1 Oblique Subjects in Urdu/Hindi

Mohanan (1994:139) lists a number of oblique subjects in Urdu/Hindi:

(1) a. omra kela kʰa ruh-i tʰ-i
   Amra.F.Nom banana.M.Sg.Nom eat stay-Perf.F.Sg be.Past-F.Sg
   ‘Amra was eating a banana.’ Urdu/Hindi

b. omra=ne kela kʰa-ya
   Amra.F=Erg banana.M.Sg.Nom eat-Perf.M.Sg
   ‘Amra ate a/the banana.’ Urdu/Hindi

c. omra=ko kela kʰa-na tʰ-a
   Amra.F=Dat banana.M.Sg.Nom eat-Inf.M.Sg be.Past-M.Sg
   ‘Amra had to eat a banana.’ Urdu/Hindi

d. omra=ko kahani yad a-yi
   Amra.F=Dat story.F.Sg.Nom memory come-Perf.F.Sg
   ‘Amra remembered a/the story.’ Urdu/Hindi

e. omra=se kela kʰa-ya nahī gu-ya
   ‘Amra could not eat the banana.’ Urdu/Hindi

f. omra=ke car bacce tʰ-e
   ‘Amra had four children.’ Urdu/Hindi

g. omra=mē bulkul doya nahī tʰ-i
   Amra.F=Loc in at all mercy.F.Nom not be.Past-F.Sg
   ‘Amra had no mercy at all.’ Urdu/Hindi

¹The sentences in (1) are adaptations of Mohanan’s original examples, (1d) has been added.
1.1 Language Background

Urdu is an Indo-Aryan South Asian language spoken in:

- Pakistan (national language) and India (one of the 22 official languages)
- World-wide due to South Asian Diaspora (U.K., U.S.A., Canada, New Zealand, etc.)

Urdu is closely related to Hindi (mainly differences in vocabulary and pronunciation plus a few structural and morphological differences).

1.2 Some Structural Features

- Urdu/Hindi displays a so-called split-ergative pattern by which subjects of verbs that are transitive or unergative appear in the ergative when the verb carries perfect morphology: (1a) vs. (1b).
- The distribution of the ergative is, however, more complicated (section 2.4).
- Agreement works as follows:
  - If the subject is unmarked (nominative), agree with the subject (1a).
  - Else, if the object is unmarked, agree with the object (1b–g).
  - Else, if the object is also marked, use default masculine singular agreement (2).

  \[
  \text{(2)} \quad \text{Amra} = \text{ne} \quad \text{roṭi} = \text{ko} \quad k^a-\text{ya}
  \]

  \[
  \begin{align*}
  \text{Amra.F=Erg} & \quad \text{bread.F.Sg=Acc} \quad \text{eat-Perf.M.Sg} \\
  \text{‘Amra ate the bread.’}
  \end{align*}
  \]

1.3 Subject Tests

- Keenan (1976) discusses a number of tests across languages for the establishment of subjecthood.
- The applicability of these tests depends on the overall structure of the individual languages.
- For the Indo-Aryan languages, the following tests are usually assumed across different stages of the language (Hook, p.c., April 2012)
  1. Antecedency of the possessive reflexive.
  2. Control of a gerundial phrase/clause.
  3. Realization of genitive case in nominalizations.
  4. Agreement with the finite verb (not always applicable).
  5. Position in clause (very seldom applicable, more a weak indication than a test).

- Hock (1990, 1991) bases his argumentation for OIA data on these tests:
1. Control of absolutes (e.g., having gone there, ...)
2. Reflexives are taken to be subject oriented
3. Word order (position in clause)

• For Hindi, Mohanan (1994) found the following tests to be reliable:

1. Anaphoric Resolution:
   - Reflexives can refer to either the grammatical or the logical subject of the clause.
   - Pronoun Obviation: Pronouns cannot refer to the subject in the minimal finite clause they are contained in.

2. Control of participial adjuncts (e.g., smiling, they opened the door; having eaten, they fell into bed)

3. Gapping in coordination

Note: In Urdu/Hindi, agreement is not a good test for oblique subjects since every instance of overt case marking blocks agreement with the case marked NP.

According to Mohanan’s tests, all of the examples in (1b–g) indeed contain oblique subjects.

1.4 Observations

• Urdu/Hindi is ultimately descended from a version of Sanskrit.

• The evidence for oblique subjects in Sanskrit is very weak (Hock 1990).

• The evidence for oblique subjects in Urdu/Hindi, on the other hand, is very strong (Mohanan 1994).

• The original case marking system of Sanskrit was lost almost entirely over the course of several centuries in MIA and has since been replaced by newly innovated case markers.

• The new case markers are primarily drawn from a set of original spatial terms.

1.5 Question

How do originally spatial terms result in oblique subject marking?
2 Basics of the Urdu/Hindi Case Marking System

Old Indo-Aryan (Vedic and Sanskrit) had an inflectional case marking system much like the sister language Latin.

<table>
<thead>
<tr>
<th>Number</th>
<th>Declension</th>
<th>Western Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>devas</td>
<td>nominative</td>
</tr>
<tr>
<td>2</td>
<td>devam</td>
<td>accusative</td>
</tr>
<tr>
<td>3</td>
<td>devena</td>
<td>instrumental</td>
</tr>
<tr>
<td>4</td>
<td>devāya</td>
<td>dative</td>
</tr>
<tr>
<td>5</td>
<td>devāt</td>
<td>ablative</td>
</tr>
<tr>
<td>6</td>
<td>devasya</td>
<td>genitive</td>
</tr>
<tr>
<td>7</td>
<td>deve</td>
<td>locative</td>
</tr>
</tbody>
</table>

Declension of Sanskrit deva- ‘god’ (adapted from Blake 2001, 64)

- The inflectional case endings eroded away and collapsed into one another in the course of Middle Indo-Aryan (MIA).
- From around 1200 on, one finds new case markers being drawn into the system in New Indo-Aryan (NIA).
- In the modern languages, the case markers are mostly clitics, some markers are inflectional.

2.1 Timeline

A. Old Indo-Aryan
   1200 BCE — 600 BCE (Vedic)
   600 BCE — 200 BCE (Epic and Classical Sanskrit)

B. Middle Indo-Aryan (Aśokan inscriptions, Pāli, Prākrits, Apabhramśa—Avahatṭha)
   200 BCE — 1100 CE

C. New Indo-Aryan (Bengali, Hindi/Urdu, Marathi and other modern North Indian languages)
   1100 CE — Present
2.2 Case Marking

Modern Urdu/Hindi Case Markers

<table>
<thead>
<tr>
<th>Marker</th>
<th>Case</th>
<th>Grammatical Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>∅</td>
<td>nominative</td>
<td>subj/obj</td>
</tr>
<tr>
<td>=ne</td>
<td>ergative</td>
<td>subj</td>
</tr>
<tr>
<td>=ko</td>
<td>accusative</td>
<td>obj</td>
</tr>
<tr>
<td></td>
<td>dative</td>
<td>subj/indirect obj</td>
</tr>
<tr>
<td>=se</td>
<td>instrumental</td>
<td>subj/obl/adjunct</td>
</tr>
<tr>
<td></td>
<td>comitative</td>
<td>source</td>
</tr>
<tr>
<td>=ka</td>
<td>genitive</td>
<td>subj, specifier</td>
</tr>
<tr>
<td>=ki</td>
<td>locative</td>
<td>subj/obl/adjunct</td>
</tr>
<tr>
<td>=ke</td>
<td>locative</td>
<td>obl/adjunct</td>
</tr>
<tr>
<td>∅/-e</td>
<td>locative</td>
<td>obl/adjunct</td>
</tr>
</tbody>
</table>

- As illustrated in the Table above, the modern case markers are not exclusively tied to a particular grammatical function (GF), but can be used in a range of different contexts.

- In Urdu/Hindi, the ergative is exceptional in this regard — however, in other NIA languages this is not true: the form used for the ergative can also mark indirect objects (datives) or obliques (mainly instrumentals).

2.3 Historical Origin of Modern Case Markers

The Locatives were derived from various lexical items with spatial semantics: e.g., mé from Sanskrit madhya ‘middle’, par from Sanskrit upari ‘on’.

The Genitive is the only case marker which inflects. This can be traced directly to its historical origin. After a fierce debate in the last century, the view espoused by Hoernle won out and was taken over by Beames (1872:285) and Kellogg (1893:129). Under this view, the genitive is analyzed as having arisen from kr.ita ‘done by’, the Sanskrit past participle of kri ‘do’ as follows. Sanskrit kr.ita > Prakrit kerita > kerai > modern Urdu/Hindi k-

The original participal inflected for agreement and the genitive case marker has not lost this property.

Instrumental se could either be connected with Sanskrit sam ‘with’ or with the locative singular noun saṇge ‘in attachment to’ (Kellogg 1893:132).

Dative/Accusative ko is generally traced to the Sanskrit locative noun kākshe ‘armpit, side’ (Kellogg 1893:130).

---

2This case marker is also similar to the dative suffix in Dravidian languages: Telugu and Tamil -ku, Malayalam -kku, Old Kannada -ke or -kke, Tulu -ku or -gu (Sahoo 2001, 38).
Ergative *ne* cannot be traced definitively within the history of Urdu/Hindi. It begins appearing around 1400 and is likely to have been the result of borrowing by language contact. For Rajasthani, Tessitori (1913, 1914) traces the dative/ergative marker *ne* back to the Apabhramśa form *kanpahī*, which in turn is related to the Sanskrit locative of *ear*, *karne*, and which was realized as *kanhaiN* (or *kanhai, kanhi, kanhali, kanñi*) in Old Rajasthani and mostly meant ‘aside, near’, but also occurs in ablative (*from*) contexts (see Butt and Ahmed 2011 for an in-depth discussion).

### 2.4 Systematic Pattern of Differential Case Marking

**Further observation:** With the exception of (1e), all of the oblique subjects in (1) in fact occur in the context of patterns of Differential Case Marking (DCM).

#### 2.4.1 Differential Object Marking

Well-known cases of DCM tend to involve Differential Object Marking (DOM) (Aissen 2003).

**Example:** Urdu/Hindi Specificity Alternation (*Nom/Acc*)

(3) a. ram gari=ko xarid-e-g-a
Ram.M.Sg.Nom car.F.Sg=Acc buy-3.Sg-Fut-M.Sg 
‘Ram will buy the car (a specific car).’ 
**Acc→Specific Object** Urdu/Hindi

b. ram gari xarid-e-g-a
Ram.M.Sg.Nom car.F.Sg.Nom buy-3.Sg-Fut-M.Sg 
‘Ram will buy a/the car.’ 
**Nom→No Information** Urdu/Hindi


#### 2.4.2 Differential Subject Alternations

Subject Alternations are a different matter: no good analyses of subject alternations (though some suggestions, see section 3).

**Nominative/Ergative Subject Alternation (cf. (1b))**


(4) a. ram khañ-a
Ram.M.Sg.Nom cough-Perf.M.Sg 
‘Ram coughed.’ 
**Nom→No Information** Urdu/Hindi

b. ram=ne khañ-a
Ram.M.Sg=Erg cough-Perf.M.Sg 
‘Ram coughed (**purposely**).’ 
**Erg→Control** Urdu/Hindi
• Most analyses assume that the verb is polysemous: 2 versions of the verb, one licenses the ergative, the other the nominative case.

• This is not elegant: there are more situations in which only the case marker differs.

Dative/Ergative Subject Alternation (Modality) (cf. (1c))

(5) a. nadya=ne zu ja-na ḥe
   Nadya.F.Sg=Erg zoo.M.Sg.Loc go-Inf.M.Sg be.Pres.3.Sg
   ‘Nadya wants to go to the zoo.’
   Erg→Control
   Urdu

b. nadya=ko zu ja-na ḥe
   Nadya.F.Sg=Dat zoo.M.Sg.Loc go-Inf.M.Sg be.Pres.3.Sg
   ‘Nadya has/wants to go to the zoo.’
   Dat→Goal
   Urdu

Possessive Subject Alternation (Temporary vs. Permanent) (cf. (1f))

(6) a. amra=ke/#ke pas car bacce ḥe
   ‘Amra has four children.’
   Urdu/Hindi

b. amra=ke pas/*ke car kulam ḥe
   ‘Amra has four pens.’
   Urdu/Hindi

Possessor of Characteristic (Temporary vs. Permanent) (cf. (1g))

(7) a. amra=mê bahot hummat ḥe
   Amra.F=Loc in much courage.F.Nom be.Pres.3.Sg
   ‘Amra is a person who is full of courage.’
   Urdu/Hindi

b. amra=ko bahot hummat ḥe
   Amra.F=Dat much courage.F.Nom be.Pres.3.Sg
   ‘Amra is feeling very courageous.’
   Urdu/Hindi

For more detailed discussions of the semantics underlying the contrasts in (6) and (7), see Mohanan (1994) and references therein, more recently Sulger (2011, 2012).

3 Current Explanations for DCM

Interestingly, most analyses of DSM turn a blind eye to the obviously systematic semantic contrasts that are being expressed and instead look to more structural factors such as markedness and indexing (e.g., Aissen 1999).
3.1 Case as Identifier of Grammatical Relations

Standard View of Case:

- Primary function is the distinguishing of grammatical relations: Subject vs. object vs. indirect object, etc.

(8) claudia puellae rosās dat
   Claudia.Nom girl.Dat.Sg rose.Acc.Pl give.Pres.3.Sg
   ‘Claudia is giving roses to the girl.’ Latin

- So when case marking is lost:
  - either new case markers are needed
  - or other parts of the grammar (e.g., agreement, word order) work to make the necessary distinctions (cf. Kiparsky’s (1987, 1988, 1997, 2001) notion of linkers)

3.2 Markedness, Indexing and Distinguishing

- In particular, notions of markedness have dominated the discussion (see Malchukov and de Swart (2009), de Hoop (2009) for a survey of the current state of the art — the discussion in this section is based on them).
- Case Markers are predicted to arise first in situations where it is difficult to distinguish agents/subjects from patients/objects, i.e., in marked situations.

3.2.1 Differential Object Marking (DOM)

- In Differential Object Marking (DOM), objects which could be mistaken for subjects (due to being animate, definite/specific, topical/emphasized, etc.) are marked.

(9) a. nadya kitab xarid-e-g-i
   ‘Nadya will buy a/the book.’ Hindi/Urdu

b. nadya kitab=ko xarid-e-g-i
   Nadya.F.Sg.Nom book.F.Sg=Acc buy-3.Sg-Fut-F.Sg
   ‘Nadya will buy a particular book.’ Hindi/Urdu

(10) a. nadya *yasin=ko mıl-e-g-i
    Nadya.F.Sg.Nom Yassin.M.Sg=Acc meet-3.Sg-Fut-F.Sg
    ‘Nadya will meet Yassin.’ Hindi/Urdu

b. *nadya yasin mıl-e-g-i
   Nadya.F.Sg.Nom Yassin.M.Sg.Nom meet-3.Sg-Fut-F.Sg
   ‘Nadya will meet Yassin.’ Hindi/Urdu

- Historical Observation/Prediction: marking of animate objects as in (10) (function: distinguish objects from subjects) is historically extended to general definiteness/specificity marking of objects (function: index a particular kind of semantic role).
3.2.2 Differential Subject Marking (DSM)

- In Differential Subject Marking (DSM) two different strategies are taken to be in operation (sometimes in conflict with one another).

1. **Distinguishing strategy**: need to distinguish subjects from objects, so mark non-prototypical subjects (i.e., subjects which could be mistaken for objects).

2. **Indexing strategy**: Identify proto-typical subjects (agents) and mark this particular semantic role.

- One typical result: *ergative* languages
  (Recall that Hindi/Urdu is split-ergative according to aspect marking)

(11) a. nadya  kitab  xarid-e-g-i
    ‘Nadya will buy a/the book.’ Hindi/Urdu

b. nadya=ne  kitab  xarid-i
    Nadya.F.Sg=Erg book.F.Sg.Nom buy-Perf.F.Sg
    ‘Nadya bought a/the book.’ Hindi/Urdu

(12) a. ram  kāś-a
    Ram.M.Sg.Nom cough-Perf.M.Sg
    ‘Ram coughed.’ Hindi/Urdu

b. ram=ne  kāś-a
    Ram.M.Sg=Erg cough-Perf.M.Sg
    ‘Ram coughed (purposefully).’  

**Predictions:**

1. Find more and different types of DCM in DSM situations than in DOM situations.


3. More DOM in accusative languages.

3.3 Problems for the Current State-of-the-Art

**Data from South Asia**

South Asian languages have the following properties:

- They form a *Sprachbund* of genetically unrelated languages (Masica 1976) — the major ones are Indo-Aryan, Dravidian and Tibeto-Burman.

- A long written history is available for Indo-Aryan and Dravidian.
• However, the synchronic variety and diachronic record of South Asian languages are rarely taken into account in formulating hypotheses about crosslinguistic structure.

• **South Asian Patterns**

  – South Asian languages include both ergative and accusative types, but the possibilities for DOM and DSM seem equal (e.g., no Dravidian language is ergative, but there are many examples of DSM in Dravidian).

  – Have many different types of DOM and DSM, contrary to expectation.

  – In particular, under the standard view, DOM is expected to be mainly asymmetric, i.e., contrasting an unmarked object with a marked one as in (9), but this is not the case.

  – No good explanation for “over-marking”, i.e., when both subject and object are marked overtly with innovated case markers (but see Wunderlich and Lakämper (2001) for a proposal within Optimality Theory).

  (13) a. nadya=ne yasin=ko mıl-a
      Nadya.F.Sg=Erg Yassin.M.Sg=Acc meet-Perf.M.Sg
      ‘Nadya met Yassin.’ Hindi/Urdu

  b. *nadya=ne yasin mıl-a
      Nadya.F.Sg=Erg Yassin.M.Sg.Nom meet-Perf.M.Sg
      ‘Nadya met Yassin.’ Hindi/Urdu

  – Examples such as (14) are questionable, even though one would expect these types of examples to be prototypical for the *Distinguishing Strategy* in DSM.

  (14) ??pattår=ne ṣija tor-a
      ‘The stone/rock broke the glass.’ (based on Mohanan 1994, 75) Hindi/Urdu

**Relational Scales**

• Furthermore, **Relational Scales/Hierarchies** as in (15) have been identified as important for the realization of arguments crosslinguistically.

  (15) Relational Scale: Subject > Non-Subject
       Animacy Scale: Human > Animate > Inanimate
       Definiteness Scale: Pronoun > Proper Name > Definite >
                             Indefinite Specific > Nonspecific

• Within Optimality Theory (OT), these scales are translated into violable constraints, which in turn are taken to govern DCM (e.g., Aissen (1999, 2003)).

• However, Bickel et al. (2012) show that the typological evidence for the language universal status of Relational Hierarchies is actually very thin (they appear to reflect areal phenomena instead).
Lexical Semantic Patterns

Lexical Semantic patterns governing the interaction between different types of verb classes and case marking patterns are rarely taken into account, though these seem to be crucial both diachronically and synchronically.

- For example, Barðdal (2004), Eythórsson and Barðdal (2005) and Barðdal (2009, 2011) have argued forcefully for a more lexical semantic approach to understanding syn-
chronic and diachronic dimensions of case marking.

- Within South Asian linguistics, see, e.g., Butt (2001), Butt and King (2004) and Butt and Ahmed (2011) for a lexical semantic approach to case.

- More specifically, see Joshi (1993), Asudeh (2001) and Deo (2003) for evidence that dative subjects in Marathi develop at different times for different verb classes.

3.4 DCM Examples from South Asia

3.4.1 DOM

Definiteness/Specificity

(16) a. aa man-aa kitaab-aa d-aa
   he.Nom I-Obl book-Obl give-Sg
   ‘He gives me the book.’ (Mirdeghan 2005) Balochi

   b. aa man-aa kitaab d-aa
   he.Nom I-Obl book.Nom give-Sg
   ‘He gives me book(s).’ (Mirdeghan 2005) Balochi

More Direct vs. Indirect

(17) a. nadya=ne yasin=ko ml-a
   Nadya.F.Sg=Erg Yassin.M.Sg=Acc meet-Perf.M.Sg
   ‘Nadya met Yassin.’ Hindi/Urdu

   b. nadya=ne yasin=se ml-a
   Nadya.F.Sg=Erg Yassin.M.Sg=Inst meet-Perf.M.Sg
   ‘Nadya met Yassin.’ Hindi/Urdu

Examples like this abound in South Asian languages (Ahmed Khan 2009).
Also consider examples as in (18).

**Static vs. Dynamic Path** (Obliques)

(18) a. us=le dilli=dek₄i kat₆mandu=samma baaṭo banaa-yo
   Pron.3.Sg=Erg Delhi=Abl Kathmandu=to street.Nom make-Past
   ‘He built a street from Delhi to Kathmandu.’ Ahmed Khan (2009)
   **static path**

   b. u dilli=baaṭa kat₆mandu=samma kud-yo
   Pron.3.Sg.Nom Delhi=Abl Kathmandu=to ran-Past
   ‘He ran from Delhi to Kathmandu.’ Ahmed Khan (2009)
   **dynamic path**

See Ahmed Khan (2009) for a range of DCM with second argument marking (not necessarily all DOM) across South Asian languages.

### 3.4.2 DSM

**Volitionality/Control**

(19) a. phrum rui-zi me-nyan-ta
   cheese rot-NonFinal Neg-accept-Impf-Mir
   ‘The cheese just keeps on getting rotten.’ (Hyslop 2010, 19)
   Kurtöp

   b. phrum-gi rui-zi me-nyan-ta
   cheese-Erg rot-NonFinal Neg-accept-Impf-Mir
   ‘The cheese is going on getting rotten (is going rotten on purpose somehow).’
   (Hyslop 2010, 19)
   Kurtöp

**Modality**

(20) a. amma kuṭṭiye adik’k-aṇam
   mother.Nom child.Acc beat-want
   ‘Mother must beat the child.’ (Butt, King and Varghese 2004)
   Malayalam

   b. ammak’k’o kuṭṭiye adik’k-aṇam
   mother.Dat child.Acc beat-want
   ‘Mother wants to beat the child.’ (Butt, King and Varghese 2004)
   Malayalam

(21) a. avan var-aam
   he.Nom come-may
   ‘He may come.’ (possibility) (Butt, King and Varghese 2004)
   Malayalam

   b. avana var-aam
   he.Dat come-may
   ‘He may come.’ (permission) (Butt, King and Varghese 2004)
   Malayalam
(22) a. ami toma=ke cai
   I.Nom you=Acc wants
   ‘I want you.’ (Klaiman 1980, 279) Bengali

   b. amar toma=ke cai
   I.Gen you=Acc wants
   ‘I need you.’ (Klaiman 1980, 279) Bengali

**Information Structure**

(23) a. tshe ozi meme-the jong-shang
   DM then grandfather-one emerge-Perf.Ego
   ‘So then an old man came out.’ (Hyslop 2010, 14) Kurtöp

   b. meme-the-gi jong-shang
   grandfather-one-Erg emerge-Perf.Ego
   ‘[An old man]EMPH came out.’ (Hyslop 2010, 14) Kurtöp

Note: Dalrymple and Nikolaeva (2011) propose that secondary topic marking is a major driving factor behind case use and innovation.

**Temporary vs. Permanent Characteristics**

(24) a. meeri-k’k’o paadaan kazhiy-illa/patt-illa
   Mary-Dat sing.Inf be.able-neg/can-neg
   ‘Mary cannot sing.’ (she is unable to sing for now) Malayalam

   b. meeri-ekkon. d. paadaan kazhiy-illa/patt-illa
   Mary-Inst sing.Inf be-able-neg/can-neg
   ‘Mary cannot sing.’ (she could never sing/she is too lazy to sing) Malayalam

(25) a. hasan=le gaari c^alaun-c^a
   Hassan=Erg car.Nom drive-NonPast.3.Sg
   ‘Hassan drives cars (that’s what he does).’ Nepali (Individual-Level/Permanent)

   b. hasan gaari c^alaun-c^a
   Hassan.Nom car.Nom drive-NonPast.3.Sg
   ‘Hassan is driving a car/cars.’ Nepali (Stage-Level/Temporary)

(26) a. raam=le (#aajaa) angreji jaan-da-c^a
   Ram=Erg today English know-Impf-NonPast.3.Sg
   ‘Ram knows English (#today).’ Nepali (Individual-Level/Permanent)

   b. raam (aajaa) angreji bol-da-c^a
   Ram today English speak-Impf-NonPast.3.Sg
   ‘Ram will speak English (today).’ Nepali (Stage-Level/Temporary)

All the Nepali data is from Poudel (2008).
3.5 Summary

- Semantic DCM is a regular part of languages across South Asia today.

- The occurrence of DCM synchronically is not easy to explain in terms of standard notions of indexing, distinguishability and marking, since quite a bit of semantic/pragmatic information seems to be at play (not all of it well understood).

- There are instances of DOM as in (17) not generally treated in the literature — I think many existing DOM patterns have been overlooked because they are not expected.

4 A Lexical Semantic Theory of Case

Current Hypothesis:

- DCM is an example of case being used to make semantic distinctions.

- Explaining DCM only in terms of grammatical markedness, indexing or distinguishability fails to take into account this primarily semantic function.

- While the primary function of case is to help identify grammatical relations, this job seems to be too “easy” (and other parts of the grammar tend to help anyway: e.g., agreement, position).

- So case marking is also used for expressing (sometimes subtle) semantic contrasts.

Contentful Case Markers

  - Case markers are not assigned by structural configurations.
  - Instead, they help determine the overall structure of the clause (cf. the notion of constructive case (Nordlinger 1998)) by carrying information with them about what grammatical functions they can mark.
  - And they carry semantic information with them that flows into the overall semantic/pragmatic analysis of the clause.

- This follows naturally from the historical origin of most case markers from adverbials, adpositions or participials — the original semantics are pressed into service.

- In particular, the precise semantic import of a case marker emerges out of a language particular system of contrasts and may therefore differ from language to language. This means that the same original source may give rise to two different case markers in the language (i.e., ergative vs. dative; Butt and Ahmed 2011).
Big Picture Prediction:

- If
  - a language has a history of semantic DCM (i.e., a history of using case to express semantic distinctions)
  - and other languages in the geographical area work similarly
- then the language is likely to innovate new case markers in order to retain a system of semantic distinctions.
- This especially includes the use of oblique subjects in situations of DSM.

5 Looking Back in Time

DCM is also an old part of at least a subset of the languages: the Indo-Aryan ones (historical material on Dravidian exists, but needs to be researched).

5.1 Old Indo-Aryan

5.1.1 Pāṇini’s Grammar of Sanskrit

- Pāṇini’s grammar of Sanskrit mentions 23 possibilities of case alternations (Katre 1987, Böhtlingk 1839–40).
- Some of these have to do with formal reasons (morphophonology).
- Some of these are governed by lexical semantics.
- Others are clearly expressing semantic distinctions.

Example:

Rule 2.3.12: The Dative and Accusative are used for verbs of movement, but the dative cannot be used if motion is an abstract one.

That is if a person named Ram goes to a village, the village can be marked either Accusative or Dative. But if only one’s thoughts “go” towards a village, the Dative cannot be used.

5.1.2 Partitive: Accusative/Genitive Alternation

(27) a. pibā somam
    drink.Imp soma.Acc
    ‘Drink soma.’ Sanskrit
    (Rgveda VIII.36.1, from Jamison 1976)

b. pibā somasya
    drink.Imp soma.Gen
    ‘Drink (of) soma.’ Sanskrit
    (Rgveda VIII.37.1, from Jamison 1976)
5.1.3 Oblique Subjects

- Sanskrit had constructions that look like they correspond to the genitive and dative subjects of Urdu/Hindi.

- Hock (1990, 1991) concludes that the genitives can indeed be considered subjects.

\[(\text{28}) \quad \text{a. mama ekaḥ putro (vartate/asti)} \]
\[
\text{my one.Nom boy.Nom is}
\]
\[
\text{‘I have one boy.’ (Hock 1991, 57) Sanskrit}
\]

\[(\text{b. mer-a ek larka he)}
\]
\[
\text{my-M.Sg one boy.M.Sg.Nom be.Pres.3.Sg}
\]
\[
\text{‘I have one boy.’ Urdu/Hindi}
\]

- However, the evidence for the dative experiencers being subjects is weak to non-existent.

- Interestingly, however, there is evidence for DCM (29) among the dative experiencer verbs.

\[(\text{29}) \quad \text{pratibhāti mā/me}
\]
\[
\text{appears/seems me.Acc/me.Gen}
\]
\[
\text{‘appears/seems to me’ (Hock 1990, 128–129) Sanskrit}
\]

- Given this situation, Hock suggests that a possible hypothesis by which “Sanskrit reflects a transitional stage in which oblique-experiencer structures were acquiring subject properties but had not yet completed that process.” (Hock 1990, 136)

- This in fact what has been found for Marathi (Joshi 1993, Deo 2003).

5.2 Case in Middle Indo-Aryan

MIA also contained semantic DCM, despite the on-going erosion of the case system.

- Agents in a participial -ta construction can be marked with either an instrumental or a genitive. Genitives are restricted to animate agents (Andersen 1986).

5.3 Summary

- Semantic DCM is an old part of at least Indo-Aryan

- Semantic DCM persists in the language even though all of the case markers currently in use were innovated from about 1200 on.

- So, even though all the surface/morphological marking changed, semantic DCM (of both DSM and DOM) has always existed in the history of Indo-Aryan.
6 From Spatial Marker to Case: *ne* and *ko* in Urdu/Hindi

6.1 Historical Development of Ergative *ne*

- Butt and Ahmed (2011) argue that the Hindi/Urdu ergative *ne* was borrowed into the language to mark agents as part of language contact around 1400 CE (cf. Beames (1872–79), Kellogg 1893, 130–132, Tessitori (1913, 1914), Montaut (2003, 2006, 2009)).

- The neighboring language most probably used the form *ne* as both a dative/accusative and an ergative (as in, e.g., Haryani or Rajasthani).

(30) \( \text{mom}=\text{ne} \quad \text{sahib}=\text{ne} \quad \text{mar-a} \)
\[
\text{Pron.1.Sg=Acc/Dat Sahib.M.Sg=Erg hit-Perf.M.Sg}
\]

‘The Sahib hit me.’ (Shirani 1987) Haryani

**Question:** Why was only the ergative use adopted in Hindi/Urdu?

6.2 Historical Development of Dative/Accusative *ko*

6.2.1 Synchronic Distribution (Ahmed 2006)

**Indirect Objects**

(31) \( \text{anjum}=\text{ne} \quad \text{saddaf}=\text{ko} \quad \text{ciṭṭi}=\text{i} \quad \text{d-i} \)
\[
\text{Anjum.F.Sg=Erg Saddaf.F.Sg=Dat letter.F.Sg.Nom give-Perf.F.Sg}
\]

‘Anjum gave Saddaf a letter.’ Urdu

**Experiencer Subjects**

(32) \( \text{nadya}=\text{ko} \quad \text{dar} \quad \text{lʊg-a} \)
\[
\text{Nadya.F.Sg=Dat fear.M.Sg.Nom be attached-Perf.M.Sg}
\]

‘Nadya was afraid.’ Urdu

**Modal Uses Expressing Obligation (Must)**

(33) \( \text{nadya}=\text{ko} \quad \text{skul} \quad \text{ja-na} \quad \text{pur-a} \)
\[
\text{Nadya.F.Sg=Dat school.F.Sg.Obl go-Inf.M.Sg fall-Perf.M.Sg}
\]

‘Nadya had to go to school.’ Urdu

**Definite/Specific and Animate Objects**

(34) a. \( \text{nadya}=\text{ne} \quad \text{gari}=\text{ko} \quad \text{cule-ya} \quad \text{hɛ} \)
\[
\text{Nadya.F.Sg=Erg car.F.Sg=Acc drive-Perf.M.Sg be.Pres.3.Sg}
\]

‘Nadya has driven that specific car.’ Urdu

b. \( \text{nadya}=\text{ne} \quad \text{yasin}=\text{ko} \quad \text{dekʰ-a} \)
\[
\text{Nadya.F.Sg=Erg Yassin.M.Sg=Acc see-Perf.M.Sg}
\]

‘Nadya saw Yassin.’ Urdu
Spatial Uses

(35) saman g\textsuperscript{4}ar=ko pa\textsuperscript{2}c-a
luggage.M.Sg.Nom house.M.Sg=Dat/Acc reach-Perf.M.Sg
‘The luggage reached the house.’ Urdu

(36) k\textsuperscript{4}mra\textsuperscript{4}r=ko hc
house.M.Sg.Nom inside=Dat/Acc be.Pres.3.Sg
‘The room is towards the inside.’ Urdu

Temporal Uses

(37) cor rat=ko a-ya
thief.M.Sg.Nom night.F.Sg=Dat/Acc come-Perf.M.Sg
‘The thief came at night.’ Urdu

Summary

- Note: a theory of markedness by which a case marker is used to contrast a given NP with other NPs (distinguishability) or identifies a particular semantic role (indexing) in the clause cannot explain the wide variety of uses for ko in Urdu.

- Presumably all these uses of ko are historically and semantically related, but how? What does specificity have to do with recipient or location semantics?

6.2.2 History of Urdu ko

Earliest Examples (1200)

- Beames (1872–79:§56) reconstructs the Urdu ko to the locative of Sanskrit kaksha ‘armpit, side’ → Old Hindi k\textacuteccha, accusative k\textacuteccham → k\textacutecchu → kо → ko.

- According to Beames, the oldest documented examples of ko come from the writer Chand around 1200 (see also Kellogg 1893:130–131).

- This is about 200 years before the appearance of the first ne forms in Old Urdu/Hindi.

- Examples of Use from Chand:
  - a gift to the Brahmins (“dative”)
  - having made obeisance to all (“dative”)
  - for the war with Prithiraaj (“purpose”)
  - He seeks one of you. (“accusative”)

- Further early examples come from Baba Farid (1173–1266), a poet who wrote in Multan, now in Pakistan.
“Dative” Uses

(38) 

farid mē janya dokh mojh ko
Farid I know grief/pain I.Obl Acc/Dat
‘Farid, I know I have grief ... (lit. grief is to/at me)’

(Baba Farid, Verse 81, from Khan (2001:226))

(39) jındu kō samjhāi
life Acc/Dat teaches
‘(it) teaches to life’

(Baba Farid, Verse 1, from Khan (2001:142))

“Accusative” Uses

(40) dhūdē diye suhag kō
seek give husband Dat/Acc
‘seeking a husband’

(Baba Farid, Verse 114, from Khan (2001:263))

(41) jind ... haddā kō korkayi
life.F.Sg bone.F.Pl Acc/Dat rattle.Perf.F.Sg
‘Life ... rattled the bones.’

(Baba Farid, Verse 1, from Khan (2001:142))

Almost Modern Examples (1800)

- The dative and object marking uses continue into modern times.

- In addition, around 1800 ko is also found systematically with directed motion verbs such as cal ‘walk/go’, poāc ‘reach’ and ja ‘go’.

- A close examination this corpus shows that ko is generally used with endpoints where there is no guarantee that they have been attained.

- Past vs. Future Tense:

(42) ek vilayat mē poāce
one city in reached
‘reached a city’ (Dehalvi 1804)

(43) ıs manzil ko kub poāco-ge
this destination Dat/Acc when reach.2-Fut.Pl
‘When will (you) reach this destination?’ (Dehalvi 1804)

- Past/Perfect vs. Past Habitual:

(44) dīli mē ga-ye
Delhi in go-Perf.M.Pl
‘(they) went to Delhi’ (Dehalvi 1804)
(45) ek qafla sodagarō ka domāf q ko ja-ta th a
one caravan.M.Sg merchants Gen.M.Sg Damascus Dat/Acc go-Impf.M.Sg be.Past.M.Sg
‘A caravan of merchants used to go to Damascus.’ (Dehalvi 1804) Old Urdu

- **Concrete vs. Abstract Endpoints:**

(46) ek gorıstan mē poāce
one graveyard in reached
‘reached a graveyard’ (Dehalvi 1804) Old Urdu

(47) apne hoq ko poāc kor
self right Dat/Acc reach having
‘after having attained one’s right’ (Dehalvi 1804) Old Urdu

- Significantly, no instance of ko is found with a ‘come’. All of these examples are in the past tense, so the endpoint was attained.

### 6.2.3 Summary

- The earliest instances of ko are documented around 1200.
- This predates the appearance of ergative ne by about 200 years.
- Early uses of ko:
  - “dative” (recipients), abstract locations (experiencers)
  - marking the object, but when the objects are not part of a bounded event, or denote unattained goals/endpoints (‘seek’).
- In the 1800s, ko appears with verbs of directed motion, but seems to be used when the attainment of the endpoint/goal is not certain or with abstract goals/endpoints.
- NB: Compare this with Pāṇini’s Rule 2.3.12.

### 6.2.4 A Semantic Explanation for the Distribution of ko

- ko is originally a postposition derived from Sanskrit kaksha ‘armpit, side’.
- It is drawn into the system of spatial postpositions to mean something like AT.
- In particular, the original meaning of ‘side’ gives a meaning of spatial proximity, but not necessarily one of total spatial coincidence.
- Clauses can either express stative or dynamic eventualities, this gives rise to different uses/interpretations of ko.
The functions of modern *ko* explained:

- Indirect Objects: endpoints on a path (which can be attained or not, i.e., the recipient might get the thing or not).
- Experiencer Subjects: abstract locations or endpoints.
- Spatial and Temporal Uses: locations or endpoints.
- Specificity: While an endpoint may not be attained, it is a specific endpoint that has been pointed to.
  - Suggestion here is that the use of *ko* to mark specific objects derives from its use to express endpoints that are abstract, but specific.
  - Note that modern *ko* does not express telicity or boundedness, which fits in with the not-necessarily attained endpoint analysis.

6.3 Ergative *ne*: A Situation of Borrowing and Blocking

- *ko* arose early as a new case marker around 1200.
- It had a spatial origin denoting a location *at* or an endpoint on a path.
- This led to a range of uses marking dative recipients, experiencer subjects and also specific objects.
- The language at this point did not use a new case marker to mark proto-typical agents (vestiges of a system whereby ergative-type agreement and oblique marking on some agents remained).
- However, when it came into contact with a neighboring language that had a case marker which spanned both the functions of agent-marking (ergative) and dative/accusative, it borrowed only the agentive/ergative function.
- The dative/accusative function was already blocked by the presence of *ko*. 
6.4 Interim Conclusions

Claims:

• The primary reason for case markers to be innovated in Indo-Aryan is to keep alive a system of semantic distinctions made by case markers.

• Approaches which try to explain the innovation of case marking only in terms of markedness are too simplex.

• The available evidence points to a situation in which systematic semantic factors are at play in the usage development of case markers.

7 Rise of Oblique Subjects

• The availability of dative experiencers does not automatically coincide with a syntactic status of dative subjects (cf. Hock (1990, 1991) on Sanskrit).

• Cole et al. (1980) argue subjecthood status is acquired in stages.

• This fits in well with the observed data for at least Marathi (Joshi 1993, Deo 2003).

**Question:** why are oblique subjects a robust part of NIA, but not of OIA?

7.1 The Verbal Complex

• As in many other languages, the verbal paradigms of OIA were affected considerably by diachronic change.

• The only surviving paradigm from OIA is an old subjunctive form, now used in the present, the subjunctive and the future (see Butt and Rizvi (2010) for details).

• Most of the verbal predication in NIA is done periphrastically, with the main verb being an old participial form.

7.2 Hypothesis

• The restructuring of the verbal paradigm meant that the coding properties of arguments could be reconfigured.

• That is, the coding system for arguments was in principle open for reorganization.

• So the correspondence between \( \theta \)-roles/semantic-roles and grammatical relations (the domain of Linking Theories, see Butt (2006) for an overview) could be reconfigured.

• This, in effect, opened up more possibilities for the licensing of oblique subjects.
• It does not imply that the lexical semantics of the underlying verb classes changed — just the linking of the thematic roles to the grammatical relations → e.g., dative experiencers can be linked to subjects in NIA, which was not possible in OIA.

8 Summary

• Urdu/Hindi make pervasive use of oblique subjects, most of them in the context of DCM.

• Sanskrit did not generally allow for oblique subjects, but did have DCM.

• The original inflectional case marking system of Sanskrit was lost in MIA and new case markers drawn from mainly spatial sources were innovated in NIA.

• Claim:
  – The new case markers were innovated primarily in order to keep semantically-based DCM alive.
  – The rise of oblique subjects is related to the expression of semantically-based DCM.
  – But mainly they arise from a reconfiguration of the alignment between semantic roles and grammatical relations.

• Optionality of expression and transitional situations as in Marathi are particularly evident for (dative) experiencers because of a set of conflicting preference constraints.

Acknowledgements

Much of the work presented here was done in collaboration with the following people as part of a project on Historical Aspects of South Asian Case Marking (Project A24 of the SFB 471, funded by the DFG): Tafseer Ahmed, Ashwini Deo, Scott Grimm and Tikaram Poudel.

References


