Introduction to the
Formal Syntax, Semantics, and Morphology of
South Asian Languages

EMILY MANETTA AND AYESHA KIDWAI

1. Overview

Languages of South Asia, including those in the Indo-Aryan, Dravidian, Austroasiatic and Tibeto-Burman families, have a deep history of significant contributions to the study of language as a whole. Indeed, Panini’s grammar of Sanskrit (ca 500BC) has been remarkably influential in the development of the formal approaches to language structure even in modern linguistics.

South Asian languages feature a range of important empirical phenomena that are widely relevant to syntactic and semantic theory such as ergativity (e.g., Pandharipande and Kachru 1977; Comrie 1984; Saksena 1985; Hook 1985; Mahajan 1990), verb finality and complex predication structures (e.g., Hook 1974; Bashir 1989; Mohanan 1994; Butt 1995), non-nominative subjects (e.g., Jayaseelan 1999, 2004; Amritavalli and Jayaseelan 2003; Amritavalli 2004; Davison 2004a, b), and scrambling (e.g., Gambhir 1981; Mahajan 1990; Dayal 1994; Kidwai 2000).

More recently, linguists working on languages of South Asia have made contributions to the analysis of wh-movement and the spectrum of wh-in-situ constructions (Dayal 1996, 2000; Fanselow and Mahajan 2000; Simpson and Bhattacharya 2003), causation and valency (Beavers and Zubair 2013), mechanisms driving case assignment and agreement processes (Butt and King 2004; Bhatt 2005; Subbarao 2012), and finiteness (see contributions to Sundaresan, Ramchand, and McFadden 2014). Scholars in the field anticipate that investigations into the many lesser-studied languages of the region have the potential to be influential in formal linguistic theory.

With this background in mind, this special issue of Linguistic Analysis brings together contributions from leading scholars in the area of South Asian morphosyntax and semantics. The goal of the volume is to report on the state of research into the core questions driving South Asian formal linguistics, and to foster wide circulation...
of the analytical ramifications of this body of research for current frameworks. In this introduction we outline important themes, both theoretical and empirical, that link the diverse contributions, and highlight key advances being made in the work reported on here. Further, we identify fruitful areas for ongoing work both within the field of South Asian linguistics and in broader, comparative contexts.

2. Contributions to the Special Issue

In what follows we have grouped the articles in this collection by the nature of their analytical focus, and have provided a short description of each. Four of the ten articles in the volume develop semantic approaches to specific particles or dependent morphemes. The authors of these articles examine topics of wide relevance, including negative polarity (Erlewine and Kotek) definiteness (Simpson and Biswas), event structure (Balusu), and coordination (Jayaseelan). Though these articles draw primarily on data from a range of South Asian languages, some are also deeply comparative (Jayaseelan’s work linking Japanese with Malayalam), or engage substantively with the crosslinguistic significance of their findings (as in the conclusion in Erlewine and Kotek).

In their article *Even-NPIs in Dharamsala Tibetan*, Erlewine and Kotek investigate the syntax and semantics of two series of negative polarity items (NPIs) in Dharamsala Tibetan. The first uses the numeral ‘one’ with the *even* particle, while the second uses a *wh*-word combined with the *even* particle, with or without ‘one’.

(1) Dharamsala Tibetan

a. (lopchuk) *chi*-ye lep-ma-song
   student one-*even* arrive-NEG-PRFV
   ‘No student/one arrived.’

b. Su-(*chi*)-ye lep-ma-song
   who-(one)-*even* arrive-NEG-PRFV
   ‘No one arrived.

Erlewine and Kotek’s proposed semantics for these two distinct NPIs is based on Lahiri’s (1998) analysis of similar constructions
in Hindi-Urdu. They propose that the *even*-particle introduces two presuppositions: one additive and one scalar. Crucially, *wh*-words in Dharamsala Tibetan have no independent indefinite use. A core innovation of their account involves employing the additive component of *even* to generate the indefinite from the *wh*-word ranging over Hamblin-style (1973) alternatives (see Crnič 2011). A major contribution of their article is a compositional semantics for *wh*-**even** NPIs of this type, which has not before now been proposed. Potentially productive results include explanations for intervention effects of even-NPIs in *wh*-interrogatives in Dharamsala Tibetan and Korean (Beck and Kim 2006), as well as the use of *wh*-**even** as a free-choice item in both Dharamsala Tibetan and Bangla (Ramchand 1996).

Andrew Simpson and Priyanka Biswas’ article *Bare Nominals, Classifiers and the Representation of Definiteness in Bangla* presents the so-called bare-noun and bare-classifier (in which the neutral order of classifier > noun is inverted) patterns, both of which may be interpreted as definite. They point out that the choice between these bare forms to express definiteness has not yet been investigated in the literature.

(2)   a. chabi Ta          b. chabi
      key   CL              key
      ‘the key’            ‘a key’, ‘keys’, or ‘the key’

Simpson and Biswas illustrate that uniqueness, identifiability, and activation level in the mind of the speaker all play a role in the choice of bare classifier vs. bare noun in Bangla. These findings in Bangla thus underscore the fact that definiteness is a multi-faceted notion whose complexity is partially masked in languages such as English which have a single lexical item for use in all instances of definite reference.

Rahul Balusu’s article *The Eventive Predicator –gaa in Telugu* provides a syntactic and semantic analysis of the Telugu morpheme –gaa, which attaches as a phrasal affix to all non-verbal predicates. This analysis is intended to go beyond previous descriptions of –gaa which established it as an adverbial or adjectival suffix alone.

(3)   raamu andam-gaa / nunna-gaa  parigetteDu  
      TELUGU
      Ramu beauty-gaa / smooth-gaa  run.pst.3msg
      ‘Ramu ran beautifully/smoothly.’
He proposes that –gaa contributes stage-level interpretation (following e.g., Adger and Ramchand 2003) when suffixed to adjectives, and that adverbs formed with –gaa have the internal structure of a small clause. Balusu analyzes –gaa as a Pred⁰ head in the terms of Ramchand’s (2008) first phase syntax, and suggests that its behavior may well be parallel to similar affixes and/or case markers in other Dravidian languages, as well as Welsh, Russian, and Finnish.

K.A. Jayaseelan’s contribution to this volume, Decomposing Coordination: The Two Operators of Coordination contrasts coordination in English to coordination in Japanese and the Dravidian language Malayalam. In Japanese and Malayalam, coordinators appear to have a “double life,” as both coordinators and operators.

(4) -uum/-oo as coordination markers:  
John-um, Bill-um, Peter-um wannu  MALAYALAM  
John-CONJ Bill-CONJ Peter-CONJ come. PAST  
‘John, Bill, and Peter came.’

(5) -uum/-oo as operators:  
aarǝ entǝ coodicc-aal-um, awan koDukk-um  MALAYALAM  
who what ask-COND-CONJ he give-FUT  
‘Whoever asks for whatever (thing), he will give (it).’

Jayaseelan’s article seeks to unify these two apparently distinct uses of –uum/-oo. He claims that English lexicalizes the concatenation operator of coordination, whereas Japanese and Malayalam lexicalize the choice function operator. The choice function operator can apply to any set-denoting expression and interpret it as a quantifier. Thus when –uum/-oo are seen on coordinands, they are actually copies of the operator (the operator itself is silent, following Szabolcsi (2013)), but when inside a quantifier, they represent the operator itself.

Four more of the articles included in this special issue make important empirical and analytical contributions to syntactic investigations already ongoing in the literature. Kush’s article on gapping in Hindi-Urdu is part of a wider new investigation into ellipsis processes in South Asian languages (see, for instance, Gribanova and Manetta, forthcoming). Hock and Ross’ work on verb constructions adds to a long research tradition in South Asian formal linguistics exploring complex predicates and serial verb constructions (e.g., Hook 1974; Mohanan 1994; Butt 1995). Bhattacharya’s work on honorificity in
Central Maghadan Prakrit languages applies Chomsky’s (2013) notion of Agree as labeling to the syntax of politeness in four markedly understudied South Asian languages. Finally, Manetta’s article, *Verb Position and Question Markers in a Verb-Second Language*, follows up on a series of works on the formal syntax of interrogatives in South Asian languages (Mahajan 1990; Dayal 1994; Simpson and Bhattacharya 2003; Manetta 2010; Slade 2012; Bhatt and Dayal 2014). We describe each of these contributions in some detail below.

In his article *Notes on Gapping in Hindi-Urdu: Conjunct Size and Focus Parallelism*, Dave Kush investigates gapping structures in which verbal material in the second conjunct of a coordinated structure may be omitted under identity with material in the first conjunct, and contrasts the properties of gapping in English and Hindi-Urdu.

(6) Manu ate a mango and Tanu ate a banana.

Manu-ne aam khaa-yaa aur Tanu-ne kela khaa-yaa.

‘Manu ate a mango and Tanu a banana.’

Based on specific properties of gapping in Hind-Urdu, he argues that the operation must apply to clause-sized conjuncts (Ross, 1967; Sag, 1976; Jackendoff, 1971; Jayaseelan, 1990), as opposed to smaller, vP-sized conjuncts (Coppock, 2001; Lin, 2002; Johnson, 2009; Toosarvandani, to appear). Kush proposes that Hindi-Urdu gapping features the ellipsis of a TP following focus movement of non-elided constituents to a TP-external position. More broadly, Kush seeks to show that since Hindi-Urdu, like English, resists gapping in embedded contexts, constraints on embedded clause gapping cross-linguistically cannot be derived from claims about conjunct size. He instead explores an alternative that relies on focus parallelism. While there is much to be developed in Kush’s approach, the basic ramifications of the empirical observations make an important contribution to the crosslinguistic study of ellipsis processes.

Tanmoy Bhattacharya’s article *Inner/Outer Politeness in Central Māgadhan Prākrit Languages: Agree as Labeling* examines honorificity in four understudied languages originating from the eastern branch of Middle Indo-Aryan: Maithili, Magahi, Angika and Kurmali. Bhattacharya begins with a useful introduction to these languages,
historical contexts, and demographic details. Of specific interest is the fact that these languages exhibit multiple agreement (agreement with multiple arguments) in honorificity.

(7) o tōra dekh-əl-thunh Maithili
    He.H you.NH.ACC/DAT see-PAST-(3h+2NH)
    ‘He(H) saw you(NH).’

Bhattacharya then goes on to argue that politeness in these languages is manifest in two distinct domains: one inner position in vP and one at the level of the entire clause. From this duality, Bhattacharya argues for two distinct types of Agree relations and that these Agree relations are also a result of the process of labeling, in line with Chomsky (2013). He suggests that in Central Māgadhan Prākrit languages, politeness layers are in fact concerned with identification of phases (vP and CP).

Hans Hock and Daniel Ross’ paper South Asian “Agreeing Verb Constructions”: Serial Verbs, Compound Verbs, Pseudocoordination, or what? identifies so-called “Agreeing Verb Constructions” (AVCs) across a wide range of South Asian languages (Munda, Dravidian, Indo-Aryan, and Tibeto-Burman), as a distinct type of complex verbal structure. Both verbs in an AVC are finite and exhibit agreement in person, gender (where applicable), and number.

(8) cel-v-ēm all-ēm Old Tamil
    go-NPST-1PL become.NEG-1PL
    ‘We will not go.’

They compare these structures to both Serial Verb Constructions and Pseudocoordination and claim that they should be accounted for not in the narrow syntax, but treated as “constructions” in the interface component (see Agbayani et al. (2011) and Molina Muñoz (2013)). Hock and Ross do not reject the minimalist, generative approach. They instead propose that constructions are Spell-out devices that operate post-syntactically. For AVCs, a construction copies the morphological form of the inflected verb to the other verb in the construction.

Emily Manetta’s paper Verb Position and Question Markers in a Verb-Second Language investigates the multiple question markers found on polar and alternative questions in the verb-second Indic language Kashmiri.
On the basis of the positioning of the two question markers with respect to the second-position verb, she argues for an approach to verb-second in Kashmiri (following Manetta 2010), in which the second position verb is always found in C regardless of clause type. The facts in Kashmiri contribute more broadly to the wider research program on the whether/Q operator and its morphological spellout (e.g., Han and Romero 2004; Beck and Kim 2006; Cable 2010), leading to the conclusion that Kashmiri marks with overt morphology multiple types of dependencies (whether polar/alterative operators or wh-dependencies) that are not overtly marked in related Indic languages.

In addition, two contributions to the special issue focus on experimental methods to inform research on formal syntax. Anna Gavarró and Maya Leela’s article *Child Relativized Minimality and the Acquisition of Word Order in Malayalam* examines the predictions that the Relativised Minimality (RM) approach to language acquisition (Friedmann, Belletti and Rizzi 2009) makes in the domain of scrambling in an SOV language like Malayalam (Dravidian). The RM approach claims that the computation of movement is delayed in children when an intervener appears between probe and goal. Gavarró and Leela’s three comprehension experiments with children aged 2-5 years yield good performance, and they thus conclude that both SOV and scrambled word orders are available to children from early on. They conclude that their results are at odds with the predictions of the RM approach, and point to early maturation of A’ movement.

The second contribution to draw on experimental work is that of Miriam Butt, Farhat Jabeen and Tina Bögel. In their article, entitled *Verb Cluster Internal Wh-Phrases in Urdu: Prosody, Syntax and Semantics/Pragmatics*, Butt et al investigate immediately post-verbal wh-constituents (as discussed in Bhatt and Dayal 2007 and Manetta 2012).

(9) k’aa tSI gatsh-kh-aa pagaah garI _y/N you go-2sgps-q tomorrow home_ ‘Will you go home tomorrow?’
In particular, they analyze verb cluster-internal wh-constituents, claiming that this word order occurs only when the verb is in focus and is motivated by prosodic concerns (as supported by their experimental evidence). Butt et al.’s account is couched in Lexical Functional Grammar (LFG) and integrates previous work on information structure (Krifka 2008) and second occurrence focus (Beaver and Velleman 2012; Büring 2015). This article represents the continuation of a larger investigation into the syntax of post-verbal constituents in verb-final languages (Bhatt and Dayal 2007; Manetta 2012; Butt 2014; Simpson and Choudhury 2015), a critical contribution currently being made to the wider field based on data from South Asian languages.

3. Summary

Overall, we view the work presented in this special issue as a snapshot of the highly productive ongoing formal morpho-syntactic and semantic research programs that engage substantively with languages of South Asia. In particular, we wish to emphasize the role that lesser-studied languages play in the contributions in this volume, and the increasing trend in applying the findings of experimental work on South Asian languages to questions of analytical importance. Finally, the diversity of theoretical frameworks employed here to capture the presented data illustrate the broad relevance of this work to a number of current approaches. It is our hope that this special issue spurs additional research, and underlines the vitality of the contribution languages of South Asia may make to theoretical linguistics quite generally.

Works Cited


