Discontinuous Genitives in Hindi/Urdu

Sebastian Sulger
Department of Linguistics
University of Konstanz
sebastian.sulger@uni-konstanz.de

Abstract

This paper discusses genitive phrases in Hindi/Urdu in general and puts a particular focus on genitive scrambling, a process whereby the basic order of constituents is changed. In Hindi/Urdu, genitive phrases may not only occur at different structural positions within the NP that they modify; under the right circumstances, they can also be found outside of the NP, yielding discontinuous structures. The theoretical challenge is to identify and formalize the linguistic constraints that govern genitive scrambling. Further, a successful computational treatment correctly attaches the genitive phrase to its head NP. I use a Lexical-Functional Grammar to solve both challenges, demonstrating that the constraints can be aptly formulated using a functional uncertainty path. Successful attachment further depends on the morphological agreement of the genitive phrase with its head. On a theoretical level, the present contribution sheds light on the possibilities of NP discontinuities in a morphologically rich language like Hindi/Urdu.

1 Introduction

Discontinuous constituents offer particular challenges for various NLP applications, such as question-answering, coreference resolution or topic modeling. This paper relates to an application that is further up the NLP toolchain: syntactic parsing. Here, the main challenges lie in:

- adapting the parser to be able to process the discontinuous structures;
- reconstruct the dependencies in the analysis, i.e., attach the discontinuous parts to their syntactic heads.

Third, from a theoretical linguistic point of view, one would also want to derive generalizations about what kinds of discontinuities are possible, and what kinds do not appear. Depending on the language studied, investigating such constraints is helpful since they can provide cross-linguistic insight into the phenomenon of discontinuity, and why it can or cannot take place.

This paper presents a study of discontinuous NPs in the morphologically-rich South Asian language Hindi/Urdu.[1] The focus is on genitive NP modifiers, which display a large deal of discontinuity. As will be seen below, in the right configurations, they may be scrambled out of their NP domain, removing them from the heads that they modify. Neither the phenomenon itself nor the configurations that allow for it have been previously discussed in the literature.

The paper contributes to solving all three of the above challenges. It discusses the empirical properties of the Hindi/Urdu genitive in general as well as genitive discontinuity, investigated by collecting data from native speakers and searching the

[1] The two languages Hindi and Urdu are so closely related that many researchers in linguistics treat them as a single language, Hindi/Urdu. Differences between Urdu and Hindi are mainly in the script (Urdu uses a version of Arabic script, while Hindi uses *Devanagari*) as well as in the vocabulary (Urdu uses more Persian and Arabic vocabulary, and Hindi evolved from Sanskrit). There are further minor differences in the phonology as well as in the derivational morphology; the syntax is almost identical.

Proceedings of DiscoNLP 2016, pages 37–46,
San Diego, California, June 17, 2016. ©2016 Association for Computational Linguistics

Hindi/Urdu Treebank (Bhatt et al., 2009) (§2,3,4). I arrive at a couple of theoretical generalizations, which can be aptly formulated via functional uncertainty within the framework of Lexical-Functional Grammar (LFG, Dalrymple (2001)). I suggest that the possibility of the genitive to appear outside its NP is a result of the rich agreement between the genitive case marker and the NP head. Finally, I describe how the Hindi/Urdu ParGram grammar (Butt and King, 2007; Bögel et al., 2009), a computational LFG grammar developed as part of the ParGram project (Sulger et al., 2013; Butt et al., 2002) and implemented in XLE (Crouch et al., 2015), is adapted to parse and correctly attach discontinuous genitives to their NPs (§5).[2] The paper concludes in §6.

2 General Description

The genitive case in Hindi/Urdu is realized using the clitic *k-*, which is attached to a possessor NP. Under the analysis of Hindi/Urdu case in Butt and King (2004), which I adapt here, all case clitics functionally head a KP (*case phrase*).[3] The genitive differs from other case clitics: it agrees in number, gender and morphological form (nominative or oblique) with the head noun, the possessum. For the feminine, there is morphological syncretism in that a single form *ki* is used throughout the feminine inflectional pattern. For the masculine, there is syncretism between the singular oblique and plural nominative and oblique. Table 1 shows the complete pattern of the clitic. In (1)–(3), the a. examples are valid NPs, displaying the correct agreement pattern.

(1) a. ram=ka mɑkan
Ram.M.SG=GEN.M.SG house.M.SG
'Ram's house'

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

c. * ram=ke mɑkan
Ram.M.SG=GEN.M.PL house.M.SG

Gender	Number	Inflection	Form
Masculine	Singular	Nominative	*ka*
		Oblique	*ke*
	Plural	Nominative	*ke*
		Oblique	*ke*
Feminine	Singular	Nominative	*ki*
		Oblique	*ki*
	Plural	Nominative	*ki*
		Oblique	*ki*

Table 1: Possible inflections of Hindi/Urdu genitive case clitic *k-*

(2) a. nina=ki beṭi
Nina.F.SG=GEN.F.SG daughter.F.SG
'Ram's car'

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

c. * nina=ke beṭi
Nina.F.SG=GEN.M.PL daughter.F.SG

(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

Within NPs, the modifying possessor phrase comes first, then the possessum (i.e., the head of the NP); this conforms to the general clausal word order in Hindi/Urdu, which is head-final (Mohanan, 1994; Butt, 1995). The position of the genitive phrase varies with respect to other NP modifiers, such as adjectives or quantifiers; see (4) for an example. NP modifiers occurring after the NP head are judged as ungrammatical by the informants; see (4c) for an example. Another example illustrating the variable word order inside the NP is shown in (5).

(4) 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
blue.F.SG Ram.M.SG=GEN.F.SG

[2] The Hindi/Urdu ParGram grammar can be tested using the INESS website at `http://iness.uib.no/`.

[3] The status of the case marker as a clitic is not of direct importance here; the interested reader is referred to Butt and King (2004) for a comprehensive discussion.

gaṛi
car.F.SG
'Ram's blue car'

c. * nili gaṛi
blue.F.SG car.F.SG
ram=ki
Ram.M.SG=GEN.F.SG

(5) a. ʊstad=ka kʊch hoʃyar
teacher.M.SG=GEN.M.SG some smart
talıb-ılm
student.M.PL
'some smart students of the teacher'

b. ʊstad=ka hoʃyar kʊch talıb-ılm

c. kʊch ʊstad=ka hoʃyar talıb-ılm

d. kʊch hoʃyar ʊstad=ka talıb-ılm

e. hoʃyar kʊch ʊstad=ka talıb-ılm

f. hoʃyar ʊstad=ka kʊch talıb-ılm

The constraint that NP modifiers have to precede their head inside the NP is corroborated by data such as in (6b) (a permutation of (6a)). Here, the genitive occurs after the NP head, *beṭe* 'sons', which is itself marked with the ergative case. The fact that (6b) is ungrammatical is a clear indication that the genitive phrases cannot be right-adjoined to the NP head.

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

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

The Hindi/Urdu genitive has a wide functional distribution: it appears on adjuncts and nominal arguments. The LFG analysis of Sulger (to appear) is assumed here, which argues for a differentiated treatment of the genitive KP in terms of the grammatical functions (GF) subject/SUBJ (7a), object/OBJ (7b) and adjunct/ADJUNCT (7c).[4]

(7) a. ram=ki tıppɑni
Ram.M.SG=GEN.F.SG comment.F.SG
'Ram's comment/criticism'

b. gaṛi=ki tɑbahi
car.F.SG=GEN.F.SG destruction.F.SG
'the car's destruction'

c. sʊrx rɑng=ki mez
red color.M.SG=GEN.F.SG table.F.SG
'the table of red color'

3 Genitive Scrambling

In addition to the variable word order inside NPs, there are examples showing that the genitive modifiers can occur outside of the NPs they modify. I will refer to this as *Genitive Scrambling*. In (8a), the genitive occurs in the canonical position inside the NP to the left of the head noun. In (8b), the genitive is scrambled outside of the subject NP to the end of the clause; still, it must be analyzed as a modifier of the head noun *dost* 'friend', since it cannot be argued to be an argument of the intransitive verb *a* 'come'.

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

b. dost ay-a
friend.M.SG.NOM come-PERF.M.SG
ram=ka
Ram.M.SG=GEN.M.SG
'Ram's friend came.' (Butt and Zinsmeister, 2009)

In (9a), the object *gaṛi* 'car' is modified by the genitive *ʊs=ki* 'her/his/its'. The genitive can be

[4] Due to space limitations, I do not go into detail regarding the treatment in Sulger (to appear). I will say, however, that the evidence includes binding of a reflexive pronoun as well as iterativity/optionality of nominal arguments vs. adjuncts. Note also that semantically, the genitive may realize various roles as a modifier, e.g., an agent in (7a), a patient in (7b), and an attribute in (7c). This semantic variety is known from many languages, including English; a semantic classification of the genitive is certainly not within the scope of this paper.

scrambled out of the object to the beginning of the clause as in (9b). From the morphosyntax, it is clear that in (9b) the feminine-inflected *ʊs=ki* 'her/his/its' modifies *gaṛi* 'car', since that is the only feminine nominal in the sentence. A very similar example is in (10).

(9) a. ram=ne
Ram.M.SG=ERG
ʊs=ki
PRON.3.SG.OBL=GEN.F.SG
gaṛi bazar=mẽ
car.F.SG.NOM market.M.SG=LOC.IN
dek^{h}-i
see-PERF.F.SG
'Ram saw her/his car in the market.' (adapted from Bögel and Butt (2013), p. 301)

b. ʊs=ki
PRON.3.SG.OBL=GEN.F.SG
ram=ne gaṛi
Ram.M.SG=ERG car.F.SG.NOM
bazar=mẽ dek^{h}-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)

(10) a. tʊm=ne kɪs=ki
you=ERG who.SG.OBL=GEN.F.SG
kɪtab xɑrid-i?
book.F.SG.NOM buy-PERF.F.SG
'Whose book did you buy?' (adapted from Bögel and Butt (2013), p. 301)

b. kɪs=ki tʊm=ne
who.SG.OBL=GEN.F.SG you=ERG
kɪtab xɑrid-i?
book.F.SG.NOM buy-PERF.F.SG
'Whose book did you buy?' (adapted from Bögel and Butt (2013), p. 301)

Genitives may also be scrambled to the right. In (11a), a permutation of (9a), the object is topicalized to the front of the clause. In (11b), the genitive phrase modifying the object is scrambled to the right and occurs after the subject. A similar example is given in (12), where *kɪs=ki* 'whose' modifies *kɪtab* 'book', but is not in the same constituent.

(11) a. ʊs=ki
PRON.3.SG.OBL=GEN.F.SG
gaṛi ram=ne
car.F.SG.NOM Ram.M.SG=ERG
bazar=mẽ dek^{h}-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)

b. gaṛi ram=ne
car.F.SG.NOM Ram.M.SG=ERG
ʊs=ki
PRON.3.SG.OBL=GEN.F.SG
bazar=mẽ dek^{h}-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)

(12) a. kɪs=ki
who.SG.OBL=GEN.F.SG
kɪtab tʊm=ne
book.F.SG.NOM you=ERG
xɑrid-i?
buy-PERF.F.SG
'Whose book did you buy?' (adapted from Bögel and Butt (2013), p. 301)

b. kɪtab tʊm=ne
book.F.SG.NOM you=ERG
kɪs=ki
who.SG.OBL=GEN.F.SG
xɑrid-i?
buy-PERF.F.SG
'Whose book did you buy?' (Bögel and Butt (2013), p. 301)

Recall that the order within NPs is head-final. As seen in (11)–(12), however, when genitives are scrambled outside of their NP, this order is not necessarily preserved. Using the terminology of Fanselow and Féry (2006), I refer to scrambled genitives that occur before their heads in the sentence as *non-inverted* scrambled genitives, and to scrambled genitives that occur after their heads as *inverted* scrambled genitives.

It is a reasonable assumption that scrambling of genitive phrases is possible since the genitive displays rich morphology which agrees with its head,

enabling speakers to identify the nominal in the sentence modified by the genitive. Fanselow and Féry (2006) identify agreement inside NPs as a main factor influencing the availability of discontinuous NPs across languages, but there are also counter-examples against this generalization; Turkish, for example, has discontinuous NPs, in spite of the absence of agreement inside nominal projections.

4 Some Preferences and Constraints

The operation of genitive scrambling does not occur without constraints. This section sums up these constraints, which serve as the empirical background for the XLE implementation of genitive scrambling as described in §5. Each of the constraints was verified by intensive consultation with at least three native speakers.

4.1 Local Attachments are Preferred

Consider (13a), which involves a topicalized object. The possessor of that object can be scrambled to the right as in (13b). In cases such as (13b), *ʊs=ki* is either a scrambled genitive modifying *gaṛi* 'car' or a canonical genitive locally attached to *bag* 'park'; the agreement morphology does not rule out either. Where the agreement morphology permits both scrambled as well as locally attached genitives, local attachments are highly preferred. Here, informants judge *ʊs=ki* 'his/her' as modifying *bag* 'park', but acknowledge that it may also modify *gaṛi* 'car'.

(13) a. ʊs=ki
PRON.3.SG.OBL=GEN.F.SG
gaṛi nadya=ne
car.F.SG.NOM Nadya.F.SG=ERG
bag=mẽ dek^{h}-i
park.F.SG=LOC.IN see-PERF.F.SG
'Her/his car, Nadya saw in the park.'

b. gaṛi nadya=ne
car.F.SG.NOM Nadya.F.SG=ERG
ʊs=ki
PRON.3.SG.OBL=GEN.F.SG
bag=mẽ dek^{h}-i
park.F.SG=LOC.IN see-PERF.F.SG
'The car, Nadya saw in her park.'
preferred over
'His/her car, Nadya saw in the park.'

The preference for local attachment is reflected in a principle well-known from cognitive science, first discussed by Kimball (1973) as the *Right Association* principle, and reformulated by Gibson (1991) as the *Recency Preference*.

4.2 Scrambling and Case

The examples above involve genitives that are scrambled out of bare NPs. Genitives may also be scrambled out of NPs that are overtly case-marked; in this case, inverted scrambled genitives are ungrammatical, and the genitive has to precede its head in the clause. Examples are shown in (14). In both sentences, *ram=ke* 'Ram's' modifies *bɑcco=ne* 'children=ERG', but since the latter is ergative-marked, the former has to precede it.

(14) a. ram=ke kɑl
Ram.M.SG=GEN.M.SG.OBL yesterday
bɑcco=ne yɪh
child.M.PL.OBL=ERG this
gana ga-ya
song.M.SG.NOM sing-PERF.M.SG
t^{h}-a
be.PAST-M.SG
'Ram's children sang this song yesterday.'

b. * bɑcco=ne kɑl
child.M.PL.OBL=ERG yesterday
ram=ke yɪh
Ram.M.SG=GEN.M.SG.OBL this
gana ga-ya
song.M.SG.NOM sing-PERF.M.SG
t^{h}-a
be.PAST-M.SG

A similar example involving a genitive scrambled from an overtly-marked object NP is given in (15): *ram=ke* 'Ram's' needs to precede its head *kʊṭṭe=ko* 'dog=ACC'.

(15) a. bɑcco=ne
child.M.PL.OBL=ERG
ram=ke kɑl
Ram.M.SG=GEN.M.SG.OBL yesterday
kʊṭṭe=ko dek^{h}-a
dog.M.SG.OBL=ACC see-PERF.M.SG
'The children saw Ram's dog yesterday.'

b. * bɑcco=ne
child.M.PL.OBL=ERG
kʊʈʈe=ko kɑl
dog.M.SG.OBL=ACC yesterday
ram=ke
Ram.M.SG=GEN.M.SG.OBL
dek^{h}-a
see-PERF.M.SG

Recall that genitive KPs modifying nominals in overtly case-marked KPs need to have oblique nominal morphology. One might assume, then, that examples such as (15b) are bad simply because there are several options for the genitive KP to modify a nominal, given the high amount of syncretism in genitive case marking for the oblique; e.g., in (15b) the genitive could modify both *bɑcco* and *kʊʈʈe*. (16) shows that this cannot be the issue. Here, the genitive can modify both nominals, being in linear precedence to both of them; cf. also (14b), which is ungrammatical, even though the agreement morphology clearly rules out any other possibilities of modification aside of *bɑcco*.

(16) ram=ke kɑl
Ram.M.SG=GEN.M.SG.OBL yesterday
bɑcco=ne kʊʈʈe=ko
child.M.PL.OBL=ERG dog.M.SG.OBL=ACC
dek^{h}-a
see-PERF.M.SG
'The children saw Ram's dog yesterday.'
or
'Ram's children saw the dog yesterday.'

4.3 Scrambling from Complement Clauses

Another constraint concerns complement clauses. None of my informants judge possessors scrambled out of finite complement clauses as grammatical; cf. the ungrammatical examples in (17). However, a majority of my informants indicate that it is grammatical to scramble genitive phrases from within non-finite complement clauses, e.g., the clause headed by the modal verb *sɑk* 'can' in (18). This is in line with the findings by Mahajan (1990), Kidwai (1999) as well as Kidwai (2000), who state that scrambling of arguments from within finite complement clauses is generally not accepted, whereas scrambling from infinite complement clauses is.

(17) * ʊs=ki
PRON.3.SG.OBL=GEN.F.SG
ram=ne kɑh-a kɪh
Ram.M.SG=ERG say-PERF.M.SG that
[nina=ne gaṛi
Nina.F.SG=ERG car.F.SG.NOM
dek^{h}-i]
see-PERF.F.SG

(18) ʊs=ki
PRON.3.SG.OBL=GEN.F.SG
ram gaṛi dek^{h}
Ram.M.SG.NOM car.F.SG.NOM see
sɑk-a
can-PERF.M.SG
'His/her car, Ram could see.'

4.4 No Scrambling out of Adjuncts

The third constraint concerning genitive scrambling is that genitive KPs may not be scrambled from within adjuncts. In (19a), *ʊs=ki* 'her/his/its' is a genitive phrase modifying *bag* 'park', which itself is locative case-marked and an adjunct to the overall clause. It is found that the possessor may not be scrambled from its NP to any other position in the clause (19b–c).

(19) a. ram=ne
Ram.M.SG=ERG
ʊs=ki
PRON.3.SG.OBL=GEN.F.SG
bag=mẽ haṭ^{h}i
park.F.SG=LOC.IN elephant.M.SG.NOM
dek^{h}-a
see-PERF.F.SG
'Ram saw an elephant in my park.'

b. * ʊs=ki ram=ne bag=mẽ haṭ^{h}i dek^{h}-a

c. * ram=ne bag=mẽ haṭ^{h}i ʊs=ki dek^{h}-a

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). It is also a well-known diagnostic for distinguishing arguments from adjuncts, as discussed by, e.g., Needham and Toivonen (2011) in an LFG setting.

4.5 No Scrambling from Deep Within

The last constraint to be discussed here indicates that it is not possible to scramble genitive phrases that are selected by nominals further down a path of grammatical functions. Consider the examples in (20a). *ʃɔhar* 'husband' is modified by a genitive SUBJ *orɑt=ke* 'the woman's'. *ʃɔhar=ki*, in turn, is an extrinsic possessor SUBJ modifying the overall object of the clause, *gaṛi* 'car'. The structure is as indicated by the bracketing in (20b). In the similar example (21), *sʊrx rang=ke* 'of red color' is an ADJUNCT modifying *mɑkan* 'house'.

(20) a. ram=ne
Ram.M.SG=ERG
orɑt=ke
woman.F.SG=GEN.M.SG.OBL
ʃɔhɑr=ki
husband.M.SG=GEN.F.SG
gaṛi dekʰ-i
car.F.SG.NOM see-PERF.F.SG
'Ram saw the woman's husband's car.'

b. ram=ne [[[orɑt=ke]$_{\text{SUBJ}}$ ʃɔhɑr=ki]$_{\text{SUBJ}}$ gaṛi]$_{\text{OBJ}}$ dekʰ-i

(21) a. nina=ne sʊrx
Nina.F.SG=ERG red
rɑng=ke
color.M.SG=GEN.M.SG
mɑkan=ka dɑrvaza
house.M.SG=GEN.M.SG door.M.SG
dekʰ-a
see-PERF.M.SG
'Nina saw the red house's door.'

b. nina=ne [[[sʊrx rɑng=ke]$_{\text{ADJUNCT}}$ mɑkan=ka]$_{\text{SUBJ}}$ dɑrvaza]$_{\text{OBJ}}$ dekʰ-a

Given such situations, consider the examples in (22)–(23). In (22a–b), *orɑt=ke* 'the woman's', the SUBJ genitive KP modifying *ʃɔhar* 'husband', cannot appear outside of the NP it is embedded in, i.e., outside the NP headed by *gaṛi* 'car', since it is embedded too far down in that NP, its GF path being (↑ OBJ SUBJ SUBJ) (starting from the main clause). (23a–b) show that the same restriction holds for attributive genitives such as *sʊrx rang=ke* 'of red color', which has the path (↑ OBJ SUBJ ADJUNCT) here.

(22) a. * orɑt=ke ram=ne ʃɔhɑr=ki gaṛi dekʰ-i

b. * ram=ne ʃɔhɑr=ki gaṛi orɑt=ke dekʰ-i

(23) a. * sʊrx rɑng=ke nina=ne mɑkan=ka dɑrvaza dekʰ-a

b. * nina=ne mɑkan=ka dɑrvaza sʊrx rɑng=ke dekʰ-a

5 XLE Implementation

This section describes the implementation of the Hindi/Urdu genitive as well as its scrambling properties and resulting discontinuities. The implementation uses the XLE grammar development platform, which includes an industrial-strength parser and generator for LFG grammars (Crouch et al., 2015).

5.1 General Setup

The lexical entry for the feminine genitive case marker *ki* is given in (24). Recall the agreement pattern of the genitive case marker in Table 1; in XLE, constraining equations can account for the requirements concerning gender, number as well as morphological form. In (24), the constraints are in the form of inside-out constraining equations, since the genitive KP may either be embedded in a SUBJ, ADJUNCT or in an OBJ f-structure inside the head noun's f-structure. The last line in (24) states that the case marker needs to be inside an f-structure that has the feature NTYPE; this ensure that the genitive only occurs as a nominal case (i.e., not on verbal arguments/adjuncts).

(24)
```
kI    K * (^ CASE) =c gen
      (({SUBJ|OBJ|ADJUNCT} ^) GEND) =c fem
      (({SUBJ|OBJ|ADJUNCT} ^) NTYPE).
```

The XLE grammar rules in (25) construct the KP and NP. (25a) states that the KP consists of an NP and an optional case marker K. (25b) states that an NP may consist of a simple pronoun or a modified noun (Nadj). In (25b), the use of the shuffle operator (,), separating the KP, AP and N nodes ensures that each of these nodes may occur in any order, thereby allowing for different word orders inside the NP. The annotation ! <h ^ (making use of the head precedence operator <h) indicates that the currently annotated c-structure node (here: KP or AP) has to precede the c-structure node of the higher-level f-structure, modeling the fact that genitives and other NP modifiers have to precede their

heads. Sample c- and f-structures for (2a) are shown in Figures 1 and 2.[5]

(25) a.
```
KP --> NP
       (K).
```
b.
```
NP --> {PRON
       |Nadj}.
```
c.
```
Nadj = KP*: (! CASE) = gen
            ! <h ^
            {@SUBJ|@OBJ|@ADJUNCT}
     ,
     AP*: @ADJUNCT
            ! <h ^
     ,
     N.
```

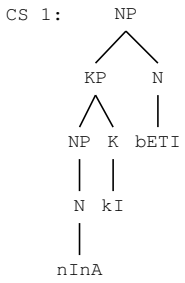

Figure 1: Hindi/Urdu NP c-structure for (2a)

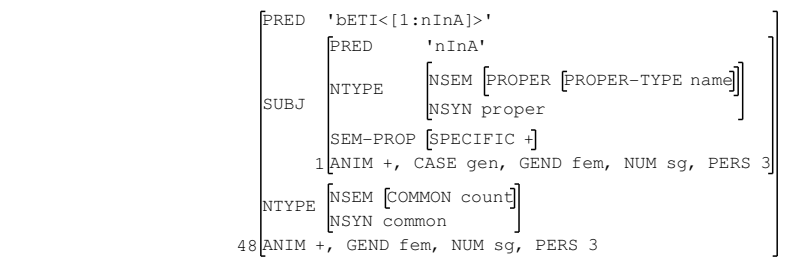

Figure 2: Hindi/Urdu NP f-structure for (2a)

5.2 Generalizing and Implementing Genitive Scrambling

The genitive scrambling facts can be formalized via a functional uncertainty path as in (26).[6] The expression is matched by a variety of paths, e.g., `SUBJ`, `OBJ`, `XCOMP SUBJ`, etc. (`XCOMP` is the grammatical function used for non-finite complement clauses). Thus, (26) describes exactly those paths that scrambled genitives may be extracted from; it does not allow for genitives scrambled from adjuncts, finite complement clauses (which are inside the `COMP` GF) or from deeper GF paths (e.g., `OBJ SUBJ`).

(26)
```
KP-SCRAMBLE-PATH = (XCOMP)
                   {SUBJ|OBJ|OBL|OBJ-GO|OBJ-TH}.
```

(27) is the XLE rule template that adds scrambled genitive KPs to the c-structure tree. Functionally, they are annotated as subjects, objects or adjuncts (lines 6–8) inside a path variable instantiated from `KP-SCRAMBLE-PATH` (line 2). Lines 3–5 check the case feature of the head noun; it is either nominative (i.e., a bear NP), in which case there is no precedence constraint, or it is not nominative (i.e., it is overtly case-marked), in which case the genitive is required to precede its head (again implemented using head precedence, see above). Finally, line 9 adds an O(ptimality)T(heory) mark to the scrambled genitive, called `attach`, which marks the analysis as non-optimal when it is in direct competition with a local attachment analysis (which does not carry the OT mark).

(27)
```
KP-SCRAMBLE = KP*: (! CASE) = gen
(^ KP-SCRAMBLE-PATH) = %PATH
{(%PATH CASE) =c nom
|(%PATH CASE) ~= nom
 ! <h %PATH}
{(%PATH SUBJ) = !
|(%PATH OBJ) = !
|! $ (%PATH ADJUNCT)
@(OT-MARK attach).
```

5.3 Testsuite Creation

To perform regression tests on the implementation, a separate testsuite file was created with examples of vanilla genitives as well as instances of genitive scrambling. The testsuite currently includes 36 grammatical and ungrammatical examples, each between two and eight words long, and has been manually constructed in close collaboration with the native speakers. All grammatical sentences are parsed successfully, while all ungrammatical sentences are ruled out.[7]

[5] The rules in the grammar are more complicated than shown here; e.g., the `Nadj` rule includes further nodes such as quantifiers, demonstratives etc. The scheme used by the Hindi/Urdu ParGram grammar for transliterating the Urdu Arabic script is described in Malik et al. (2010) as well as Bögel (2012).

[6] Documentation for the implementation of functional uncertainty in XLE is at `http://ling.uni-konstanz.de/pages/xle/doc/notations.html#N4.1.5`.

[7] One ungrammatical sentence in fact times out with the default XLE timeout setting of 30 seconds, pending investigation.

Given the ambiguity of the genitive discussed in Section 2, all sentences yield ambiguous parse results. As an example, reconsider (13b). The sentence is part of the testsuite and yields two optimal as well as two unoptimal solutions. Under the two optimal readings, *ʊs=ki* 'his/her' locally modifies *bag* 'park' as a subject or an adjunct; under the two unoptimal readings, *ʊs=ki* 'his/her' is a scrambled genitive subject or adjunct modifying *gaṛi* 'car'.

XLE does not display unoptimal solutions by default; the developer/annotator can select the unoptimal solution(s) by clicking the OT mark that controls the (dis)preference. Figure 3 shows the optimal solution where *ʊs=ki* 'his/her' is a subject, while Figure 4 shows the corresponding unoptimal solution, i.e., the scrambled genitive analysis.[8]

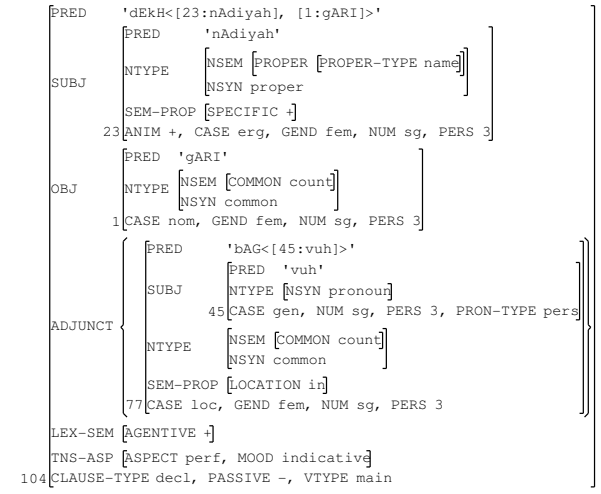

Figure 3: Hindi/Urdu NP f-structure for (13b)

6 Summary

The paper describes Hindi/Urdu genitives in general and its scrambling properties in particular. I take a detailed look at the empirical distribution of this phenomenon, including its syntactic constraints, and formulate a generalization using LFG. The generalization is implemented in the Hindi/Urdu ParGram grammar using XLE.

[8]The c-structures are not shown here due to space limitations. In the c-structure corresponding to the f-structure in Figure 3, the genitive attaches below the NP headed by *bag* 'park', while in the c-structure for Figure 4, the genitive attaches to the clausal node, resulting in a flat structure.

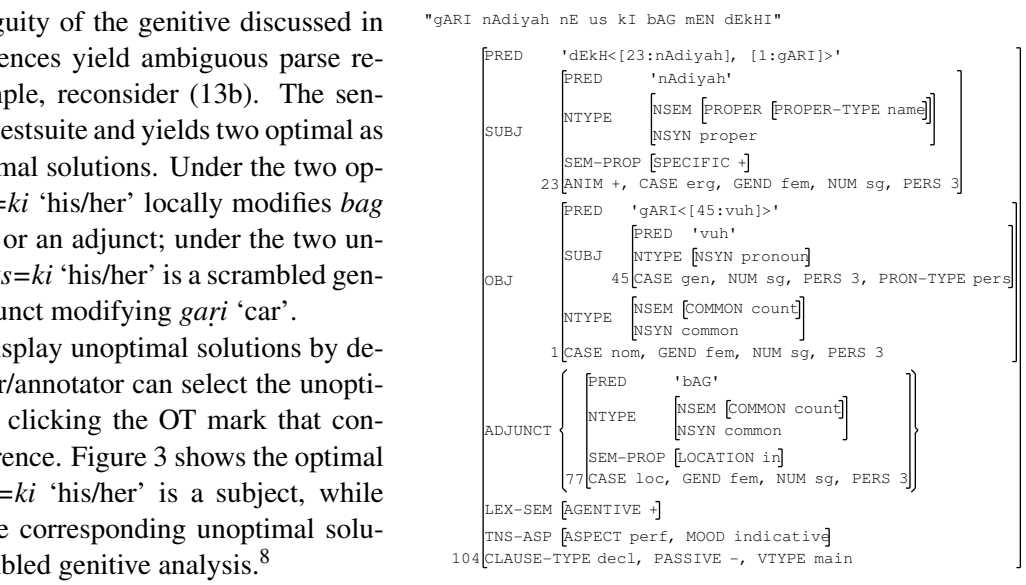

Figure 4: Hindi/Urdu NP f-structure for (13b)

Future theoretical work includes a comparison with other morphologically-rich languages. An initial investigation has shown that scrambling data in Turkish, as discussed by e.g. Kornfilt (2003), are similar, but display a constraint called the "barrier constraint" by Chomsky (1986), which rules out possessors that occur directly right-adjoined to arguments; the constraint does not exist in Hindi/Urdu. Since ParGram includes a Turkish grammar (Çetinoglu, 2009), a comparison of the annotations necessary to cover the genitive scrambling facts would be interesting.

Acknowledgments

This work is supported by a Nuance Foundation grant on *Tense and Aspect in Multilingual Semantic Construction*. I would like to thank the native speakers who have provided me with judgments; in alphabetical order, these are Qaiser Abbas, Tafseer Ahmed, Rajesh Bhatt, Miriam Butt, Farhat Jabeen, Asad Mustafa, Ghulam Raza and Ashwini Vaidya.

References

Rajesh Bhatt, Bhuvana Narasimhan, Martha Palmer, Owen Rambow, Dipti Misra Sharma, and Fei Xia. 2009. A Multi-Representational and Multi-Layered Treebank for Hindi/Urdu. In *Proceedings of the Third Linguistic Annotation Workshop*, pages 186–189, Sun-

tec, Singapore, August. Association for Computational Linguistics.

Tina Bögel and Miriam Butt. 2013. Possessive Clitics and Ezafe in Urdu. In Kersti Börjars, David Denison, and Alan Scott, editors, *Morphosyntactic Categories and the Expression of Possession*, pages 291–322. John Benjamins.

Tina Bögel, Miriam Butt, Annette Hautli, and Sebastian Sulger. 2009. Urdu and the Modular Architecture of ParGram. In *Proceedings of the Conference on Language and Technology 2009 (CLT09)*. Center for Research in Urdu Language Processing (CRULP).

Tina Bögel. 2012. Urdu – Roman Transliteration via Finite State Transducers. In *Proceedings of FSMNLP'12*, pages 25–29.

Miriam Butt and Tracy Holloway King. 2004. The Status of Case. In Veneeta Dayal and Anoop Kumar Mahajan, editors, *Clause Structure in South Asian Languages*, pages 153–198. Kluwer.

Miriam Butt and Tracy Holloway King. 2007. Urdu in a Parallel Grammar Development Environment. *Language Resources and Evaluation: Special Issue on Asian Language Processing: State of the Art Resources and Processing*, 41:191–207.

Miriam Butt and Heike Zinsmeister. 2009. ESSLLI 2009 Course on Case, Scrambling and Default Word Order. Course material.

Miriam Butt, Helge Dyvik, Tracy Holloway King, Hiroshi Masuichi, and Christian Rohrer. 2002. The Parallel Grammar Project. In *Proceedings of the COLING-2002 Workshop on Grammar Engineering and Evaluation*, pages 1–7.

Miriam Butt. 1995. *The Structure of Complex Predicates in Urdu*. Dissertations in Linguistics. CSLI Publications.

Özlem Çetinoglu. 2009. *A Large Scale LFG Grammar for Turkish*. Ph.D. thesis, Sabanci University.

Noam Chomsky. 1986. *Barriers*. MIT Press.

Dick Crouch, Mary Dalrymple, Ronald M. Kaplan, Tracy Holloway King, John T. Maxwell III, and Paula Newman, 2015. *XLE Documentation*. Palo Alto Research Center.

Mary Dalrymple. 2001. *Lexical Functional Grammar*, volume 34 of *Syntax and Semantics*. Academic Press, New York.

Gisbert Fanselow and Caroline Féry. 2006. Prosodic and Morphosyntactic Aspects of Discontinuous Noun Phrases: a Comparative Perspective. Manuscript, University of Potsdam.

Edward Gibson. 1991. *A Computational Theory of Human Linguistic Processing: Memory Limitations and Processing Breakdown*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA.

Ayesha Kidwai. 1999. Word Order and Focus Positions in Universal Grammar. In Georges Rebuschi and Laurice Tuller, editors, *The Grammar of Focus*, pages 213–244. John Benjamins.

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

John Kimball. 1973. Seven Principles of Surface Structure Parsing in Natural Language. *Cognition*, 2:15–47.

Jaklin Kornfilt. 2003. Scrambling, Subscrambling, and Case in Turkish. In Simin Karimi, editor, *Word Order and Scrambling*. Blackwell Publishing.

Anoop Kumar Mahajan. 1990. *The A/A-Bar Distinction and Movement Theory*. Ph.D. thesis, MIT.

Muhammad Kamran Malik, Tafseer Ahmed, Sebastian Sulger, Tina Bögel, Atif Gulzar, Ghulam Raza, Sarmad Hussain, and Miriam Butt. 2010. Transliterating Urdu for a Broad-Coverage Urdu/Hindi LFG Grammar. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC 2010)*, pages 2921–2927.

Tara Mohanan. 1994. *Argument Structure in Hindi*. Dissertations in Linguistics. CSLI Publications.

Stephanie Needham and Ida Toivonen. 2011. Derived Arguments. In Miriam Butt and Tracy Holloway King, editors, *Proceedings of the LFG11 Conference*, pages 401–421. CSLI Publications.

John Robert Ross. 1967. *Constraints on Variables in Syntax*. Ph.D. thesis, MIT.

Sebastian Sulger, Miriam Butt, Tracy Holloway King, Paul Meurer, Tibor Laczkó, György Rákosi, Cheikh Bamba Dione, Helge Dyvik, Victoria Rosén, Koenraad De Smedt, Agnieszka Patejuk, Özlem Çetinoglu, I Wayan Arka, and Meladel Mistica. 2013. ParGramBank: The ParGram Parallel Treebank. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 550–560, Sofia, Bulgaria, August. Association for Computational Linguistics.

Sebastian Sulger. to appear. *Modeling Nominal Predications in Hindi/Urdu*. Ph.D. thesis, University of Konstanz.