

# Urdu Patterns and Theoretical Perspectives

Miriam Butt & Heike Zinsmeister

ESLLI 2009  
Bordeaux

July 28th, 2009

This is where we are so far:

- Looked at how case works crosslinguistically.
- Case helps with identifying participant arguments, but also does extra work (semantic contrasts, e.g. boundedness, specificity, tense/aspect).
- Three strategies for identifying participant arguments (Kiparsky):
  - position
  - case marking
  - agreement with grammatical relations
- Most languages employ a mixture, with greater emphasis on one of the strategies.

## German Patterns:

- German has case marking, though mostly on the articles and with much syncretism in the paradigms.
- German allows for scrambling of major constituents.
- German clauses are organized as follows  
[*Vorfeld*] V2 [*Mittelfeld*] VC [*Nachfeld*]
- The main scrambling of participant arguments among each other happens in the *Mittelfeld*.
- Factors governing the scrambling: animacy, case marking, pronoun or not

## Next:

- Compare patterns and strategies from Urdu (also an Indo-European language) to get a better idea of what is possible crosslinguistically.
- See what theoretical approaches have to say about the patterns and strategies in German and Urdu.

# Urdu Case Markers

Urdu case is marked via a set of clitics (Butt and King 2005).

## *The Urdu Case Markers*

<b>Clitic</b>	<b>Case</b>	<b>Gram. Func.</b>	<b>Morph. Effect</b>
∅	nominative	subj/obj	none
ne	ergative	subj	oblique marking on NP
ko	accusative	obj	oblique marking on NP
	dative	subj/ind. obj	oblique marking on NP
se	instrumental	subj/obl/adjunct	oblique marking on NP
k-	genitive	subj (infinitives) specifier	agrees with head noun none
mẽ/par/tak/∅	locative	obl/adjunct	oblique marking on NP

The Urdu Ergative basically marks agents, but generally only with perfect morphology.

- (1) a. ram                      gaṛi                      cāla-ta                      (hε)  
Ram.M.Sg.Nom car.M.Sg.Nom drive-Impf.M.Sg be.Pres.3.Sg  
‘Ram drives a car.’  
b. ram=**ne**                      gaṛi                      cāla-yi                      (hε)  
Ram.M.Sg=Erg car.M.Sg.Nom drive-Perf.M.F.Sg be.Pres.3.Sg  
‘Ram has driven a/the car.’

This is a very usual pattern in languages of the world, called *split-ergativity*.

You do, however, also get semantic alternations.

- Volitionality with Unergatives.

- (2) a. ram                      k<sup>h</sup>ās-a  
Ram.M.Sg.Nom cough-Perf.M.Sg  
‘Ram coughed.’  
b. ram=**ne**                      k<sup>h</sup>ās-a  
Ram.M.Sg=Erg cough-Perf.M.Sg  
‘Ram coughed (purposefully).’

- Modality (Must vs. Want) with Infinitives

- (3) a. **nadya=ko**              zu                      ja-na              hε  
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
- b. **nadya=ne**              zu                      ja-na              hε  
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

The *ko* marker (originally a spatial marker) has multiple functions.

**Accusative:** marks specificity and is obligatory on animate objects.

- (4) a. nadya=ne      gaṛi      cala-yi      hε  
Nadya.F.Sg=Erg car.F.Sg.Nom drive-Perf.F.Sg be.Pres.3.Sg  
'Nadya has driven a car.'
- b. nadya=ne      gaṛi=**ko**      cala-ya      hε  
Nadya.F.Sg=Erg car.F.Sg=Acc drive-Perf.M.Sg be.Pres.3.Sg  
'Nadya has driven the car.'
- c. nadya=ne      **yasin=ko**      mar-a      hε  
Nadya.F.Sg=Erg Yassin.M.Sg=Acc hit-Perf.M.Sg be.Pres.3.Sg  
'Nadya has hit Yassin.'

As a dative, *ko* can mark canonical recipients, but also subjects of psych verbs.

(5) anjum=ne          saddaf=ko          citt<sup>hi</sup>          d-i  
Anjum.F.Sg=Erg Saddaf.F.Sg=Dat letter.F.Sg.Nom give-Perf.F.Sg  
'Anjum gave Saddaf a letter.'

(6) a. nadya=ko          ḍar          lag-a  
Nadya.F.Sg=Dat fear.M.Sg.Nom be attached-Perf.M.Sg  
'Nadya was afraid.'

b. nadya=ko          kahani          yad          a-yi  
Nadya.F.Sg=Dat story.F.Sg.Nom memory come-Perf.F.Sg  
'Nadya remembered the story.'

- The unmarked argument is generally called “nominative” (alternatively, absolutive).
- It is only used for subjects and objects.
- It is used to contrast with dative (experiencers) and ergative (agents) on subjects and accusative (specificity, animacy) on objects.

# Urdu Agreement Patterns

- If the subject is unmarked, the verb agrees with the subject.
- If the subject is marked and the object is unmarked, then agree with the object.
- If both subject and object are marked, then show default masculine singular agreement.

- (7) a. nadya                      gaṛi                      cāla-ti                      hε  
Nadya.F.Sg.Nom car.F.Sg.Nom drive-Impf.F.Sg be.Pres.3.Sg  
'Nadya drives a car.'
- b. nadya=ne                      gaṛi                      cāla-yi                      hε  
Nadya.F.Sg=Erg car.F.Sg.Nom drive-Perf.F.Sg be.Pres.3.Sg  
'Nadya has driven a car.'
- c. nadya=ne                      gaṛi=ko                      cāla-ya                      hε  
Nadya.F.Sg=Erg car.F.Sg=Acc drive-Perf.M.Sg be.Pres.3.Sg  
'Nadya has driven the car.'



Although in principle, major constituents can scramble fairly freely with respect to one another, you get examples of what has been called “Word Order Freezing” (Mohanan 1994).

- (10) boṭal                      kēci                      toṛ-ti                      hε  
bottle.F.Sg.Nom scissors.F.Sg.Nom break-Impf.F.Sg be.Pres.Sg  
'The/A bottle breaks the/a scissors.'  
**NEVER:** The/A scissors breaks the bottle.

So: **STRONG** preference for the first item to be the Subject.



# Role of Information Structure

Butt and King (1997) claim the following for Urdu/Hindi Clauses:

- First constituent: Topic [-New,+Prom(inent)]
- Immediately Preverbal Constituent: Focus [+New,+Prom]
- Stuff after the verb: Backgrounded Information [-New,-Prom]
- Everything else: Completive Information [+New, -Prom]

(12) nadya kahā=se a rah-i hε?  
Nadya.F.NOM where=from come Stat-FSG be.PRST.SG  
'Where is Nadya coming from?'

(13) [nadya to]]<sub>T</sub> ab<sup>hi</sup> [tōfi]<sub>CI</sub> [bazar=se]<sub>F</sub> xarid  
Nadya.F.NOM indeed just now toffee.F.NOM market.M=from buy  
rah-i t<sup>h</sup>-i  
Stat-FSG be-PST.FSG  
'Nadya was just buying toffee AT THE MARKET.'

Example for backgrounded information:

- (14) to apne ap=se ek vada kar li-ya t<sup>h</sup>-a [māĩ=ne]<sub>Back</sub>  
so self self=from one vow make take-PF.MSG be-PST.MSG I=ERG  
'So I made a vow to myself.' (Dilwale Dulhania Le Jayenge)



- Mohanan (1994) points out that there is a strong dispreference against two case markers of the same kind being immediately adjacent to one another.
- This is a very surfacy constraint that is not expected from a deep theoretical linguistic point of view.
- Mohanan analyzes this as a form of the Obligatory Contour Principle (OCP)
- OCP (Leben 1973): prohibition against adjacent identical entities — first formulated with respect to phonological phenemona (Tone/Intonation).

## ***ko* with an animate object (required)**

- (16) ram=ne            baccõ=ko            samhal-a  
Ram.M.Sg=Erg child.M.Pl.Obl=Acc take care of.Perf.M.Sg  
'Ram took care of the children.'

## ***ko* as marker of psych subject (required)**

- (17) ram=ko            xat            lik<sup>h</sup>-na            paṛ-a  
Ram.M.Sg=Dat letter.M.Sg.Nom write-Inf.M.Sg fall-Perf.M.Sg  
'Ram had to write a letter.'

## **Combining These Results in a Dispreferred Utterance**

- (18) ??ram=ko            baccõ=ko            samhal-na            paṛ-a  
Ram.M.Sg=Dat child.M.Pl.Obl=Acc take care of-Inf.M.Sg fall-Perf.M.Sg  
'Ram had to take care of the children.'

But if you put stuff between the two *ko*-marked arguments, then the sentence is fine.

(19)

ram=ko            kal            baccõ=ko            samhal-na            paṛ-a  
Ram.M.Sg=Dat yesterday child.M.PI.Obl=Acc take care of-Inf.M.Sg fall-Perf.M.Sg  
'Ram had to take care of the children yesterday.'

(20)

ram=ko            [pause] baccõ=ko            samhal-na            paṛ-a  
Ram.M.Sg=Dat            child.M.PI.Obl=Acc take care of-Inf.M.Sg fall-Perf.M.Sg  
'Ram [pause] had to take care of the children.'

The same can be shown with *se* (Instrumental/Comitative) marked arguments (Mohanani 1994).

What do Theories of Case have to say about the patterns in Urdu and German?

- Case Marking?
- Word Order?
- Information Structure?
- Identification of Participant Arguments?

Theories surveyed (see Butt 2006 for more complete discussion):

- GB/Minimalism
- Lexical-Functional Grammar (LFG)
- Linking Theories (e.g., Kiparsky, Role and Reference Grammar (RRG))

All theories assume some kind of a default word order which is almost completely determined by an underlying semantic preference.

- Thematic Role Hierarchy (LFG, other Argument-structure approaches)
- Semantic/Thematic Roles encoded hierarchically in a tree structure (e.g., GB/Minimalism, First Phase Syntax)
- Logical Order of Arguments according to Semantic Decomposition.

# Word Order and Thematic Roles

Various different proposals for Thematic Role Hierarchies exist (see Levin and Rappaport 2004 for an overview).

- LFG (based on Kiparsky's evidence from idioms):  
*agent* > *beneficiary* > *recipient/experiencer* > *instrument* > *theme/patient* > *location*
- Grimshaw (1990):  
(Agent (Experiencer (Goal/Source/Location (Theme))))
- Siewierska (1993)  
*agent* > *patient* > *recipient* > *benefactive* > *instrumental* > *spatial* > *temporal*
- Jackendoff (1990), based on semantic decomposition:  
*Actor* > *Patient/Beneficiary* > *Theme* > *Source/Goal/ReferenceObject* > *IdentificationalGoal/ReferenceObject*

# Word Order and Semantic Decomposition

The RRG approach as an example for semantic decomposition.

<b>Aktionsart</b>	<b>LS Representation</b>	<b>Example</b>
State	<b>predicate'</b> (x,y)	<b>be'</b> (Pat,[lawyer]) <i>Pat is a lawyer.</i>
	<b>predicate'</b> (x)	<b>shattered'</b> (glass) <i>The glass is shattered.</i>
Activity	<b>do'</b> (x,[predicate(x)])	<b>do'</b> (children,[ <b>cry'</b> (children)]) <i>The children cried.</i>
	<b>do'</b> (x,[predicate(x,y)])	<b>do'</b> (Dana,[ <b>eat'</b> (Dana,pizza)]) <i>Dana ate pizza.</i>
Achievement	INGR <b>predicate'</b> (x)	INGR <b>popped'</b> (bubble) <i>The bubble popped.</i>
	INGR <b>predicate'</b> (x,y)	INGR <b>see'</b> (Kim,message) <i>Kim noticed the message.</i>
Accomplishment	BEC <b>predicate'</b> (x)	BEC <b>melted'</b> (snow) <i>The snow melted.</i>
	BEC <b>predicate'</b> (x,y)	BEC <b>know'</b> (Leslie,Korean) <i>Leslie learned Korean.</i>

# Linking Theories

Linking Theories (LFG, RRG, Kiparsky, etc.) assume that

- There is some kind of semantic predicate-argument encoding (thematic roles, Proto-Roles, semantically decomposed structures).
- This semantic predicate-argument encoding is related to grammatical relations via a *linking relation*.
- Different theories formulate different heuristics/rules to accomplish this linking.
- Items providing information about linking could be:
  - thematic roles/lexical semantic decomposition (see RRG) — all theories use this.
  - case marking (some theories use this, for others case simply spells out the relation)
  - agreement (some theories use this, for others it is an epiphenomenon)
  - position (some theories use this, for others it is an epiphenomenon)

LFG assumes a feature correspondence between grammatical relations and thematic roles (or proto-roles).

(21)	<b>Grammatical Functions</b>	<b>Features</b>
	SUBJ	$[-r, -o]$
	OBJ	$[-r, +o]$
	$OBJ_{\theta}$	$[+r, +o]$
	$OBL_{\theta}$	$[+r, -o]$

(22)	<b>Features</b>	<b>Grammatical Functions</b>
	$[-o]$	SUBJ, OBL
	$[+o]$	OBJ, $OBJ_{\theta}$
	$[-r]$	SUBJ, OBJ
	$[+r]$	$OBJ_{\theta}$ , $OBL_{\theta}$

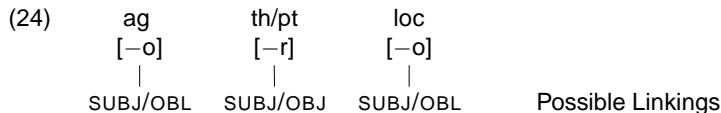
(23) **Classification of Thematic Roles**

Patientlike roles:  $[-r(\text{restricted})]$

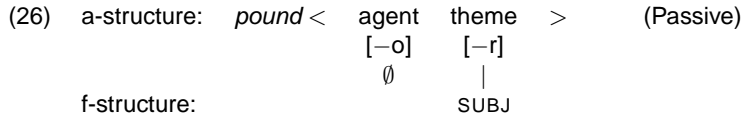
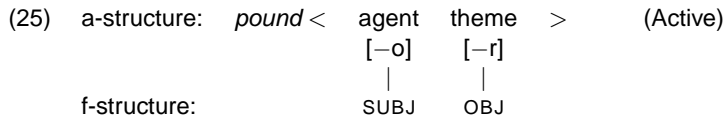
Secondary patientlike roles:  $[+o(\text{bjective})]$

All others:  $[-o]$

# Linking in LFG



Mapping/Linking Principles and some Wellformedness Conditions then apply to provide the actual linking.



Case Marking and further semantic factors (e.g., aspect) have been integrated into LFG's Linking Theory, but agreement or position have not.

# Word Order and Hierarchical Semantic Relations

In GB/Minimalism, the assumption has been that semantic relations are coded structurally/hierarchically within a VP/vP.

Example for English Dative Alternation from Beck and Johnson (2004)

- (27) a. Kim gave a bone to the dog.  
b. Kim gave the dog a bone.

The structural positions are interpreted semantically.

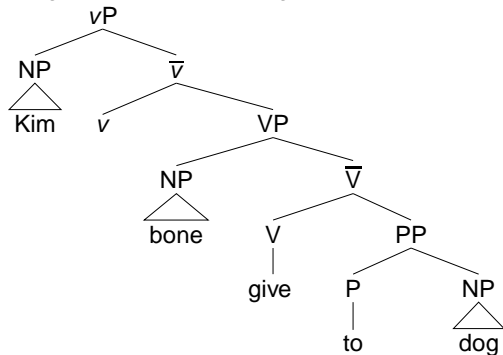
(28) Example from Adger (2003:139)

- a. NP daughter of vP → interpreted as agent
- b. NP daughter of VP → interpreted as theme
- c. PP daughter of  $\bar{V}$  → interpreted as goal

# Word Order and Hierarchical Semantic Relations

Example for English Dative Alternation from Beck and Johnson (2004)

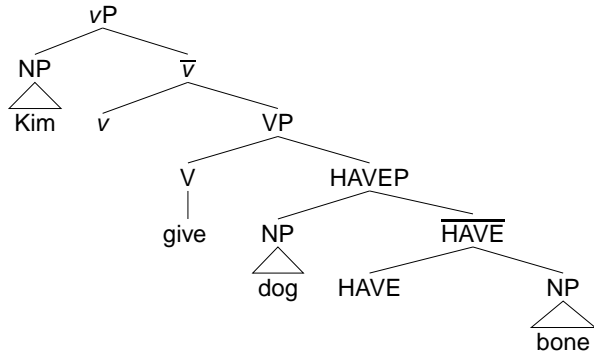
(29) Kim gave a bone to the dog. (First Phase)



# Word Order and Hierarchical Semantic Relations

Example for English Dative Alternation from Beck and Johnson (2004)

(30) Kim gave the dog a bone. (First Phase)



In GB/Minimalism, Scrambling is taken to be related to various factors (different approaches assume different reasons):

- 1 Need to show agreement (movement to an agreement licensing position)
- 2 Need to receive case (movement to a case assigning position)
- 3 Need to be interpreted in a certain manner: Aspect, Specificity, Topic/Focus, so movement to a relevant position.

Framework allows for a very large number of nodes, which each get interpreted in a certain manner.

Word Order follows from this (to some extent), as does case marking (spell out of a given set of information/feature bundle).

Different Word Orders follow from:

① Information Structure Packaging, e.g., Butt and King (1997) for Urdu

- First constituent: Topic [-New,+Prom(inent)]
- Immediately Preverbal Constituent: Focus [+New,+Prom]
- Stuff after the verb: Backgrounded Information [-New,-Prom]
- Everything else: Completive Information [+New, -Prom]

② Special Interpretation of some Syntactic Positions

Example: Immediately preverbal position can be incorporated for generic readings: *book-selling*.

③ Prosodic Needs (position of clitics/weak pronouns)

# Case, Agreement and Position — Linking Theories

## Agreement

- LFG: Agreement is uninteresting, a well-formedness checker which helps identify grammatical relations (SUBJ, OBJ, I.OBJ).
- Other Linking Theories: Agreement is one *linker*, helps to identify which semantic role goes with which grammatical relation.

## Position

- Kiparsky's Linking Theory: Position is one *linker*, helps to identify which participant goes with which grammatical relation.
- LFG and Other Linking Theories: Position has to do with other factors (e.g., information packaging).

## Case

- LFG: Case can be integrated into Linking Theory (Butt and King 2003, 2005)
- Kiparsky's Linking Theory: Case is one *linker*, helps to identify which participant goes with which grammatical relation.
- Other Linking Theories: Case is a realization of a given linking relation.

Within LFG, arguments have been put forward for *Constructive Case*.

- The idea is that the case markers themselves contribute information to syntactic and semantic analysis of the clause.
- Argumentation on the basis of Wambaya (Australian, Nordlinger 1998) and Urdu (Butt and King 1991).
- Note that in many Urdu examples, the only difference between two clauses is the case marker – it therefore seems sensible to locate the meaning difference with the case marker.

- (31) a. nadya=ne          garī                  cala-yi                  hε  
Nadya.F.Sg=Erg car.F.Sg.Nom drive-Perf.F.Sg be.Pres.3.Sg  
'Nadya has driven a car.'
- b. nadya=ne          garī=**ko**                  cala-ya                  hε  
Nadya.F.Sg=Erg car.F.Sg=Acc drive-Perf.M.Sg be.Pres.3.Sg  
'Nadya has driven the car.'

Sample LFG Entry for the Urdu dative/accusative *ko*

(32) *ko*

Possibility 1    ( $\uparrow$  CASE) = ACC  
                  (OBJ  $\uparrow$ )  
                  ( $\uparrow_{sem-str}$  SPECIFICITY) = +

Possibility 2    ( $\uparrow$  CASE) = DAT  
                  (GOAL  $\uparrow_{arg-str}$ )  
                  (SUBJ  $\uparrow$ )  $\vee$  (OBJ<sub>go</sub>  $\uparrow$ )

# Choosing Between Options

Generally, Linking Theories allow for a flexible range of *Linking Options* (e.g., the dative shift in English).

For example, in LFG, there is generally more than one option to choose from.

(33)

ag	th/pt	loc	
[-o]	[-r]	[-o]	
SUBJ/OBL	SUBJ/OBJ	SUBJ/OBL	Possible Linkings

**Choices** are governed by:

- morphosyntactic factors (e.g. passives, applicatives)
- semantic factors (e.g., aspect, affectedness – see English dative shift)
- information structure (topic, focus)

**Example:** Locative Inversion in Chicheŵa (Bresnan and Kanerva 1989, Bresnan 1994) is sensitive to focus status of the first element.

- (34) a. a-lendô-wo      a-na-bwér-á                      ku-mu-dzi  
2-visitor-2 those 2 SB-REC PST-come-IND 17-3-village  
'Those visitors came to the village.'  
(visitors=subject) Chicheŵa
- b. ku-mu-dzi      ku-na-bwér-á                      a-lendô-wo  
17-3-village 17 SB-REC PST-come-IND 2-visitor-2 those  
'To the village came those visitors.'  
(village=subject) Chicheŵa



- So far, all the choices/options were governed by structural factors.
- These factors represent a binary decision space.
- However, more and more evidence is being amassed that word order additionally is a matter of preference based on
  - weight/heaviness of the constituent (syntactic/prosodic) (Bresnan et al. 2007, Wasow 2002)
  - Animacy (e.g., Müller 1999 for German)
- These types of preferences can be incorporated into theories of syntax/case via the integration of *Optimality Theory*.

- Optimality Theory allows for a collection of violable constraints.
- These constraints are ranked in terms of the importance (how bad it is to violate them).
- The ranking of the constraints can differ across languages.
- The inventory of constraints are supposed to be
  - drawn from our system of Universal Grammar
  - based on processing/comprehension preferences (which can conflict with one another, cf. the bidirectional OT developed by Blutner)

- Knowing about preferences not only provides a better handle within theoretical linguistics on the observed data.
  - Helps with understanding different word order patterns (e.g., German)
  - Helps with dispreferences about two identical case markers (Case OCP).
- Within computational linguistics, knowing about preferences can help with:
  - disambiguation (you have 200 readings for a parse, what's the best one?)
  - generation (you have 200 possible word orders corresponding to an abstract representation, which is the one you should pick?)

- But identifying preferences is difficult to do with a small body of data.
- **Hence:**
  - corpus linguistic work becomes relevant — statistical evaluations of huge corpora can identify preferences (Wednesday&Thursday)
  - psycholinguistic work can help identify processing/production preferences (Friday)

# References I

- Adger, D. (2003). *Core Syntax: A Minimalist Approach*. Oxford University Press, Oxford.
- Beck, S. and Johnson, K. (2004). Double objects again. *Linguistic Inquiry*, **35**(1), 97–124.
- Blake, B. (2001). *Case*. Cambridge University Press, Cambridge. Second Edition.
- Bresnan, J. (1994). Locative inversion and the architecture of universal grammar. *Language*, **70**(1), 72–131.
- Bresnan, J. and Kanerva, J. (1989). Locative inversion in Chicheŵa: A case study of factorization in grammar. *Linguistic Inquiry*, **20**, 1–50.
- Butt, M. (2006). *Theories of Case*. Cambridge University Press, Cambridge.
- Butt, M. and King, T. H. (1997). Null elements in discourse structure. Ms., <http://ling.uni-konstanz.de/pages/home/butt>.
- Butt, M. and King, T. H. (2003). Case systems: Beyond structural distinctions. In E. Brandner and H. Zinsmeister, editors, *New Perspectives on Case Theory*, pages 53–87. CSLI Publications, Stanford.
- Butt, M. and King, T. H. (2005). The status of case. In V. Dayal and A. Mahajan, editors, *Clause Structure in South Asian Languages*. Springer Verlag, Berlin.
- Deo, A. (2003). Valency change and case marking: Marathi dative experiencers. Handout from PIONEER Workshop on 'Case, Valency and Transitivity'.
- Joshi, S. (1993). *Selection of Grammatical and Logical Functions in Marathi*. Ph.D. thesis, Stanford University.
- Mohanan, T. (1994a). *Argument Structure in Hindi*. CSLI Publications, Stanford.
- Mohanan, T. (1994b). Case OCP in Hindi. In M. Butt, T. King, and G. Ramchand, editors, *Theoretical Perspectives on Word Order Issues in South Asian Languages*. CSLI, Stanford, CA.
- Nordlinger, R. (1998). *Constructive Case: Evidence from Australian Languages*. CSLI Publications, Stanford.
- Ramchand, G. (2008). *Verb Meaning and the Lexicon: A First Phase Syntax*. Cambridge University Press, Cambridge.