

Towards a Computational Semantic Analyzer for Urdu

Annette Hautli Miriam Butt

Department of Linguistics, University of Konstanz

9th Workshop on Asian Linguistic Resources, IJCNLP '11

Motivation

- ① Advances in the computational processing of Urdu
- ② Increasing amount of lexical resources for Urdu available

Motivation

- ① Advances in the computational processing of Urdu
- ② Increasing amount of lexical resources for Urdu available

Task

Gathering information from various resources and putting them together to form one coherent resource for Urdu.

Motivation

- ① Advances in the computational processing of Urdu
- ② Increasing amount of lexical resources for Urdu available

Task

Gathering information from various resources and putting them together to form one coherent resource for Urdu.

Challenge

What formalism can we employ that puts this information together? And what are the particular challenges with respect to Urdu?

Taking stock

- Urdu is still a language with comparably few linguistic resources
- Syntactic parsers:
 - ▶ Treebank-based PCFG parser (Abbas, 2002)
 - ▶ Urdu dependency parser trained with MaltParser (Ali and Hussain, 2010)
 - ▶ Urdu ParGram grammar based on LFG (Butt and King 2004, Bögel et al. 2009)
- Lexical resources:
 - ▶ Emille corpus (Baker et al., 2004)
 - ▶ “*Experiences in Building Urdu Wordnet*” (Adeeba and Hussain, 2011)
 - ▶ Urdu WordNet based on Hindi WordNet (Ahmed and Hautli, 2009)
 - ▶ Automatic collection of Urdu multiwords (Hautli and Sulger, 2011)
 - ▶ Development of a lexical resource for Urdu verbs

Taking stock

- Urdu is still a language with comparably few linguistic resources
- Syntactic parsers:
 - ▶ Treebank-based PCFG parser (Abbas, 2002)
 - ▶ Urdu dependency parser trained with MaltParser (Ali and Hussain, 2010)
 - ▶ Urdu ParGram grammar based on LFG (Butt and King 2004, Bögel et al. 2009)
- Lexical resources:
 - ▶ Emille corpus (Baker et al., 2004)
 - ▶ “*Experiences in Building Urdu Wordnet*” (Adeeba and Hussain, 2011)
 - ▶ Urdu WordNet based on Hindi WordNet (Ahmed and Hautli, 2009)
 - ▶ Automatic collection of Urdu multiwords (Hautli and Sulger, 2011)
 - ▶ Development of a lexical resource for Urdu verbs

The Urdu ParGram grammar

- Parser based on the formalism of Lexical Functional Grammar (Bresnan and Kaplan 1981) run on the development platform XLE (Crouch et al. 2011)
- The Urdu ParGram grammar as part of an international effort to create parallel grammars for different languages

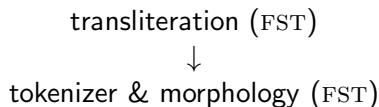
The Urdu ParGram grammar

- Parser based on the formalism of Lexical Functional Grammar (Bresnan and Kaplan 1981) run on the development platform XLE (Crouch et al. 2011)
- The Urdu ParGram grammar as part of an international effort to create parallel grammars for different languages

transliteration (FST)

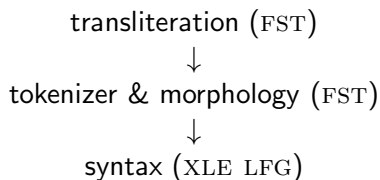
The Urdu ParGram grammar

- Parser based on the formalism of Lexical Functional Grammar (Bresnan and Kaplan 1981) run on the development platform XLE (Crouch et al. 2011)
- The Urdu ParGram grammar as part of an international effort to create parallel grammars for different languages



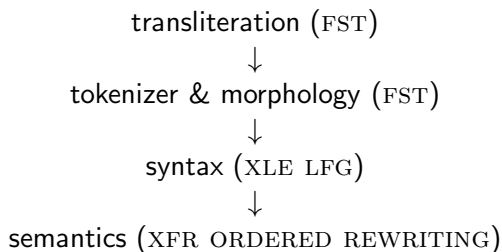
The Urdu ParGram grammar

- Parser based on the formalism of Lexical Functional Grammar (Bresnan and Kaplan 1981) run on the development platform XLE (Crouch et al. 2011)
- The Urdu ParGram grammar as part of an international effort to create parallel grammars for different languages



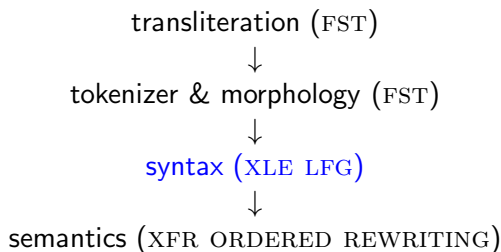
The Urdu ParGram grammar

- Parser based on the formalism of Lexical Functional Grammar (Bresnan and Kaplan 1981) run on the development platform XLE (Crouch et al. 2011)
- The Urdu ParGram grammar as part of an international effort to create parallel grammars for different languages



The Urdu ParGram grammar

- Parser based on the formalism of Lexical Functional Grammar (Bresnan and Kaplan 1981) run on the development platform XLE (Crouch et al. 2011)
- The Urdu ParGram grammar as part of an international effort to create parallel grammars for different languages

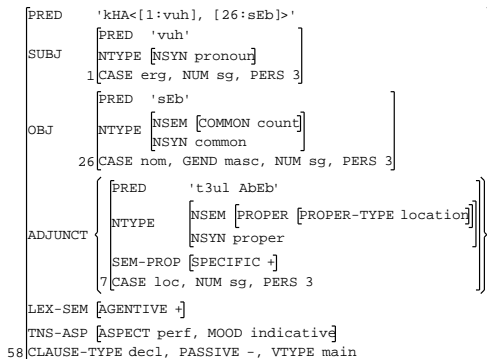
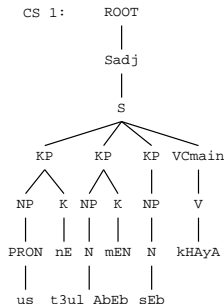


The Urdu ParGram grammar

اس نے تل آبیب میں سیب کھایا

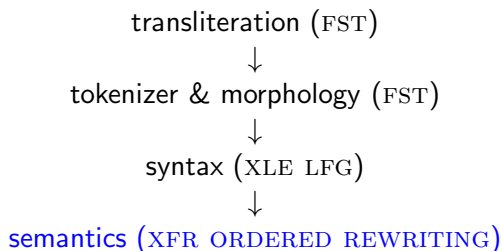
us nE t3ul AbEb mEN sEb kHAYa
 he Erg Tel Aviv in apple eat.Perf.F.Sg
 'He/She ate an apple in Tel Aviv.'

"us nE t3ul AbEb mEN sEb kHAYa"



The Urdu ParGram grammar

- Parser based on the formalism of Lexical Functional Grammar (Bresnan and Kaplan 1981) run on the development platform XLE (Crouch et al. 2011)
- The Urdu ParGram grammar as part of an international effort to create parallel grammars for different languages



The XFR rewrite system

- Rewriting and flattening of f-structure facts by rewrite rules (Crouch and King, 2003)
 - ▶ `SUBJ(%1,%2) ==> subj(%1,%2) .`
- Each clause is embedded in a context where predications are true or false
- Allows for the incorporation of lexical resources such as WordNet and VerbNet using a database interface

The XFR rewrite system

What information would we like to get from a semantic representation?

اس نے تل آبیب میں سیب کھایا

us nE t3ul AbEb mEN sEb kHAyA

he Erg Tel Aviv in apple eat.Perf.F.Sg

'He/She ate an apple in Tel Aviv.'

The XFR rewrite system

What information would we like to get from a semantic representation?

اس نے تل آبیب میں سیب کھایا

us nE t3ul AbEb mEN sEb kHAYa

he Erg Tel Aviv in apple eat.Perf.F.Sg

'He/She ate an apple in Tel Aviv.'

1. What predications hold in the context of the sentence?

The XFR rewrite system

What information would we like to get from a semantic representation?

اس نے تل آبیب میں سیب کھایا

us nE t3ul AbEb mEN sEb kHAYA
he Erg Tel Aviv in apple eat.Perf.F.Sg
'He/She ate an apple in Tel Aviv.'

1. What predications hold in the context of the sentence?
2. What are the thematic roles of the grammatical functions?

The XFR rewrite system

What information would we like to get from a semantic representation?

اس نے تل آبیب میں سیب کھایا

us nE t3ul AbEb mEN sEb kHAYA
he Erg Tel Aviv in apple eat.Perf.F.Sg
'He/She ate an apple in Tel Aviv.'

1. What predications hold in the context of the sentence?
2. What are the thematic roles of the grammatical functions?
3. What is the lexical information contained in the sentence?

The XFR semantics

1. What predications hold in the context of the sentence?

اس نے تل آبیب میں سیب کھایا

us nE t3ul AbEb mEN sEb kHAyA
he Erg Tel Aviv in apple eat.Perf.F.Sg
'He/She ate an apple in Tel Aviv.'

```
context_head(t,kHA:25),  
in_context(t,role(subj,kHA:25,vuh:1)),  
in_context(t,role(obj,kHA:25,sEb:21)),  
in_context(t,role(mod,kHA:25,'t3ul AbEb':7)).
```

The XFR semantics

2. What are the thematic roles of this sentence?

- Development of a lexical resource for Urdu verbs in the style of VerbNet
 - ▶ Assignment of thematic roles to the grammatical functions
 - ▶ *kHA* ‘to eat’: SUBJ → Agent
 OBJ → Patient
 - ▶ VerbNet information is stored in a database which can be accessed by the XFR system
 - ▶ The XFR rules replace the grammatical functions with the thematic roles from the database
- Locational information is available from the f-structure representation and directly put into the semantic representation

The XFR semantics

اس نے تل آیب میں سیب کھایا

us nE t3ul AbEb mEN sEb kHAyA
he Erg Tel Aviv in apple eat.Perf.F.Sg
'He/She ate an apple in Tel Aviv.'

context_head(t,kHA:25),
in_context(t,role('Agent',kHA:25),vuh:1),
in_context(t,role('Patient',kHA:25),sEb:21),
in_context(t,role('Location',kHA:25),'t3ul AbEb':7),

The XFR semantics

3. What is the lexical information contained in the sentence?

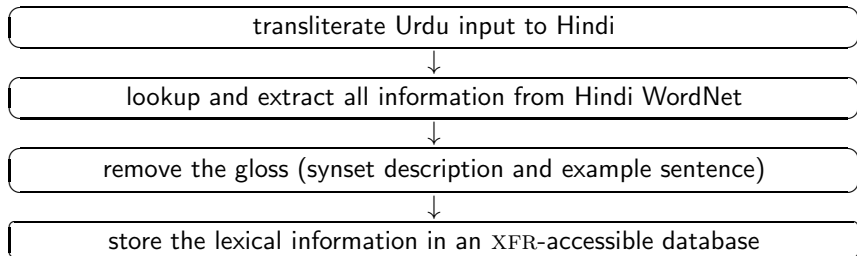
اس نے تل آبیب میں سیب کھایا

us nE t3ul AbEb mEN sEb kHAyA
he/she Erg Tel Aviv in apple eat.Perf.F.Sg
'He/She ate an apple in Tel Aviv.'

- *kHA* 'to eat': ingestive verb where the agent consumes an eatable object
- *us* 'he/she': living thing that performs the eating event
- *sEb* 'apple': fruit that is the object of consumption
- *t3ul AbEb* 'Tel Aviv': location
- *mEN* 'in': indicates that the event takes place in Tel Aviv

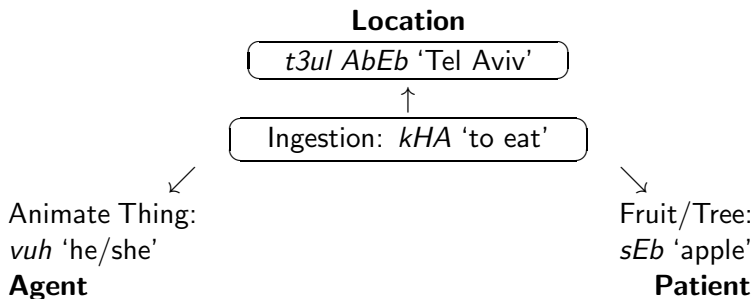
The XFR semantics

- The lexical information in our system comes from Urdu WordNet, which is build on the basis of Hindi WordNet (Ahmed and Hautli 2010)



The XFR semantics

Inclusion of all resources:



The XFR semantics

Treatment of spatial expressions (Ahmed, 2010):

The XFR semantics

Treatment of spatial expressions (Ahmed, 2010):

اس نے تل آبیب میں سیب کھایا

us nE t3ul AbEb mEN sEb kHAyA
he Erg Tel Aviv in apple eat.Perf.F.Sg
'He/She ate an apple in Tel Aviv.'

```
context_head(t,kHA:25),  
in_context(t,role('Agent',kHA:25),vuh:1),  
in_context(t,role('Patient',kHA:25),sEb:21),  
in_context(t,role('Location',kHA:25),location:100),
```

The XFR semantics

Treatment of spatial expressions (Ahmed, 2010):

اس نے تل آبیب میں سیب کھایا

us nE t3ul AbEb mEN sEb kHAyA
he Erg Tel Aviv in apple eat.Perf.F.Sg
'He/She ate an apple in Tel Aviv.'

```
context_head(t,kHA:25),  
in_context(t,role('Agent',kHA:25),vuh:1),  
in_context(t,role('Patient',kHA:25),sEb:21),  
in_context(t,role('Location',kHA:25),location:100),  
in_context(t,role('figure',location:100,vuh:1)),  
in_context(t,role('ground',location:100,'t3ul AbEb':7)),  
in_context(t,role('configuration','t3ul AbEb':7,in)).
```

The XFR semantics

Treatment of modality:

The XFR semantics

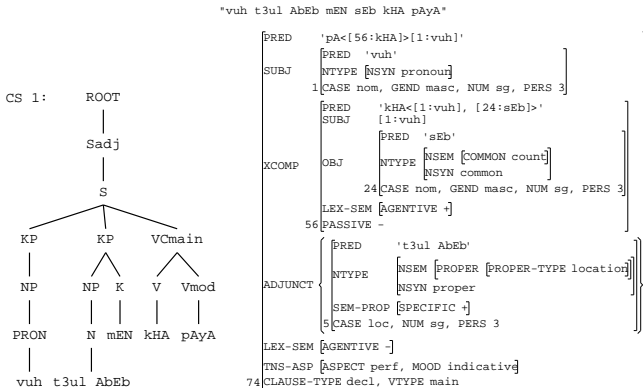
Treatment of modality:

- Expression of modality mostly constructionally (Bhatt et al. 2011)

us t3ul AbEb mEN sEb kHA pA-yA

he Tel Aviv in apple eat find-Perf

'He/She was able to eat an apple in Tel Aviv.'



The XFR semantics

Treatment of modality:

- Expression of modality in Urdu mostly constructionally

us t3ul AbEb mEN sEb kHA pA-yA
he Tel Aviv in apple eat find-Perf
'He/She was able to eat an apple in Tel Aviv.'

```
context_head(t,pA:6),  
context_head(ctx(kHA:25),kHA:25),  
in_context(t,ABIL(pA,ctx(kHA:25))),  
in_context(t,role(Holder_Of_Obligation,pA:6),vuh:1),  
in_context(ctx,role('Agent',kHA:25,vuh:1)),  
in_context(ctx,role('Patient',kHA:25,sEb:21)),  
in_context(ctx,role('Location',kHA:25,'t3ul AbEb':7)).
```

Challenges

An extreme case - EAT expressions in Hindi/Urdu (Hook and Pardeshi, 2009):

- Employment of 'eat' in idiomatic expressions
- About 160 EAT expressions for Hindi/Urdu
- Variety of uses due to loan translations from Persian

Challenges

h2asan=nE kEk=kO kHAyA

h2asan cake eat.Perf.Sg.Masc

'Hasan ate the cake.'

eat=⟨ Agent, Patient ⟩

inqilAbl fikar zang kHA jAEgl

revolutionary thought rust eat go.Fut.Fem.Sg

'Revolutionary thinking will gather rust.'

eat (gather rust) =⟨ Patient, Theme ⟩

is sAl=kl mandl sheyar-bAzAr kHA gAyl

this year slowdown stockmarket eat go.Perf.Fem.Sg

'This year's slowdown wrecked (lit. devoured) the stock market.'

eat (wreck) =⟨ Agent, Theme ⟩

Tasks ahead

- Further develop the lexical resources for Urdu
- Define concrete application areas for the semantic representation and adjust it accordingly
- Develop evaluation standards for semantic representations and run large-scale experiments
 - ▶ Hindi TreeBank could provide some semantic information
- Work on the theoretical semantic analysis of the language

Summary

- The XFR rewrite system is an adequate way of combining various resources in one tool
 - ▶ Lexical information from WordNet
 - ▶ Verb frames from a verb resource
- Based on a detailed syntactic analysis, the semantic representation can go deeper, e.g.
 - ▶ spatial expressions
 - ▶ modality constructions

Thank you!