Possessive Clitics and Ezafe in Urdu

Final Version October 2010

Tina Bögel and Miriam Butt

University of Konstanz

Abstract

In this paper we compare two ways of expressing possession in the Indo-Aryan language Urdu. While the genitive case marker can be analyzed as a clitic in a relatively straightforward way, the *ezafe* construction poses a challenge when it comes to its classification as either a phrasal affix or clitic. Samvelian (2007) analyzes Persian *ezafe* as a phrasal affix that is generated within the morphological component, rejecting a postlexical analysis. After taking a look at the data for both constructions, we challenge Samvelian's view of *ezafe* and explore the possibilities for the interplay of phonology, morphology and syntax to resolve the tension between the lexical/affixal properties of clitics and their behavior as an independent syntactic item. In addition to the syntactic representation, we invoke postlexical prosodic phonology to cover all the properties of clitics in general and *ezafe* in particular. Thus, we show that it is not necessary to distinguish between phrasal affixes and clitics.

1 Introduction

In the Indo-Aryan language Urdu, the notion of possession can be expressed in several different ways.[1] Two very common possibilities of expressing possession are the genitive case marker *k-* ((1a)) and the *ezafe* construction ((1b)), a loan construction from Persian.

(1) a. yɑsin=ki   gari
         yasin.M.Sg=Gen.F.Sg car.F.Sg
         'Yassin's car'   Urdu

     b. sahɪb=e   tɑkʰt
         owner.M.Sg=Ez throne.M.Sg
         'The owner of the throne'   Urdu

While the genitive case marker allows for a relatively straightforward analysis as a clitic (Butt and King 2004), the *ezafe* construction poses some problems as to its classification as either an affix or a clitic. In the case of Persian *ezafe*, Samvelian (2007) has argued that the *ezafe* is a phrasal affix, which is generated within the nominal morphology. However, this analysis does not account for all of the inherent properties of *ezafe* elegantly. By exploring the different aspects and properties of the Urdu *ezafe* construction, we argue against its treatment (and the treatment of other clitics) within the morphology. We compare the data on the *ezafe* to the properties of the genitive *k-* and discuss both constructions with respect to the morphology-syntax-prosody interface and the discussion of clitics and phrasal affixes in general.

The paper is structured as follows. We first introduce our architectural framework in section 2. Some necessary background on other morphosyntactic properties of Urdu is provided in section 3. The genitive construction and data arguing for an analysis of the genitive case marker as a clitic are introduced in section 4. Section 5 provides a short overview over the discussion of the Persian *ezafe* before turning to the Urdu *ezafe* and analyzing the data with respect to the phrasal affix/clitic distinction. After evaluating the

properties of the *ezafe* construction, we come to the conclusion that the *ezafe* should be analyzed as a clitic. In order to show how the syntactic and prosodic properties of *ezafe* can be modelled straightforwardly given our architectural assumptions, we describe a concrete implementation in section 6.
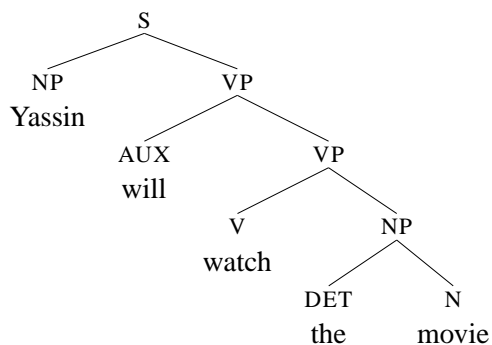
2   Grammar Architecture Assumed

This section provides a short architectural overview of the framework we assume for the discussion on possessive clitics in Urdu. We provide this overview in order to be maximally clear about our assumptions. Moreover, we model our analysis concretely in terms of a computational implementation in order to be maximally clear about which module of grammar models which particular properties.

   We assume Lexical-Functional Grammar (LFG) (Bresnan 1982, 2001, Dalrymple 2001, Asudeh and Toivonen 2009) as our theoretical framework. LFG is an inherently modular grammar that views language as being modelled via different dimensions of structure, each involving its own rules, concepts and forms. Its core is a syntactic component consisting of c(onstituent)-structure, which encodes the basic constituents, their linear precedence and the hierarchy of the elements, and f(unctional)-structure, which abstracts away from the surface realization and models grammatical relations, functional information and other dependencies. A simple (and simplified) example is shown in (2).

(2)   a. Yassin will watch the movie.

   b. **c-structure**   c. **f-structure**

$$
\begin{bmatrix}
\text{PRED} & \text{'watch}\langle\text{SUBJ},\text{OBJ}\rangle\text{'} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'Yassin'} \end{bmatrix} \\
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'movie'} \\ \text{SPEC} & \text{the} \end{bmatrix} \\
\text{TENSE} & \text{future}
\end{bmatrix}
$$

   Apart from these two core syntactic representations, LFG allows for other components as well. The overall architecture of LFG is known as a *projection architecture*, since the differing modules of grammar are related to one another through well-defined mathematical projections. The ϕ-projection, for example, relates the f-structure to the c-structure (the c-structure projects the f-structure). The mathematical inverse of this projection can also be taken, so that the c-structure can be related to the f-structure (starting from an f-structure, information about the corresponding c-structure nodes can be obtained).

   Beyond c- and f-structure, several other types of projections (or modules) have also been proposed in the literature. Chief among these are a s(emantic)-structure (Halvorsen and Kaplan 1988), which models the semantic interpretation of a sentence, and i(nformation)-structure (King 1997), which encodes information about notions such as topic and focus (see also Bresnan 2001). For our purposes, namely, the analysis of clitics, the p(rosodic)-structure (Butt and King 1998) is also of relevance.

   The morphological module is taken to be independent of the syntactic representations. That is, it is taken to provide the word forms that make up the terminal nodes in the syntactic tree. LFG generally adheres to the *Lexical Integrity Principle* (e.g., Bresnan 1982, Dalrymple 2001), which states that the syntax does not "reach" into the morphology, but that words are built up by an independent set of rules before forming the terminal nodes of the syntax. There are, of course, issues to be resolved such as the

treatment of items like *we'll* or compounds, but by and large, the *Lexical Integrity Principle* is adhered to quite strictly within LFG.[2]

The morphological component can also be conceived of as a projection (Sadler and Spencer 2002). In this paper, we work within the architecture developed with the ParGram (Parallel Grammars) project. This is a loose alliance of researchers within the LFG community, who build large-scale, robust computational grammars using common underlying principles and common technology (Butt et al. 1999, 2002) as a specific and testable instantiation of the LFG architecture. For the implementation of these grammars (including the Urdu grammar that forms the basis for the analysis of possessive clitics within this paper), the grammar development platform XLE (Crouch et al. 2010) is used.

ParGram assumes a specific architecture to realize the theoretical aspects of LFG in a concrete and testable way within the computational grammars. In a ParGram grammar, the morphological component is implemented via a finite-state machine (Beesley and Karttunen 2003) and is related to the syntax via a well-defined relation.[3] The morphological component relates lemma forms and grammatical information (e.g., about number, gender, person) in an abstract form to concrete, inflected word forms. See section 6.1 for a concrete illustraton.

We illustrate the architecture assumed by us in Figure 1.
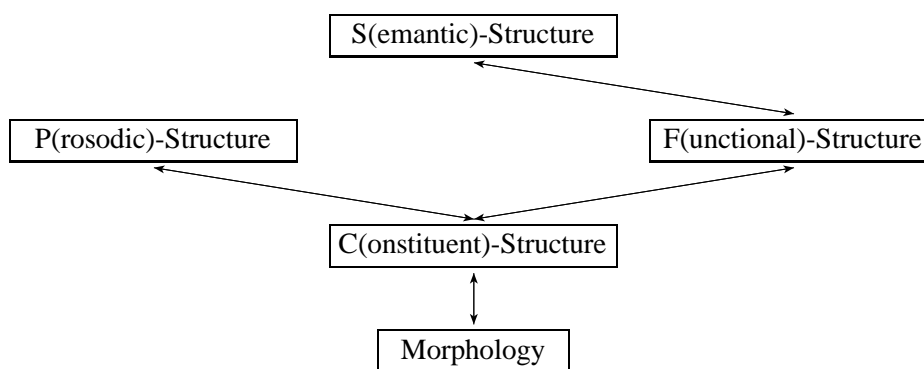

Figure 1: Overall Grammar Architecture within ParGram

Information from the morphological component is related to the syntactic component via a well-defined relation. Similarly, the f-structure and c-structure are related to one another via LFG's projection architecture and in a further abstraction away from the overt realization of the clause (i.e., the morphology and the c-structure), the semantics are projected from the f-structure using the information stored therein, while the prosodic information is projected away from the c-structure and is modelled in terms of an independent p(rosodic)-structure (see Bögel et al. 2009, 2010 for newer work on this).

Note that all the arrows in Figure 1 are bidirectional. When a clause is processed (parsed), we begin with the string, identify the words, morphologically analyze the words and arrange these into a syntactic tree. The syntactic tree provides information about the functional structure of the clause and that in turn provides information about its meaning. When a clause is to be produced (generated), one begins with the semantic structure, decides what functional structure and syntactic tree could correspond to it and then inflects the individual words accordingly in the morphological component. The bidirectionality of the arrows also allows for give-and-take between components. A certain syntactic tree can only be feasible

---

[2]There are cases which have been taken to challenge the Principle of Lexical Integrity. Broadwell (2008) and Wescoat (2009) both propose a relaxation of the principle and introduce the notion of *lexical sharing* for Turkish suspended affixation and Udi person markers. But see Bögel (2010) for a discussion of a similar problem with regard to Pashto endoclitics which seeks to adhere to the Principle of Lexical Integrity and which proposes a solution involving a division of labor between a prosodic and a syntactic component, much as in this paper.

[3]Indeed, Karttunen (2003) shows that the concrete finite-state implementation is equivalent to the basic architecture assumed by *Paradigm-Function* or *Realizational Morphology* (Stump 2001), with the difference that the finite-state version is mathematically better understood and allows for parsing/processing as well as generation/production.

if the morphological analysis provides the right information, but on the other hand, a word like *walks* is always ambiguous from the point of view of the morphology and all possibilities will be offered up to the syntax, which can then serve to disambiguate since the noun *walks* vs. the verb *walks* will only be feasible in certain, different, syntactic configurations.

Also note that we only show those parts of the architecture that are relevant to the discussion here (see O'Connor (2004), Mycock (2006), Asudeh and Toivonen (2009) for detailed discussions of more elaborate and also slightly differing architecture versions within LFG). In particular, one would want to assume a connection between prosody and semantics, but this should be mediated by i-structure (i.e., the prosody and semantics of topic/focus, etc.).

After having established the basic architecture we assume, the following sections first provide some background on Urdu and then go on to describe two possessive constructions in Urdu: the genitive construction and the *ezafe* construction. For each of these constructions we provide data on their functions and morphosyntactic properties and show a possible analysis in the above architecture.

3   Urdu: Agreement and Case System

Urdu is a language spoken mainly in Pakistan and India that is structurally more or less identical to Hindi. The major difference between the two languages is the script in that Urdu uses a version of the Arabic script, but Hindi is written in Devanagari. Furthermore, the originally Indic vocabulary of Urdu has been significantly enriched by borrowings from Arabic, Persian, while Hindi relies more heavily on Sanskrit (both have borrowed from English).

Urdu is a free word order language (major constituents can reorder freely) which conforms to a mostly head-final pattern (some complement clauses are head initial). In terms of agreement, the verb agrees with either the nominative subject or the nominative (unmarked) object in gender, number and person (in person only with the auxiliaries and the future form). In (3a) for example, the verb *boli* 'speak' takes the feminine form in order to agree with the feminine subject *nadya*. However, this ability is blocked if there is no nominative argument. In (3b), the ergative case marker follows the subject, semantically marking the action of the intransitive clause as being volitional.[4] In this case, when there is no available nominative subject or object to agree with, the verb takes the morphological default form, which is in case of Urdu the masculine form of the verb (*bola* 'speak').

(3)   a.   nadya   bol-i
           nadya.F.Sg=Nom speak-Perf.F.Sg
           'Nadya spoke.'   Urdu

      b.   nadya=ne   bol-a
           nadya.F.Sg=Erg speak-Perf.M.Sg
           'Nadya spoke.' (consciously/volitionally)   Urdu

The same can be shown for transitive verbs, where the object is available for agreement when the subject is non-nominative as in (4a). Here, the verb agrees with the only noun that is in the nominative case: *kıtab* 'book'. When the object is also non-nominative, the verb again reverts to the default masculine singular (4b).

(4)   a.   nadya=ne   kıtab   kɑmr-e=mẽ   dekʰ-i
           nadya.F.Sg=Erg book.F.Sg.Nom room-M.Sg.Obl=Loc see-Perf.F.Sg
           'Nadya saw a/the book in the room.'   Urdu

      b.   nadya=ne   kıtab=ko   kɑmr-e=mẽ   dekʰ-a
           nadya.F.Sg=Erg book.F.Sg=Acc room-M.Sg.Obl=Loc see-Perf.M.Sg
           'Nadya saw a specific book in the room.'   Urdu

---

[4]For further case alternations and the semantic interpretation of case markers see Butt and King (2004).

The examples in (3) and (4) also illustrate a very wide-spread phenomenon in Urdu, namely the use of case alternations to express semantic contrasts. In (3) the contrast is between volitionality and non-volitionality, in (4) the accusative *ko* expresses specificity. In general, the core grammatical relations like subject and object can be marked with a number of different case alternations in the form of case clitics that follow the noun and that have an effect on the semantic interpretation of the clause (Mohanan 1994, Butt and King 2004).

Urdu has several case markers, the basic ones are nominative (which is not realized overtly), ergative (*ne*), accusative (*ko*), dative (*ko*), instrumental (*se*), genitive (*k-*) and locative (*mẽ/par/tak/ø*) (Butt and King 2004, 157). All of the overt case markers require an oblique marking on the noun they mark. This is illustrated above in (4) with respect to *kamre* 'room', which is the oblique form of *kamra*. The morpheme *-e* generally represents the oblique, *-a* masculine singular and *-i* feminine singular.

4 The Genitive Case Marker

Among the case markers in Urdu, the genitive is an exception in that it is the only one which inflects for number and gender. All the other case markers show no inflection. The genitive is considered to originate from a past participial form of Sanskrit *kar* 'do' (e.g., Kellogg 1893, 129) — this would explain why it still shows number and gender agreement. The genitive furthermore agrees with the head element of the genitive construction in gender and number (see section 4.2), a property that is also not inherent to any of the other case markers (cf. Payne 1995).

Another remarkable feature of genitives in Urdu is that the genitive phrase can occur outside of the noun phrase it modifies and can indeed occur at quite a distance from its head noun. Although this property is not immediately relevant for the discussion in this paper, we provide some examples in order to fulfill a request for more information by one of the reviewers.

4.1 Function and Use of the Genitive

The genitive case marker in Urdu is used for several different constructions, four of which will be introduced in the following examples (based on Platts 1909, 250–254) in order to provide an overview over the possible functions:

(5) The genitive can be used to express *relations* such as kinship:

    nadya=ka   beṭa
    nadya.F.Sg=Gen.M.Sg son.M.Sg
    'Nadya's son'   Urdu

(6) The *subjective genitive* is used to indicate the subject of a verbal noun:

    mohan=ka   bʰag-na
    Mohan.M.Sg=Gen.M.Sg run-Inf.M.Sg
    'Mohan's running away'   (Platts 1909, 253)   Urdu

(7) The *objective genitive* in contrast to (6) denotes the object of the action, feeling or notion.

    dusrõ=ka   ɣam
    other.Pl=Gen.M.Sg sorrow.M.Sg
    'The sorrow of others'   (Platts 1909, 253)   Urdu

(8) The *possessive genitive* denotes a thing that is owned by a possessor:

    nadya=ki   kıtab
    Nadya.F.Sg=Gen.F.Sg book.F.Sg
    'Nadya's book'   Urdu

The Urdu genitive encompasses the uses of 'have' as well as 'belong', both of which are verbs which Urdu does not posess. For example *Nadya has one leg* or *This book belongs to Nadya* are both expressed via the genitive in Urdu.

(9) a. nadya=ka   ek paõ   hɛ
       Nadya.F.Sg=Gen.M.Sg one foot.M.Sg be.Pres.3.Sg
       'Nadya has one foot.'   Urdu

   b. ye   kıtab   nadya=ki   hɛ
       this.Sg.Nom book.F.Sg.Nom Nadya.F.Sg-Gen.F.Sg be.Pres.3.Sg
       'This is Nadya's book.'   Urdu

Raza (2010) also identifies several attributive uses involving material, price, size, height, color, weight and age. An example specifying the color and material of an object is provided in (10).

(10) [sʊrx rang=ki]   [lɑkṛi=ki]   mez
     red   color.M.Sg=Gen.F.Sg wood.F.Sg=Gen.F.Sg table.F.Sg
     'a table of red color and of wood'   Urdu

As far as we can tell, the complete possible ranges of use of the genitive in Urdu reveal no big surprises. The uses are consonant with an abstract sense of possession. Indeed, more generally, as is the case with English *have*, the Urdu genitive is used to express a relationship of some sort between two entities. It is also used to mark agents in verbal nouns, but this is also not an unusual property.

4.2 Morphosyntactic Properties

In contrast, the morphosyntactic properties of the genitive are interesting in that the case marker inflects for gender and number with the head noun of the construstion. This behavior is unique within the family of Urdu case markers. In (11), a typical genitive construction is illustrated:

(11) pakıstan=ki   hʊkumɑt
      Pakistan=Gen.F.Sg government.F.Sg
      'The government of Pakistan'   Urdu

Urdu is generally a head-final language and the genitive case also conforms to this pattern. In example (11), the modifying noun is *pakıstan* and the head noun is *hʊkumɑt* 'government'. The case marker *ki* agrees with this head noun in gender and number. This pattern is also true for complex nested constructions as in (12), where the first genitive marker *ki* agrees with *ami* 'mother' and the second genitive marker *ke* agrees with *kʊtte* 'dog' (as does the adjective *kale* 'black') (Payne 1995).

(12) [[yɑsin=ki]   ami]=ke   kal-e   kʊtt-e
      Yassin.M.Sg=Gen.F.Sg mother.F.Sg=Gen.M.Pl black-Nom.M.Pl dog-Nom.M.Pl
      'Yassin's mother's black dogs'   Urdu

Butt and King (2004) argue that case markers in Urdu should be analyzed as prosodically deficient clitics that rely on a host to their left. As one possible test, they adduce coordination. Following Miller (1992) and Zwicky (1987), who propose coordination as one criterion to distinguish affixes and clitics, Butt and King (2004) note that Urdu case markers have scope over noun coordinations.

(13) [nadya   or yɑsin]=ki   ami=ne   hɑs-a
      Nadya.F.Sg and Yassin.M.Sg=Gen.F.Sg mother.F.Sg=Erg laugh-Perf.M.Sg
      'Nadya and Yassin's mother laughed.'   Urdu

In (13), the genitive case marker *ki* scopes over the coordinated structure and agrees with the head noun *ami* 'mother'. This would not be possible with any inflectional affix in Urdu. Instead, affixes have to attach to every single conjunct of the coordination ((14)).

(14) a. [[[lɑṛk]-ɪyã] or [[cɪṛ]-ɪyã]] bol-ẽ-g-i
          girl.F.Pl.Nom and bird.F.Pl.Nom speak-3.Pl-Fut-F.Pl
          'The girls and the birds will speak (make noises).'   Urdu

      b. *[lɑṛk   or cɪṛ]-ɪyã   bol-ẽ-g-i
          girl.F.Pl.Nom and bird.F.Pl.Nom speak-3.Pl-Fut-F.Pl
          'The girls and the birds will speak (make noises).'   Urdu

The conclusion from this coordination test is that case markers attach to phrases and therefore seem to be syntactically placed, rather than morphologically.

Another test given in Butt and King (2004) is the inclusion of other clitics in between the case marker and the nominal host. The focus clitics *hi/bʰi* 'only/also' for example, may be placed between the case marker and the noun ((15a)).

(15) a. bɑcc-õ=hi=ka   kʰana
          child-M.Pl.Obl=Foc=Gen.M.Sg food.M.Sg
          'The CHILDREN's food'   Urdu

      b. *bɑcc=hi-õ=ka   kʰana
          child=Foc-M.Pl.Obl=Gen.M.Sg food.M.Sg
          'The CHILDREN's food   Urdu

In (15a), the focus clitic *hi* separates the case marker from its nominal prosodic host. This is not possible with an inflectional affix, as demonstrated in (15b), where the focus clitic cannot intervene between the nominal stem and the plural oblique suffix *-õ*. The construction becomes ungrammatical.

4.3 Syntactic Distribution

As illustrated above in (12), Urdu allows nested genitives. However, as also already illustrated by (10), Urdu in addition allows for multiple genitives modifying a single head noun. These are not hierarchically organized, but both independently modify the head noun and can be moved about freely, as shown in (16).

(16) a. [ʊs=ki]   [skul=ki]   gari
          Pron.3.Sg.Obl=Gen.F.Sg school.M.Sg=Gen.F.Sg car.F.Sg
          'His/her car that is also the school's'   Urdu

      b. [skul=ki]   [ʊs=ki]   gari
          school.M.Sg=Gen.F.Sg Pron.3.Sg.Obl=Gen.F.Sg car.F.Sg
          'His/her car that is also the school's'   Urdu

The meaning of (16) is that there is a car which belongs to a school and which a certain person (him/her) has been given the use of. What (16) definitely does not mean is: *His school's car.* In order to express this, the genitive on 'his' needs to agree with 'school', as in (17), thus giving rise to a nested structure.

(17) [[ʊs=ke]   skul]=ki   gari
      Pron.3.Sg.Obl=Gen.M.Obl school.M.Sg=Gen.F.Sg car.F.Sg
      'His/her school's car'   Urdu

As Raza (2010) shows, multiple genitives within complex Urdu NPs pose a problem in that it is not always easy to resolve which head each of the genitives modifies. The syntactic distribution of genitives within the Urdu NP is quite free. Indeed, not only is the distribution within an NP quite free, genitives are also able to appear outside of the NP in which they are licensed.

An example has already been given in (9b), where the genitive marked *Nadya* appears to the left of its head ('book'). Clearer examples are provided in (18).

(18) a. **gari**   nadya=ne   **ʊs=ki**   bazar=mẽ   dekʰ-i
      car.F.Sg.Nom Nadya.F.Sg=Erg Pron.3.Sg.Obl=Gen.F.Sg market.M.Sg=in see-Perf.F.Sg
      'His/her car, Nadya saw in the market.' ('car' is topicalized)   Urdu

   b. **kɪtab**   tʊm=ne **kɪs=ki**   xɑrid-i?
      book.F.Sg.Nom you=Erg who.Sg.Obl=Gen.F.Sg buy-Perf.F.Sg
      'Whose book did you buy?'   Urdu

   c. lɑɽka,   **jɪs=ki**   nadya=ne   xarid-i   **kɪtab**
      boy.M.Sg.Nom RelPron.Sg.Obl=Gen.F.Sg Nadya.F.Sg=Erg buy-Perf.F.Sg book.F.Sg
      'the boy whose book Nadya bought'   Urdu

Not all languages allow this kind of separation between the possessor and the pronominal genitives (English does not, for example; see Szabolcsi (1983) for seminal work on this issue). For West Flemish, Haegeman (2004) argues that instances of remote possessors are not derived by movement, but are instances of possessor doubling. It is doubtful that this analysis can be carried over to the Urdu facts, but an investigation of the full set of Urdu facts still needs to be done (as far as we are aware). Note that the long distance dependency (as it would be called in LFG) between the possessor and the possessed can often be resolved unambiguously via agreement, as is the case in (18). But agreement cannot be relied upon to always produce an unambiguous resolution of the long distance dependency, as the masculine morphology not only also serves as a default, but is furthermore ambiguous between plural and singular in the oblique.

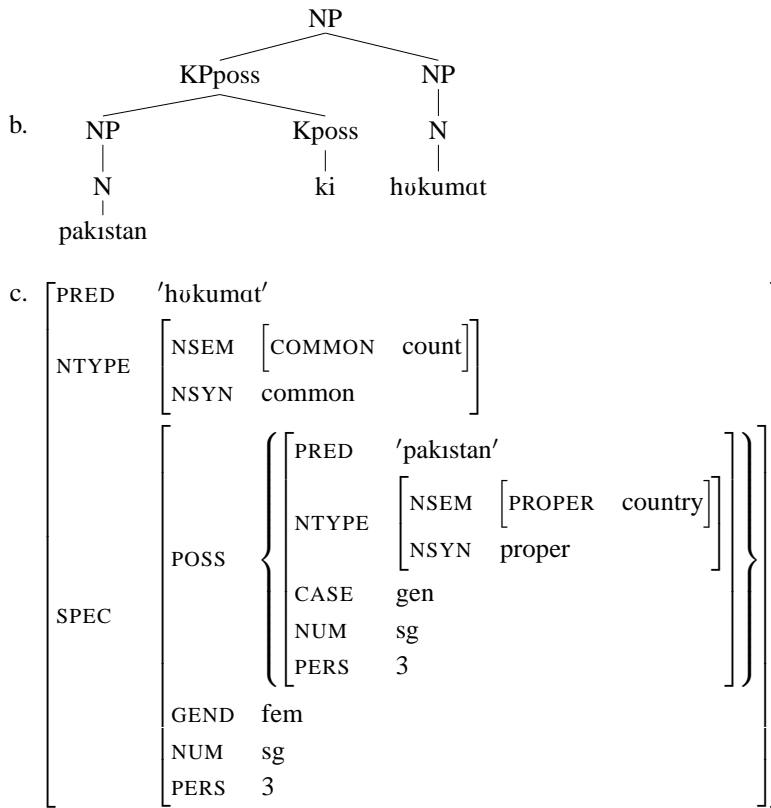## 4.4 Analysis of the Genitive Case Clitic

The morphosyntactic facts provided in section 4.2 support an analysis of the genitive case marker as a clitic. Its placement is phrasal rather than lexical; it therefore seems to belong in the domain of syntax. One consequence of the *Lexical Integrity Principle* (e.g., Dalrymple 2001, Bresnan 1982) is that the syntax does not determine morphological structure: words are assumed to be built by different rules than those of syntax.[5] Bound morphemes can therefore not appear independently in the phrase structure and are therefore also not represented with an independent terminal node. However, since case clitics seem to be independent functional items that are placed by the syntax, Butt and King (2004), also working within LFG, represent these clitics with an independent terminal node. They introduce them as functional heads of a KP as illustrated in the general scheme in (19):

(19)   KP → NP K

LFG's modular framework allows for a thorough analysis of the genitive case clitic. The two basic representations for the syntax are the c-(onstituent) structure, which encodes the basic constituency structure and linear hierarchy of the elements, and the f-(unctional) structure, which abstracts away from the surface positions and models grammatical relations, functional information and other dependencies. In the following example, the f- and c-structure of the genitive construction in (20) are displayed.

---

[5]In the ParGram grammars this difference is also realized as one of formal power. The morphological component is finite state, while the syntax allows for recursive rules.

(20) a. pakıstan=ki   hʊkumɑt
       Pakistan=Gen.F.Sg government.F.Sg
       'The government of Pakistan'   Urdu

b.

```
                        NP
                 ┌──────────────┐
              KPposs            NP
           ┌──────────┐         │
          NP        Kposs       N
          │           │         │
          N           ki      hʊkumɑt
          │
       pakıstan
```

c.

$$
\begin{bmatrix}
\text{PRED} & \text{'hʊkumɑt'} \\
\text{NTYPE} & \begin{bmatrix} \text{NSEM} & \begin{bmatrix} \text{COMMON} & \text{count} \end{bmatrix} \\ \text{NSYN} & \text{common} \end{bmatrix} \\
\text{SPEC} & \begin{bmatrix} \text{POSS} & \left\{ \begin{bmatrix} \text{PRED} & \text{'pakıstan'} \\ \text{NTYPE} & \begin{bmatrix} \text{NSEM} & \begin{bmatrix} \text{PROPER} & \text{country} \end{bmatrix} \\ \text{NSYN} & \text{proper} \end{bmatrix} \\ \text{CASE} & \text{gen} \\ \text{NUM} & \text{sg} \\ \text{PERS} & 3 \end{bmatrix} \right\} \\ \text{GEND} & \text{fem} \\ \text{NUM} & \text{sg} \\ \text{PERS} & 3 \end{bmatrix}
\end{bmatrix}
$$

In the c-structure representation in (20b), the genitive case clitic *ki* is represented as the functional head of a KPposs.[6] Together with the noun *pakıstan*, it forms the specifier within an NP whose head noun is *hʊkumɑt*. The genitive also agrees in gender and number with this head noun. This is modelled at f-structure via feature unification (not visible as such in (20c)). The f-structure also encodes *hʊkumɑt* as the head noun of the construction and *pakıstan* as its specifier SPEC. More particularly, as a possessive POSS specifier.

In terms of prosodic alignment, none of the genitive constructions introduced in this section pose a serious problem. Any relation between two nouns that is expressed by the genitive receives the following basic prosodic and syntactic bracketing:

(21)   prosodic and syntactic bracketing for *mohan=ki bılli* 'Mohan's cat'.

syntactic bracketing	[[[mohan]$_{\text{NP}}$=[ki]$_{\text{K}}$]$_{\text{KP}}$	[bıll-i]$_{\text{NP}}$]$_{\text{NP}}$
prosodic bracketing	((mohan)$_\omega$=ki)$_\varphi$	((bıll-i)$_\omega$)$_\varphi$
	Mohan.M.Sg=Gen.F.Sg	cat-F.Sg
	'Mohan's cat'	

---

[6]One reviewer asks why the genitive clitic must be a functional head. The question is justified in that LFG would in principle allow us to analyze genitive and other case-marked phrases in a way that does not introduce the case marker as a head that projects a phrase. The analysis shown here was proposed by Butt and King (2004) for two reasons: 1) case marked phrases have a slightly different distribution in Urdu than nominative (unmarked) ones and a KP vs. NP distinction allows this to be modelled fairly straightforwardly; 2) case in Urdu works in a "constructive" manner. That is, as argued for by Nordlinger (1998) for Australian languages, case markers in Urdu help construct the functional analysis of a clause. That is, they are not mere feature bundles which reflect grammatical information. Rather, Butt and King (2004) posit a lexical semantic analysis of case in Urdu by which case markers have lexical entries and contribute quite a bit of syntactic and semantic information to the clause.

9

The case clitic always shares a mother node with the modifying noun, which at the same time is also its prosodic host. The syntactic and prosodic modules are therefore aligned, with XPs generally corresponding to prosodic phrases as assumed in the majority of the prosodic literature (cf. Selkirk 1995b).

5   The *Ezafe* Construction

We now turn to the Urdu *ezafe* construction. This is a loan construction from Persian, where it originated from an Old Iranian relative pronoun -*hya*. In Middle Iranian, the pronoun evolved into *y/i* and became specialized as a device for nominal attribution. With respect to modern Persian, the *ezafe* construction has been extensively discussed (Samiian 1983, 1994, Ghomeshi 1997, Kahnemuyipour 2000, Holmberg and Odden 2008, Larson and Yamakido 2008, Samvelian 2007). Some researchers have viewed the *ezafe* as being a clitic, while others classify it as part of nominal morphology.

A Persian example for the *ezafe* construction is provided in (22), which illustrates that Persian *ezafe*, in contrast to the generally head-final nature of the language and NPs in particular, allows the head noun to be initial, with modifiers licensed to its right. Modifiers of the head noun can include adjectives, nouns and some PPs and each of these in turn can function as the host for a further *ezafe*-construction. This stacking of *ezafe* is also illustrated by (22).

(22) [in ketâb]-e [kohne]-ye [bi   arzeš]-e maryam
     this book-Ez ancient-Ez without value-Ez Maryam
     'this ancient worthless book of Maryam's'   (Samvelian 2007, 606)   Persian

The Urdu construction functions similarly; however, in Urdu it is connected to forms of high literacy (i.e., poetry, novels, newspaper texts) and is therefore more restricted in use than its Persian counterpart. Some simple examples are shown in (23). In example (23a), the modifier is a noun (*kʰɪrad* 'wisdom'), while in (23b), the head noun is modified by an adjective (*bʋland* 'high'). The constructions show the peculiarity already described for the Persian *ezafe* construction in (22): while Persian and Urdu both generally are head-final languages, the *ezafe* construction is head-initial.

(23) a. arbab=e   kʰɪrad
          owner=Ez wisdom
          'wise person'   (Platts 1909, 99)   Urdu

      b. sɑda=e   bʋlɑnd
          voice=Ez high
          'a high voice'   (Delacy 2003, 100)   Urdu

The Urdu *ezafe* is prosodically incorporated into the head noun to its left — as in Persian, it is always pronounced as a unit with the head. While attached prosodically to the head on its left, the *ezafe* simultaneously licences a modifier to its right. This stands in contrast to the Urdu genitive construction examined in the previous section, which was seen to conform to the head-final pattern typical for Urdu. The genitive marker also prosodically attaches to the unit to its left, however it also simultaneously licenses it and forms a syntactical-functional unit with it. While *ezafe* and the genitive thus appear to be similar from a functional perspective, their prosodic and morphosyntactic realization differs considerably.

5.1   Persian *Ezafe* — Morphology or Syntax?

The combination of properties of the Persian *ezafe* construction have led to a wide discussion about its morphosyntactic status. Ghomeshi (1997) for example, analyses the *ezafe* as a phonological linker within X-bar theory. Her conclusion is that the *ezafe* never attaches to phrases but selects as its domain X⁰'s or bare (lexical) heads. Samvelian (2007), working within Head-driven Phrase Structure Grammar (HPSG),

challenges this view by introducing new data with respect to Persian *ezafe*. She demonstrates that the *ezafe* can attach to phrases and proposes that the *ezafe* is a *phrasal affix* attaching to nominal heads and marking them morphologically as expecting a modifier.[7] Unlike Anderson (2005, 1992), who assumes phrasal affixes to come into play postlexically, Samvelian, following proposals made by Zwicky (1987) and Miller (1992), analyses the *ezafe* as part of word level morphology and distinguishes clitics and phrasal affixes on this basis.[8]

In contrast to word-level inflectional affixes, which attach directly to their host and cannot be separated from them, phrasal affixes in Persian appear on the right edge of nominal constituents (usually non-maximal projections) and do not bear lexical stress. They are attached after the word-level affixes and cannot separate these from their hosts, as shown in (24). This is, in fact, a typical property of clitics (Zwicky and Pullum 1983, Criterion F).

(24) in   pesar-hâ-ye/*pesar-ye-hâ ahmaq
      this boy-Pl-Ez/boy-Ez-Pl   silly
      'these silly boys'   (Samvelian 2007, 619)   Persian

Samvelian also shows that the *ezafe* has wide scope over coordination. However, in contrast to what is generally assumed in the literature (e.g., Zwicky and Pullum 1983), she does not take this as a proof for a clitic status of the *ezafe*. Instead, she invokes arguments supporting the classification of the *ezafe* as a phrasal affix which, in her conception, is situated in the morphological component rather than being introduced postlexically as originally formulated by Anderson (1992). Her main argument involves data showing that the *ezafe* and other elements like the indefinite article *-i*, which she also analyses as a phrasal affix, seem to be in complementary distribution with respect to one another. Samvelian suspects an involvement of the *Haplology Criterion*,[9] which she considers to be a non-last-level phonological process. In order for the *ezafe* and other phrasal affixes to conform to the Haplology Criterion, both must be generated within the same level. According to her argumentation, the *ezafe* then cannot be introduced postlexically and must be generated within the morphology.

Bögel et al. (2008) challenge this assumption. First of all, the group of affixes Samvelian compares with the *ezafe* seem to belong to a totally different class. Not all of them have scope over coordination and show a much more promiscuous behavior with regard to their host. This weakens the argument of complementary distribution, because these elements will, in case of an *ezafe* being present, simply appear somewhere else. Furthermore, it is not clear why phonological processes like the *Haplology Criterion* or complementary distribution should be restricted to the prelexical morphological module.

Consider the illustration of haplology in (25) with respect to the English plural and genitive *'s* (for which Zwicky discussed haplology in 1987).

(25) a. The dog's   bones
             dog.Sg=Poss

   b. The dogs'   bones
             dog.Pl=Poss

---

[7]Other papers on Persian *ezafe* do not consider the prosodic and morphosyntactic statue of *ezafe* in depth and are therefore not discussed in this context. An exception is Larson and Yamakido (2008), who assume that the *ezafe* is a clitic that attaches to the left while forming a constituent with its complement to the right. Although they present no discussion of their assumption, we believe that this view is essentially right.

[8]The term *phrasal affix* was originally coined by Anderson (1992) in order to point out the parallelism of distribution and function between morphological affixes and clitics. However, phrasal affixes are clearly part of the postlexical (hence not morphological) component for Anderson.

[9]Definition of haplology: One syllable is deleted in the case of two identical or similar syllables, see for example Zwicky (1987).

In (25a), the possessive marker *'s* is not subject to haplology; in (25b) however, the case is different. Here, *dog* also contains the plural marker *-s* and the construction is therefore subject to haplology — the possessive *-s* has been deleted in the presence of a regular plural (cf. Zwicky 1987). It is not clear, however, that this process has to take place in the morphological component. The possessive marker in English poses similar problems to its classification as a clitic or affix as the *ezafe*. It cannot straightforwardly be described as being part of the morphological component; instead, there is evidence that the *'s* is distributed within the syntax. However, if the possessive marker is distributed within the syntax and the plural marker is clearly part of the morphological component, Samvelian's argument that the *ezafe* and the indefinite *-i* have to be generated within the same level because they are subject to haplology does not hold.

In this context, Anderson (2005) makes an interesting proposal. He reviews the argumentation surrounding the English phenomenon in (25) and suggests that the morphemes/clitics be integrated at the level of the syllable, as shown in (26) (Anderson 2005, 93).

(26)
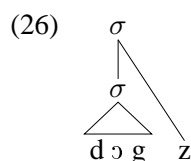
Thus, haplology is seen as applying within the phonological/prosodic component which syllabifies and groups syllables into feet in order to arrive at a prosodic structure of the clause. Under this view, analyzing a phenomenon as haplology does not automatically translate into an argument for placing it within the morphological component.

Another argument Samvelian (2007) makes is based on the question of access to word-level properties. Samvelian assumes that access to word-level properties can only take place within the morphological module. Since the Persian *ezafe* is restricted to nominal heads, it needs access to the part-of-speech class that is involved. However, this kind of part-of-speech information is generally needed to ensure syntactic well-formedness. For example, consider the placement of adverbs in a noun phrase: *very big dog* vs. *big very dog*. The adverb must have access to word-level properties of the things around it in order to ensure the right order. Alternatively, one could argue that the right order is guaranteed by a set of syntactic rules, but then this alternative could also just as well apply to the distribution of *ezafe*. Samvelian's argument of access to word-level properties therefore does not hold either.

Nevertheless, within HPSG, Samvelian comes to the conclusion that the *ezafe* should be generated within the morphological component. After considering the data for the Urdu *ezafe* construction in the following section, we come up with a different analysis of *ezafe*.

5.2 Urdu *Ezafe*

In the case of the Urdu *ezafe* construction, establishing clitic status is not as straightforward as with the genitive case marker. In this section we therefore take a closer look at Urdu *ezafe*, which, as already mentioned, is a loan construction from Persian. Persian was the language of the Mughal Court for several centuries and heavily influenced the language of the courtiers and the poets of the court. Urdu *ezafe* is mainly found in the high/literary language and in the newspapers. It remains productive, but modern speakers show a tendency to not use the construction (Schmidt 1999, 247) as part of the normal spoken language. In particular, *ezafe* constructions must be formed with words of Persian origin (which limits its range of applicability).

5.2.1 Basic Properties

The head of an *ezafe* construction is usually a noun, but as shown in (27) some adpositions are also allowed. These appear to be exactly those adpositions that have been argued to be noun-like by Samiian (1994) (see

also Larson and Yamakido 2008), presumably because they are locative expressions derived from former nouns. Full PPs are also allowed, as shown in (28).

(27) a. zer=e   zɑmin
          under=Ez earth.F.Sg
          'under the earth'   Urdu

      b. zer=e   hɑrasɑt
          under=Ez arrest
          'under arrest'   Urdu

(28) bɪh tɑrɑf=e   ʃɪmal
      to   direction.F.Sg=Ez north
      'to the direction of north'   Urdu

The modifier may be a noun as in (27) or (28) or an adjective as in (29b) and (30). As in Persian, stacking of *ezafe* constructions is allowed, although stacked constructions are not found as frequently as in Persian. Some examples are shown in (29).

(29) a. dʊʃmɑn=e   ɑrbab=e   vɑfa
          enemy.M.Sg=Ez people=Ez fidelity.M.Sg
          'the enemy of the people of fideliy'   (Asadullah Khan Ghalib)   Urdu

      b. hɑva=e   dor=e   me=e   xʊʃgɑvar
          air.F.Sg=Ez period=Ez wine=Ez pleasant
          'the air of the period of a pleasant wine'   (Haider Ali Atash)   Urdu

Most commonly, the heads and modifiers involved in *ezafe* constructions tend to be single words. The question thus arises whether XPs are in principle possible to the left and to the right of an *ezafe*. (28) already showed that the head could be a PP, not just a P. (30) shows that the head can in principle also allow the full structure of an NP.

(30) [ek bɑhʊt accʰ-a   [divan=e   am]]
      one very   good-M.Sg hall.M.Sg=Ez private
      'a very good private hall of audience'   Urdu

Similarly, the modifier can also be more complex. This is illustrated in (31) with a coordinated modifier 'star and moon'.

(31) pɑrcɑm=e [sɪtara   o   hɪlal]
      flag=Ez   star.M.Sg and crescent moon
      'the flag of the star and crescent'   (Pakistani National Anthem)   Urdu

5.2.2 Headedness/Alignment of the *Ezafe*

We have asserted that the Urdu *ezafe* construction is head-initial. This assertion can be verified via agreement facts. In (32), the adjective *baṛ-i* 'big' agrees in gender and number with the head noun *vadi* 'valley', which in case of (32a) is the initially positioned head of an *ezafe* construction.

(32) a. [**vadi**=ɛ   sɪndʰ]   bɑhʊt **baṛ-i**   hɛ
          valley.F.Sg=Ez Indus.M.Sg very   big-F.Sg be.Pres.3.Sg
          'The Indus valley is very big.'   Urdu

b. **vadi** bɑhʊt **baɽ-i** hɛ
valley.F.Sg very big-F.Sg be.Pres.3.Sg
'The valley is very big.' Urdu

In a genitive construction, on the other hand, agreement identifies the final noun as the head. In (33), the genitive clitic *ki* and the adjective *baɽ-i* 'big' both agree with the head noun *vadi* 'valley', which is the final element of the genitive construction.

(33) [sɪndʰ=**ki** **vadi**] bɑhʊt **baɽ-i** hɛ
Indus.M.Sg=Gen.F.Sg valley.F.Sg very big-F.Sg be.Pres.3.Sg
'The valley of the Indus/Sindh is very big.' Urdu

Semantically, (33) is almost identical to (32a) in that both constructions express a relation of possession. Syntactically, however, they clearly differ with respect to where the head is placed. (33) is head-initial, (32a) is head-final.

5.2.3 Separability

Section 4.3 showed that the genitive can be separated from its head and can appear at quite a distance from it. The Urdu *ezafe* construction does not display syntactic independence as the genitive, and generally the parts of an *ezafe* construction are not separable. However, Ghulam Raza (p.c.) points to examples as in (34) where the modifier can also appear before the head+*ezafe* so that both (34a) and (34b) are possible.

(34) a. mɑrd=e xub
man.M.Sg=Ez much
'a good man'

b. xub mɑrd=e
much man.M.Sg=Ez
'a good man'

These types of patterns are apparently mainly found in the older Urdu literature and at this point it is not clear to us how productive patterns as in (34) really are. However, that the parts making up an *ezafe* construction are in principle separable is shown by examples like (35), which a parenthetical intervenes between the head+*ezafe* and the modifier.[10]

(35) bɪh tɑrɑf=e mɑslɑn ʃɪmal
to direction.F.Sg=Ez like north
'to the direction of, say, north' Urdu

5.2.4 Semantics

The *ezafe* construction is not restricted to expressions of possession. Like the genitive construction (section 4), the *ezafe* has a much wider spectrum of meaning than the expression of possession. This can be seen quite clearly with respect to adjective modifiers. For example, in (36) the adjective *azam* 'great' does not stand in a possession relation to the 'minister', but simply modifies the head noun in an attributive manner.

(36) vɑzir=e azɑm
minister.M.Sg=Ez great
'prime minister' Urdu

---

[10]Rajesh Bhatt and one of the reviewers point out that the more restricted scrambling possibilities for *ezafe* as compared to the genitive construction may be related to the absence of agreement inside the *ezafe* construction vs. the existence of agreement within a genitive construction. We concur that this is a likely explanation which should be explored further.

Noun-noun constructions also do not necessarily all express a possessive relation. In (37), the head noun *talıb* 'seeker' is modified by the noun *ılm* 'knowledge'. In this case, the head noun clearly does not "possess" the modifying noun — the interpretation of this construction would rather be "seeker of knowledge" instead of "possessor of knowledge".

(37)   talıb=e   ılm
            seeker.M.Sg=Ez knowledge.M.Sg
            'student'   Urdu

Both the genitive and *ezafe* thus appear to be used to express a very vague relationship between two elements. One of those relationships is that of possession. In trying to understand the semantics of *ezafe* in Urdu, the closest analog seems to us to be that of compounding. That is, the *ezafe* allows for diverse set of semantic relations and mainly expresses that there is a relationship between the two elements. Note that compounding in Urdu appears to be very limited, but whether there is a connection between this and the heavy use of genitive and/or *ezafe* constructions is a topic for further research.

5.2.5   Urdu *Ezafe* — Clitic or Phrasal Affix?

As mentioned before, the question whether the *ezafe* should be considered to be a phrasal affix that is generated within the morphology (as proposed by Samvelian 2007) or whether the *ezafe* is a clitic has been extensively discussed within the community for Persian. If the *ezafe* turns out to be a clitic, then there remains the question as to how to analyse all of its morphosyntactic, semantic and prosodic properties.

In what follows, for the sake of completeness, we go through the CRITERIA proposed by Zwicky and Pullum (1983) to distinguish affixes and clitics with respect to *ezafe*.

> CRITERION A: Clitics can exhibit a low degree of selection with respect to their hosts,
> while affixes exhibit a high degree of selection with respect to their stems.

The *ezafe* is not fussy about different kinds of noun stems/classes. It does, however, require nouns of Persian origin to its left (these still form a number of distinct classes). It is thus more like an affix in that the degree of selction is fairly high.

> CRITERION B: Arbitrary gaps in the set of combinations are more characteristic of affixed words
> than of clitic groups.

When looking at *ezafe* constructions, there are no arbitrary gaps.

> CRITERION C: Morphophonological idiosyncrasies are more characteristic of affixed words
> than of clitic groups.

With respect to the *ezafe*, there are no unexpected phonological forms. Regardless of what sound the noun to the left of it ends in (vowels, different types of consonants), the *ezafe* is always expressed as an *-e*. CRITERION C therefore favors a clitic analysis.

> CRITERION D: Semantic idiosyncrasies are more characteristic of affixed words
> than of clitic groups

Semantic idiosyncrasies can not be observed with *ezafe*.

> CRITERION E: Syntactic rules can affect affixed words, but cannot affect clitic groups.

This CRITERION basically states that clitics and their hosts are not treated as a syntactic unit in contrast to affixed words. One possible syntactic operation to evaluate the behavior of the *ezafe* is coordination — and, just as the accusative case clitic in (39), the *ezafe* in (38) takes scope over a whole noun conjunction.

(38) [ye [mal   o daulɑt]]=e dʊnya
    this material and wealth=Ez world.F.Sg
    'this material and wealth of the world'   Urdu
    (from *zarb-e-kaleem* by Muhammad Iqbal)

(39) [mal   or daulɑt]=ko kʊma-o
    material and wealth=Acc earn-Imp.Rude
    'Earn/gather material and wealth!'   Urdu

A normal morphological affix is not able to do so, it has to attach to every single conjunct (this already has been demonstrated in (14b)).

> CRITERION F: Clitics can attach to material already containing clitics but affixes cannot.

In the case of the *ezafe*, the CRITERION F does not offer a clear distinction between a clitic and an affixal analysis as with the genitive case marker. The observation that the focus clitic *hi* can intervene between the stem and genitive case marker (repeated in (40)), does not extend to *ezafe*, as shown in (41).

(40)   bɑcc-õ=hi=ka   kʰana
      child-M.Pl.Obl=Foc=Gen.M.Sg food.M.Sg
      'The CHILDREN's food'   Urdu

(41) a. *xʊrak=hi=e   ʃer
        food.M.Sg=Foc=EZ lion.M.Sg
        'The lion's FOOD'   Urdu

      b. *xʊrak=e=hi   ʃer
        food.M.Sg=Ez=Foc lion.M.Sg
        'The lion's FOOD'   Urdu

      c. xʊrak=e   ʃer=hi
        food.M.Sg=Ez lion.M.Sg=Foc
        'The LION's food' / 'The lion's FOOD'   Urdu

In (41a), the focus clitic *hi* cannot intervene between the *ezafe* and the stem. However, unlike with the genitive, the focus clitic is not allowed anywhere within the construction, as can be seen in (41b). The only possibility for a placement of the focus clitic is after the whole *ezafe* construction, where its position is neutral — it can either focus *ʃer* 'lion' or *xʊrak* 'food'. Therefore, the position of the focus clitic does not necessarily give evidence for the *ezafe*'s classification as a clitic. But it does not rule out the classification either, since it is not possible for the clitic to attach directly after the *ezafe*.

Considering all of these facts, we take the *ezafe* to be a clitic, even though it is not such a clear case as the genitive case marker. As argued in section 5.1, Samvelian's (2007) argumentation for the *ezafe* being analyzed in the morphological component does not hold. We see no reason for the *ezafe* not to come into play postlexically and assume that Samvelian's analysis is partly motivated by the architecture assumed by her chosen framework, HPSG. We contend that within the modular architecture of LFG, a postlexical analysis follows naturally. Therefore, we now provide a concrete analysis of the Urdu *ezafe* in the following section, based on the Urdu XLE grammar being developed in Konstanz.

6   A Modular Analysis of Urdu *Ezafe*

Given our analysis of the *ezafe* as a clitic, we now address the question of how to analyse its morphosyntactic, semantic and prosodic properties. We implemented the *ezafe* as part of our on-going work on building a computational grammar of Urdu within the LFG-based ParGram project.

LFG is an inherently modular theory of grammar. As already discussed in section 2, we adhere to the *Principle of Lexical Integrity*, which restricts the formation of words to the morphological module. The *ezafe*, being a clitic, is therefore treated within the syntactic domain, which interacts with prosody and postlexical phonology. Prosodic phrasing is part of Prosodic Phonology (Selkirk 1984, 1986, Nespor and Vogel 1986, Selkirk 1995b), which we view as an additional module of the grammar. In our implementation we model this via a *prosodic projection p* (Butt and King 1998).[11]

Prosody is of great interest in a clitic analysis of the *ezafe*. Clitics are considered to be "little words", which, being prosodically deficient, depend on a prosodic host. They are phrased together with another prosodic word as part of the prosodic phrasing (Inkelas and Zec 1990, Selkirk 1995a). Phrasing clitics prosodically together with their host is therefore part of the prosodic component, but not the part of the morphological component. In the following sections, a brief description of each of the LFG modules involved in the analysis of an *ezafe* construction is given.

6.1 The Morphological Component

The morphological module contains a list of Urdu lexical stems plus a set of patterns which determine how these stems can be inflected. Thus, the lexicon does not contain full forms, but forms words dynamically. It also allows for more than one morphological analysis per word form when this is relevant. An example is shown in (42), where two inflected forms of the verb *likʰ* 'write' are analyzed.

(42)   Surface form and analysis of *likʰti* 'is writing' and *likʰa* 'wrote, write'

surface form	likʰti	likʰa
analysis	likʰ+Verb+Impf+Fem+Sg	likʰ+Verb+Perf+Masc+Sg
		likʰ+Verb+Caus+Imp+2P+Sg+Rude

The morphology operates in what is generally known as the "two-level" manner. That is, a surface form is related to an abstract analysis which contains the underlying lemma (stem or root) plus a sequence of abstract *tags* (e.g., Fem, Sg), which are independent of any particular syntactic or morphological framework. These tags represent the morphological analysis and descriptively encode what is deemed to be important information about this word form (designing the tags is part of the linguistic analysis that needs to be done in building up the morphological module).

Our concrete implementation of the Urdu morphological module uses the finite state techniques developed by Beesley and Karttunen (2003), as is standard for ParGram grammars. These finite-state techniques are not limited to concatenation (as used to be the case), but allow an insightful treatment of non-concatenative morphology like reduplication, infixation, or stem interdigitation (Beesley and Karttunen 2000). For details on the Urdu morphological analyzer, see Bögel et al. (2007).

Using finite state techniques for morpho(phono)logical analysis is not only computationally extremely efficient, it has also been shown to be basically equivalent to at least one theoretical perspective of morphology, namely that of Paradigm-Function Morphology or Realizational Morphology (Stump 2001, Karttunen 2003). Thus, while we here present a very concrete implementation, the architecture cannot be said to be a mere implementational artefact. Rather, it is a concrete modeling of ideas that are found in the theoretical morphological literature.

Within the morphological component, the head noun and the modifier of the *ezafe* construction each receive their own, independent morphological analyses. For a construction as in (43) or (44), the nouns *hʊkumat* 'government' and *pakɪstan* receive the morphological analyses shown in (45).

---

[11]Note that we are currently developing an alternative architecture for the phonology-syntax interface than the one assumed here (Bögel et al. 2009, 2010). The alternative architecture assumes a model whereby the prosodic component and the syntactic component are independent modules in their own right, but where the prosodic representations and the syntactic representations are lined up with one another (amount of prosodic interruptions of syntactic constituents is kept as minimal as possible) via a set of Optimality-Theoretic style constraints. This seems to do better justice to the phenomena observed at the prosody-syntax interface. However, a thorough discussion would lead us too far afield and is also not wholly relevant to this paper.

(43) hʊkumɑt=e   pakıstan
     government.F.Sg=Ez Pakistan
     'Pakistan's government'   *Urdu*

(44) pakıstan=ki   hʊkumɑt
     Pakistan=Gen.F.Sg government.F.Sg
     'The government of Pakistan'   *Urdu*

(45)   `hʊkumɑt +Noun +Persian +Fem +Sg`
     `pakıstan +Noun +Country`

In our particular architecture, these abstract morphological tags are passed on to the LFG grammar, running on the XLE grammar development platform, a software developed at the Palo Alto Research Center (PARC, Crouch et al. 2010). The tags associated with the word forms that form the terminal nodes in the syntax are interpreted by the grammar according to its needs and expectations. For example, the tag `+Sg` when associated with a noun is associated with the information that the noun is of number singular (in LFG terms: (↑NUM) = sg). If the same tag is encountered in conjunction with a verb (`+Verb`) and the language being analyzed is English, for example, then this information is interpreted as saying that the subject of the clause needs to be singular (in LFG terms: (↑SUBJ NUM) = sg). For details as to how this morphology-syntax interface works, see Kaplan et al. (2004).[12] In our grammar, the `+Persian` along with a `+Arabic` tag is used to constrain the distribution and morphological paradigms for words of Persian and Arabic origin (certain words allow only certain prefixes, the *ezafe* construction is constrained to appear mainly with words of Persian origin, etc.).

The *ezafe* and the genitive clitics are not analyzed as affixal morphemes. Instead, we treat them as independent lexical elements. The content of the lexical entries is shown in (46).

(46) a.   e   EZ   (↑ MOD MOD-TYPE) = ezafe.

      b.   ki   Kposs   (↑ CASE) = gen
                                 (↑ GEND) = fem
                                 (↑ NUM) = sg.

These entries state that there is an *ezafe* that has the Part-of-Speech EZ and carries the functional information that there is a modification (MOD) and that the type of this modification (MOD-TYPE) is of the *ezafe* type. We analyse the functional contribution of *ezafe* as a type of modifier of the head noun in order to express the rather wide set of relations the modifier can bear with respect to the head noun.

For the genitive, we show the feminine singular version for ease of exposition. The inflected versions of the *k-* genitive are actually created/analyzed within the morphological module, but the end result is an inflected version. In this case, the fully inflected form contains the information that this is a genitive and that it is feminine and singular.

6.2 C- and F-structure Analysis

As already described in section 4.4 for the genitive case clitic, the c(onstituent)-structure and the f(unctional)-structure form the core of LFG syntax. With respect to the f-structure, we posit the analysis shown in (47). Being the head noun, *hʊkumɑt* 'government' is encoded as the main predicate (PRED) of the construction. This main predicate contains a modifying constituent MOD (*pakıstan*), which is licensed by the *ezafe* at c-structure. The nature of this relationship is expressed via the modification type: MOD-TYPE ezafe.

With respect to the c-structure, there are several differing possibilities. One possibility is to analyze the *ezafe* and the modifier it licenses as forming a constituent and excluding the head (48a). Another possibility is to assume a tripartite structure as in (48b) (nothing legislates against ternary trees in LFG per se). Yet another structure would group the *ezafe* together with its head and exclude the modifier (48c).

---

[12]This conception/architecture is in fact quite similar to what Sadler and Spencer (2002) propose in terms of projections.

(47)
$$
\begin{bmatrix}
\text{PRED} & \text{'h\textupsilon kum\textalpha t'} \\[4pt]
\text{NTYPE} & \begin{bmatrix} \text{NSEM} & \begin{bmatrix} \text{COMMON} & \text{count} \end{bmatrix} \\ \text{NSYN} & \text{common} \end{bmatrix} \\[10pt]
\text{MOD} & \left\{ \begin{bmatrix} \text{PRED} & \text{'pak\textsci stan'} \\[4pt] \text{NTYPE} & \begin{bmatrix} \text{NSEM} & \begin{bmatrix} \text{PROPER} & \text{country} \end{bmatrix} \\ \text{NSYN} & \text{proper} \end{bmatrix} \\[4pt] \text{CASE} & \text{gen} \\ \text{NUM} & \text{sg} \\ \text{PERS} & 3 \\ \text{MOD-TYPE} & \text{ezafe} \end{bmatrix} \right\} \\[10pt]
\text{GEND} & \text{fem} \\
\text{NUM} & \text{sg} \\
\text{PERS} & 3
\end{bmatrix}
$$

(48) a. NP [Ez XP]

    b. [NP Ez XP]

    c. [NP Ez] XP

Larson and Yamakido (2008) assume a general structure as in (48a) and we favor this structure as well (this is what is currently implemented in the Urdu grammar). The *ezafe* construction exhibits the following properties that we must take into account: 1) the head must always be nominal (or a PP), the XP modifiers can be nominal or adjectival (and possibly also prepositional, as in Persian); 2) the head must be available to be agreed with by elements outside of the *ezafe* construction; 3) the presence of the *ezafe* licenses the modifying XP; 4) the structure can be recursive; 5) it looks like the modifying XP can in principle appear before the head+*ezafe*.

Possibility (48c) models property 5 well, but embeds the functional head of the construction fairly far down, as shown in the recursive sketch of this possibility in (49). This is problematic because the head is difficult to access for agreement purposes and it is also difficult to state a constraint that just when the XP+*ezafe* is initial, the XP is restricted to be nominal (or a PP). The licensing of the modifying XP also becomes a matter of stating a long distance dependency between the *ezafe* and the modifying constituent that must be propagated up and down through various levels of the tree.[13]

---

[13]Again, a reviewer asks why the *ezafe* should be considered to project a tree structure. In this case, the answer is that it is immaterial to us what the intermediate nodes are called. We could use Y′ez/YPez or just Y′/YP instead of Ez′/EzP. That is, the tree in (49) could also look as follows:
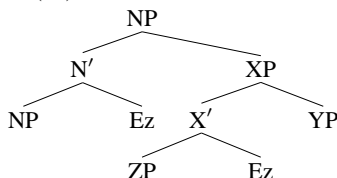

We chose to follow a standard X-bar schema for ease of exposition. Generally LFG allows a fairly free nomenclature and branching of c-structures. This is because the functional information is represented at f-structure, not c-structure. The c-structure is meant to capture linear order and the hierarchical arrangement of constituents.

(49) Recursive Version of (49c)

```
                    EzP
            _____|_____
          Ez'              EzP
       ____|____        ____|____
     NP        Ez     Ez'        XP
                   ____|____
                 XP        Ez
```
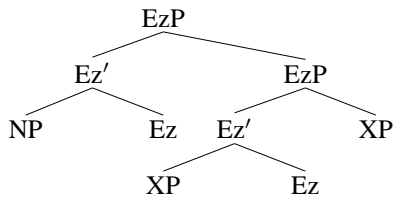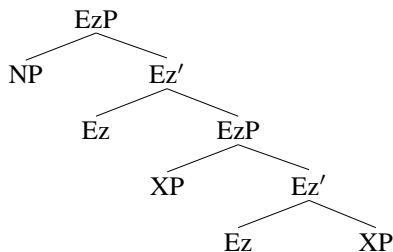
Another possibility is the structure in (48b). This could play out as shown in (50) for a recursive structure. In this structure, the first NP of the *ezafe* construction is easily identifiable and thus easily accessible for agreement and for a statement of the constraint that it has to be nominal. The modifier is straightforwardly introduced by the ternary structure, which includes the *ezafe* as a sister (so no issue with long distance dependencies) and the modifier XP/EzP could be free to appear outside of the *ezafe* construction as constituent.

(50) Recursive Version of (48b)

```
            EzP
        _____|_____
      NP   Ez    EzP
             _____|_____
            XP    Ez    XP
```
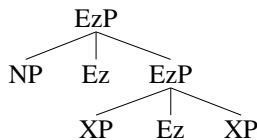
The same advantages and disadvantages are offered by (48c), with the exception that: 1) on the positive side the *ezafe* can be interpreted as licensing the XP modifier in its complement position; 2) on the negative side it is harder to see how the modifier XP could act as a constituent independently of the *ezafe*, i.e. property 5 is not accounted for well (but then, it is as yet unclear to us whether it is still part of the modern language).

(51) Recursive Version of (48a)

```
            EzP
        _____|_____
      NP         Ez'
              ____|____
            Ez        EzP
                  _____|_____
                 XP        Ez'
                       ____|____
                     Ez        XP
```

Given the data adduced for Urdu *ezafe* so far, the structures in (48a) and (48b) are thus both viable options. (48a) assumes that the *ezafe* forms a constituent with the complement it introduces and this could be motivated by its historical source as a relative clause. (48b), in contrast, has the structure typical for coordinations in LFG. If one saw *ezafe* as a kind of coordinator or linker, rather than the licenser of a complement, then (48b) would be more appropriate.

For the purposes of this paper, we remain agnostic between (48a) and (48b). Both illustrate the same point with respect to the prosody, which we discuss in the next section. Before discussing the prosodic analysis, however, we here briefly respond to a request by reviewers to situate the syntactic analysis of Urdu *ezafe* with respect to recent proposals for Persian/Iranian *ezafe* by den Dikken and Singhapreecha (2004), Larson and Yamakido (2008) and Larson (2010).

den Dikken and Singhapreecha (2004) posit an analysis of *Predicate Inversion* for several types of *of*-phrases across languages. They argue that the various 'of's must be analyzed as linkers that are inserted when a small clause is inverted due to information-structural demands. The word order effect of this inversion is then undone due to remnant movement. They include an analysis of Persian *ezafe*. Extending

their analysis to Urdu, the derivation for a simple example would be: 1) Small clausse: [NP AP]; 2) Predicate Inversion triggered by informational structural needs and giving rise to topic/focus structure: [AP ezafe NP]; 3) Remnant Movement (plus head movement of *ezafe*): NP ezafe AP. However, den Dikken and Singhapreecha (2004) provide no evidence for the informational structural effect triggering Predicate Inversion in Persian. That is, there is no evidence that the topic/focus readings they associate with Predicate Inversion exist at all in Persian. We also cannot find any evidence for them with respect to Urdu *ezafe* and are furthermore troubled by the fact that there is no non-inverted version of the ezafe construction, as one would expect given the proposed derivation.

In contrast, Larson and Yamakido (2008) analyze Persian *ezafe* as a case marker that functions like a "generalized genitive preposition" within the DP. They base their argumentation and data on Samaiian's (1994) original proposal that *ezafe* should be understood as a type of case marker. However, Samvelian (2007) adduces new data which argues against Samaiian's (1994) original conclusion and Samvelian dismisses the case marker theory as untenable. Samvelian's paper is not cited by Larson and Yamakido so presumably they were not aware of this work.

Under Larson and Yamakido's (2008) analysis, *ezafe* is used to license modifiers in their base position within a DP. As already mentioned, Larson and Yamakido analyze the *ezafe* as forming an XP with its complement, but also recognize its status as a clitic which cliticizes to an element to its left. We find Larson and Yamakido's analysis interesting, but to be convinced by it, we would need to see how the analysis of *ezafe* as a case marker fits in with an analysis of the Persian case system as a whole and with respect to the Persian nominal genitive in particular. Larson and Yamakido also do not discuss how the indefinite Persian morpheme *-i* is dealt with within the structure of the proposed DP, nor does it become clear how quantifiers, etc. are dealt with within the recursive ezafe DP structure.

We also find Larson's (2010) proposal that Chinese *de* be analyzed as a type of "reverse ezafe" very interesting. Larson (2010) and Larson and Yamakido (2008) draw a connection between the expression of Persian *ezafe*, Chinese *de* and English *of* and possessive *'s*. We agree that there is presumably a connection, but also believe that the similarity lies not in Chinese *de* and Persian *ezafe* being case licensers/markers, but in the expression of vague relations between two entities in general. Urdu (and Persian) do not contain the verb 'have' nor does Urdu make much use of compounding, another device to express a range of relations between two elements, among them possession. A crosslinguistic study exploring the use of genitives, *ezafe*-type markers, the presence or absence of 'have' and the productivity of compounding should prove to be instructive and should be able to take Larson's comparative work further (cf. also the ideas of Ritter and Rosen (1997), Belvin and den Dikken (1997), Harley (1998) on English *have* as a generalized linker).

In sum, we see no evidence in Urdu for the *Predicate Inversion* analysis proposed by den Dikken and Singhapreecha (2004) and, as was shown in the paper, case marking and *ezafe* display a very different morphosyntactic distribution in Urdu, rendering Larson's proposal that the *ezafe* should be analyzed as a type of case marker implausible. We believe that *ezafe* does function as a type of linker between two entities. How this kind of vague linking relation expressed by *ezafe* fits into the larger typology of linking by nominal genitives, 'have', or compounding remains to be determined.

6.3 Prosody

In a phrase like (43), namely, *hukumat e pakıstan*, the intonational break is after the *ezafe* and never before it. Furthermore, as shown in section 5.2.5, nothing can intervene betwen the *ezafe* and the head noun. We therefore conclude that the *ezafe* is prosodically attached to the word on its left. In the previous section, we settled on two possible syntactic structures for Urdu *ezafe*. Neither of these structures groups the *ezafe* together with its head in a constituent and we are thus faced with an misalignment between the syntax and the prosody of the *ezafe* construction.

However, given LFG's modular architecture, we are not forced to align prosodic and syntactic constituency. Rather, a modular architecture expects mismatches across modules. In our implementation, we assume that major syntactic phrases generally correspond to p(rosodic)-phrases. This is in line with most

algorithms assumed for a mapping between syntax and prosody (e.g., see Selkirk 1995b). Elements like *ezafe*, which cannot form an independent p(rosodic)-word, must cliticize onto a host element. The *ezafe* is an enclitic and cliticizes to a host to its left.

In our implementation, we follow Butt and King's (1998) original proposal for integrating a p(rosodic)-structure into standard LFG. The p-structure is systematically related to the c-structure representation, but the prosodic constituency is allowed to systematically differ from the syntactic constituency.

Neither of the syntactic representations discussed positively in section 6.2 group the *ezafe* with the head in the syntactic representation. However, in terms of prosodic constituency, the *ezafe* is grouped together with the head, which is its prosodic host. This is shown in Figure 2.[14] As can be seen, the construction is analyzed as a prosodic phrase which consists of two prosodic words: *pakıstan* and *hʋkumat* 'government'. The *ezafe*, being a prosodically deficient clitic, is not an independent prosodic word (P-WORD). At p-structure, it is only registered as [CL-FORM ezafe], where CL stands for "clitic" and is encoded as having been incorporated into the domain of the P-WORD *hʋkumat*. The basic prosodic bracketing, represented by Figure 2 is also illustrated in (52) via the more usual form of prosodic bracketing.
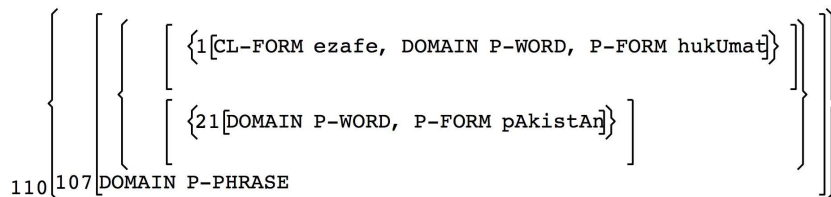
$$
110\left[107\begin{array}{l} \left\{ \left[\left\{ \left[\{1\left[\text{CL-FORM ezafe, DOMAIN P-WORD, P-FORM hukUmat}\right]\} \right] \right. \right. \\ \left. \left. \left[\{21\left[\text{DOMAIN P-WORD, P-FORM pAkistAn}\right]\} \right] \right\} \right] \right\} \\ \text{DOMAIN P-PHRASE} \end{array}\right]
$$

Figure 2: p-structure analysis of *hʋkumat e pakıstan*

(52) $((\text{hʋkumɑt})_\omega \text{ e})_\omega (\text{pakıstan})_\omega)_\varphi$

We have now successfully modelled all the properties of the *ezafe* construction and conclude that the analysis in terms of the independent modules of morphology (this includes lexical phonological processes), syntax and postlexical prosody provide exactly the right results for Urdu *ezafe*.

Under our analysis, Urdu *ezafe* can be considered to be a phrasal affix in the sense of Anderson in that it is a clitic which attaches to phrases, whose function is analogous to that played by morphological affixes and whose distribution is constrained. Most importantly, its prosodic properties are dealt with postlexically and not in the morphological component. This stands in contrast to Samvelian's idea that phrasal affixes be dealt with within a morphological component.

7 Conclusion

In our paper we have described two possessive constructions in Urdu. One is a genitive which is formed with a case marker that, unlike the other case markers in Urdu, inflects for number and gender. The other is an *ezafe* construction borrowed into Urdu from Persian. The two constructions differ in that the genitive respects the usual head-final pattern of the language while the *ezafe* displays a head-initial pattern. The genitive case marker is contained in a constituent/unit with the word to its left with respect to both syntax and prosody. The better syntactic analyses of the *ezafe* constructions, on the other hand, show a mismatch between prosody and syntax: while *ezafe* incorporates prosodically into the word on its left, it does not simultaneously need to form a syntactic constituent with it.

We argued that both the genitive case marker and the *ezafe* should be analyed as clitics and that the properties of *ezafe* can be modelled straightforwardly within LFG's modular architecture without needing to fall back on an otherwise unmotivated analysis by which *ezafe* is introduced as a morphological affix

---

[14]The structure is shown here exactly as it is produced by our computational implementation. Capital letters in the Urdu words signal long vowels, the numbers are a grammar internal bookkeeping device and refer to the corresponding nodes at c-structure.

in the morphological component. The prosodic clitic properties of both the genitive marker and *ezafe* are modelled postlexically at p(rosodic)-structure, their syntactic properties are modelled at c(onstituent)-structure and f(unctional)-structure.

This analysis contrasts with the analysis proposed by Samvelian (2007) for Persian *ezafe*. She sees *ezafe* as a *phrasal affix* that is generated within the morphology but attaches to phrases. We argued against her analysis of *ezafe*, contending that the Persian data is compatible with a view of Persian *ezafe* as a clitic whose prosodic properties are due to postlexical phonology.

Indeed, our analysis is closer to the original formulation by Anderson (1992), who sees a phrasal affix as a clitic whose prosodic distribution is handled postlexically. For him the "affix" part of phrasal affix encodes his idea that the distribution and function of morphological affixes is mirrored by the distribution and function of some clitics, which he sees as the "morphology of phrases". This morphology of phrases also includes phenomena like the English possessive *'s*, which is generated postlexically under his analysis. The notion of phrasal affixes as elements with special syntactic distribution and a function that is analogous to that played by morphological affixes works well with our understanding of the Urdu possessive clitics.

In closing, we note that what Anderson's idea of phrasal affixes also seems to be pointing to is a diachronic process whereby today's clitics generally tend to end up being the morphological affixes of tomorrow. Thus, his analogy between the morphology of words and the morphology of phrases appears to really be about a diachronic relationship.

References

Anderson, Stephen. 1992. *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Anderson, Stephen. 2005. *Aspects of the Theory of Clitics*. Oxford: Oxford University Press.

Asudeh, Ash and Ida Toivonen. 2009. Lexical-Functional Grammar. In B. Heine and H. Narrog, eds., *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press. In Press.

Beesley, Kenneth and Lauri Karttunen. 2003. *Finite State Morphology*. Stanford: CSLI Publications.

Beesley, Kenneth R. and Lauri Karttunen. 2000. Finite-state non-concatenative morphotactics. In *SIGPHON-2000. Fifth Workshop of the ACL Special Interest Group in Computational Phonology*. Centre Universitaire, Luxembourg.

Belvin, Robert and Marcel den Dikken. 1997. *There* happens *to, be, have*. *Lingua* 101:151–183.

Bögel, Tina. 2010. Pashto (endo-)clitics in a parallel architecture. In M. Butt and T. H. King, eds., *Proceedings of LFG10*. Stanford: CSLI Publications.

Bögel, Tina, Miriam Butt, Annette Hautli, and Sebastian Sulger. 2007. Developing a finite-state morphological analyzer for Urdu and Hindi. In *Proceedings of the Sixth International Workshop on Finite-State Methods and Natural Language Processing*. Potsdam.

Bögel, Tina, Miriam Butt, Ronald M. Kaplan, Tracy H. King, and John T. Maxwell III. 2009. Prosodic phonology in LFG: a new proposal. In *Proceedings of LFG09*. Stanford: CSLI Publications.

Bögel, Tina, Miriam Butt, Ronald M. Kaplan, Tracy H. King, and John T. Maxwell III. 2010. Second position and the prosody-syntax interface. In *Proceedings of LFG10*. Stanford: CSLI Publications.

Bögel, Tina, Miriam Butt, and Sebastian Sulger. 2008. Urdu ezafe and the morphology-syntax interface. In *Proceedings of LFG08*. Stanford: CSLI Publications.

Bresnan, Joan. 1982. *The Mental Representation of Grammatical Relations*. Cambridge: The MIT Press.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell.

Broadwell, George Aaron. 2008. Turkish suspended affixation is lexical sharing. In *Proceedings of LFG08*. Stanford: CSLI Publications.

Butt, Miriam, Helge Dyvik, Tracy Holloway King, Hiroshi Masuichi, and Christian Rohrer. 2002. The Parallel Grammar project. In J. Carroll, N. Oostdijk, and R. Sutcliffe, eds., *Proceedings of the Workshop on Grammar Engineering and Evaluation held as part of the 19th International Conference on Computational Linguistics (COLING)*, pages 1–7. Taipei, Taiwan.

Butt, Miriam and Tracy Holloway King. 1998. Interfacing phonology with LFG. In *Proceedings of LFG98*. Stanford: CSLI Publications.

Butt, Miriam and Tracy Holloway King. 2004. The status of case. In V. Dayal and A. Mahajan, eds., *Clause Structure in South Asian Languages*, pages 153–198. Dordrecht: Kluwer Academic Publishers.

Butt, Miriam, Tracy Holloway King, Maria-Eugenia Niño, and Frédérique Segond. 1999. *A Grammar Writer's Cookbook*. Stanford: CSLI Publications.

Crouch, Dick, Mary Dalrymple, Ron Kaplan, Tracy King, John Maxwell, and Paula Newman. 2010. *XLE Documentation*. On-line documentation, Palo Alto Research Center (PARC).

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

Delacy, Richard. 2003. *Teach Yourself Beginner's Urdu Script*. London: Hodder Education.

den Dikken, Marcel and Pornsiri Singhapreecha. 2004. Complex noun phrases and linkers. *Syntax* 1(7):1–54.

Ghomeshi, Jila. 1997. Non-projecting nouns and the ezafe construction in Persian. *Natural Language and Linguistic Theory* 15(4):729–788.

Haegeman, Liliane. 2004. DP periphery and clausal periphery: possessor doubling in West Flemish? Relating nominal periphery to clausal periphery. In D. Adger, C. de Cat, and G. Tsoulas, eds., *Peripheries*, pages 211–240. Dordrecht: Kluwer Academic Publishers.

Halvorsen, Per-Kristian and Ronald Kaplan. 1988. Projections and semantic description in Lexical-Functional Grammar. In *Proceedings of the International Conference on Fifth Generation Computer Systems*, pages 1116–1122.

Harley, Heidi. 1998. You're having me on: Aspects of *have*. In J. Guéron and A. Zribi-Hertz, eds., *La grammaire de la possession*, pages 195–226. Paris: Université Paris X - Nanterre.

Holmberg, Anders and David Odden. 2008. The noun phrase in Hawrami. In S. Karimi, V. Samiian, and D. Stilo, eds., *Aspects of Iranian Linguistics*. Newcastle: Cambridge Scholars Publishing.

Inkelas, Sharon and Draga Zec, eds. 1990. *The Phonology-Syntax Connection*. Stanford: CSLI Publications.

Kahnemuyipour, Arsalan. 2000. Persian ezafe construction revisited: Evidence for modifier phrase. In *The 2000 Annual Conference of the Canadian Linguistic Association*. Ottawa: Cahiers Linguistique d'Ottawa.

Kaplan, Ron, John T. Maxwell III, Tracy Holloway King, and Richard Crouch. 2004. Integrating finite-state technology with deep LFG grammars. In *Proceedings of the Workshop on Combining Shallow and Deep Processing for NLP (ESSLLI)*.

Karttunen, Lauri. 2003. Computing with realizational morphology. In A. Gelbukh, ed., *Computational Linguistics and Intelligent Text Processing*, vol. 2588, pages 205–216. Heidelberg: Springer Verlag. Lecture Notes in Computer Science.

Kellogg, S. H. 1893. *Grammar of the Hindi Language*. Delhi: Munshiram Manoharlal Publishers Pvt. Ltd. Second Edition, reprinted 1990.

King, Tracy Holloway. 1997. Focus domains and discourse function structure. In M. Butt and T. H. King, eds., *Proceedings of LFG97*. Stanford: CSLI Publications.

Larson, Richard. 2010. Chinese as a reverse Ezafe language. *Juyanxue Luncong* In Press.

Larson, Richard and Hiroko Yamakido. 2008. Ezafe and the deep position of nominal modifiers. In L. McNally and C. Kennedy, eds., *Adjectives and Adverbs: Syntax, Semantics and Discourse*. Oxford: Oxford University Press.

Miller, Phillip. 1992. *Clitics and Constituents in Phrase Structure Grammar*. New York: Garland.

Mohanan, Tara. 1994. *Argument Structure in Hindi*. Stanford: CSLI Publications.

Mycock, Louise. 2006. *The Typology of Constituent Questions: A Lexical-Functional Grammar Analysis of wh-questions*. Ph.D. thesis, University of Manchester.

Nespor, Marina and Irene Vogel. 1986. *Prosodic Phonology*. Dordrecht: Foris.

Nordlinger, Rachel. 1998. *Constructive Case: Evidence from Australian Languages*. Stanford: CSLI Publications.

O'Connor, Rob. 2004. *Information Structure in Lexical-Functional Grammar: The Discourse-Prosody Correspondence in English and Serbo-Croation*. Ph.D. thesis, University of Manchester.

Payne, John. 1995. Inflecting postpositions in Indic and Kashmiri. In F. Plank, ed., *Double Case: Agreement by Suffixaufnahme*. Oxford: Oxford University Press.

Platts, John. 1909. *A Grammar of the Hindustani or Urdu Language*. London: Crosby Lockwood and Son, republished in 2002 by Sang-e-Meel Publications, Lahore.

Raza, Ghulam. 2010. Analyzing the structure of urdu NPs with multiple genitives. In *Proceedings of the Conference on Language and Technology 2010 (CLT10)*. Islamabad: National Language Authority.

Ritter, Elizabeth and Sara T. Rosen. 1997. The function of *have*. *Lingua* 101:295–321.

Sadler, Louisa and Andrew Spencer. 2002. Syntax as an exponent of morphological features. In G. Booij and J. van Marle, eds., *Yearbook of Morphology 2001*, pages 71–96. Dordrecht: Kluwer Academic Publishers.

Samiian, Vida. 1983. *Origins of Phrasal Categories in Persian: an X-bar Analysis.*. Ph.D. thesis, University of California.

Samiian, Vida. 1994. The ezafe construction: Some implications for the theory of X-bar syntax. In M. Marashi, ed., *Persian Studies in North America*, pages 17—41. Bethesda: Iranbooks.

Samvelian, Pollet. 2007. A (phrasal) affix analysis of the Persian Ezafe. *Journal of Linguistics* 43:605–645.

Schmidt, Ruth Laila. 1999. *Urdu – an Essential Grammar*. London: Routledge.

Selkirk, Elisabeth. 1984. *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge: MIT Press.

Selkirk, Elisabeth. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3:371–405.

Selkirk, Elisabeth. 1995a. The prosodic structure of function words. In J. Beckman, L. W. Dickey, and S. Urbanczyk, eds., *Papers in Optimality Theory*, pages 439–70. Amherst, MA: GLSA Publications. Also in *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, ed. James L. Morgan and Katherine Demuth, 187–214: Lawrence Erlbaum Associates.

Selkirk, Elisabeth. 1995b. Sentence prosody: intonation, stress and phrasing. In J. Goldsmith, ed., *The Handbook of Phonological Theory*, pages 550–569. Oxford: Blackwell.

Stump, Gregory T. 2001. *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

Szabolcsi, Anna. 1983. The possessor that ran away from home. *Linguistic Review* 1:89–102.

Wescoat, Michael T. 2009. Udi person markers and lexical integrity. In *Proceedings of LFG09*. Stanford: CSLI Publications.

Zwicky, Arnold. 1987. Suppressing the Z's. *Journal of Linguistics* 23(1):133–148.

Zwicky, Arnold and Geoffrey K. Pullum. 1983. Clitization vs. inflection: English n't. *Language* 59:502–513.