

Hindi-Urdu Syntax Issues

Miriam Butt

Universität Konstanz

Workshop on Hindi-Urdu Treebank, IIIT Hyderabad
January 2008

1 Background

2 Basic Urdu Facts

3 Case and Semantics

4 Complex Predicates

5 Urdu Ezafe

Urdu ParGram Grammar

- Approach to Urdu syntax are based on the grammar architecture of Lexical-Functional Grammar (even if there is no formal analysis).
- LFG allows for several interacting levels of analysis.
- The core ones:
 - c(onstituent)-structure (surface word order, constituency)
 - f(unctional)-structure (grammatical relations, functional information)
- Levels I have played with:
 - a(rgument)-structure
 - i(nformation)-structure
 - p(rosodic)-structure
 - m-structure (a place to put your wellformedness checking junk)

Areas Spent Time On

- Complex predicates (argument merger, case)
- Case (clitics, semantics)
- Agreement
- Tense/Aspect (semantics)
- Information-Structure and Word Order (Topic/Focus)
- Ezafe Construction (clitics, morphology-syntax interface)

ParGram Grammars

All of the work done has been implemented (more or less completely) in an LFG Grammar of Urdu that follows the ParGram guidelines.

- ParGram = Parallel Grammars
- Try to keep analyses as parallel as possible across languages.
- Establish a common set of features.
- Use a common set of methodologies.
- Languages so far: English, German, French, Japanese, Chinese, Norwegian, Welsh, Malagasy, Turkish, Bahasa Indonesian.

Urdu ParGram Grammar

- Currently the grammar uses an ASCII transliteration.

Urdu ParGram Grammar

- Currently the grammar uses an ASCII transliteration.
- An Urdu script to ASCII converter has been written (outsourced to CRULP), but remains to be integrated.

Urdu ParGram Grammar

- Currently the grammar uses an ASCII transliteration.
- An Urdu script to ASCII converter has been written (outsourced to CRULP), but remains to be integrated.
- The grammar contains implementation of linguistic analyses that have interested me, but needs to be much broader and more robust.

Urdu ParGram Grammar

- Currently the grammar uses an ASCII transliteration.
- An Urdu script to ASCII converter has been written (outsourced to CRULP), but remains to be integrated.
- The grammar contains implementation of linguistic analyses that have interested me, but needs to be much broader and more robust.
- Like all ParGram grammars, the syntax interfaces with a FST morphology — all basic inflections are covered.

Urdu ParGram Grammar

- Currently the grammar uses an ASCII transliteration.
- An Urdu script to ASCII converter has been written (outsourced to CRULP), but remains to be integrated.
- The grammar contains implementation of linguistic analyses that have interested me, but needs to be much broader and more robust.
- Like all ParGram grammars, the syntax interfaces with a FST morphology — all basic inflections are covered.
- **New Grant:** 3 years, 1 postdoc and 2 Phds. Mission is to build a large and robust grammar as well as contribute some resources, including a treebank (along the lines of the PARC 700).

Word Order

- The basic word order is SOV.
- The language is generally head-final, with the following exceptions:
 - complementizers (embedded clauses can be either head-initial or head-final)
 - correlative and relative clauses
 - Ezafe Construction
 - (Others?)
- Word Order seems to be conditioned primarily by *information-structure*.

Basic Clause Structure

- The Urdu ParGram grammar assumes a very flat clause structure.
- The verbal complex forms a constituent (it is separated from the rest of the clause via an intonational break, constitutes its own intonational phrase and scrambles mostly as a unit).
- Each major NP and PP forms a constituent — they can all be scrambled among each other.

Basic Clause Structure — Grammatical relations

- Grammatical relations (SUBJ, OBJ, OBJ_theta, OBL) as well as adjuncts are represented at f-structure.
- The set of subjecthood tests established by Mohanan (1994) are used to test for subjects.
- The types of objecthood tests described in (Dalrymple 2001) are used to test for objects.

Demo Ditransitive

Agreement

Agreement is with the highest nominative (phonologically null argument that is subject or object) argument.

- If the subject is nominative, the verb agrees with the subject.
- If the subject is non-nominative and the object is nominative, then the verb agrees with the subject.
- Otherwise the verb shows masculine singular agreement (default).

Agreement

GB/MP-type analyses try to draw deep conclusions from agreement patterns (e.g., relate them to case, word order, specificity). However, agreement patterns in South Asian languages are very varied and do not seem to correlate with anything deep.

- Agreement can be blocked by case marking, but is not always (Nepali).
- Agreement can be with only subjects (Marathi).
- There can be no agreement at all (Bengali).
- Agreement can be with person, number (old patterns or newly innovated via cliticized pronouns) or gender (from participles turned active verbs).

Agreement

- Some Hindi dialects are (or have) given up on agreement (I know of no Urdu ones).
- In standard Urdu/Hindi there is agreement according to number and person or person and gender.
 - The future inflection is the only form showing all (and marking person doubly).
 - Present and subjunctive (and imperative) show person and number agreement.
 - Everything else, including the genitive, does number and gender.

All these patterns are just done by stipulation in the Urdu grammar (wellformedness checking via unification).

Tense/Aspect

The tense/aspect system of Urdu/Hindi has been described extensively. However, I find all of the existing literature confusing as it takes a “top-to-bottom” or constructionist approach: looks at a clause and then labels it.

In fact, while the tense/aspect system appears to be complex and not easily described within standard formal semantics, there are only a very few basic building blocks (Butt and Rizvi 2008) that seem to be combined systematically and restrictively.

Tense/Aspect: Interesting bits

- There is only one verb which shows present tense: *ho* 'be'.
- The future is written differently in Urdu and Hindi. Hindi treats it all as an inflection, Urdu treats the *g-aa* in *cal-e-g-aa* as a separate word.

Tense/Aspect: Interesting bits

There appear to be two different versions of *rah* 'stay' as a tense/aspect auxiliary. One as a progressive marker and one as a marker of further iteration.

(1)

anjum adnan=ko mar rah-i hε
 Anjum.F.Sg.Nom Adnan.M.Sg=Acc hit stay-Perf.F.Sg be.Pres.3.Sg
 'Anjum is hitting Adnan.'

(2)

kutta b^hōk-t-a rah-a
 dog.M.Sg.Nom bark-Impf-M.Sg stay-Perf.M.Sg
 'The dog kept on barking.'

Tense/Aspect: Interesting bits

Infinitives marked with case have temporal meanings.

(3)

mem-sahiba cai **bana-ne=ko** t^h-ĩ
 Madam.F.Nom tea.F.Nom make-Inf.Obl=Acc be.Past-F.PI
 'Madam was just about to make tea.' (Glassman 1986:233)

The same is true for infinitives marked with the =*vala* 'one' construction

(4)

kıfti dūb-ne=**vali** hε
 boat.F.Sg.Nom drown-Inf.Obl=one.F.Sg be.Pres.3.Sg
 'The boat is about to sink.' (Lit. 'The boat is a sinking one.')

(Schmidt 1999:139)

Case

Case is related to tense/aspect and modality in Urdu/Hindi:

- When the finite verb carries perfect morphology and the verb is an unergative, then the subject can be ergative.
- When the finite verb carries perfect morphology and the verb is an agentive (di)transitive, then the subject must be ergative (with the exception of *laa* 'bring')
- Expressions of obligation (must) often correlate with the dative.
- Expressions of desire (want) often correlate with the ergative.

Case

Case is intimately connected with the expression of lexical semantics as well as semantic information at the clausal level.

■ Lexical Semantics

- unergatives vs. unaccusatives (demo)
- psych verbs
- volitionality vs. affectedness

Case

■ Clausal Semantics

- Well-known use of *ko* to mark specificity (demo).
- Case used to express modal meanings (demo 'wants to' vs. 'has to go to the zoo').

As a result, the Urdu grammar currently has quite a lot of lexical semantic and clausal semantic features that the other grammars don't have.

Case

Where does the information come from? — The case markers themselves.

- Lexical Semantic approach to case markers:
 - They each get their own entry.
 - They specify the grammatical relation(s) they are compatible with.
 - Information about the types of meanings they trigger is encoded in the lexicon and in the rules.
 - Formally done via inside-out functional uncertainty (see also Nordlinger 1998).

Definiteness

- Just an indefinite and demonstratives are used, no definite articles.
- Hewson and Bubenik (2006) see a correlation between the use an reinnovation of case marking and the absence of determiners.
- If a language starts introducing determiners, it's case marking system degrades to nothingness.
- There should be some way to capture this correlation.

The Most You Can Do

There are various types of complex predicates in Urdu/Hindi and they can be stacked (but not randomly).

- (5) tara=ne amu=ko bacce=se hat^{hi}
 Tara=Erg Amu=Dat child.Obl=Inst elephant.M.Sg.Nom
 [pinc kar-va le-ne di-ya]
 pinch do-Caus take-Inf.Obl give-Perf.M.Sg
 'Tara let Amu have the elephant pinched (by the child).'

- All of the verbal stuff forms one **monoclausal** predicate (one SUBJ, one OBJ θ , one OBL, one OBJ, no embeddings).

Tests for Complex Predication

Some tests for monoclausality in Urdu are:

- Verb agreement (agrees with object if subject is marked/non-nominative)
- Anaphora Resolution
- Control

Verb Agreement

If you change the gender of the object, then the verb agrees with it, even though it is not a “semantic” argument of the finite verb (i.e., is not licensed directly by the finite verb).

- (6) tara=ne amu=ko bacce=se **billi**
 Tara=Erg Amu=Dat child.Obl=Inst cat.F.Sg.Nom
 pinc kar-va le-ne **di**
 pinch do-Caus take-Inf.Obl give.Perf.F.Sg
 ‘Tara let Amu have the elephant pinched (by the child).’

Argument Structure Approach

- Traditional Theoretical Analyses of Complex Predication within LFG rely heavily on Argument Structure Composition and Linking Principles which map from Argument Structure to Grammatical Functions.
- This theoretical approach has not been taken up readily within computational linguistics — no broad, consistent and therefore implementable consensus or proposal among linguists.

Doing without Argument Structure: The Restriction Operator

- LFG Grammars are now instead using the Restriction Operator introduced by Kaplan and Wedekind (1993) to manipulate f-structures.
- Note that this is conceptually derivational, but formally actually still monotonic (nothing gets thrown away).

The Restriction Operator and Syntactic Complex Predicates

- Butt, King, and Maxwell (2003) introduce (and implement) a solution which combines:
 - the Restriction Operator
 - the ability to *integrate* argument structures to form a **complex PRED.**
- The Restriction Operator can thus manipulate the subcategorization frame of a complex predicate.

The Restriction Operator and Syntactic Complex Predicates

- Composition of two argument structures, but a monoclausal f-structure (Butt 1995).

- In (7b): main verb ‘cough’ and light verb ‘give’.
- The “lettee” (Nadya) is a dative marked OBJ_{θ} .

(7) a. nAdyA kHANS-I
 Nadya.Nom cough-Perf.F.Sg
 ‘Nadya coughed.’

b.

| | | | |
|---------------------------|-----------|---------------|----------------|
| yassIn=nE | nAdyA=kO | kHANS-n-E | dl-yA |
| Yassin=Erg | Nadya=Dat | cough-Inf-Obl | give-Perf.M.Sg |
| ‘Yassin let Nadya cough.’ | | | |

The Urdu Permissive

An example with a transitive main verb ('make').

(8) a. nAdyA=nE gHar banA-yA
 Nadya=Erg house.Nom make-Perf.M.Sg
 'Nadya made a house.'

b.

yassIn=nE nAdyA=kO gHar banA-n-E dl-yA
 Yassin=Erg Nadya=Dat house.Nom make-Inf-Obl give-Perf.M
 'Yassin let Nadya make a house.'

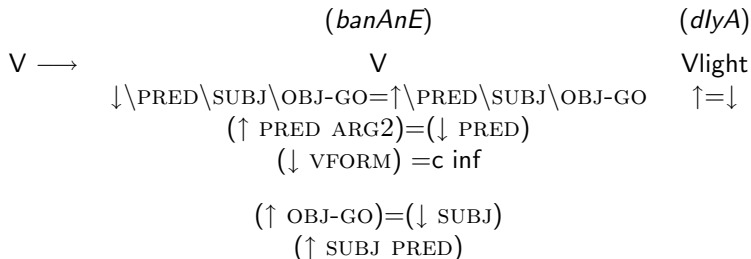
- From the perspective of a restriction analysis, the permissive:
 - “adds” a new subject
 - “demotes” the other verb’s subject to a dative-marked indirect object
- Sample lexical entries for ‘give’ ((9)) and ‘make’ ((10)).
 - (9) $(\uparrow \text{ PRED}) = 'dE < (\uparrow \text{ SUBJ}), \% \text{ PRED2} >'$
 - (10) $(\uparrow \text{ PRED}) = 'banA < (\uparrow \text{ SUBJ}), (\uparrow \text{ OBJ}) >'$

- Restriction allows f-structures and predicates to be manipulated in a controlled and detailed fashion.
- The restriction operator can be applied to an f-structure with respect to a certain feature in order to arrive at a restricted f-structure which does not contain that feature.
- Example: Restrict out the embedded subject of the composed PRED.

(11) PRED 'dE<SUBJ,'banA<SUBJ,OBJ>'>'

- The analysis of complex predicates uses restriction as part of the f-structure annotations on phrase structure rules.

(12)



Permissive Restricted F-structures

■ Final complex f-structure:

- the predicates *dE* 'give' and *banA* 'make' have been composed;
- the "embedded" SUBJ 'Nadya' has been restricted out as part of the composition.

(13)

| | |
|---------|-------------------------------|
| PRED | 'dE<SUBJ,'banA<OBJ-GO,OBJ>'>' |
| SUBJ | [PRED 'Yassin'] |
| OBJ-GO | [PRED 'Nadya'] |
| OBJ | [PRED 'gHar'] |
| TNS-ASP | [ASP perf, TENSE pres] |

C-structure Role in Restriction

- The Urdu CP analysis and Wedekind and Oersnes's analysis of the Danish passive crucially rely on the **syntactic, phrase-structure compositional** aspects of these phenomena.
- There must be a **c-structure node** on which to put the restriction annotation that alters the valency of the verb, creating the final f-structure.

The Urdu Causative and Restriction

- The Urdu causative is created morphologically by affixation.
- **Question:** Can Restriction be extended to provide a uniform analysis of valency changing operations?

An Example

- Agentive transitives have an instrumental causee.

(14) a.

yassin=nE paodA kAT-A
 Yassin=Erg plant.M.Nom cut-Perf.M.Sg
 'Yassin cut the plant.'

b.

nAdyA=nE yassin=sE paoda kaT-**vA**-yA
 Nadya=Erg Yassin=Inst plant.M.Nom cut-**Caus**-Perf.M.Sg
 'Nadya had the plant cut by Yassin.'

F-structures for Causatives

- The basic f-structure for a non-causative verb

$$(15) \left[\begin{array}{l} \text{PRED} \quad 'cut<SUBJ,OBJ>' \\ \text{SUBJ} \quad [\text{PRED} 'Yassin'] \\ \text{OBJ} \quad [\text{PRED} 'plant'] \end{array} \right]$$

- The basic f-structure for the resulting causative

$$(16) \left[\begin{array}{l} \text{PRED} \quad 'cause<SUBJ,'cut<OBL,OBJ>'> \\ \text{SUBJ} \quad [\text{PRED} 'Anjum'] \\ \text{OBJ-GO} \quad [\text{PRED} 'Yassin'] \\ \text{OBJ} \quad [\text{PRED} 'plant'] \end{array} \right]$$

Causative Morphology

- The structures in (15) and (16) can be related via the same type of restriction rules used to analyze complex predicates.
- Restriction must take place within the formation of the lexically causativized verb.
- The integration of finite-state style morphologies (Beesley and Karttunen 2003) into the LFG architecture provides a way to do this.

- The Urdu morphology associates a surface form with a lemma and a set of morphological tags.

- (17) a. likHA \Leftrightarrow likH +Verb +Perf +Masc +Sg
b. likHAyA \Leftrightarrow likH +Verb +Caus1 +Perf +Masc +Sg
c. likHvAyA \Leftrightarrow likH +Verb +Caus2 +Perf +Masc +Sg

C-structure Rules for Causatives with Restriction

- The lemma and morphological tags are parsed by c-structure sublexical rules (Kaplan et al. 2004).
- The sublexical rules are formally identical to standard c-str rules.
- The *+Caus* tag provides a phrase-structure locus for the restriction operator.

Causatives via Restriction and Predicate Composition

■ Causative annotated sublexical c-structure rule

$$\begin{array}{l}
 (18) \quad V \rightarrow \qquad \qquad \qquad V_{\text{stem}} \qquad \qquad \qquad \text{CauseMorph} \\
 \qquad \qquad \qquad \downarrow \backslash \text{PRED} \backslash \text{SUBJ} = \uparrow \backslash \text{PRED} \backslash \text{SUBJ} \qquad \qquad \qquad \uparrow = \downarrow \\
 \qquad \qquad \qquad (\downarrow \text{SUBJ}) = \{ (\uparrow \text{OBJ-GO}) \\
 \qquad \qquad \qquad \qquad \qquad \qquad | (\uparrow \text{OBJ}) \\
 \qquad \qquad \qquad \qquad \qquad \qquad | (\uparrow \text{OBL}) \} \\
 \\
 \qquad \qquad \qquad (\uparrow \text{ PRED ARG2}) = (\downarrow \text{ PRED})
 \end{array}$$

Comparison: Minimalism and the Restriction Operator

- A comparison of a recent Minimalist Approach (Ramchand 2007) with the computational implementation via the Restriction Operator shows up remarkable similarities.
- But within LFG, the Linking vs. the Restriction Operator Approach are very different.

Demo

Demo

increasingly complex structures.

Urdu Ezafe

Urdu contains the following constructions, which are generally written as one or two words in the grammar books (Schmidt 1999, Platts 1909).

(19)

hukuumat-e-paakistaan
government-Ez-Pakistan

'the government of Pakistan'

Schmidt (1999:246)

Urdu

(20)

fahr-e bagdaad
city-Ez Bagdad

'the city of Bagdad'

Platts (1909:62)

Urdu

This is a hallmark of Urdu, but not of Hindi and Hindi-oriented grammars do not tend to discuss it.

Usual Urdu NP Syntax

This construction is strange since Urdu NP syntax usually conforms to the head-final pattern that is (almost) pervasive in this SOV language.

(21)

eek laal gaarii

one red car.F.Sg

'one/a red car'

Urdu

(22)

paakistaan=kii hukuumat

Pakistan=Gen.F.Sg government.F.Sg

'Pakistan's government Schmidt (1999:246)

Urdu

Urdu Ezafe

hukuumat-e-paakistaan

government-Ez-Pakistan

'the government of Pakistan'



Urdu Ezafe — A Borrowing from Persian

- The Urdu construction in fact has been borrowed from Persian (e.g., Platts 1906), where it is known as the Ezafe (from Arabic *iḍāfa* ‘adjunction/addition’) construction.

Urdu Ezafe — A Borrowing from Persian

- The Urdu construction in fact has been borrowed from Persian (e.g., Platts 1906), where it is known as the Ezafe (from Arabic *iḍāfa* ‘adjunction/addition’) construction.
- The Urdu term: *izaafat*.

Urdu Ezafe — A Borrowing from Persian

- The Urdu construction in fact has been borrowed from Persian (e.g., Platts 1906), where it is known as the Ezafe (from Arabic *idāfa* ‘adjunction/addition’) construction.
- The Urdu term: *izaafat*.
- The Ezafe construction tends to be part of the high/literary language since Persian was the language of the Mughal court and mostly influenced the language of the courtiers and poets.

Urdu Ezafe — A Borrowing from Persian

- The Urdu construction in fact has been borrowed from Persian (e.g., Platts 1906), where it is known as the Ezafe (from Arabic *idāfa* ‘adjunction/addition’) construction.
- The Urdu term: *izaafat*.
- The Ezafe construction tends to be part of the high/literary language since Persian was the language of the Mughal court and mostly influenced the language of the courtiers and poets.
- It remains productive today, but modern speakers show a tendency to leave it out (Schmidt 1999:247).

Urdu Ezafe — A Borrowing from Persian

- The Urdu construction in fact has been borrowed from Persian (e.g., Platts 1906), where it is known as the Ezafe (from Arabic *idāfa* ‘adjunction/addition’) construction.
- The Urdu term: *izaafat*.
- The Ezafe construction tends to be part of the high/literary language since Persian was the language of the Mughal court and mostly influenced the language of the courtiers and poets.
- It remains productive today, but modern speakers show a tendency to leave it out (Schmidt 1999:247).
- In the Urdu Arabic-based script there are four different diacritic ways of rendering the Ezafe -e — and like most vowel diacritics it is often left out.

Properties of Urdu Ezafe: Headedness

- As in Persian, the syntactic head is the noun to the very left.
- Evidence comes from verb agreement.
- In (23) and (24) the adjective agrees with the leftmost noun.

(23)

vaadii=e sind^h bahut **baṛ-ii** hai
 valley.F.Sg=Ez Indus.M.Sg very big-F.Sg be.Pres.3.Sg
 'The Indus valley is very big.'

(24)

daryaa=e sind^h bahut **baṛ-aa** hai
 river.M.Sg=Ez Indus.M.Sg very big-M.Sg be.Pres.3.Sg
 'The Indus river is very big.'

Properties of Urdu Ezafe: Constituency

Unlike Persian:

- One does not have the interesting recursive property of Persian — as far as we have been able to determine, there is only ever one Urdu Ezafe complement.
- Complements of Urdu Ezafe cannot be phrasal.
- But the head of the Ezafe construction can be phrasal.

(25)

[ye baṛ-aa diivaan]=e aam
 this big-M.Sg hall of audience=Ez public
 'this public hall of audience'

(26)

[[har baṛ-e diivaan]=e aam]=mẽ
 every big-Obl hall of audience=Ez public=in
 'in every big public hall of audience'

Properties of Urdu Ezafe: Constituency

We assume that the Ezafe forms a constituent with the complement it introduces — motivated by its historical source as a Relative Clause.

Phrase-Structural Analysis

$$\text{NP}_{\text{ez}} \longrightarrow \text{NP EzP}$$

$$\text{EzP} \longrightarrow \text{Ez} \{ \text{N} \mid \text{A} \}$$

However, we could just as well assume a ternary-branching tree:

$$\text{NP}_{\text{ez}} \longrightarrow \text{NP Ez} \{ \text{N} \mid \text{A} \}.$$

Properties of Urdu Ezafe: Status of Ezafe

- As with case clitics, the Urdu Ezafe can take scope over a coordination.

(27)

[maal or daulat]=ko kumaa-o
 material and wealth=Acc earn-Imp.Rude
 'Earn/gather material and wealth!'

(28)

[maal or daulat]=e dunyaa
 material and wealth=Ez world
 'the material and wealth of the world' (from Iqbal)

- In Urdu, morphological inflections are not able to do this.
- We therefore treat the Urdu Ezafe as a **clitic**.

LFG Analysis of Urdu Ezafe

LFG is an inherently modular theory of grammar.

- The Principle of Lexical Integrity means that Word Formation is taken to be the domain of *Morphology*.
- Anything beyond the Word Level is the provenance of *Syntax* (build phrases out of words).
- *Postlexical and Prosodic Phonology* interacts with Syntax.
- *Lexical Phonology* interacts with the Morphology.

LFG Analysis of Urdu Ezafe

- *Clitics* are “little words” — attaching a clitic to a word is **not** the business of Morphology.
- However, *clitics* tend to not be able to form prosodic words on their own, which is why they are phrased together with another prosodic word as part of the prosodic phrasing (prosodic phonology).
- The introduction/placement of “special” clitics (cf. Zwicky 1977) like the Ezafe, which have a given syntactic distribution, is the business of Syntax.

Phrase-Structural Analysis

$$\text{NP}_{\text{Ez}} \longrightarrow \text{NP EzP}$$

$$\text{EzP} \longrightarrow \text{Ez} \{ \text{N} \mid \text{A} \}$$

LFG and Prosodic Phonology

- Crucially, the *prosodic phrasing* is the business of *prosodic phonology* (Selkirk 1984, 1986, Nespor and Vogel 1986, Selkirk 1995).
- In our analysis, we therefore assume that we have a module of grammar that is the prosodic phonology.
- In our implementation, we model this (as best as we can) via the *prosodic projection* $p::$ (Butt and King 1998).

Morphological Analyzer

- In our implementation, we use a finite-state morphological analyzer (Beesley and Karttunen 2003) for the autonomous morphology module.
- The Urdu finite-state morphological analyzer is being developed at Konstanz (Bögel et al. 2007).
- The Morphological Analyzer takes a surface form and provides an analysis in terms of abstract tags.
- These abstract tags are fed into the LFG grammar and are associated with f-structural information where appropriate (Kaplan et al. 2004).
- Example: $+Sg \implies (\uparrow NUM) = sg$

Morphological Analyzer

Sample FST Morphology Output of Ezafe and Nouns:

```
analyzing {sher|panjAb}  
{sher|panjAb}+Noun+Unmarked+Masc+Sg+Nom
```

→ **Nouns are present in the Morphological Analyzer**

```
analyzing {e}  
e "+Token"
```

→ Ezafe is not dealt with by the Morphological Analyzer (rather, it is treated as part of the syntax in the LFG Grammar)

LFG Analysis of Urdu Ezafe

- The lexical entry for Ezafe looks like this:

Lexical Entry for Ezafe

e EZ * (^ MOD MOD-TYPE) = ezafe

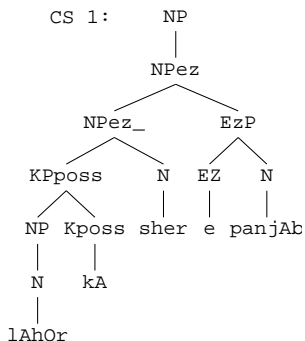
- It is analyzed as introducing a MOD(ifier) along the lines of parts of compound words (cf. the English ParGram grammar).
- The feature MOD-TYPE signals the type of the modification (which is ezafe and different from compounding).
- The phrase-structural analysis makes sure that a modifying Noun or Adjective follows after the Ezafe:

EzP \rightarrow EZ {N | A}

C-Structure Analysis

(29)

fer=e panjaab
 lion=Ez Punjab
 'A/The lion of Punjab'



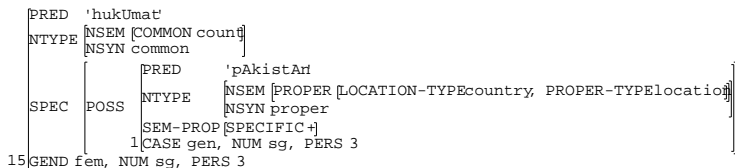
LFG Analysis of Urdu Ezafe

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.

Possessors vs. Ezafe

- We differentiate between the Ezafe Construction and the "standard" ParGram analysis of possessive NP constructions in terms of SPEC [POSS]:
- SPEC [POSS] Analysis of "government of Pakistan"

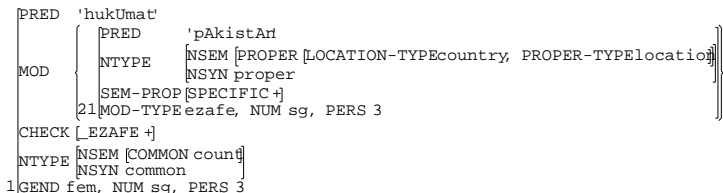
"pAkistAn kI hukUmat"



Possessors vs. Ezafe

■ Ezafe Construction of “government of Pakistan”

"hukUmat e pAkistAn"

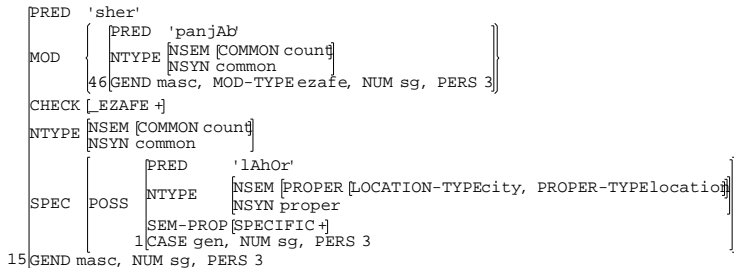


Possessors vs. Ezafe

(30)

laahor=kaa fer=e panjaab
 Lahore=Gen.M.Sg lion.M.Sg=Ez Punjab
 'Lahore's lion of Punjab'

"lAhOr kA sher e panjAb"



References I

- Anderson, Stephen R. 1992. *A-Morphous Morphology*. Cambridge: Cambridge University Press.
- Anderson, Stephen R. 2005. *Aspects of the Theory of Clitics*. Oxford: Oxford University Press.
- Beesley, Kenneth and Lauri Karttunen. 2003. *Finite State Morphology*. Stanford, CA: 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.
- Butt, Miriam and Tracy Holloway King. 1998. Interfacing phonology with LFG. In M. Butt and T. H. King, eds., *On-line Proceedings of the LFG98 Conference*. CSLI Publications. University of Queensland, Brisbane.
- Butt, Miriam and Tracy Holloway King. 2005. 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 and Jafar Rizvi. 2008. Tense and aspect in urdu. In P. Cabredo-Hofherr and B. Laca, eds., *Layers of Aspect*. CSLI Publications.
- Dalrymple, Mary. 2001. *Lexical Functional Grammar*. New York: Academic Press.
- Gomeshi, Jila. 1997. Non-Projecting Nouns and the Ezafe Construction in Persian. *Natural Language and Linguistic Theory* 15(4):729–788.
- Hewson, John and Vit Bubenik. 2006. *From Case to Adposition: The development of configurational syntax in Indo-European Languages*. Amsterdam: John Benjamins.
- Holmberg, Anders and David Odden. 2005. The Noun Phrase in Hawramani. Presented at the *First International Conference on Aspects of Iranian Linguistics*.
- Kahnemuyipour, Arsalan. 2000. Persian Ezafe Construction Revisited: Evidence for Modifier Phrase. In *The 2000 Annual Conference of the Canadian Linguistic Association Conference*, pages 173–185.

References II

- Kaplan, Ronald M., John T. Maxwell III, Tracy H. King, and Richard Crouch. 2004. Integrating finite-state technology with deep LFG grammars. In *Proceedings ESSLLI, Workshop on Combining Shallow and Deep Processing for NLP*.
- Larson, Richard and Hiroko Yamakido. 2005. Ezafe and the deep position of nominal modifiers. Presented at the *Barcelona Workshop on Adjectives and Adverbs*.
- Miller, Philip H. 1992. *Clitics and Constituents in Phrase Structure Grammar*. New York: Garland.
- Nespor, Marina and Irene Vogel. 1986. *Prosodic Phonology*. Dordrecht: Foris.
- Nordlinger, Rachel. 1998. *Constructive Case: Evidence from Australian Languages*. Stanford: CSLI Publications.
- Platts, John T. 1909. *A Grammar of the Hindustani or Urdu Language*. London: Crosby Lockwood and Son. Republished in 2002 by Sang-e-Meel Publications, Lahore.
- 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*. New York: Routledge.
- Selkirk, Elizabeth. 1984. *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, MA: The MIT Press.
- Selkirk, Elizabeth. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3:371–405.
- Selkirk, Elizabeth. 1995. The prosodic structure of function words. In J. Beckman, L. Dickey, and S. Urbanczyk, eds., *Papers in Optimality Theory*, pages 439–470. Amherst. University of Massachusetts Occasional Papers in Linguistics 18.
- Zwicky, Arnold. 1977. On clitics. Bloomington: Indiana University Linguistics Club.
- Zwicky, Arnold. 1987. Suppressing the z's. *Journal of Linguistics* 23(1):133–148.