Introduction to Morphology, Syntax and Lexical Semantics of Hindi and Urdu

Dipti Misra Sharma
<dipti@iiit.ac.in>
LTRC, IIIT, Hyderabad
India

Dec 8, 2012
COLING 2012

Outline

- Introduction
  - Some facts about Hindi and Urdu
- Linguistic properties
- Morphology
- Some basic Syntax
- Lexical semantics
Hindi: Some facts

- A major language of Indo-Aryan family
- Official language of 11 Indian states: Uttar Pradesh, Uttarakhand, Bihar, Delhi, Jharkhand, Chhattisgarh, Himachal Pradesh, Haryana, Rajasthan and Madhya Pradesh
- Also spoken outside India in Fiji, Mauritius, Guyana etc
- Number of speakers who returned Hindi as their mother tongue (in India): 422,048,642 (41.03%) (2001, Census of India report)
- A large population in India who speak Hindi as their second language
- Script: Devanagari – a syllabic script

Urdu: Some facts

- An Indo-Aryan language
- Evolved in India around eight-tenths centuries from 'khariboli', a dialect spoken in and around Delhi
- Significant borrowings from Arabic and Persian
- It was also known as 'rekhta' (mixed language)
- Official language of Pakistan
- Official language of ?? states of India
- Also spoken in Fiji, Bangladesh etc
- Number of speakers in India 51,536,111 (5.01%) (2001, Census of India report)
- Script Perso-Arabic
Hindi-Urdu (Hindustani)

- Hindi and Urdu are mutually intelligible
- Linguists consider them as two registers of the same language
- Similar in grammatical structures
- Differ in vocabulary, particularly in the formal written varieties
- A mixed variety of the two is used as a lingua franca in India and is also known as Hindustani

Some Basic characteristics of Hindi/Urdu

- Hindi/Urdu have relatively free word order
- The unmarked word order in both the languages is subject-object-verb (SOV).
- Auxiliary verbs follow the main verb.
- Nouns are followed by postpositions.
- Adjectives precede the nouns they modify.
- In Urdu, sometimes, adjectives follow the noun (ezafe constructions).
- Large use of participles, complex predicates, and causatives.
- Reduplication and echo-compounding are productively used in Hindi/Urdu (in fact almost all the Indian languages).
Morphology

Hindi and Urdu have following morphological properties

- Grammatical gender: masculine and feminine
- Number: singular and plural
- Person: first, second and third
- Case: direct, oblique and vocative
- Adjectives inflect for number, gender and case
  - Some adjectives do not decline

Nouns

- Nouns in Hindi/Urdu are inflected for number and case
  - Gender
    All nouns have inherent gender: pankhaa (fan.masc), lataa (creeper.fem), ghar (house.masc)
  - Number
    Singular: pankhaa (fan), lataa (creeper), ghar (house)
    Plural: pankhe (fans), lataeM (creepers), ghar (houses)
  - Case
    The case roles in Hindi are normally marked by postpositions. However, Hindi nouns reflect two cases morphologically, Direct and Oblique
Case

- **Direct** nouns are in nominative and are not followed by a postposition. They occur denoting subject and/or object.
  
  \[\text{LaRkaa} \text{ aayaa} \] ‘the boy came’
  \[<\text{ladkaa},n,\text{m.sg.3.d.,}> <\text{a},v,\text{m.sg.3.},\text{yaa},>\]
  
  \[\text{laRkii} \text{ aayii} \] ‘the girl came’
  \[<\text{ladkii},n,\text{f.sg.3.d.,}> <\text{a},v,\text{f.sg.3.},\text{yaa},>\]

- **Oblique** nouns are objects of a postposition such as ne (erg), ko (acc,dative), se (instr), meM (loc), par (loc), and kaa (gen).
  
  \[\text{laRke ne roTii khaayii} \] ‘the boy ate bread’
  \[<\text{laRkaa},n,\text{m.sg.3.obl.,}> <\text{roTi},n,\text{f.sg.3.d.,}> <\text{khaa},v,\text{f.sg.3.},\text{yaa},>\]
  
  \[\text{laRke ne roTii ko zamiin se uThaayaa} \] ‘the boy picked the bread from the floor’
  \[<\text{laRkaa},n,\text{m.sg.3.obl.,}> <\text{roTi},n,\text{f.sg.3.obl.,}> <\text{zamiin},n,\text{f.sg.3.obl.,}> <\text{uThaa},v,\text{m.sg.3.},\text{yaa},>\]
Pronouns

Morphologically, like nouns, the pronouns also inflect for number and case.

Sg - dir : yaha (this), vaha (that), jo (who/which), kaun (who.interro), kyaa (what) and kuch (some)

Sg – Obl : isa (this), asa (that), jisa (which), kisa (who/which.interro), koi (someone), and kisi (someone)

Pl - dir : ye (these), ve (those), jo (who.pl)

Pl – Obl

a. except before ne (erg) : ina (these), una (those), jina (whoever), kina (who.interro), kinhii (who.indef)

b. : inho (these), unho (those), jinho (who), kinho (who) and kinhii (some people.indef)

Pronouns (Contd…)

- Before ‘ne’: maiM (I) and tuu (you.sg) don’t change. For example, maiMne (I.erg) and tuune (you.erg)

- Before other postpositions: maiM (I) and tuu (you.sg) change to the oblique forms, mujh (me) and tujh (you.sg). Thus, mujhko (to me) and tujhko (to you.sg)

- Before all postpositions: ham (we), tum (you.sg/pl) and aap (you-hon) don’t change form.

- maiM (I), tuu (you), ham (we), and tum (you) don’t attach to kaa (gen) postposition. Instead they have irregular forms meraa (my), teraa (your), hamaaraa (our), and tumhaaraa (your)
Adjectives

- Morphologically, an adjective is inflected for gender, number, and case as it agrees with the following noun.
- Postpositions are attached only to the nouns. Adjectives preceding these nouns also have oblique form. *acche lākke ne* 'good boys erg'
- The transformations that an adjective (eg 'acchaa 'good') undergoes with regard to number, gender and case are given below

<table>
<thead>
<tr>
<th>Case</th>
<th>Ditect</th>
<th>Oblique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Sg</td>
<td>Pl</td>
<td>Sg</td>
</tr>
<tr>
<td>Masc</td>
<td>acchaa</td>
<td>acche</td>
</tr>
<tr>
<td>Fem</td>
<td>acchii</td>
<td>acchii</td>
</tr>
</tbody>
</table>

Verbs

- Verbs in Hindi/Urdu are inflected for tense, aspect, mood (TAM) and the agreement features of gender, number and person.
- Tense, aspect and mood are mostly expressed by auxiliaries in Hindi/Urdu. Thus, only certain moods, aspects and tense are marked in the verb forms.

Given below are some examples of various verb forms from Hindi/Urdu:

<table>
<thead>
<tr>
<th>Root</th>
<th>Infinitive</th>
<th>Habitual</th>
<th>Perfective</th>
<th>Causative</th>
<th>Causative</th>
</tr>
</thead>
<tbody>
<tr>
<td>ro</td>
<td>ronaa</td>
<td>rotaa</td>
<td>royaa</td>
<td>rulaa</td>
<td>ruvaa</td>
</tr>
<tr>
<td>cry</td>
<td>to cry</td>
<td>cry.hab</td>
<td>cried</td>
<td>cause someone to cry</td>
<td>make someone to cry</td>
</tr>
</tbody>
</table>
Auxiliaries

- Auxiliaries mark Tense, Aspect and Modality information on verbs

(a) *bacce khaanaa khaa rahe haiM*
    Children_d meal eat prog be.pl.pres
    'The children are having a meal'

(b) *bacce khaanaa khaate rah sakte haiM*
    Children_d meal eat_nf prog ablit be.pl.pres
    'The children can continue to have their meal'

- Auxiliaries also carry the gender, number and person information

Postpositions

- Postpositions largely mark the case relations

*baccoM ne raat meM mez se khaanaa le liyaa*
  *children_obl erg night in table ablat food take refl.pst*

- Hindi also has compound postpositions
Compound Post-positions

*Compound postpositions* are formed by connecting the postpositions *ke*, *kii*, and *se* with other words as follows:

<table>
<thead>
<tr>
<th>Compound postposition</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>ke amusaar</em></td>
<td>‘according to’</td>
</tr>
<tr>
<td><em>ke alaavaa</em></td>
<td>‘in addition to’</td>
</tr>
<tr>
<td><em>ke kaaran</em></td>
<td>‘because of’</td>
</tr>
<tr>
<td><em>ke dvaaraa</em></td>
<td>‘through’</td>
</tr>
<tr>
<td><em>ke saamne</em></td>
<td>‘in front of’</td>
</tr>
<tr>
<td><em>ke liye</em></td>
<td>‘for’</td>
</tr>
<tr>
<td><em>kii ora/taraf</em></td>
<td>‘towards’</td>
</tr>
<tr>
<td><em>kii taraf</em></td>
<td>‘like’</td>
</tr>
<tr>
<td><em>kii jagah</em></td>
<td>‘in place of’</td>
</tr>
<tr>
<td><em>se baahar</em></td>
<td>‘out of’</td>
</tr>
<tr>
<td><em>se pahle</em></td>
<td>‘before’</td>
</tr>
</tbody>
</table>

Urdu Specific Features

Prepositions in Urdu

'ezafe' in Urdu
Urdu has Prepositions

- Unlike Hindi, Urdu has prepositions as well.
- Some Urdu examples with prepositions are:
  (a) *qabāl* ‘before’
      
      qabāl az_ayn (qabāl azeen) qabāl az_ waqt
      ‘before’ ‘from_this’ ‘before’ ‘from_time’
      ‘before this’ ‘before time’

  (b) *dar* ‘in,inside, amidst’
      
      dar_ayn asnah (dariin asnah)
      ‘in_this’ ‘moment’

  (c) *az* ‘from,since,for’
      
      az raah e hamdardi, az sare nau, khaarij az bahes
      ‘from’ ‘way’ ‘empathy’ ‘from’ ‘beginning’ ‘new’ ‘beyond’ ‘from’ ‘discussion’
      ‘for the sake of courtesy’

  (d) *ta* ‘to,until,till’
      
      ta waqt
      ‘until’ ‘time’

'ezafe' in Urdu

- Urdu has what is referred to as 'ezafe'.
- Normally marks a genitive but is not restricted to genitive alone

**EZ: N+Adj**

nasl-e-insani
‘race-of-humanity’

lamha-e-aakhar
‘The last moment’

**EZ: Adj+N**

qabil-e-rahem
‘qualified for sympathy’

**EZ: Adj+Adj**

qabil-e-qubul
‘qualified-for-acceptance’

**EZ: N+N**

daar-e-hukumat
‘period-of-rule’

hukumat-e-hind
‘government-of-India’
Reduplication: A morphological processes

- Words belonging to various categories can be reduplicated
- These expressions are often hyphenated
- Reduplication has various morphological functions depending on the lexical category which is reduplicated. For example,
  - Nouns: it adds the sense of 'every',
  - Verbs: it brings the sense of adverbial participle
  - Adjectives and adverbs: it adds intensity

- Hindi has three types of reduplication: full, partial and redundant
- Reduplication is highly productive in these languages

Full Reduplication

If the word is $X$ then its reduplicated form is $X-X$.

- \textit{raam-raam} `Ram-Ram' (proper noun)
- \textit{bacca-bacca} 'child-child' (common noun)
- \textit{garam-garam} `hot-hot' (adjective)
- \textit{dhiire-dhiire} `slowly-slowly' (adverb)
- \textit{jaa-jaa} `go-go' (verb)
- \textit{naa-naa} `not-not' (negative particle)
- \textit{kyaa-kyaa} `what-what' (question word)
- \textit{jaate-jaate} `going-going' (participle)
Partial Reduplication (Echo words)

- In partial reduplication, an expression \(X\) is repeated partially.
- Only a part of a given word is repeated which gives the meaning of ‘\(X\) etc.’.
- In Hindi/Urdu the first consonant of \(X\) is replaced by \(v\) -.

For example, khaanaa-vaanaa ‘food-etc.’

- vaanaa is not a valid word in Hindi.

Some more examples of partial reduplication in Hindi are:

- jaanaa ‘going’ → jaanaa-vaanaa ‘going etc.’
- aalooy ‘potato’ → aalooy-vaalooy ‘potato etc.’
- aisaa ‘like this’ → aisaa-raisaa ‘like this etc.’

There are also examples which do not fall in this pattern.
The meaning of such words changes substantially and does not have the sense of ‘etc.’

- bhaag ‘run’ → bhaagambhaag ‘rush’
- jhuuTh ‘lie’ → jhuuth-muuTh ‘just like that (without meaning it)’
- dekh ‘see’ → dekhii-dekhaa ‘in imitation’

Some Basic Syntax

- Hindi and Urdu are both relatively free word order SOV languages.
- For case marking, Hindi primarily uses postpositions.
- The verb agrees either with subject or with object.
- Adjective agrees with the noun it modifies.
Simple Transitive

trans-1: आतिफ़ किताब पढ़ेगा
Atif kitab paRhegaa
Atif book.f read.m.sg.fut
‘Atif will read the book’

trans-2: आतिफ़ ने किताब पढ़ी
Atif ne kitaab paRhii
Atif erg book.f read.f.sg.pst
‘Atif read the book’

trans-3: आतिफ़ को किताब पढ़नी पड़ी
Atif ko kitaab paRhnii paRii
Atif dat book.f read.f.inf compel.f.pst
‘Atif had to read the book’

Intransitive: Unergative

Unerg-1: आतिफ़ सोएगा
Atif soyegaa
Atif sleep.m.sg.fut
‘Atif will sleep’

unerg-2: *आतिफ़ ने सोया
*Atif ne soyaa
Atif erg sleep.m.sg.pst
‘Atif slept’

unerg-3: आतिफ़ को सोना पड़ेगा
Atif ko sonaa paRegaa
Atif dat sleep.inf compel.fut
‘Atif will have to sleep’
Intransitive: Unaccusative

unacc-1: दरवाज़ा खुलेगा
       darvaazaa khulegaa
       door.m.sg.d open.m.sg.fut
       'The door will open'

unacc-2: *दरवाज़े ने खुला
       *darvaaze ne khulaa
       door.m.sg.obl erg open.pst
       'The door opened'

unacc-3: दरवाज़े को खुलना पड़ेगा
       darvaaze ko khulnaa paRegaa
       door.m.sg.obl dat open.inf compel.fut
       'The door will have to open'

 Existential

exist-1: उस कमरे में चूहे हैं
        us kamre meM cuuhe haiM
        that room in rats be.pres.pl
        'There are rats in that room'
**Dative Subject**

unacc-4: कल रात बादलो में चौंद दिखा
kal raat baadaloM meM caaMd dikhaa
yesterday night clouds in moon see(unacc).pst
'Yesterday night, the moon was seen behind the clouds'

dat-subj-1: कल रात बादलो में मुझको चौंद दिखा
kal raat baadaloM meM mujhko caaMd dikhaa
yesterday night clouds in me.dat moon see(unacc).pst
'Yesterday night, I saw the moon behind the clouds'

**Ditransitive**

ditrans-1: राम मोहन को किताब देगा
raam mohan ko kitaab  degaa
Ram Mohan dat book.f give.m.sg.fut
'Ram gave a book to Mohan'

ditrans-2: राम ने मोहन को किताब दी
raam ne mohan ko kitaab dii
Ram erg Mohan dat book.f give.f.sg.pst
'Ram gave a book to Mohan'
Complement Clause

compl-cl-1: राम जानता है कि सीता बेर से आएगी
raam jaantaa hai ki siita der se aayegii
Ram know.hab.m.sg be.sg.pres that Sita late part come.f.sg.fut

'Ram knows that Sita will arrive late'

Relative Clause

rel-cl-1: मेरी बहन जो दिल्ली में रहती है कल आ रही है
merii bahan jo dillii meM rahtii hai kal aa
My sister who Delhi in stay.hab.f.sg be.sg.pres tomorrow come rahii hai
prog.f.sg be.sg.pres

'My sister who stays in Delhi is coming tomorrow'
Relative Clause

rel-cl-2: मैंने वह किताब जो तुमने दी थी पढ़ ली

maiMne vah kitaab jo tumne dii thii paRh lii
I.erg that book.f which you.erg give.f.sg.pst be.f.sg.pst read refl.f.sg.pst
'I have read the book which you gave me'

rel-cl-3: मैंने वह किताब पढ़ ली जो तुमने दी थी

maiMne vah kitaab paRh lii jo tumne dii thii
I.erg that book.f read refl.f.sg.pst which you.erg give.f.sg.pst be.f.sg.pst
'I have read the book which you gave me'

rel-cl-4: जो किताब तुमने दी थी वह मैंने पढ़ ली

jo kitaab tumne dii thii vah maiMne paRh lii
which book.f you.erg give.f.sg.pst be.f.sg.pst that I.erg read refl.f.sg.pst
'I have read the book which you gave me'

Complex Predicate

compl-pred-1: राम रवि की प्रतीक्षा कर रहा था

raam ravi kii pratikshaa kar rahaa thaa
Ram Ravi gen wait do prog.m.sg be.m.sg.pst
'Ram was waiting for Ravi'

compl-pred-2: राम रवि को याद कर रहा था

raam ravi ko yaad kar rahaa thaa
Ram Ravi acc remember do prog.m.sg be.m.sg.pst
'Ram was remembering Ravi'
Causatives

Unerg-1: आतिफ़ सोईगा
Atif soyegaa
Atif sleep.m.sg.fut
‘Atif will sleep’

causative-1: आया ने आतिफ़ को सुलाया
aayaa ne Atif ko sulaayaa
maid erg Atif acc sleep.caus.pst
‘The maid caused the child to sleep’

causative-2: माँ ने आया से आतिफ़ को सुलाया
maaN ne aayaa se Atif ko sulvaayaa
mother erg maid by Atif acc sleep.caus.pst
‘The mother made the maid to cause the child to sleep.’

Lexical Semantics

Semantic properties of certain verb types seem to affect the case selection for certain arguments. For example,

- Experiencer verbs
  - The experiencer argument takes dative case
  
    राम को बुखार है, राम को चाँद दिखाया, राम को दुख है
    raam ko bukhaar hai, raam ko caand dikhaa, raam ko dukh hai
    ‘Ram’ ‘dat’ ‘fever’ ‘be.pres’, ‘Ram’ ‘dat’ ‘moon’ ‘see-unc.pst’, ‘Ram’ ‘dat’ ‘sorrow’ ‘be.pres’

- Participatory verbs
  - The second argument of the participatory verbs takes ‘se’ postposition
  
    राम रवि से करो राम से शादी करेगी, सीता राम se शादी करेगी,
    raam ravi se carca se, siitaa raam se shaadi karegii,
    ‘Ram’ ‘Ravi’ ‘to’ ‘discussion’ ‘do.m.sg.fut’, ‘Sita’ ‘Ram’ ‘with’ ‘marriage’ ‘do.f.sg.fut’,

    रवि मोहन से मिलेगा
    ravi mohan se milegaa
    ‘Ravi’ ‘Mohan’ ‘with’ ‘meet.m.sg.fut’
References