}|\text{@OBJ}|\text{@ADJUNCT}\} \\
& , \\
& \text{AP}*: \! \text{@ADJUNCT} \\
& \! < h \hat{\phantom{h}} \\
& , \\
& \text{N}.
\end{align*}
\end{equation}
Sample analysis

Figure: Hindi/Urdu NP c- and f-structures for *nina ki beći* 'Nina’s daughter'
Functional uncertainty path models all possible attachment paths of genitive modifiers

\( XCOMP \) is the function used in LFG for non-finite complement clauses (e.g., modals, English \textit{to}-infinitive clauses)

\[
(20) \quad \text{KP-SCRAMBLE-PATH} = (XCOMP) \{\text{SUBJ}|\text{OBJ}|\text{OBL}|\text{OBJ-GO}|\text{OBJ-TH}\}.
\]
Generalizing and implementing genitive scrambling II

- **KP-SCRAMBLE** template attaches genitive KPs
- Stores their head paths to local variable \(\%PATH\)
- **CASE** feature of \(\%PATH\) is either nominative (i.e., bare NP) or not (i.e., overtly case-marked); in latter case, c-structure node of genitive (\(!\)) has to precede c-structure node of head (\(\%PATH\))

\[(21)\]  

\[
\text{KP-SCRAMBLE} = \text{KP}*: (\! \text{CASE}) = \text{gen} \\
(\sim \text{KP-SCRAMBLE-PATH}) = \%PATH \\
\{ (\%PATH \text{ CASE}) = c \text{ nom} \} \\
\{ (\%PATH \text{ CASE}) \sim= \text{ nom} \} \\
\{ <h \%PATH} \} \\
\{ (\%PATH \text{ SUBJ}) = \! \} \\
\{ (\%PATH \text{ OBJ}) = \! \} \\
\{ \!$ (\%PATH \text{ ADJUNCT}) \} \\
\@(\text{OT-MARK attach}).\]
Generalizing and implementing genitive scrambling III

@(OT-MARK attach) makes sure that non-local attachment are treated as non-optimal solutions by XLE over competing local attachments.

(22) \[ KP-SCRAMBLE = KP*: (! \text{CASE}) = \text{gen} \]
\[ (\sim KP-SCRAMBLE-PATH) = \%PATH \]
\[ \{(%PATH \text{CASE}) = c \text{ nom} \]  
\[ |(%PATH \text{CASE}) \sim = \text{nom} \]  
\[ ! <h \%PATH} \]
\[ \{(%PATH \text{SUBJ}) = ! \]  
\[ |(%PATH \text{OBJ}) = ! \]  
\[ |! \$ (%PATH \text{ADJUNCT}) \]
\[ @(OT-MARK attach). \]
Testsuite creation

- Testsuite created to perform regression tests on the grammar
- Currently includes 36 examples of ungrammatical and grammatical instances of genitives & genitive scrambling
- Manually constructed in close collaboration with native speakers
- 5 treebank examples included
- All testsuite items parsed successfully (3.34 CPU secs total for 28 grammatical items)
Conclusion

Outline

1. Introduction
2. Data
3. Constraints on genitive scrambling
4. LFG/XLE implementation
5. Conclusion
Concluding remarks

- First discussion of genitive scrambling in the literature yet
- Empirical findings from Hindi/Urdu suggest that it is indeed rich morphological agreement that drives NP discontinuity
- How does the language fit in a larger typological picture?
- Future work includes comparison with Turkish
  - Scrambling data are similar
  - But: “barrier constraint” (cf. [Chomsky, 1986]), ruling out possessors directly right-adjoined to their heads
  - Constraint not present in Hindi/Urdu
- Implementation demonstrates generalizations can be modeled efficiently in LFG & XLE
Concluding remarks

Thank you!
References I


Possessive Clitics and Ezafe in Urdu.

Urdu and the Modular Architecture of ParGram.

Bresnan, J. (2001).
Lexical-Functional Syntax.
Blackwell Publishing.

The Status of Case.

Urdu in a Parallel Grammar Development Environment.
References II

ESSLLI 2009 Course on Case, Scrambling and Default Word Order.
course material.

*Barriers*.
MIT Press.

*XLE Documentation*.
Palo Alto Research Center.

*Lexical Functional Grammar*, volume 34 of *Syntax and Semantics*.

Prosodic and morphosyntactic aspects of discontinuous noun phrases: a comparative perspective.
Manuscript, University of Potsdam.

*A Computational Theory of Human Linguistic Processing: Memory Limitations and Processing Breakdown*.

Word order and focus positions in universal grammar.
References III

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

Seven principles of surface structure parsing in natural language.
*Cognition, 2:15–47.*

*The A/A-Bar Distinction and Movement Theory.*
PhD thesis, MIT.

Derived Arguments.

Analyzing the Structure of Urdu NPs with Multiple Genitives.
In *Proceedings of the Conference on Language and Technology 2010 (CLT10).* Center for Research in Urdu Language Processing (CRULP).

Ross, J. R. (1967).
*Constraints on Variables in Syntax.*
PhD thesis, MIT.
NP Modifiers have to Precede their Heads

- Constraint: NP modifiers have to precede their heads inside the NP
- Corroborated by data such as in (23)

(23) a. [[nadya=ke do beṭe]_{NP=ne}KP gaṛi=ko
Nadya.F.SG=GEN.M.PL two son.M.PL=ERG car.F.SG=ACC
cala-yi hε
drive-PERF.F.SG be.PRES.3.SG
‘Nadya’s two sons have driven the car.

b. * [do beṭe nadya=ke]_{NP=ne}KP gaṛi=ko
two son.M.PL Nadya.F.SG=GEN.M.PL=ERG car.F.SG=ACC
cala-yi hε
drive-PERF.F.SG be.PRES.3.SG
Another example

(24) \text{gaši} \quad \text{nadya=ne} \quad \text{vs=ki}
\text{car.F.SG.NOM} \quad \text{Nadya.F.SG=ERG} \quad \text{PRON.3.SG.OBL=GEN.F.SG}
\text{bag=mē} \quad \text{dek}^{h}\text{-i}
\text{park.F.SG=LOC.IN} \quad \text{see-PERF.F.SG}

‘The car, Nadya saw in her park.’

preferred over

‘His/her car, Nadya saw in the park.’
"gARI nAdiyah nE us kI bAG mEN dEkHI"

Figure: Optimal Hindi/Urdu f-structure
"gARI nAdiyah nE us kI bAG mEN dEKHI"

Figure: Non-optimal Hindi/Urdu f-structure