Overview

• Introduction to the nature of syntactic representations. (Rambow, 15 minutes)
• Introduction to the morphology, syntax, and lexical semantics of Hindi and Urdu. (Sharma, 40 minutes)
• The morphological representation for Hindi and Urdu, including encoding issues, tokenization, part-of-speech tags, and morphological representation. (Sharma and Rambow, 20 minutes)
• The dependency representation (DS) for Hindi and Urdu syntax: principles, representation, and examples. (Sharma, 25 minutes)
• The lexical semantic representation (PB) for Hindi and Urdu: principles, representation, and examples. (Vaidya, 25 minutes)
• The phrase structure representation (PS) for Hindi and Urdu syntax: principles, representation, and examples. (Rambow, 25 minutes)
• Sample initial experiments in Hindi and Urdu NLP using the HUTB. (Sharma and Rambow, 15 minutes).
Phrase Structure Representation

Owen Rambow
CCLS, Columbia University
rambow@ccls.columbia.edu
Phrase Structure (PS) Representation in the Hindi and Urdu Treebanks

• Devised by Rajesh Bhatt, University of Massachusetts, Amherst
  – Assisted by Annahita Farudi and Owen Rambow
• Developed in conjunction with DS and PB
• Inspired by Chomskyan tradition
Background for PS

• Chomskyan program:
  – Motivated by claims about language acquisition in children
  – Develop a theory of syntax such that syntax of a language can be explained by
    • Language-universal principles
    • Language-specific parameters

• PS for Hindi inspired by Chomskyan program, but not following any specific Chomskyan approach
Basic Principles of PS

• PS represents relation between **lexical predicate-argument structure** (interface to lexicon) and **surface word order** (interface to phonology and semantics, roughly speaking)

• These two levels are related by derivations:
  – Words and constituents move and leave **traces**
    • Transformational grammar

• Monostratal representation

• Not unlike English Penn Treebank!
Specific Assumptions about Representation Made by PS

- Phrase structure
- Notion of lexical heads with projections (X-bar theory, sort of) and associated functional projections
  - Nouns with postpositions
  - Verbs with auxiliaries and complementizers (*ki*)
- Binary branching
  - Theoretical reasons
  - To be different from DS
Basic Transitive Clause (1)

- There are two privileged positions in the verbal projection, corresponding usually to DS’s k1 and k2
Basic Transitive Clause (2)

• The representation is maintained when we have an ergative construction
Intrasitive Clause: Unergative

- PS makes a distinction between unergative and unaccusative
- In unergative, there simply is no object
Intrasitive Clause: Unaccusative

- Argument starts in lower position (because of lexical semantics), and moves to higher position (because higher position has no occupant)
Existentials

- Existential *ho `be’* is unaccusative (because agent-free), and location is an adjunct.
Ditransitive

- The recipient is introduced as adjoined to the VP-Pred: a fixed, but not structural position.
Putting it All Together: Dative Subjects

**Dikhaa** is interpreted semantically as a ditransitive: someone makes something appear to someone.

- Since the agent is absent, the lower argument raises to the higher position (like unaccusative).
- The dative beneficiary is base generated in the fixed dative position (adjoined to VP-Pred) and then scrambles elsewhere.
Complement Clauses with *ki*

राम जानता है * ki सीता देर से आतेगी
Relative Clause

'Ι have read the book which you gave me'
Complex Predicate

राम रवि को याद कर रहा था
raam ravi ko yaad kar raha thaa
Ram Ravi acc remember do prog.m.sg be.m.sg.pst
'Ram was remembering Ravi'
Causative