THE STRUCTURE OF COMPLEX PREDICATES IN URDU

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF LINGUISTICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By
Miriam Jessica Butt
December, 1993

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Peter Sells
(Principal Advisor)

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Charles A. Ferguson

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Paul Kiparsky

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Veneeta S. Dayal
(Rutgers University)

Approved for the University Committee on Graduate Studies:

I dedicate this dissertation to my brother, Yassin Mischal Butt.

Abstract

Constructions variously described as complex predicates, compound verbs, composite predicates, or serial verbs range across quite an impressive number of differing expressions in differing languages. As such, it is not immediately obvious how to arrive at a consistent formulation of the differences, if there are any, among them, and how to provide a unifying analysis of complex predicates crosslinguistically.

The South Asian language Urdu employs a wide variety of complex predicates. I conduct a detailed examination of two differing Urdu complex predicates: the *permissive* and the *Aspectual* complex predicates. The Urdu permissive brings into focus the essential problem complex predicates pose for theories of syntax. The permissive is a complex predicate formed by the combination of two semantic heads in the syntax (not the lexicon). These semantically distinct heads correspond to a single syntactic predicate, which may be discontinuous. I show that this discontinuity at phrase structure does not affect the status of the permissive as a complex predicate. The problematic aspect for theories of syntax is thus the question of how to represent the fact that a complex predicate may behave both like a syntactically complex structure with respect to certain phenomena, and like a syntactically simple structure with respect to other phenomena. Within Lexical-Functional Grammar (LFG), this mismatch in semantic and syntactic information is easily represented in terms of independent levels of representation which are related to one another by a theory of linking. However, LFG as originally formulated does not allow for the existence of a semantically complex, but syntactically discontinous single head.

Recent work intent on solving this problem in a variety of theoretical frameworks has tended towards an argument structure approach. In particular, Alsina (1993) formulates an argument structure account of Romance causatives within LFG which allows for the argument structure composition of two discontinuous heads in the syntax. I follow Alsina in proposing an analysis of complex predicate formation at argument structure. However,

rather than moving towards a progressively minimalistic and abstract argument structure, which does not explicitly contain thematic role or other semantic information, as proposed in Grimshaw (1990), S. Rosen (1989), Ritter and S. Rosen (1993), and Alsina (1993), I take up the kind of argumentation found in Van Valin (1990), for Role and Reference Grammar (RRG), and propose an elaborated argument structure based on Jackendoff's (1990) theory of Conceptual Semantics. Urdu Aspectual complex predicates provide evidence for an elaborated level of argument structure. An Aspectual complex predicate is well-formed only if constraints on semantic properties such as volitionality and inception/completion are met: a main verb negatively specified for one of these domains cannot combine with a light verb positively specified for the same domain.

In conclusion, this dissertation presents an in-depth examination of the structure and properties of two differing Urdu complex predicates, the permissive and the Aspectual complex predicates. I formulate a unifying theory of complex predicate formation and in the process address issues concerned with argument structure, linking, and case marking. Finally, I show that the theory of complex predicates I present not only allows a successful account of Romance restructuring verbs and Japanese *suru* 'do', but can also be used as a firm base of comparison for an analysis of serial verb constructions.

Acknowledgements

Having put to paper everything I could about complex predicates within the constraints of time, space, knowledge, and wisdom inherent in a dissertation, I find that I do not want to turn my back on the subject in weary relief. Instead, I still seem to harbor a deep seated belief that a complete understanding of complex predicates is bound to yield the question to the answer of Life, the Universe, and Everything. K.P. Mohanan was my first mentor in linguistics and the first person to introduce me to complex predicates. His delight in exploring alternative, inside-out ways of looking at things left me confused at first, but ultimately taught me very much about linguistic and scientific methodology. My linguistic debt to Mohanan is vast, and I would like to thank him for his time and patience with me.

When Mohanan left Stanford for warmer climes, Peter Sells inherited several Mohanan waifs and proceeded to firmly guide us through the process of writing a dissertation. The particular genius of Peter Sells for me consisted of the fact that, no matter how garbled my thoughts were, he could recognize the essence of what I was trying to get at in no time, and sort out my jumbled thoughts into an organized system. In addition to this, Peter Sells was unique in the amount of effort and care that went into the process of advising me: he was always there to respond to desperate cries for help, and always had a good joke to crack as well. Working with Peter Sells was an invaluable part of this dissertation.

The dissertation could also not have been written without the input from Joan Bresnan, Charles Ferguson, Paul Kiparsky, Veneeta Srivastav, and Tom Wasow. Joan Bresnan was a member of my committee until my final year, her sabbatical year. A meeting with Joan Bresnan is representative of perhaps the most challenging aspect of my graduate school career. I would leave her office either in complete devastation, resolved not to become a linguist at all, or beaming with exultation because the intense questioning, examination, and discussion of ideas and linguistic issues at large had been so deeply satisfying and engaging.

Charles Ferguson gave generously of his time, taught me very much about South Asian literature I was not aware of, and deceptively mildly insisted that I do things properly. This dissertation would not be as thorough without his input. Paul Kiparsky and Veneeta Dayal joined the committee as the dissertation was already well under way, but had an immediate and pronounced effect. Paul Kiparsky attacked issues of morphology and lexical semantics I had not considered before. Veneeta Dayal took a close look at my data and claims, pushed me on them steadily, and streamlined my argumentation in the process. Tom Wasow, who served on the Oral Exam committee, generously commented on my work in precise detail. His input allowed a more intelligent presentation of my material. Finally, I would like to thank Yoshiko Matsumoto for serving as the chair of my Oral Exam committee.

Alex Alsina, Mary Dalrymple, Henriëtte de Swart, and Tara Mohanan all read parts of the dissertation and provided me with detailed and insightful comments time and time again. In particular, I could not have done without the many discussions with Mary Dalrymple about the architecture of LFG, and the many exchanges of ideas with Tara Mohanan about the nature of complex predicates and the structure of Hindi or Urdu. In this respect, Alice Davison at the University of Iowa had a most stimulating effect as well. Many sections in this dissertation arose from a need to address the questions and issues she raised.

I also owe many thanks to the more general linguistic community at Stanford. Many discussions with Lynn Cherny, Michael Inman, Michio Isoda, Chris Manning, John McWhorter, Yo Matsumoto, and Chris Piñón kept me on my toes and provided me with interesting material to think about. Veena Dwivedi, although at the University of Massachusetts at Amherst, must also be included among this list, as she was the one person I could always most immediately engage in discussions on the structure of Hindi/Urdu.

Jennifer Cole, Tracy King, and Gillian Ramchand are the fellow graduate students without whose friendship and support I could not have done. From the time Tracy King and I organized our first conference together in our second year, it was clear that we made a great team. As a team, we went on to organize another conference, and to co-author a paper. In this spirit, we might have co-authored a dissertation, except for the fact that she benightedly insisted on working on Russian. Many aspects of this dissertation have benefitted from Tracy's input, and some of the discussion on case marking is based on joint work.

Gillian Ramchand was the most delightful peer I could have had in graduate school. Whether it was over breakfast, over an A's game, or during a session of late night TV

watching, she was always ready to engage in a discussion of linguistic issues, or explain a particular problem I had not been quite able to come to grips with as yet. Since she and I were interested in many of the same issues, and sometimes even some of the same languages, Gillian is perhaps the person I learned the most from at Stanford.

I owe many thanks to Jennifer Cole. Besides being a constant source of irreverent linguistic jokes which would keep me giggling despite myself, she was the person whose advice on how to organize a written document, prepare a presentation, or LaTeX something more efficiently was the most detailed, yet comprehensive, succint, and useful. She was also the person who firmly put me on the path towards this dissertation, and kept me there in times when I felt most insecure.

I would also like to thank Aditi Lahiri, who generously provided me with a place to work at the University of Konstanz. At just the right time, her good advice and support instilled the sense I needed that my work was indeed worth the while.

Many thanks go to my mother, Ilse Achilles, who provided much of my financial support during my last year at Stanford. The fact that I knew I could fall back on her gave me the necessary room for a timely completion of the dissertation.

Finally, Gina Wein, Michelle Murray, Kyle Wohlmut, and Emma Pease are the people who provided much of the technical support and know-how that goes into the writing of a dissertation. I could not thank them enough for their help with the paperwork, for keeping the printers and computers going, for answering formatting questions, restoring lost files, and for being great people to work with. Kyle Wohlmut in particular deserves special thanks for keeping me sane towards the bitter end, as he was the one person left I could go drink beer with. Even better, I could go do a spot of archery with him, shoot for beer, and then go drink the beer.

Contents

Chapter 1

Introduction

Constructions variously described as complex predicates, compound verbs, composite predicates, or serial verbs range across quite an impressive number of differing expressions in differing languages. As such, it is difficult to form a clear cut conception of what a complex predicate, a serial verb, a compound verb, or even a complement construction really is. It is also not immediately obvious how to arrive at a consistent formulation of the differences, if there are any, among them, and to provide a unifying analysis of complex predicates crosslinguistically. For example, do Chinese resultatives (Li 1990, Huang 1992), serial verbs (Sebba 1987, Baker 1989, Durie 1993), or compound verbs in South Asian languages (Vale 1948, Hook 1974, T. Mohanan 1990, Ramchand 1990, etc.) all have the same structure as Romance restructuring verbs or causatives (Aissen and Perlmutter 1983, Davies and C. Rosen 1988, S. Rosen 1989, Alsina 1993) or the complex predicate with *suru* 'do' in Japanese (Grimshaw and Mester 1988, Sells 1989, Matsumoto 1992)?

The South Asian language Urdu[1] employs a wide variety of complex predicates. In fact, complex predicates are generally pervasive throughout South Asia. Consequently, grammarians and linguists working on South Asian languages have a long tradition of describing and analyzing complex predicates (for example, Vale 1948, Kachru 1966, Bahl 1974, Hook 1974, Masica 1976, Hook 1991). In this dissertation, I conduct a detailed examination of two differing Urdu complex predicates: the *permissive* and the *Aspectual* complex predicates. I propose an analysis of complex predicates based primarily on the

[1]Urdu is spoken primarily in Pakistan and northern parts of India. It is structurally almost identical to Hindi, one of the official languages of India. The two languages are so closely related that some researchers refer to them as Hindi-Urdu.

Urdu data, and then go on to extend the analysis suggested by Urdu to complex predicates found in Romance and Japanese. I also review some of the facts known about serial verbs, and suggest a possible approach to formulating a clear differentiation between serial verbs and complex predicates.

A close examination of the structural and semantic properties of the Urdu permissive and Aspectual complex predicates indicates that the definition of a complex predicate in (1) accurately characterizes the common structural properties of two otherwise very different complex predicates (see also T. Mohanan 1993c).

(1) **Definition of a Complex Predicate:**

• The argument structure is complex (two or more semantic heads contribute arguments).
• The grammatical functional structure is that of a simple predicate. It is *flat*: there is only a single predicate (a *nuclear* PRED) and a single subject.
• The phrase structure may be either simple or complex. It does not necessarily determine the status of the complex predicate.

Evidence from the Urdu permissive clearly shows that the defining characteristic of a complex predicate is that a complex argument structure corresponds to a simple predicate. The Urdu permissive furthermore brings into focus the essential problem complex predicates pose for theories of syntax. The permissive is a complex predicate formed by the combination of two semantic heads in the syntax (not the lexicon). The semantically distinct heads combine into a single syntactic head which may be discontinuous. This discontinuity at phrase structure does not affect the status of the permissive as a complex predicate.

The problematic aspect for theories of syntax is thus precisely the question of how to represent the fact that a complex predicate may behave both like a syntactically complex structure with respect to certain phenomena, and like a syntactically simple structure with respect to other phenomena. Within Lexical-Functional Grammar (LFG), which encodes the syntactic structure of an expression in terms of several distinct levels of representation, it is possible to show that complex predicates must be simple with respect to grammatical functions (relations), but may be either simple or complex with regard to c-(onstituency) structure (phrase structure).

Since Cattell (1984) carefully laid out the issues and problems surrounding Composite Predicates in English, syntactic tools which allow a characterization of some complex predicate constructions have been developed in several syntactic frameworks. Within

Government-Binding (GB), for example, a theory of *head-to-head movement* was proposed (Baker 1988). Within Relational Grammar (RG), *clause union* allows the combination of two predicates into one (Aissen and Perlmutter 1983, Davies and C. Rosen 1988). However, these kinds of approaches cannot provide quite the necessary tools for a successful analysis of the Urdu permissive. The permissive in contrast with the superficially similar *instructive* in Urdu shows that the former is clearly a complex predicate, while the latter must be analyzed as a complement construction. Despite these clear differences between the two constructions, however, they are exactly parallel in terms of scrambling, negation, and coordination. This indicates that a purely phrase structural account, such as head-to-head movement, will not suffice for an analysis of the permissive. A clause-union approach, while not couched in exclusively phrase structural terms, cannot make a clear distinction between differing types of clause-union: the complex predicate type of clause-union cannot be easily distinguished from the clause-union involved in an infinitival complement construction (the instructive for example).

More recently, therefore, treatments of complex predicates have tended towards an argument structure approach. This is true for research done in both GB and LFG on Romance, Japanese, and South Asian languages (for example, S. Rosen 1989, Grimshaw and Mester 1988, T. Mohanan 1990, Matsumoto 1992, Alsina 1993). While LFG initially allows a straightforward characterization of complex predicates, as given in the definition in (1), LFG as originally formulated (see Bresnan 1982b) did not provide for the possibility of discontinuous heads of a complex predicate at phrase structure. The Urdu permissive clearly shows that such a possibility must be allowed for.

Alsina (1993) formulates an argument structure account of Romance causatives within LFG which allows for the argument structure composition of two discontinuous heads in the syntax. I follow Alsina in proposing an analysis of complex predicate formation at argument structure. However, rather than moving towards a progressively minimalistic and abstract argument structure, which does not explicitly contain thematic role or other semantic information, as proposed in Grimshaw (1990), S. Rosen (1989), Ritter and S. Rosen (1993), and Alsina (1993), I take up the kind of argumentation found in Van Valin (1990), for Role and Reference Grammar (RRG), and propose an elaborated argument structure based on Jackendoff's (1990) theory of Conceptual Semantics. The elaborated argument structure approach is motivated by the semantic restrictions on complex predicate formation needed to account for the Urdu Aspectual complex predicates. The two parameters I identify in

particular are volitionality (conscious choice) and inception/completion. I argue that a minimalistic approach to argument structure does not allow an adequate characterization of these parameters, and, in turn, precludes an adequate analysis of complex predicate formation with regard to Urdu Aspectual complex predicates.

The dissertation proceeds as follows. Chapter 2 provides some background to Urdu, and sketches the assumptions I make as to its structure. In particular, as Urdu is a morphologically ergative language, and my treatment of case marking is not necessarily standard, but follows T. Mohanan (1990, 1993a), I provide a brief discussion of ergativity and case. I also follow T. Mohanan (1990) in the assumption that Urdu has a flat clause structure, and therefore briefly discuss issues of configurationality. In the final section of Chapter 2, I provide a brief sketch of LFG, the theoretical framework in which I couch most of my argumentation.

Chapter 3 provides an in-depth look at the Urdu permissive. I first establish that, while both the permissive and the instructive are formed through a combination of a finite predicate with an infinitive, the permissive is a complex predicate, while the instructive consists of a matrix verb and an infinitival complement. I then show that the two constructions nevertheless are exactly parallel in terms of scrambling, negation, and coordination. This establishes that complex predicate formation must take place at argument structure. The remainder of the chapter then provides additional insights into the structure of the permissive by examining Urdu infinitival complements in general.

Chapter 4 introduces the Urdu Aspectual complex predicates. I first provide an account of the structure of Aspectual complex predicates, and show that they cannot be analyzed as compound verbs which are formed in the lexicon, or as constructions involving auxiliaries (see Hook 1974). I then go on to show that Aspectual complex predicate formation is governed by the semantic notions of volitionality (conscious choice) and inception/completion. An exploitation of these notions allows an almost complete analysis of possible and impossible complex predicate formations.

An analysis within LFG for the permissive and the Aspectual complex predicates is then developed in Chapters 5 and 6. Chapter 5 first introduces an elaborated argument structure based on Jackendoff's (1990) theory of Conceptual Semantics, formulates a theory of complex predicate formation, and shows how Aspectual complex predicate formation can easily be accounted for within an elaborated argument structure approach. Chapter 6 concentrates on the problems of linking presented by the Urdu permissive and presents a

detailed account of the linking between c-structure, f(unctional)-structure, and a(rgument)-structure. Primarily, I make use of case marking information in addition to the algorithm for relating c-structure to f-structure originally formulated in Bresnan (1982b).

In Chapter 7, I show that my analysis of complex predicates is easily extendable to complex predicates of other languages discussed in the literature. In particular, I provide an analysis of Japanese *suru* 'do', and of auxiliary selection in Italian restructuring verbs. On the basis of my analysis of complex predicates, I also go on to suggest possible means of clearing up some of the terminological confusion that has resulted in the literature because of the fact that no clear definition of a complex predicate has been available. In particular, the same construction has sometimes been analyzed either as a complex predicate or a serial verb (Marathi). I propose a set of criteria which clearly distinguish between serial verbs and complex predicates.

Finally, in Chapter 8, I present the conclusions to be drawn, and the implications a satisfactory analysis of complex predicates presents for theories of syntax. I maintain that only those constructions which meet the criteria in (1) can be analyzed as complex predicates, and that complex predicates must be distinguished from serial verbs, compound verbs, or complement constructions in order to allow for a unifying analysis of complex predicate formation crosslinguistically. Furthermore, I propose that the key to a successful analysis of complex predicates is a treatment of argument structure processes, but that the recent trend towards more abstract argument structure representations will not allow an adequate characterization of complex predicate formation. Rather, an elaborated argument structure approach such as the one proposed here must be employed.

Chapter 2

Relevant Background Issues

2.1 Introduction

This chapter first lays out the relevant assumptions I make about the structure of Urdu, and then goes on to give a brief sketch of LFG, the theoretical framework I utilize for my analysis of complex predicates. For the most part, I take the view of Hindi/Urdu laid out in T. Mohanan (1990), which is not always in accordance with other approaches. It is therefore worthwhile to lay out my assumptions clearly with regard to case and configurationality in order to avoid possible confusion. In the interests of maximum clarity, I also begin by providing a brief description of the particular dialect of Urdu I draw the bulk of my data from.

2.2 Dialect Studied

Most of the data in this dissertation are drawn from the dialect of Urdu spoken in Lahore, Pakistan. The remainder of the data are culled from grammars or linguistic articles about either Hindi or Urdu. The two languages Hindi and Urdu are very closely related, so much so that some researchers refer to one language of Hindi-Urdu (e.g., Davison 1991a,b). In fact, Ernest Bender has written two grammars (Bender 1967a,b), one for Urdu and one for Hindi, which are almost completely identical, down to the most minor section heading. The only differences one can readily discern are in terms of the vocabulary.

There are, of course, various dialects of both Hindi and Urdu, which differ somewhat structurally. For example, in some regions of India a dialect of Hindi is spoken which does

not exhibit any verb agreement. This kind of a dialect can also be found in Lahore, but it is not extremely widespread. I therefore take there to be a number of varying dialects of both Hindi and Urdu, which are quite closely related to one another. In fact, some dialect of Hindi may be closer to the Lahori dialect of Urdu I investigate, than the version of Urdu spoken in Lucknow, India.[1]

Despite the documented similarities between Hindi and Urdu, native speakers who are not linguists will often insist on making a sharp distinction between the two languages. These differences seem to be motivated entirely by political and religious differences between India and Pakistan, and Hindus and Muslims. The interested reader can find a good description of the subleties involved in the regional and social language differences in Gumperz (1971). Gumperz looks at various levels of communal relations in India and discusses the varied use of language at these levels in some detail. There seem to be four easily identifiable levels of relations (Gumperz 1971:81).

> 1) the all-India level, defined by the subcontinent; 2) the region, defined by literary language and distinctive caste patternings; 3) the subregion, defined by certain spoken dialects, and showing some peculiar cultural distinctiveness and characteristic castes; and 4) the local level, which may constitute either a single village or a group of villages tied by common kinship or other social ties.

At these levels, Gumperz makes distinctions between local vernaculars, argots used for trade at the subregional level, subregional dialects in urban centers. In particular he notes that the typical urban dweller will be able to demonstrate a good command of standard Hindi, but will additionally usually be familiar with English as an argot and command one or another of the sub-regional dialects. Monolingualism in South Asia is thus a rarity. It should be noted that while Gumperz's work is oriented towards India, the general picture he presents can be applied to Pakistan as well. The main difference is that the complicating factors of caste and a multitude of different religions are not found in Pakistan, so the picture is a little simpler than that described for India.

Furthermore, unlike dialects of Urdu spoken in India, the Urdu spoken in Lahore has been very isolated since the partitioning of the British Empire of India into Pakistan and India. In fact, it is very difficult to find native, monolingual, speakers of Urdu in Lahore.

[1]The Urdu spoken in Lucknow is held to be the representative of *pure* Urdu.

Most of the Lahori population has grown up speaking either Punjabi or English (depending on such factors as class and education) as a first language, and Urdu as a second language. However, with the increasing push towards Urdu as the language of government, education and prestige, as opposed to English, the situation has changed subtly within the last generation. During some research on child language acquisition I conducted in 1989, it became very clear that parents whose first language was Punjabi (the language of low prestige) were addressing their children only in Urdu, but would switch to Punjabi when addressing peers. The children are thus growing up as native speakers of Urdu, with a passive knowledge of Punjabi. Whether this practice of child-rearing is causing Punjabi to affect the dialect of Urdu in Lahore remains a subject for investigation (for an initial examination Hindi-Punjabi code-switching in Delhi, see Gumperz (1971)). Within the scope of this dissertation, I will merely point out the cases where the data is only found in the Lahori dialect (as far as I have been able to determine at any rate).

I chose my informants so that there would be as little influence from Punjabi as possible in their speech. This was motivated by the fact that Punjabi is more similar to Urdu than English, or some of the other languages common in Lahore, and that therefore a structural influence from Punjabi seemed more likely than from other languages. I used two main informants, and always checked the data I extracted from them against other, occasional, informants from Lahore, the existing literature on the phenomenon, and speakers of Hindi. All of the data in this dissertation has thus been confirmed by a source other than my two main informants. The informants both grew up speaking English and Urdu in the house, with English as the more primary language. They initially employed both Urdu and English within the framework of their education, but then later switched to an American High School, where Urdu was taught as a subject, but not generally used. In the home of one of the informants, Farsi was also a language occasionally employed, while in the home of the other, Punjabi was sometimes used. However, the use of Punjabi was not part of the general communicative practice among the family members.

2.3 Case

2.3.1 The Morphology

Table (1) shows the inventory of case clitics in Urdu. T. Mohanan (1990) divides the Hindi/Urdu case system up into three distinct morphological realizations: case clitics, postpositions, and inflected stem forms. Postpositions, such as *ke-liye* 'for', *ke-pas* 'at', and *ke-pic*^{h}*e* 'behind', are generally used with adjuncts. They surface in some of the examples I use, but are otherwise not of any particular significance for this dissertation.

(1)

Case Clitic	**Function**
∅	Nominative
ne	Ergative
ko	Accusative
ko	Dative
se	Instrumental
kaa/kii/ke (M)/(F)/(Obl)	Genitive
mẽ	Locative (in)
par	Locative (on, at)
tak	Locative (toward)

The nominative, also referred to as *direct* case (Kachru 1980), is phonologically null. The stem forms of nominative, or direct arguments, are never inflected. This is in contrast to adjuncts and non-nominative arguments. The stem forms of nouns ending in *-aa* (mostly masculine nouns) are always inflected when they are used as adjuncts or non-nominative arguments. Table (3), based on T. Mohanan (1990:80) shows the possible inflections of a representative masculine noun *baccaa* 'child'.

(2)

Function	**Singular**	**Plural**
Nominative	baccaa	bacce
Oblique	bacce	baccõ
Vocative	bacce	bacco

The *oblique* form of the stem is used when a noun is followed by a case clitic, or when that noun is used as a locative. This is illustrated in (3). In (3a), the subject *baccaa* 'child' is unmarked and glossed as nominative. In (3b) it appears in the ergative. Notice that here the form of 'child' is the oblique *bacce*. Example (3c) demonstrates the use of the oblique form *ḍaakxaane* 'post-office' in a locative context, where the presence of an overt case clitic is optional. Example (3d) repeats (3c), but the clitic *tak* 'toward' is realized here.

(3) a. **baccaa** ṭiivii dek^{h} rah-aa hai
child.M=Nom TV.M-Nom see stay-Perf.M.Sg be.Pres.3.Sg
'The child is watching TV.'

b. **bacce=ne** aaj ṭiivii dek^{h} liy-aa hai
child.M.Obl=Erg today TV.M-Nom see take-Perf.M.Sg be.Pres.3.Sg
'The child has already watched TV today.'

c. baccaa aaj **ḍaakxaane** jaa rah-aa hai
child.M=Nom today post office.M.Obl go stay-Perf.M.Sg be.Pres.3.Sg
'The child is going to the post office today.'

d. baccaa aaj **ḍaakxaane=tak** jaa rah-aa hai
child.M=Nom today post office.M.Obl=toward go stay-Perf.M.Sg be.Pres.3.Sg
'The child is going to the post office today.'

The observant reader may have noticed that in (3a) and (3b) above, the object *tiivii* 'TV' is also glossed as nominative.[2] I thus take there to be two nominative arguments in (3a). It should be noted that this is not in keeping with the traditional terminology established by the bulk of the literature on ergativity (some references are: Dixon 1979, Levin 1983, Marantz 1984, Van Valin 1981, Slobin 1993), which makes a distinction between Nominative-Accusative and Ergative-Absolutive systems (but also see Van Valin 1990, Bobaljik 1992, Bittner and Hale 1993 for a somewhat different perspective). In a sentence like (3a), therefore, the object would traditionally be glossed as accusative. In (3b), it would be glossed as absolutive. The contrast between (3a) and (3b) would prompt an analysis of Urdu as a *split-ergative* language. However, as discussed in the next section, although Urdu makes uses of an ergative case clitic, it is not a split-ergative language in the traditional sense of the word.

2.3.2 The Ergative

The defining characteristic of ergative languages is taken to be that they group the subjects of intransitive verbs (S) together with the objects (O) of transitive verbs (Dixon 1979, Van Valin 1990). This may manifest itself either in terms of overt morphological case marking

[2]Abbreviations used are as follows. F = feminine; M = masculine; Erg = ergative; Nom = nominative; Gen = genitive; Dat = dative; Acc = accusative; Inst = instrumental; Loc = locative; Inf = infinitive; Obl = oblique; Perf = perfect; Impf = imperfect; Stat = stative; Pres = Present; Subj = Subjunctive; Pron = pronoun; Sg = singular; Pl = plural. A '-' indicates a morpheme boundary, while a '=' separates a clitc from a lexical item.

(morphological ergativity), or in terms of differing syntactic behaviour of the two groups (syntactic ergativity). In accusative languages such as English, on the other hand, subjects of intransitive verbs are grouped together with subjects of transitive verbs (A). Ergative languages thus may be considered to be 'marked' in the sense that they do not treat all subjects alike, but align intransitive subjects with objects. From the perspective of a theory such as Government-Binding, whose original conception was based primarily on accusative languages like English and Romance, and where case marking is associated structurally with subject/object position, ergative languages present quite a challenge. A relatively simple analysis of ergative systems within this framework postulates that an ergative system is underlyingly the *inverse* of an accusative system (Marantz 1984).

This type of analysis predicts that, since an ergative system is the inverse of an accusative system, a child learning an ergative language must be faced with quite a challenge. Research on ergative languages within child language acquisition has thus mostly been directed at efforts to uncover the "learnability problem" (e.g., Fortescue 1985, Ochs 1985, Pye 1990, Schiefflin 1985, Slobin 1993). However, as is evident from the literature, concisely summarized in Van Valin (1993), and confirmed by my own research on the acquisition of the Urdu ergative (Butt 1991), ergative systems are in fact acquired easily and quite quickly.

As Van Valin (1990, 1993) points out, since the overwhelming number of ergative systems are split-ergative in some way or form, the interesting question for child language acquisition is actually how children are able to master a case marking split conditioned by aspect (e.g., Georgian), agentivity (e.g., Tsova-Tush), or pronominal forms versus full NPs (e.g., Punjabi) so quickly. The learnability problem for him thus does not reside in the mere existence of ergativity, but in the question of how the pattern of split ergativity is acquired.

Indeed, other analyses of ergative systems have attempted to provide a more detailed understanding of how the ergative case is conditioned.[3] Mahajan (1990), although mostly concerned with questions of movement in Hindi, must provide an analysis of the ergative as well. He proposes that noun phrases be allowed to have two cases: Structural Case and Inherent Case. An ergative argument then would be analyzed as being structurally nominative, since it is the subject, but inherently ergative. Bittner and Hale (1993), on the other hand, in their examination of Inuit, Warlpiri, and Hindi propose that Case is

[3]I only discuss Mahajan (1990) and Bittner and Hale (1993) here, and very briefly at that, as they are directly relevant for Hindi/Urdu. There are, of course, many other relevant approaches to ergativity. See, for example, Silverstein (1976), Levin (1983), Johns (1992).

assigned in particular Case-binding configurations. An ergative system under this analysis arises when Infl Case-binds the VP-internal subject, but the verb does not Case-bind the object. The particular structural configurations posited for each of the languages are motivated by syntactic phenomena such as agreement, control, and anaphora. This allows a distinction to be drawn between Inuit as a *raising ergative* language, and Warlpiri as as *transparent ergative* language, and an account of agreement, binding, and scope differences found between the two languages follows quite nicely from the proposed analysis. Hindi is presented as a further type of ergative system, a *domain-restricting* type, which accounts for the split ergative pattern conditioned by aspect. However, the data on Hindi does not do justice to the intricacies of the actual ergative pattern the language presents.

In fact, Van Valin (1990) argues that a structural account inherently fails to do justice to split ergative systems conditioned by parameters such as aspect or agentivity. Van Valin proposes to account for various phenomena of *split intransitivity* in terms of Role and Reference Grammar (RRG), which allows a semantic, rather than a structural, analysis of split ergativity, as well as the Italian auxiliary selection facts. In particular, Van Valin argues that the semantic parameters of inherent lexical aspect (Aktionsart) and agentivity underlie split ergativity crosslinguistically.

It should be noted here that Garrett (1990) provides conclusive historical evidence from Anatolian and Papuan languages, which shows that the development of an ergative system cannot be conditioned by semantic factors, but must be motivated by processes of reanalysis of unproductive or rare instrumental markers in null-subject clauses. However, the fact that the development of an ergative system cannot be motivated semantically does not preclude the ergative marker gaining a semantic significance in the aftermath of reanalysis. In fact, this is essentially how I propose to view the Urdu ergative. It is well known (see Garrett 1990 for a concise description) that the Indo-Iranian tense/aspect splits arose through the deverbal adjective *-*to*-. It had a passive interpretation with transitive verbs, but an active one with intransitives. The *-*to*- first began to be used as a marker of the perfective participle. Ultimately, the verbs it appeared with were then reanalyzed as having active voice. Thus, noun phrases marked as instrumental in a passive construction were reanalyzed as ergative subjects of an active sentence and a split-ergative pattern based on transitivity and perfectivity resulted (Dixon 1979). However, language change in Urdu has not stopped there. The data below show that Urdu deviates from a split-ergative pattern. I propose that the correlation of ergativity with transitivity/perfectivity is being reanalyzed as a correlation

between ergativity and volitionality. As discussed in Chapter 4, this is particularly evident from data involving Aspectual complex predicates. It would therefore seem likely that the development of complex predicates from originally biclausal constructions has provided the locus of change from an transitivity marking system to a volitionality marking system (Garrett, p.c.). This hypothesis of course remains a topic for further research, and I limit myself here to demonstrating that the Urdu ergative case-marking system indeed deviates from a simple split-ergative system.

As mentioned, Hindi/Urdu has been described synchronically as a split ergative language (Dixon 1979, Bittner and Hale 1993), where the split is exclusively conditioned by transitivity and perfective aspect. However, discrepancies from the expected split ergative pattern have also long been noted in the literature.[4] In particular, Pandharipande and Kachru (1977) have marched Hindi through the phenomena used in Dixon (1979) to identify ergative systems, and found that Hindi does not fit the definition of a split ergative system as articulated in Dixon. An ergative system under the terminology established by Dixon is one which groups subjects of intransitive verbs (S) with objects of transitive verbs (O) with regard to a number of syntactic phenomena such as case marking or verb agreement. On the other hand, an accusative system is one which treats subjects of intransitive verbs (S) and subjects of transitive verbs as belonging to same group. Pandharipande and Kachru show on the basis of evidence from verb agreement, past participial modification, relativization, and some other phenomena that S is not always aligned with O, but that S behaves like A some of the time. In other words, for some syntactic processes, the subject status of a given NP, whether it be ergative or not, is relevant, while for some other processes the particular case marking on a given NP is relevant.[5]

In the remainder of this section, I describe some of the data which elucidate the ergative pattern found in Urdu The data show that the ergative must have, subsequent to its reanalysis from an instrumental, been invested with semantic content. Khan (1987), T. Mohanan (1990), and Butt and King (1991) all argue that the ergative must be analyzed as a marker of agentivity or volitionality in Urdu/Hindi (also see Van Valin (1990) and Holisky (1987) for similar argumentation on Tsova-Tush and Acehnese, and Fortescue and Lennert Olsen (1993) for acquisition data from West Greenlandic).

[4]For example, see W.S. Allen (1951) for one of the first lucid descriptions of the Hindi ergative pattern.

[5]It should also be noted here, that this is true for case marking as well. As will be demonstrated, the ergative marker can appear both on S and and A, the nominative on S and O. There is thus no consistent grouping of S and O vs. A in terms of case marking.

There are several immediate exceptions to the postulated robust split between transitive/perfective/ergative versus intransitive/perfective/nominative. For example, several intransitive verbs exist which require an ergative subject. A representative example is the intransitive *cʰĩ̃ĩ̃k* 'sneeze' in (4). Some other verbs of this type, which would be classified as *unergative* under the Unaccusative Hypothesis (Perlmutter 1978, Burzio 1986) are *nahaa* 'bathe', *bol* 'speak' and *khããns* 'cough'.[6]

(4) aamraa=ne cʰĩ̃ĩ̃k-aa
Amra.F=Erg sneeze-Perf.M.Sg
'Amra sneezed.'

While the kind of data shown in (4) could be treated as instances of lexical exceptions, the data in (5) are harder to account for under the view that the ergative is fulfilling a purely structural function. The intransitive verb *ciikʰ* 'scream' in (5) can appear with either a nominative, or an ergative subject. When it appears with an ergative subject, there is a sense of purposefulness, or volitionality, that is not present with the nominative subject: the action of screaming could have been involuntary in (5a), but not in (5b). Examples of other intransitive verbs like this are *has* 'laugh' and *ro* 'cry'.

(5) a. aamraa ciikʰ-ii
Amra.F=Nom scream-Perf.F.Sg
'Amra screamed.'

b. aamraa=ne ciikʰ-aa
Amra.F=Erg scream-Perf.M.Sg
'Amra screamed on purpose.'

Lahori Urdu also yields the minimal pair in (6), which provides further evidence for a volitional sense of the ergative. The particularly striking point about this construction is that it cannot be construed as perfective in any way or form. The ergative in (6b) is therefore clearly not strictly correlated with the appearance of perfective morphology in

[6]There is exactly one exception to the paradigm that all subjects of transitive verbs are marked as ergative in the perfective. The exception is the transitive *laa* 'bring'. It must always take a nominative subject, no matter what the aspectual morphology is. Kachru (1980) suggests that this verb may actually have been derived from the complex predicate *le aa* 'take come', which is in use today as well. As will be shown in Chapter 4, it is always the second verb, the light verb, which determines the case marking of the subject. In *le aa* 'take come', the *aa* is the light verb, and it always requires a nominative subject. Interestingly, though, the verb *laa* 'bring' can now often be heard with an ergative subject in the perfective. It would thus seem that the one exception is now being made to conform to the paradigm after all. Under an account which posits the ergative as a marker of volitionality, such a reanalysis is expected.

Urdu.[7] In (6a) the infinitive construction with a dative subject (for arguments that the dative NP is indeed a subject, see T. Mohanan 1990) expresses an obligation on the part of the subject. This construction is in fact very much like some modal constructions in Urdu. The sentence in (6b) differs from (6a) only with respect to the case on the subject, but (6b) expresses a desire on the part of the subject *Anjum*. Again, the ergative here seems to be denoting volitionality in some sense.

(6) a. anjum=ko xat lik^{h}-naa hai
Anjum.F=Dat letter.M=Nom write-Inf.M be.Pres.3.Sg.
'Anjum has to write a letter.'

b. anjum=ne xat lik^{h}-naa hai
Anjum.F=Erg letter.M=Nom write-Inf.M be.Pres.3.Sg
'Anjum wants to write a letter.'

More evidence along the same lines comes from N-V complex predicates investigated by T. Mohanan (1990). In (7a) the remembering event can be involuntary, but in (7b), the action of remembering can only be agentive. For some further discussion of examples such as (6) and (7), see Butt and King (1991).

(7) a. anjum=ko kahaanii yaad ay-ii
Anjum.F=Dat story.F=Nom memory come-Perf.F.Sg
'Anjum remembered the story. (Memory came to Anjum)'

b. anjum=ne kahaanii yaad k-ii
Anjum.F=Erg story.F=Nom memory do-Perf.F.Sg
'Anjum remembered the story.'

Urdu thus clearly indicates that the ergative is not used according to the traditional notion of ergativity, but is rather used to mark volitional agents. It may be objected that there are instances of transitive perfective sentences where the action may be non-volitional, but where the subject is nevertheless marked with an ergative. In fact, transitive, perfective sentences seldom denote a non-volitional act. Non-volitionality is more generally expressed

[7]This construction is not generally found in dialects of Urdu/Hindi, but is widespread in Lahori Urdu. It is possible to speculate that the construction in (6b) is in some way a result of Punjabi influence on Urdu, as the Punjabi dative is *nu*. The Punjabi version of (6a) thus may have allowed the development of (6b) as a viable construction. However, Y. Kachru (p.c.) notes that the contrast in (6) is not found in a dialect of Hindi spoken in Delhi, where a strong Punjabi influence on Hindi has been documented (Gumperz 1971). Thus, the influence of Punjabi cannot necessarily be made to account for the extended use of the ergative in Urdu. Rather, I would postulate that in Lahori Urdu, the ergative has gained in semantic significance, and become firmly established as a marker of volitionality.

by dative subject constructions, or complex predicates, where the light verb requires a nominative subject. Furthermore, as suggested by Mahajan (1990), the ergative indeed has two functions: it is both a purely grammatical marker of a transitive paradigm (Structural Case), and a semantic marker of volitionality (Inherent Case). The complex interaction of these two functions will be explored in some more detail in the course of analyzing Urdu Aspectual complex predicates.

2.3.3 Dative and Accusative *ko*

The last set of examples in the previous section also illustrated the use of the dative *ko*. T. Mohanan (1990, 1993a) argues that the dative in Hindi appears only on goals, whether they be spatial, as in (8a), or abstract as in (7a). However, because the dative *ko* in (8a) is homophonous with the accusative *ko* in (8b), the two cases have often been treated as one and the same. For example, Mahajan's (1990) recent analysis of Hindi scrambling phenomena presupposes that every instance of *ko* must be treated as inherent dative case. On the other hand, within T. Mohanan's (1990) approach to arguments in Hindi, a dative *ko* is carefully distinguished from an accusative *ko*.

(8) a. anjum=ne **saddaf=ko** kitaab d-ii
Anjum.F=Erg Saddaf.F=Dat book.F=Nom give-Perf.F.Sg
'Anjum gave Saddaf a/the book.'

b. adnaan=ne **roṭii=ko** paka-yaa
Adnan.M=Erg bread.F=Acc cook-Perf.M.Sg
'Adnan made a particular/the bread.'

The disagreement with regard to the analysis of datives and accusatives evident in the literature is not a recent phenomenon. Allen (1951) very clearly argued against some prevalent ideas of the time that every instance of the Hindi *ko* should simply be analyzed as a dative. And in fact, it can be very clearly shown that although the dative *ko* and the accusative *ko* are homophonous, they fulfill two distinct functions and appear in complementary distribution.[8] For one, as (9) shows, the dative *ko* is never optional, while the accusative *ko* is. Compare the sentences in (9) with the sentences in (8). Example (9a), a

[8]Kiparsky (p.c.) notes that the distributional facts are also consistent with an analysis which treats *ko* as a single case marker which fulfills two distinct functions, depending on where it is realized. This approach differs in perspective from the position I argue for in this section, but in principle there are no concrete identifiable differences. The position I wish to argue against here, is that the *ko* represents a single case marker which can be identified with a single, unifying function in all its realizations.

version of (8a) without a *ko* on the indirect object is bad, while (9b) is good even without a *ko* on the direct object *roṭii* 'bread'.

(9) a. *anjum=ne **saddaf** kitaab d-ii
Anjum.F=Erg Saddaf.F book.F=Nom give-Perf.F.Sg
'Anjum gave Saddaf a/the book.'

b. adnaan=ne **roṭii** paka-yii
Adnan.M=Erg bread.F cook-Perf.F.Sg
'Adnan made a/the bread.'

A clear distinction can thus be made between the *ko* of indirect objects and the *ko* of direct objects. The optionality of *ko* on direct objects has been remarked on frequently and it has been noticed by several people that there is a correlation between animacy, definiteness, and the appearance of *ko* (e.g., Allen 1951, Masica 1976, Comrie 1981, Mahajan 1990, T. Mohanan 1990, Singh 1993). Allen (1951:70) furthermore remarks on the contrast between the 'definite' in (8b) and the 'indefinite' in (9b): "... that these terms cover a variety of subtleties in usage, and the translations by means of the English indefinite and definite articles must be considered convenient rather than accurate." Work by Porterfield and Srivastav (1988) and Dayal (1992) in their elucidation of the complicated semantics of bare NPs such as *roṭii* 'bread' in (9b) indeed show that they cannot simply be interpreted as 'indefinite'. Furthermore, I argue in Butt (1993b) that the Urdu accusative *ko* in (9b), analogous to the Turkish accusative (see Enç 1991), must be a marker of *specificity*, rather than definiteness. This can be demonstrated by the set of sentences in (10). The sentence in (10a) sets up a context which requires that the direct object *murɣii* 'chicken' be interpreted as nonspecific. The sentence in (10b), where the *murɣii* 'chicken' is not marked with *ko*, is perfectly good within the context of (10a): 'chicken' here can be interpreted as nonspecific. However, in (10c), *murɣii-ko* 'chicken-Acc' can only receive a specific interpretation, and the sentence is therefore strange within the context of (10a). The only viable interpretation of (10c) is one in which it must be assumed that the cook already had a particular chicken in mind, which he had wanted to cook, so he went out and bought it.

(10) a. adnaan aaj raat=kii salen ke-liye murɣii
Adnan.M=Nom today night.F=Gen.F curry for chicken.F=Nom
cah-taa
want-Impf.M.Sg
t^{h}aa
be.Past.3rd.Sg
'Adnan wanted chicken for tonight's curry.'

b. us=ke xansaame=ne bazaar=se **murɣii**
pron=Gen.Obl cook.M.Obl=Erg market.M=from chicken.F=Nom
xariid-ii
buy-Perf.F.Sg
'His cook bought a chicken from the market.'

c. #us=ke xansaame=ne bazaar=se **murɣii=ko**
pron=Gen.Obl cook.M=Erg market.M=from chicken.F=Acc
xariid-aa
buy-Perf.M.Sg
'His cook bought a particular chicken from the market.'

In conclusion, then, although the dative *ko* and the accusative *ko* are homophonous in Urdu/Hindi, they appear on different grammatical relations and are governed by differing semantic conditions. The accusative *ko* is a marker of specificity on direct objects, while the dative *ko* indicates the semantic notion of goal and can appear on both subjects and indirect objects. I thus distinguish the dative *ko* from the accusative *ko* in the remainder of this dissertation.[9]

2.3.4 Direct/Structural and Inherent Case

One of the last issues to be discussed is the treatment of unmarked direct objects, such as *roṭii* 'bread' in (9b) above as 'nominative'. T. Mohanan (1993a) presents detailed argumentation in favor of analyzing unmarked direct object NPs as nominative, rather than accusative, and I will not recapitulate her argumentation here. Rather, I examine some of the theoretical issues which motivate the differential labeling of unmarked direct object NPs and establish the approach to case/Case I rely on in this dissertation.

Mahajan (1990) proposes that argument NPs in Hindi may have both structural and inherent Case. Structural Case is assigned in SPEC AGRP, SPEC IP, or in the complement to V position. The inherent Case of an argument is specified in the lexical entry of a

[9]As Smith (1992) points out for Icelandic and Germanic in particular, there is a historical connection between dative and accusative case in that one appears to change into the other. Furthermore, it appears to be very common across languages that the dative and accusative cases are homophonous (Kiparsky, p.c.). While these intriguing facts may point towards the need of a unifying analysis of datives and accusatives, the issue is beyond the scope of this dissertation. For my purposes, it suffices that in Urdu the dative and accusative can be clearly differentiated synchronically.

particular verb form. A close examination of the system reveals that the NPs which have only structural Case are exactly the unmarked NPs. These unmarked NPs are the ones I propose to label as nominative, and in fact nominative arguments in Urdu can only ever be either subjects or direct objects.

The argument NPs with overt case morphology within Mahajan's system are all treated as having inherent Case, although they may also be assigned structural Case if they are in the appropriate structural position. This dual system of Case assignment always applies to, for example, direct objects marked with *ko*, or ergative subjects, exactly because these NPs both function as subjects or direct objects, and are overtly marked with case clitics.

Zaenen, Maling and Thráinsson (1985), in their analysis of Icelandic case, differentiate between grammatical, idiosyncratic, and semantic case. Grammatical or structural case is determined on the basis of syntactic information only (e.g., all subjects receive nominative case), while idiosyncratic or quirky case must be stipulated by each verb in its lexical entry. The notion of semantic case is discussed only briefly, and is only intended to apply to obliques and adjuncts. Grammatical case in this framework is thus roughly equivalent to structural Case, while idiosyncratic or quirky case is equivalent to inherent Case.

The notion of *semantic case* has no equivalent in Mahajan's (1990) treatment of Hindi. Although inherent Case is considered to be semantic in some form, the idea behind *semantic case* is that the case marking of a particular NP can be predicted from semantic regularities in the language, and need not be stipulated separately in the lexical entry of each verb.[10] T. Mohanan (1990) shows that the case marking in Hindi is in fact predictable on the basis of semantic information (also see K.P. Mohanan 1989b for a more general discussion). Dative case, for example, always only appears on goals, whether they be abstract or concrete. However, semantic information must interact with syntactic information in the determination of case marking on subjects and objects (see Butt and King 1991). The ergative, for example, is a marker of volitionality, but it may only appear on subjects. Similarly, the accusative *ko* is a marker of specificity, but can only appear on direct objects.

This interaction of semantic and syntactic information is what would seem to be at the heart of Mahajan's proposal that argument NPs may in fact have both inherent and structural Case. As any overtly case-marked NP is deemed to have inherent Case within Mahajan's proposal, it follows that what I label as 'nominative' corresponds exactly to those

[10]Note that the notion of semantic, as opposed to quirky, case is not necessarily incompatible with Mahajan's approach – he simply has not made use of it.

NPs which are *only* assigned structural Case. The use of nominative in T. Mohanan (1990) and here could thus also be termed *direct* Case, as in Ramchand (1993), for example. The nominative in effect, is the case which conveys no semantic information, and is assigned purely on the basis of syntactic information. Despite the difference in the labeling of such 'direct' Case marked NPs, the proposals in Mahajan (1990) and the approach presented here are quite similar in this respect. However, rather than relegating the semantic content of case clitics to the status of unpredictable quirky case that must be specified in the lexicon, I propose to make use of the semantic regularities governing case in Urdu, and formulate a linking mechanism within LFG in Chapter 6, which relates a(rgument)-structure information to a c(onstituent)-structure representation on the basis of semantic information provided by the case clitics.

2.4 Configurationality

As with case marking in Urdu/Hindi, the literature on Hindi/Urdu is divided with regard to the issue of configurationality.[11] Following, T. Mohanan (1990), I assume that Urdu is nonconfigurational and has the basic sentence structure shown in (11), i.e., NPs and a single $\overline{\text{V}}$ may appear in any order in a given clause. However, most of the literature focusing on theoretical syntax assumes that Hindi/Urdu is configurational (for some recent work, see Davison 1991a, Dwivedi 1993, Mahajan 1990, Srivastav 1991c). Hindi/Urdu is a language with relatively free word order, constituency tests do not clearly delimit a VP, and structural subject/object asymmetries are diffult to identify. The treatment of Hindi/Urdu as configurational thus seems to be motivated almost entirely by the underlying requirement in Government-Binding theory that all languages must be configurational. To date, there has not been a rigorous examination of the assumption of configurationality for Hindi/Urdu, although there has been some work examining the free word order/scrambling phenomena (Gambhir 1981, Mahajan 1990).

(11) $\text{S} \rightarrow \text{NP}^*, \quad \overline{\text{V}}$

[11]Note that I use the labels Urdu and Hindi in varying order. Given the existing political tensions, this is meant to prevent one of the languages as being taken more primary than the other. Most of the literature I refer to has Hindi as the main focus of investigation. However, given the close relationship between Urdu and Hindi, the results obtained for Hindi generally apply to Urdu and vice versa. I therefore generally refer to Hindi first when I base my assertions on literature describing mostly Hindi, and use Urdu first when I refer to phenomena where actual Urdu data are available.

Mahajan (1990) in fact argues that the relatively free word order of Hindi/Urdu can be accounted for naturally within a configurational approach. He identifies three types of movement: Argument Shift, Adjunction to XP, and X^0 Shift.[12] These three types of movement essentially have the effect of the phrase structure rule in (11) above. The rule in (11) would appear to be preferable in the sense that maximum simplicity is a desirable achievement within linguistic theory. However, a configurational D-Structure, to which the three types of movement apply, is not intended to a provide only an account of constituency or movement. Rather, such phenomena as anaphora, verb agreement, Case assignment, and wh-movement are analyzed crosslinguistically in terms of D-Structure principles and constraints. Mahajan (1990), for example, accounts for verb agreement, Wh-movement, and anaphora in Hindi. The question then is whether a nonconfigurational account or a configurational account is able to provide a better analysis of these phenomena.

K.P. Mohanan (1982) argues that Malayalam is a nonconfigurational language and uses several tests to support his proposal. Data from phenomena such as scrambling, clefting, pronominal non-coreference and quantifier scope are easily accounted for by positing a flat structure in Malayalam. Furthermore, K.P. Mohanan argues that there are no subject/object asymmetries in Malayalam. I could duplicate K.P. Mohanan's argumentation for Urdu and conclude that Urdu must of necessity be nonconfigurational. However, as Speas (1990) works through K.P. Mohanan's arguments, and claims that a configurational analysis of Malayalam is justified after all, the duplication of K.P. Mohanan's arguments serves no immediate purpose here.

The argumentation in Speas (1990) instead demonstrates that the issue of configurationality must enter a different level of discussion. The relatively free word order, and the lack of a clear VP constituent would seem to argue against a configurational approach. However, Mahajan (1990) has shown how free word order can be achieved through various types of movement, which also serve to obscure the underlying presence of a VP node. T. Mohanan (1990) presents an account of Hindi anaphora which can be very simply stated in terms of subject orientation and linear precedence. However, accounts of Hindi anaphora have also been formulated within configurational approaches (Gurtu 1985, Mahajan 1990, Dayal 1993). Similarly, the fact that overt Wh-movement is not obligatory in Hindi/Urdu can be accounted for quite simply under a nonconfigurational approach, but configurational

[12]Gambhir (1981) presents a detailed discussion on discourse factors affecting word order alternations, but these are not of immediate concern here.

analyses have also been proposed (Mahajan 1990, Srivastav 1991c).[13]

The essential question with regard to configurationality thus no longer seems to be whether or not evidence for a VP in terms of constituency can be marshaled. Clearly, even when languages like Warlpiri or Malayalam show very scanty evidence, if any, for a VP node, an argument for configurationality can be made on the basis of other factors. One major factor is evidence for subject/object asymmetries (Speas 1990). If configurationality is defined in terms of a distinction between subjects and objects with regard to some phenomena, then all languages must indeed be configurational within a Government-Binding approach, since grammatical relations such as subject and object are always defined configurationally. Morphological features such as case and agreement are instantiations of the underlying structure. Some languages may have morphological surface instantiations, some may not. Within a framework such as LFG, however, grammatical relations are encoded at the level of f(unctional)-structure, a level of representation which is linked to phrase structure (c-structure), but is not defined by it. Under such a view of grammar, a configurational distinction between subjects and objects is not needed, and therefore the motivation for analyzing the language as configurational is lessened.

The crux of the configurationality issue addressed in Speas (1990), it seems to me, is how grammatical relations are defined/represented. A nonconfigurational structure would pose many difficulties for a theory like Government-Binding, because subjects and objects are defined configurationally. This is true even within the VP-internal approach adopted by Speas and Mahajan. It therefore makes no sense to argue for nonconfigurationality within Goverment-Binding. In fact, it becomes clear in Speas and Mahajan, that issues of constituency or word order are increasingly immaterial to the D-Structure. These issues are sorted out at S-Structure or PF. However, since arguments of constituency or word order are also put forth to argue for one type of D-Structure over another, the issue at the moment is quite confused. The following quote from Guilfoyle, Hung, and Travis (1992:394) in their discussion of phenomena sensitive to either agent NPs or topic NPs serves as a telling illustration.

> Tagalog has fairly free word order, and has been claimed to be a nonconfigurational language (Carrier-Duncan 1985). Nevertheless, it seems that it must

[13]Data from ellipsis are often cited to support the existence of a VP, and relevant data could be constructed for Urdu as well. However, see Dalrymple (1991) for a discussion of why ellipsis is not a reliable indicator of syntactic structure, but should instead be analyzed as a semantic phenomenon.

> be configurational at some level if we are to explain the constraints on Wh-questions, relative clauses, and cleft structures described in Section 2. We will assume that at D-structure the order of constituents is Agent-V-Theme-X, ...

In fact, Guilfoyle, Hung, and Travis (1992) propose two subject positions, SPEC of IP and SPEC of VP, to account for the phenomena at hand. Significant is that that the SPEC of IP position is reserved for purely structural phenomena like extraction and quantifier float, while the SPEC of VP is associated with *theta-sensitive* properties such as binding and control. In a theory like LFG, which encodes grammatical relations at a level entirely separate from phrase structure, the phenomena explained in terms of SPEC of VP would be treated at a-strucure, while the phenomena sensitive to the SPEC of IP position would be analyzed in terms of c-structure (see Kroeger (1993) for a discussion of Tagalog within an LFG approach). LFG thus clearly separates out two different sets of information and avoids the kind of confusion found in Mahajan (1990), for example, between issues of word order and grammatical relations.

Since I present my analysis within LFG, I do not depend on the assumption of a configurational D-structure in order to be able to deal with the relevant phenomena. Furthermore, phenomena such as word order, anaphora, and verb agreement can be given a more pleasing account from a nonconfigurational point of view. For the purposes of this dissertation, I therefore assume a flat matrix clause structure for Urdu. However, note that languages argued to have a flat matrix clause structure, are also sometimes analyzed as containing embedded VPs. An example is Icelandic, where embedded clauses are analyzed as containing VPs (Kaplan and Zaenen 1989). Furthermore, Sells (1990) has shown that although in Japanese the evidence for or against a configurational treatment is conflicting, nonfinite *-te* complements provide decisive evidence for the existence of a VP. As the analogs of Japanese nonfinite *-te* type clauses have the distribution of NPs in Urdu (see Chapter 3), the types of auxiliary constructions in Icelandic presented in Kaplan and Zaenen (1989) have no analog, and no other types of embedded clauses are examined in any detail in this dissertation, I have no firm conclusion to offer as to the existence of an embedded VP in Urdu.

2.5 Theoretical Framework

2.5.1 General

As already mentioned, the theoretical framework within which I frame my examination of the structure of complex predicates is Lexical-Functional Grammar (LFG). LFG was first developed in the late 1970s by Joan Bresnan and Ronald Kaplan. The most comprehensive description of early LFG is found in Bresnan (1982b), where theoretical motivations and detailed treatments of a number of languages, ranging from Russian and Icelandic to Malayalam, are given. LFG proposed that syntactic phenomena across languages must be factored apart across separate but interacting levels of representation. In particular, grammatical relations were given an independent status at f(unctional)-structure, which was held to be constrained, but not exclusively determined, by the level of c(onstituent)-structure. Furthermore, much of the work ascribed to transformations at that time was postulated to take place in the lexicon. The realization of passivization and the dative alternation, for example, was performed in the lexicon (Bresnan 1978).

Essentially, LFG consists of three major modules: the lexicon, the syntax, and the semantics. Word formation takes place in the lexicon, and lexical items are listed with the phonological, semantic, morphological, and syntactic information particular to them. These lexical items enter into the syntax and form sentences or phrases according to the constraints and principles which hold at the levels of f-structure and c-structure. The semantics of a phrase, such as quantifier scope for example, are represented at s(emantic)-structure, which is taken to be projected from the syntax. While the actual organization of the lexicon is not well understood, a good amount of work has been done exploring the properties of the syntactic representations.

2.5.2 C-structure and F-structure

The level of c-structure relies on the principles associated with a version of X′-theory outlined in Bresnan (1982a).[14] However, c-structure is taken to be variable across languages. A configurational language like English thus might be represented by the phrase structure rule in (12), while a nonconfigurational language like Warlpiri (Simpson 1983) would be represented by (13).

[14]For more recent theoretical developments in this area see Kroeger (1993).

(12) S $\longrightarrow$ NP VP

(13) S $\longrightarrow$ X Aux X^*

Phrase structure rules encode dominance as well as precedence relations, and describe syntactic entities with the usual types of symbols. Major category types are as shown in Table (14), taken from Bresnan (1982a:295).

(14)

type	0	1	2
category	V	V′	V″ (VP)
	P	P′	P″ (PP)
	N	N′	N″ (NP)
	A	A’	A″ (AP)
		S	S′

The type 0 categories are *lexical categories*, the type 1 and 2 categories are *projections* of the lexical categories. Type 2 categories are, of course, furthermore *maximal projections*. There is also a fundamental distinction between S and S′ and the other categories. The former are *exocentric* because they are not projections of a lexical category, while the latter are *endocentric* (see Kroeger 1993 for a more detailed exposition). In addition to these major categories, the phrase structure allows for *minor categories* such as *Det* and *Comp* as well. These categories are described as null or degenerate in the sense that type 1 or 2 categories do not project from them. Bresnan (1982a) further asserts that some languages may simply lack some of the categories. So, for example, the proposal that Hindi/Urdu may simply lack the category VP is easily expressed within LFG, and is in fact taken to be a natural occurrence within the greater scheme of phrase structure.[15]

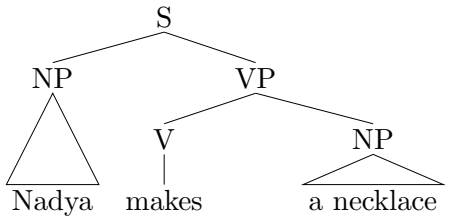
In contrast, the level of f-structure is taken to be *deep* in the sense that it reflects universal, crosslinguistically invariant, underlying principles of syntax. Thus, a simple transitive sentence *Nadya makes a necklace* in both Urdu and English will have the same fundamental f-structure shown in (15), but the corresponding c-structures for English and Urdu respectively might be as in (16) and (17).

[15]This view of phrase structure contrasts sharply with the fundamental concepts underlying X′ Theory in Government-Binding (e.g., Grimshaw 1991), and in fact allows great freedom of expression at c-structure. Recently, the relatively unconstrained nature of c-structure has been recognized as undesirable, and efforts are being made towards the development of a more constrained system (Bresnan, p.c.). For already existing relevant work see Kroeger (1993) and King (1993).

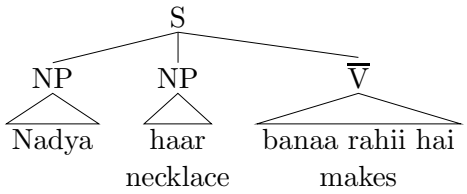
(15)

$$\begin{bmatrix} \text{PRED} & \text{`make} < \underline{\quad}, \underline{\quad} > \\ \text{TENSE} & \text{PRES} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`Nadya'} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`necklace'} \\ \text{DEF} & - \end{bmatrix} \end{bmatrix}$$

(16)



(17)

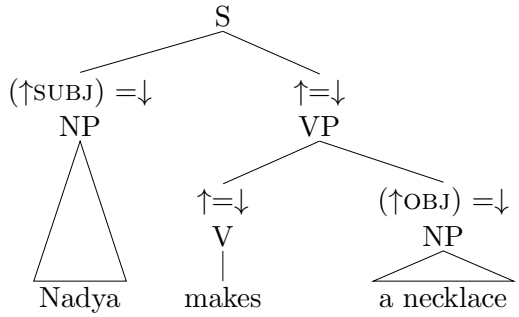


The theory is constructed so that it is possible for a given f-structure to have differing c-structure realizations even within the same language. This is an especially useful feature for languages with relatively free word order like Malayalam (K.P. Mohanan 1982), or Warlpiri (Simpson 1983), as the various permutations of a given sentence may have various differing (but base-generated) c-structures, which all correspond to the same f-structure. The f-structure can thus be said to represent the 'internal' (or deep) structure of a sentence, while the c-structure represents the 'external' (or surface) structure.

An f-structure like (15) above is an attribute value matrix, which essentially describes the application of a function to an attribute like SUBJ, to arrive at the value of *Nadya*. This functional nature is made use of in the mapping from c-structure to f-structure and is also responsible for the denotation of the notions *subject*, *object*, etc., as *grammatical functions*, as opposed to the more traditional *grammatical relations*. Each node in a given c-structure is taken to correspond to a particular piece of the f-structure. The function ϕ relates c-structure representations to f-structures. In the English tree in (16), for example, the subject NP corresponds to SUBJ at f-structure, the object NP to OBJ, and the V, VP,

and S nodes all correspond to the entire f-structure in (15). The mapping procedure is thus not one-to-one, but rather many-to-one (for more detailed and mathematically more precise descriptions of the system see, for example, Bresnan (1982b) and Kaplan (1987)). The underlying idea is that information from heads 'flows' upwards. In effect, information from differing nodes is collapsed, or unified, into a single f-structure piece in the process of mapping from c-structure to f-structure (see Andrews and Manning (1993) on advantages and disadvantages of such collapsing of information). The mapping procedure is instantiated by *annotations* on c-structure. The annotated version of (16) is shown in (18).

(18)



Heads are generally annotated with ↑=↓, while the notation (↑SUBJ) =↓ signifies that the subject of the mother node, the S, is the material contained in the annotated node. Note that lexical items like *Nadya* or *makes* enter the syntax inflected and with all of the information specified in their lexical entries. The lexical entry for *makes*, for example, is shown in (19). Since word formation takes place in the lexicon, the lexical entries for *make* and the tense inflection *-s* have already been combined in (19).

(19) makes (↑PRED) = 'make<____, ____>'
(↑TENSE) = PRES
(↑PERS) = 3
(↑NUM) = SG

The PRED entry for *make* shows that this verb has two arguments. Alsina (1993) interprets this notation to stand for the semantics of the verb, with the two argument slots merely abstracted out. I will have more to say about the representation of argument structure in Chapters 5 and 6.

The precise algorithm needed to map from a c-structure like (19) to the corresponding f-structure in (16) and the mathematical motivations have been described elsewhere, and I will not repeat them here. In particular, see Kaplan and Bresnan (1982) for a detailed presentation of the algorithm.

A nice feature of this mapping procedure is that it allows for discontinuous constituents in languages like Warlpiri. A node can simply be annotated as containing some piece of the f-structure OBJ, and the information will unify in the process of mapping from c-structure to f-structure. If such discontinuous constituents form an ill-formed sentence, i.e., contain conflicting specifications in their lexical entries, the information will be inconsistent at f-structure, and the sentence is recognized as ill-formed. Besides this checking for consistency, the principles of *Coherence* and *Completeness* check on the well-formedness of a given sentence at f-structure.[16] While the phrase structure rules of a language may either over- or undergenerate, as exemplified by the sentences in (20) and (21) for English, the Principles of Completeness and Coherence identify such generations as ungrammatical (Kaplan and Bresnan 1982:211).

(20) *The girl handed.

(21) *The girl fell the apple the dog.

The f-structure representation which would correspond to the sentence in (20) would not be *complete* because all the arguments governed by the predicate *hand* are not present. The example in (21) is ruled out because the corresponding f-structure representation would not be *coherent*: the governable functions OBJ (*the apple*) and OBJ2 (*the dog*) would be represented at f-structure, but are not actually subcategorized for by the predicate *fall*. A formal statement of the principles is given below (Kaplan and Bresnan 1982:211–212).

Principle of Completeness: An f-structure is *locally complete* if and only if it contains all the governable grammatical functions that its predicate governs. An f-structure is *complete* if and only if it and all its subsidiary f-structures are locally complete.

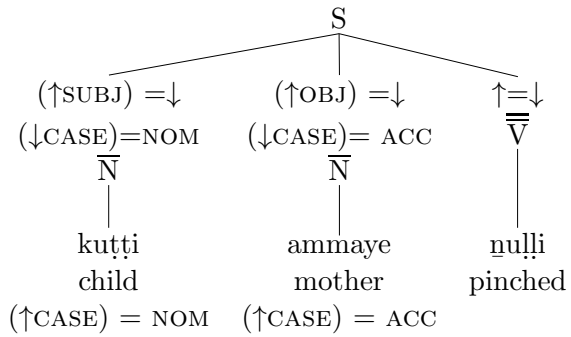
Principle of Coherence: An f-structure is *locally coherent* if and only if all the governable grammatical functions that it contains are governed by a local predicate. An f-structure is *coherent* if and only if it and all its subsidiary f-structures are locally coherent.

Finally, it should be noted that while the annotations on c-structure trees in terms of $\uparrow=\downarrow$, etc., may seem arbitrary and stipulative at first glance, they are actually taken

[16] A given c-structure is, of course, only well-formed if it conforms to the phrase structure rules of a given language.

to be motivated by independent principles (Alsina 1993) internal to the structure of a particular language. In English, for example, the subject must always be the NP which is the sister to the VP. The object is the NP which is the sister of the V. Heads simply receive the ↑=↓ annotation. In free word order languages like Malayalam, on the other hand, the annotations on f-structure must follow principles imposed by the system of case in the language (K.P. Mohanan 1982). Subjects, for example, must always be nominative (unmarked). An example of an annotated c-structure for Malayalam is given in (22).

(22)



The annotations on the $\overline{\text{N}}$s can be assigned at random, but if the CASE specifications on the $\overline{\text{N}}$ do not match the case of the lexical item, then the corresponding f-structure cannot be well-formed. Annotations on c-structure are thus always in accordance with, and derived from, the particular requirements of a given language.[17] However, while K.P. Mohanan's (1982) proposal for Malayalam yields the right results, I argue later that the assignment of annotations at c-structure is better motivated through the medium of argument structure information, which interacts with the case assignment principles of a free word order language like Urdu or Malayalam. This will be discussed in detail in Chapters 5 and 6.

2.5.3 Argument Structure

In early LFG, at the time of K.P. Mohanan's (1982) examination of Malayalam, an independent level of argument structure, which interacted in a significant way with f-structure or c-structure representations had as yet not been worked out. Although all predicates were assumed to have an argument structure, the semantic arguments of a predicate like *make*, the *maker* and *thing made*, stood in a one-to-one relationship with grammatical functions such as SUBJ and OBJ. This was represented as in (23) (Bresnan 1982b).

[17] Warlpiri can also be accounted for within this approach (Simpson and Bresnan 1982).

(23)
	(SUBJ)	(OBJ)	(grammatical functions)
	|	|	(lexical assignment of grammatical functions)
'make(	arg1	arg2)'	(predicate argument structure)

While predicate argument structure was given an independent status (based on work by Grimshaw (1979)), not much was done with it. In subsequent work, a theory which motivated the assignment of particular arguments to grammatical functions was developed. This Lexical Mapping Theory (LMT) has been articulated in L. Levin (1986), Bresnan and Kanerva (1989), Alsina and Mchombo (1989), Bresnan and Moshi (1990), Alsina (1990), and Bresnan and Zaenen (1990).

Linking principles are stated with the help of the argument structure hierarchy shown in (24) and the a(rgument)-structure features [+/−o] and [+/−r].

(24) **Thematic Hierarchy:**
$agent < beneficiary < experiencer/goal < instrument < patient/theme < locative$

The a-structure features [+/−o] and [+/−r] stand for *object* and *restricted* respectively. They constrain the way in which θ-roles at a-structure are mapped onto grammatical functions. Different types of θ-roles are intrinsically specified for some of these features. Essentially, patientlike roles are [−r], secondary patientlike roles are [+o], and all other roles are [−o]. This ensures that agents ([−o]) are nonobjective, and are usually mapped to SUBJ, while themes and patients are mapped to OBJ. In addition to these *Intrinsic Role Classifications*, some *Default Role Classifications* apply. These Default rules assign [−r] to the highest thematic role, where highest is determined according to the thematic hierarchy in (24), and all other roles are assigned [+o]. Furthermore, a feature [+r] cannot be assigned to a θ-role already specified for [−r]: a clash of features is not allowed.

Figure (25) shows the correspondences between feature specification and grammatical functions. A θ-role with the features [−r,−o] corresponds to a SUBJ, a θ-role with [−r,+o] to an OBJ, etc. An OBJ_θ is a restricted object, for example, an indirect object. The θ subscript indicates that obliques and restricted objects tend to be associated with restricted semantics. A fine point to note is the following. While it may appear that grammatical functions like SUBJ, OBJ, etc. exist as primitive notions within the theory, a given grammatical function, a SUBJ for example, is actually nothing more and nothing less than the features [−r,−o]. Grammatical functions thus are not independent of the features, but are instead defined

and therefore also constrained by them. That is, differing kinds of grammatical functions other than the ones displayed in (25) could not simply be stipulated to exist.

(25)

Grammatical Function	**Features**
SUBJ	[−r, −o]
OBJ	[−r, +o]
OBJ_θ	[+r, +o]
OBL_θ	[+r, −o]

In more recent work, Alsina (1993) proposes a differing approach to mapping between a-structure and f-structure as part of his examination of Romance causatives. Romance causatives necessitate the composition of argument structures in the syntax, rather than in the lexicon. Alsina (1993) posits a process of *Predicate Composition*, which allows the argument structures of at least two predicates to combine into a single argument structure in the syntax. However, once the composition of argument structures is allowed to take place in the syntax, the mapping from a-structure to grammatical functions must also be also be allowed to take place in the syntax, rather than being restricted to the lexicon, as had been assumed in the formulation of LMT. Alsina also points out that LMT is too restrictive and cannot provide a good account of syntactically ergative languages like Dyirbal, or languages with 'inverse' argument structures. Furthermore, he judges the features [+/−o] and [+/−r] not to be sufficiently motivated, and instead proposes to make a distinction between [+/−subj] (subjects) and [+/−obl] (terms vs. nonterms). With the help of these distinctions, and an argument structure which draws on Dowty's (1991) notion of proto-roles, Alsina formulates an alternative theory of linking, the Functional Mapping Theory (FMT).

Romance causatives bring up many of the same fundamental problems as the Urdu complex predicates I examine in the next two chapters. My analysis of complex predicates is therefore very close to Alsina's (1993) treatment of Romance causatives. In particular, the notion of *Predicate Composition* is almost identical. However, I argue that the type of Aspectual complex predicates in Urdu provide evidence for a more elaborated a-structure than is provided by either θ-roles, or a version of Dowty's proto-roles, and propose a level of a-structure based on Jackendoff's (1990) theory of Conceptual Semantics. Given the elaborated level of a-structure I propose, the mapping procedures formulated within LMT are actually more immediately compatible with my proposal than Alsina's FMT. I therefore rely on a version of LMT within this dissertation, but require LMT to apply in the syntax,

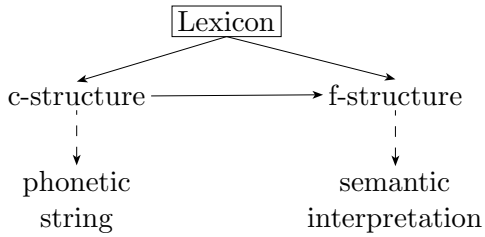
not in the lexicon. Correspondingly, I will simply be referring to the mapping principles I utilize in Chapters 5 and 6 as *Mapping Theory* (MT).

2.5.4 Summary

The above sketch of the properties of LFG is very brief, but it describes the basic notions needed for an examination of Urdu complex predicates. Much of it should also become clearer in the next few chapters. For more comprehensive discussion on both the theoretical implications and details of LFG, I refer the reader to Bresnan (1982b), Sells (1985), Kaplan (1987), Alsina (1993) and the various articles pertaining to LMT mentioned above.

LFG also contains a level of s(emantic)-structure, which has been worked on in Halvorsen (1983), Halvorsen (1987), Halvorsen and Kaplan (1988), Kaplan (1987), Dalrymple, Lamping and Saraswat (1993), etc. However, although I will be investing the a-structure with more detailed semantic information than is possible with θ-role representations, I will not be concerned with s-structure per se and will therefore not provide any details describing this level of representation here. Instead, I provide the following picture of the overall organization of the grammar (based on Alsina (1993) and Sells (1985)).

(26)



This is essentially the organization of grammar I follow in this dissertation. The examination of Urdu complex predicates will make some reformulations necessary. I give the level of a-structure, which is not included in the picture in (26), an explicit status, formulate principles which map a-structure information to c-structure, and reformulate some of the linking principles between c-structure and f-structure.

Essentially, the fundamental problem posed for theories of syntax by Urdu and Romance complex predicates is that they contain two heads, which must combine into a single, discontinuous, head in the syntax. This problem poses a challenge for unification based theories like LFG or HPSG (Head-Driven Phrase Structure Grammar), as well as a structure based theory like GB. However, the organization of grammar within LFG allows a very insightful examination of the structure of complex predicates precisely because it factors out

differing syntactic phenomena into independent levels of representation. This will become immediately evident through the examination of the Urdu permissive in the next chapter.

Chapter 3

The Permissive

3.1 Introduction

An example of the Urdu permissive is shown in (1). Both the infinitive predicate *banaa-ne* 'make', and the finite verb *di-yaa* 'let' contribute to the overall argument structure of the clause. While the argument *Saddaf* is both the 'lettee' and the maker, i.e., is shared by the predicates *de* 'let' and *banaa* 'make' in a semantic sense, the argument *Anjum* is contributed only by *de* 'let', and *haar* 'necklace' is an argument of *banaa* 'make'. An example of the permissive with the intransitive infinitive *jaa* 'go' is shown in (2).

(1) anjum=ne saddaf=ko haar **banaa-ne** **di-yaa**
Anjum.F=Erg Saddaf.F=Dat necklace.M=Nom make-Inf.Obl give-Perf.M.Sg
'Anjum let Saddaf make a necklace.'

(2) anjum=ne saddaf=ko **jaa-ne** **di-yaa**
Anjum.F=Erg Saddaf.F=Dat go-Inf.Obl give-Perf.M.Sg
'Anjum let Saddaf go.'

I demonstrate that the permissive must be analyzed as a complex predicate. With regard to agreement, control, and anaphora the behavior of the permissive exactly parallels the behavior of a simple clause. Interestingly enough, though, the two predicates of the permissive in (2) are separable. The Urdu permissive is thus an example of a complex predicate formed by discontinuous heads.

A detailed look at the permissive reveals that there must be two possible constituent structures. The two possibilities are sketched in (3). In (3a) the two predicates form a

constituent, while in (3b) the permissive seems to consist of a matrix verb which takes an embedded complement, *haar banaa-ne* 'to make a necklace'.

(3) a. anjum=ne saddaf=ko haar [**banaa-ne** **di-yaa**]
Anjum.F=Erg Saddaf.F=Dat necklace.M=Nom make-Inf.Obl give-Perf.M.Sg
'Anjum let Saddaf make a necklace.'

b. anjum=ne saddaf=ko [haar **banaa-ne**] **di-yaa**

Furthermore, a construction I dub the *instructive* shows evidence for exactly the same duality at constituent structure, but it cannot be analyzed as a complex predicate. On the basis of data from these two constructions, I argue that complex predicate formation must take place at argument structure, and not at phrase structure. In this chapter, I present an initial analysis which makes sense of the data by positing different f-structures, but identical c-structures for the two constructions. In the course of presenting the analysis, it will become clear that an independent level of argument structure is actually needed for a comprehensive account of complex predicate formation. Before I proceed on to an argument structure analysis, however, I take a detailed look at the permissive and other infinitive constructions in Urdu in this chapter. This allows a better understanding of the general structure of Urdu infinitive clauses, and paves the way for the complete analysis of complex predicate formation presented in Chapters 5 and 6.

3.2 Evidence for a Complex Predicate Analysis

3.2.1 General

The example of the Urdu permissive given in (1) is repeated here in (4). With regard to verb agreement, anaphora and control, the infinitive *banaa-ne* 'make' and the finite verb *di-yaa* 'let' function as a single unit. In this regard, it contrasts with the superficially very similar instructive in (5).

(4) anjum=ne saddaf=ko haar **banaa-ne** **di-yaa**
Anjum.F=Erg Saddaf.F=Dat necklace.M=Nom make-Inf.Obl give-Perf.M.Sg
'Anjum let Saddaf make a necklace.'

(5) anjum=ne saddaf=ko haar **banaa-ne=ko** **kah-aa**
Anjum.F=Erg Saddaf.F=Dat necklace.M=Nom make-Inf.Obl=Acc say-Perf.M.Sg
'Anjum told Saddaf to make a necklace.'

Notice that the only immediately apparent difference, besides the different finite verbs, between the two constructions in (4) and (5) is the presence of the clitic *ko* on the infinitive *banaa-ne* 'make' in (5). Despite this surface similarity, the agreement, anaphora, and control facts presented in the next sections clearly differentiate between the two constructions.

3.2.2 Agreement

The agreement facts for simple sentences in Urdu are illustrated in (6). The basic pattern (see T. Mohanan 1990 for details) is that the verb agrees with its highest nominative argument.[1] When there is no nominative argument in the clause, the verb is inflected with the default masculine singular *-aa*.

(6) a. **anjum** xat **lik^{h}-tii** hai
Anjum.**F**=Nom letter.M=Nom write-Impf.**F**.Sg be.Pres.3.Sg
'Anjum writes a letter.'

b. anjum=ne **xat** **lik^{h}-aa** hai
Anjum.F=Erg letter.**M**=Nom write-Perf.**M**.Sg be.Pres.3.Sg
'Anjum wrote a letter.'

c. anjum=ne **ciṭṭ^{h}ii** **lik^{h}-ii** hai
Anjum.F=Erg note.**F**=Nom write-Perf.**F**.Sg be.Pres.3.Sg
'Anjum wrote a note.'

d. anjum=ne ciṭṭ^{h}ii=ko **lik^{h}-aa** hai
Anjum.F=Erg note.F=Acc write-Perf.**M**.Sg be.Pres.3.Sg
'Anjum wrote the note.'

In (6a) the verb *lik^{h}tii* 'write' agrees in gender and number with the nominative feminine subject *Anjum*. When the case on the subject is non-nominative, as in (6b) and (6c), the verb does not agree with the subject, but with the nominative object.[2] In (6d) both the subject and the object have overt case so the verb agrees with neither.

Agreement in Urdu is clause-bounded. This is demonstrated by the sentences in (7), which contain an embedded participial adverbial headed by *kar* 'having'. This participial adverbial must always be controlled by the subject of the matrix clause (see Davison 1985).

[1]The "highest" here makes reference to a hierarchy of grammatical functions. For example, subject is higher than object.

[2]While the nouns *xat* 'letter' and *c^{h}iṭṭ^{h}ii* in (6) have a large overlap in meaning, I consistently gloss *xat* as 'letter' and *c^{h}iṭṭ^{h}ii* as 'note' in the interests of maximum clarity.

The pattern of verb agreement for the sentences in (7) shows that the matrix verb only agrees with one of the nominative matrix arguments. If there are no nominative arguments in the matrix clause, the verb carries the default inflection *-aa*. Thus, the matrix predicate *diyaa* 'gave' does not agree with the embedded nominative argument *ṭofii* 'candy' in (7c).

(7) a. **anjum** [naan xariid kar] ciṭṭʰii saddaf=ko
Anjum.F=Nom bread.M=Nom buy having note.F=Nom Saddaf.F=Dat
de-gii
give-Fut.F.Sg
'Anjum, having bought bread, will give Saddaf the note.'

b. anjum=ne [naan xariid kar] **ciṭṭʰii** saddaf=ko
Anjum.F=Erg bread.M=Nom buy having note.F=Nom Saddaf.F=Dat
d-ii
give-Perf.F.Sg
'Anjum, having bought bread, gave Saddaf the note.'

c. anjum=ne [**ṭofii** xariid kar] ciṭṭʰii=ko kamre=mẽ
Anjum.F=Erg toffee.F=Nom buy having note.F=Acc room-in
saddaf=ko
Saddaf.F=Dat
di-yaa
give-Perf.M.Sg
'Anjum, having bought candy, gave Saddaf the note in the room.'

The patterns of agreement in (6) and (7) contrast a simple sentence and a sentence with an embedded participial adverbial, and provide a basis of comparison for the behavior of the Tell Construction and the permissive. Crucially, in (8c) the finite matrix verb *kahaa* 'told', as in (7c), does not agree with the nominative feminine *ciṭṭʰii* 'note'. This indicates that *ciṭṭʰii* 'note' cannot be a matrix argument. Rather, it must be the argument of the embedded predicate *likʰ* 'write'.

(8) a. **anjum** saddaf=ko [xat likʰ-ne]=ko **kah-tii**
Anjum.F=Nom Saddaf.F=Dat letter.M=Nom write-Inf.Obl=Acc say-Impf.F.Sg
hai
be.Pres.3.Sg
'Anjum tells Saddaf to write the letter.'

b. anjum=ne saddaf=ko [xat likʰ-ne]=ko **kah-aa**
Anjum.F=Erg Saddaf.F=Dat letter.M=Nom write-Inf.Obl=Acc say-Perf.M.Sg
'Anjum told Saddaf to write the letter.'

c. anjum=ne saddaf=ko [**ciṭṭʰii** likʰ-ne]=ko **kah-aa**
Anjum.F=Erg Saddaf.F=Dat note.F=Nom write-Inf.Obl=Acc say-Perf.M.Sg
'Anjum told Saddaf to write the note.'

The data for the permissive in (9), on the other hand, exactly parallel the data for the simple sentences in (6). Examples (9b) and (9c) provide the crucial data. In (9b) the object *xat* 'letter' is masculine, and so is the inflection on the verb. It could be argued that the masculine *-aa* on the verb is the default agreement marker, however, this argument is immediately refuted by (9c). Here the nominative object *ciṭṭʰii* 'note' is feminine, and so is the inflection on the verb. Since the verb agrees with the object, the object cannot be embedded, but must be a matrix argument. The permissive thus behaves as if it is a single clause headed by a single predicate, as in (6), and not as if it contains an embedded constituent, as in (7).

(9) a. **anjum** saddaf=ko xat likʰ-ne **de-gii**
Anjum.F=Nom Saddaf.F=Dat letter.M=Nom write-Inf.Obl give-Fut.F.Sg
'Anjum will let Saddaf write a letter.'

b. anjum=ne saddaf=ko **xat** likʰ-ne **di-yaa**
Anjum.F=Erg Saddaf.F=Dat letter.M=Nom write-Inf.Obl give-Perf.M.Sg
'Anjum let Saddaf write a letter.'

c. anjum=ne saddaf=ko **ciṭṭʰii** likʰ-ne **d-ii**
Anjum.F=Erg Saddaf.F=Dat note.F=Nom write-Inf.Obl give-Perf.F.Sg
'Anjum let Saddaf write a note.'

The permissive and the instructive thus differ with respect to agreement. As will be seen, the agreement facts presented here remain constant under scrambling. The permissive and the Tell Construction have exactly the same scrambling possibilities, but no matter which possibility is examined, the agreement facts remain the same. This is an important point as it demonstrates that the status of the permissive as a complex predicate cannot be straightforwardly represented at the level of phrase structure.

3.2.3 Control

Example (10) illustrates that the subject of the participial adverbial headed by *kar* 'having' is always controlled by the matrix subject. It is only *Anjum*, the subject, who opens the door in (10), and never *Saddaf*, the object.

(10) **anjum=ne**$_i$ saddaf=ko$_j$ [____$_{i,*j}$ darvaazaa k^{h}ol kar] andar
Anjum.F=Erg Saddaf.F=Acc door.M=Nom open having inside
bula-yaa
call-Perf.M.Sg
'Anjum, having opened the door, called to Saddaf to come in.'

Example (11) illustrates an instance of the instructive in combination with the participial adverbial *darvaazaa k^{h}ol kar* 'having opened the door'. Example (11) differs from the simple sentence in (10) in that both the matrix subject *Anjum* and the indirect object *Saddaf* are possible controllers of the participial adverbial.

(11) **anjum=ne**$_i$ **saddaf=ko**$_j$ [____$_{i,j}$ darvaazaa k^{h}ol kar] samaan=ko
Anjum.F=Erg Saddaf.F=Dat door.M=Nom open having luggage.M=Acc
kamre=mẽ rak^{h}-ne=ko
room.M=in put-Inf.Obl=Acc
kah-aa
say-Perf.M.Sg
'Anjum told Saddaf to put the luggage in the room, after having opened the door.'

Recall from the previous section on agreement that the Tell Construction appears to contain an embedded infinitive. The data in (11), then, is not surprising. Since *Anjum* is the subject of the matrix verb, and *Saddaf* controls the subject of the embedded infinitive *rak^{h}ne* 'to put', there are two possible controllers for the participial adverbial.

The permissive again differs from the instructive. Example (12) is exactly parallel to the simple case in (10). As in (10), the object *Saddaf* cannot be a possible controller of the participial adverbial in (12). This indicates that *Saddaf* in the permissive is not acting as a controller of the embedded infinitive's (*rak^{h}ne* 'to put') subject. In effect, there is no embedded subject contributed by the infinitive predicate.[3]

(12) **anjum=ne**$_i$ saddaf=ko$_j$ [____$_{i,*j}$ darvaazaa k^{h}ol kar] samaan=ko
Anjum.F=Erg Saddaf.F=Dat door.M=Nom open having luggage.M=Acc
kamre=mẽ rak^{h}-ne
room.M=in put-Inf.Obl

[3] An anonymous reviewer of Butt (1993a), which presents this data, pointed out that I could have used less complicated examples. However, as the reviewer points out, the control facts in Urdu/Hindi are also influenced by linear precedence and pragmatic factors. In particular, the NP immediately preceding the participial adverbial tends to be preferred as the controller. That is, in (11), the dative NP *Saddaf* is the preferred controller, but if the participial adverbial had been placed between the subject *Anjum* and the dative NP *Saddaf*, the interpretation of *Saddaf* as a possible controller is strongly dispreferred. The examples here were chosen to control for this.

di-yaa
give-Perf.M.Sg
'Anjum, having opened the door, let Saddaf put the luggage in the room.'

The permissive again exactly parallels the behavior of a simple predicate, while the instructive behaves as if it contained a matrix verb and an embedded infinitive. Although linear precedence factors cause differing weightings as to which possible controller is preferred, the control facts essentially remain constant under scrambling as well: the dative NP *Saddaf* in the permissive never becomes a possible controller, while it always is in the instructive.

3.2.4 Anaphora

A final piece of evidence for the view that only the permissive is a complex predicate comes from anaphora. The Urdu reflexive *apn-aa* 'self' can only take a subject as its antecedent. The antecedent of the pronominal *us=kaa*, on the other hand, cannot be a subject.[4] T. Mohanan (1990) states this restriction at the level of f-structure: the antecedent for a pronominal cannot be an f-structure subject within the same minimal domain of predication; the antecedent of the reflexive must be an f-structure subject within the same minimal domain of predication (also see Dalrymple (1990) for a treatment of anaphora within LFG). The examples in (13) illustrate this for a simple sentence.[5] In (13a) the reflexive *apn-ii* can only be coreferent with the subject *Anjum*. The pronoun *us=kii* in (13b), on the other hand, can have anything but the subject as an antecedent.

(13) a. **anjum=ne**$_i$ adnaan=ko$_j$ **apn-ii**$_{i,*j}$ gaaṛii=mẽ dek^{h}-aa
Anjum.F=Erg Adnan.M=Acc self-F car.F=in see-Perf.M.Sg
'Anjum saw Adnan in her (Anjum's) car.'

b. anjum=ne$_i$ **adnaan=ko**$_j$ **us=kii**$_{*i,j,k}$ gaaṛii=mẽ dek^{h}-aa
Anjum.F=Erg Adnan.M=Acc pron=Gen.F car.F=in see-Perf.M.Sg
'Anjum saw Adnan in his (Adnan's or somebody else's) car.'

Contrasting the behavior of the permissive with that of the Tell Construction once again clearly demonstrates that the permissive is a complex predicate while the instructive is not.

[4]This simple generalization is, of course, not all there is to Urdu anaphora, but is sufficient for my purposes. For more detailed discussions of Hindi/Urdu anaphora see Gurtu (1985), Subbarao (1984), Davison (1990), Harbert and Srivastav (1988), Dayal (1993), etc.

[5]Both the reflexive *apn-aa* and the pronominal *us-kaa* agree in number and gender with the noun modified.

In (14a) the reflexive *apn-ii* cannot be coreferent with the matrix subject *Anjum*.[6] Since *Adnan* controls a subject within a the minimal domain of predication that contains the reflexive, i.e., controls the subject of the embedded infinitive *calaane* 'to drive', it is the antecedent for *apn-ii* 'self'.

(14) a. anjum=ne$_i$ **adnaan=ko**$_j$ [**apn-ii**$_{?*i,j}$ gaaṛii calaa-ne=ko]
Anjum.F=Erg Adnan.M=Dat self-F car.F=Nom drive-Inf.Obl=Acc
kah-aa
say-Perf.M.Sg
'Anjum told Adnan to drive self's (Adnan's) car.'

b. **anjum=ne**$_i$ **adnaan=ko**$_j$ [**us=kii**$_{i,j,k}$ gaaṛii calaa-ne=ko]
Anjum.F=Erg Adnan.M=Dat pron=Gen.F car.F=Nom drive-Inf.Obl=Acc
kah-aa
say-Perf.M.Sg
'Anjum told Adnan to drive his/her car.'

Unlike as in (13b), the pronominal *us=kii* in (14b) can refer either to the matrix subject *Anjum*, or to the matrix indirect object *Adnan*. The pattern in (14) contrasts with the pattern for simple sentences in (13); it therefore follows that the instructive cannot be analyzed as a simple clause. The permissive in (15), on the other hand, once again behaves like a simple clause.

(15) a. **anjum=ne**$_i$ adnaan=ko$_j$ **apn-ii**$_i, *j$ gaaṛii calaa-ne
Anjum.F=Erg Adnan.M=Dat self-F car.F=Nom drive-Inf.Obl
d-ii
give-Perf.F.Sg
'Anjum let Adnan drive self's (Anjum's) car.'

b. anjum=ne$_i$ **adnaan=ko**$_j$ **us=kii**$_{*i,j,k}$ gaaṛii calaa-ne
Anjum.F=Erg Adnan.M=Dat pron=Gen.F car.F=Nom drive-Inf.Obl
d-ii
give-Perf.F.Sg
'Anjum let Adnan drive his car.'

Just as in (13a), the *apn-ii* 'self' in (15a) can only be coreferent with the subject *Anjum*. And in (15b), just as in (13b), the pronominal *us=kii* cannot be coreferent with the subject

[6]Some speakers allow the reflexive to refer to the matrix subject under certain conditions. Harbert and Srivastav (1988) suggest that the data can be accounted for systematically by a distinction between argument and adjunct infinitive complements.

Anjum. Possible antecedents for *us=kii* are either the non-subject argument *Adnan*, or another person specified in a previous utterance. Thus, the permissive again behaves as if it is monoclausal, while the instructive does not.

3.2.5 Conclusion

This section has shown that data from verb agreement, control and anaphora point to an analysis of the permissive as a complex predicate. The permissive is structurally a simple clause, while the instructive is complex. That is, the two predicates of the permissive function as a single predicate with a single argument structure. The instructive, on the other hand, has a matrix verb which takes an embedded complement. The next section demonstrates that this difference cannot easily be expressed phrase structurally: the phrase structure properties of the permissive and the instructive are the same, despite the fact that one is a complex predicate while the other is not.

3.3 Phrase Structure Ambiguities

This section examines data from scrambling, negation and coordination and shows that with respect to these processes, the permissive and the instructive behave exactly alike. Furthermore, not only do the permissive and the instructive pattern the same, the data actually seem to indicate contradictory phrase structures. However, if one is willing to separate out grammatical function information from constituent information, as is done in LFG, the seemingly conflicting data in this section can be accounted for by assuming that both the permissive and the instructive have two possible, differing, c-structure realizations.

3.3.1 Scrambling

Recall that word order in Urdu is relatively free. T. Mohanan (1990) assumes that Hindi (Urdu) has a flat structure and explains the scrambling phenomena she encounters as follows: only direct daughters of S can scramble freely. I follow T. Mohanan (1990) in assuming that Urdu has a flat structure. The possible permutations of a simple Urdu sentence are shown in (16). The three elements in the sentence (*Anjum*, *Saddaf* and *dek*^{h}*aa* 'see') can appear in any order.

(16) a. [$_{NP}$ anjum=ne] [$_{NP}$ saddaf=ko] [$_{\overline{V}}$ dek^{h}-aa]
Anjum.F=Erg Saddaf.F=Acc see-Perf.M.Sg
'Anjum saw Saddaf.'

b. [saddaf=ko] [anjum=ne] [dek^{h}-aa]

c. [anjum=ne] [dek^{h}-aa] [saddaf=ko]

d. [dek^{h}-aa] [anjum=ne] [saddaf=ko]

e. [dek^{h}-aa] [saddaf=ko] [anjum=ne]

f. [saddaf=ko] [dek^{h}-aa] [anjum=ne]

As demonstrated in (17), lexical items cannot in general freely scramble out of or within phrasal constituents.[7] In (17a), the $\overline{\text{V}}$ contains three elements: a main verb *banaa*, an aspect marker *rahii*, and an auxiliary *hai*. As (17b-f) show, any attempt to scramble the three elements within or out of the $\overline{\text{V}}$ produces an ill-formed result. Although not demonstrated here, the same is true for items contained within an NP.

(17) a. [$_{N}P$anjum] [$_{N}P$haar] [$_{\overline{\text{V}}}$ **banaa rah-ii**
Anjum.F=Erg necklace.M=Nom make Stat-Perf.F.Sg
hai]
be.Pres.3.Sg
'Anjum is making a necklace.'

b. *anjum haar [**rah-ii banaa hai**]

c. *anjum haar [**rah-ii hai banaa**]

d. *anjum [**hai**] haar [**banaa rah-ii**]

e. *anjum [**rah-ii hai**] haar [**banaa**]

f. *anjum [**banaa**] haar [**rah-ii hai**]

g. anjum [**banaa rah-ii hai**] haar

[7] It is possible to scramble out of finite embedded clauses as well. However, such 'long distance scrambling' is not the same type of scrambling observed in (16), or in the remainder of this thesis. Mahajan (1990) and Srivastav (1991c) distinguish topicalization from scrambling. Topicalization is claimed to be able to occur across a clause boundary, while scrambling can only take place within a clause. Gurtu (1985) in earlier work makes a similar distinction. Also see Dwivedi (1993) for a detailed examination of topicalization in Hindi.

The only well-formed scrambled sentence is shown in (17g). In conjunction with the ill-formed examples in (17b-f), (17g) shows that it is only possible to scramble the entire $\overline{\text{V}}$ within a finite clause. Under the assumption that scrambling depends on constituency, the data in (16) and (17) follow from the phrase structure I posit, and the generalization that only direct daughters of S can scramble.

3.3.1.1 Instructive

Data from agreement, anaphora, and control indicated that the Tell Construction in (18a) should be analyzed as a matrix verb which takes an infinitive complement. Given these data, the constituency structure roughly indicated in (18a) is predicted for the Tell Construction. And indeed, as (18b) and (18c) show, the infinitive complement *ciṭṭʰii likʰne=ko* 'to write a note', does scramble as a constituent. Also notice that the agreement facts remain constant under scrambling: the matrix verb *kahaa* 'said' in (18) never agrees with the embedded object *ciṭṭʰii* 'note'.

(18) a. [anjum=ne] [saddaf=ko] [ciṭṭʰii **likʰ-ne]=ko**
Anjum.F=Erg Saddaf.F=Dat note.F=Nom write-Inf.Obl=Acc
[**kah-aa**]
say-Perf.M.Sg
'Anjum told Saddaf to write a note.'

b. anjum=ne **kah-aa** saddaf=ko [ciṭṭʰii **likʰ-ne**]=**ko**

c. anjum=ne [ciṭṭʰii **likʰ-ne**]=**ko** saddaf=ko **kah-aa**

However, as (19) shows, the infinitive and the finite verb can also scramble as a unit. Since the instructive is not a complex predicate, there would seem to be no justification of this scrambling behavior.

(19) a. anjum=ne saddaf=ko [**likʰ-ne=ko kah-aa**] ciṭṭʰii

b. anjum=ne [**likʰ-ne=ko kah-aa**] ciṭṭʰii saddaf=ko

It could be argued that the assumption that scrambling is a test for constituency is wrong. Perhaps it is the case that in sentences involving infinitive complements, anything can appear anywhere. The data in (20), however, show that this is not the case. If anything were allowed to appear anywhere, there is no explanation for why the sentences in (20) are ungrammatical.

(20) a. *anjum=ne saddaf=ko **lik^{h}-ne=ko** ciṭṭ^{h}ii **kah-aa**

b. *anjum=ne saddaf=ko ciṭṭ^{h}ii **kah-aa lik^{h}-ne=ko**

c. *anjum=ne **kah-aa** ciṭṭ^{h}ii saddaf=ko **lik^{h}-ne=ko**

A close examination of (18)–(20) reveals that scrambled versions of the instructive are only well-formed under two circumstances: 1) if the infinitive complement is a constituent, as in (21a); 2) if the two predicates form a constituent, as in (21b).[8]

(21) a. [anjum=ne] [**ciṭṭ^{h}ii lik^{h}-ne=ko**] [saddaf=ko] [kah-aa]

b. [anjum=ne] [**lik^{h}-ne=ko kah-aa**] [saddaf=ko] [ciṭṭ^{h}ii]

As the next section shows, the permissive patterns exactly like the instructive.

3.3.1.2 Permissive

Example (22a) illustrates the constituency structure expected for the permissive as a complex predicate. Since the two predicates in the permissive were seen to function as a single predicate, they are expected to form a unit at phrase structure. However, as the data in (22)–(24) show, the scrambling pattern for the permissive is exactly the same as that of the instructive in (18)–(20). And as with the instructive, the permissive agreement facts are not affected by scrambling: the matrix verb *dii* 'let' always agrees with the nominative object *ciṭṭ^{h}ii* 'note'. This shows that *ciṭṭ^{h}ii* 'note' must be a matrix object.

(22) a. [anjum=ne] [saddaf=ko] [ciṭṭ^{h}ii] [**lik^{h}-ne**
Anjum.F=Erg Saddaf.F=Dat note.F=Nom write-Inf.Obl
d-ii]
give-Perf.F.Sg
'Anjum let Saddaf write a note.'

b. anjum=ne **d-ii** saddaf=ko [ciṭṭ^{h}ii **lik^{h}-ne**]

c. anjum=ne [ciṭṭ^{h}ii **lik^{h}-ne**] saddaf=ko **d-ii**

In (22b) and (22c), the infinitive *lik^{h}ne* 'to write' scrambles as a unit with *ciṭṭ^{h}ii* 'note', the argument it contributes to the complex predicate. On the other hand, as (23) shows, the

[8] Although all the scrambling possibilities are not represented here, this observation is borne out by the other possible scrambled versions of the sentence.

two predicates *lik*^{h}*ne dii* 'let write' can be also scrambled together as a unit. The pattern in (23) is what would be expected of a complex predicate. The pattern in (22) is not.[9]

(23) a. anjum=ne saddaf=ko [**lik**^{h}**-ne d-ii**] ciṭṭ^{h}ii

b. anjum=ne [**lik**^{h}**-ne d-ii**] ciṭṭ^{h}ii saddaf=ko

As with the instructive, it is not the case that any item can simply appear anywhere in a sentence. The examples in (24) are ill-formed precisely because neither the two predicates, nor the infinitive and its argument form a unit.

(24) a. *anjum=ne saddaf=ko **lik**^{h}**-ne** ciṭṭ^{h}ii **d-ii**

b. *anjum=ne saddaf=ko ciṭṭ^{h}ii **d-ii lik**^{h}**-ne**

c. *anjum=ne **d-ii** ciṭṭ^{h}ii saddaf=ko **lik**^{h}**-ne**

Just as for the instructive, then, there seem to be two possible c-structure realizations for the permissive. Data from negation and coordination in the next sections provide further evidence that there are two possible c-structures for the permissive and the instructive.

This type of 'constituency paradox' has also been noticed for German infinitival complements. In particular, the German *lassen* 'let' has long been problematic because it is ambiguous in its "syntactic behavior with respect to mono- and bisententiality" (McKay 1985:12). In particular, *lassen* 'let' patterns differently from other predicates taking infinitival complements in terms of reflexivization, negation, extraposition, and clitic movement

[9]Dayal (p.c.) points out that the examples in (22) and (21) may also be interpreted as purposives, in which a note is given to *Saddaf* for the purpose of writing. The permissive reading nevertheless remains possible as well, and is in fact the preferred interpretation in examples which pragmatically disfavor a purposive reading. Consider the data in (i) and (ii).

i.

Adamjee=sahib	apn-e	bet-e=ko	har	saal
Adamjee=sir	self-Obl	child-M.Obl=Dat	every	year
ek	nay-aa	makaan	**banaa-ne**	**de-taa**
one	new-M.Sg	house.M=Nom	make-Inf.Obl	give-Impf.M.Sg
hai				
be.Pres.3.Sg				

'Mr. Adamjee lets his son build a new house every year.'

ii.

Adamjee=sahib	[**de-taa**	**hai**]	apn-e	bet-e=ko
Adamjee=sir	give-Impf.M.Sg	be.Pres.3.Sg	self-Obl	child-M.Obl=Dat
har	saal	ek	nay-aa	makaan
every	year	one	new-M.Sg	house.M=Nom
banaa-ne				
make-Inf.Obl				

'Mr. Adamjee lets his son build a new house every year.'

(McKay 1985). I would suggest that the constituency paradox found for German infinitival complements can be accounted for by positing two possible c-structure realizations, and that the particular properties of *lassen* 'let' can be attributed to the fact that it participates in complex predicate formation, just as the Urdu permissive *de* 'let' does.

3.3.2 Negation

The negative *nahĩĩ* can be used both for phrasal and clausal negation in Urdu/Hindi (see T. Mohanan (1992) and Dwivedi (1991) for details). The clausal, or sentential, negation illustrated in (25) is the one relevant here.

(25) [anjum] [haar] [nahĩĩ [banaa rah-ii
Anjum.F=Nom necklace.M=Nom not make Stat-Perf.F.Sg
hai]]
be.Pres.3.Sg
'Anjum is not making a necklace.'

T. Mohanan's analysis, which I adopt, is that the negative appears to the left within a $\overline{\text{V}}$ in sentential negation. For the instructive, the prediction is that it should be possible to negate either of the two predicates. This prediction is borne out. The *nahĩĩ* 'not' can negate either the matrix verb, as in (26a), or it can negate the embedded infinitive, as in (26b). It can also be ambiguous as to which predicate is being negated. This is illustrated in (26c).

(26) a. anjum saddaf=ko [haar **banaa-ne**]=**ko** nahĩĩ
Anjum.F=Nom Saddaf.F=Dat necklace.M=Nom make-Inf.Obl=Acc not
kah-egii
say-Fut.F.Sg
'Anjum will not tell Saddaf to make a necklace.'

b. anjum [haar nahĩĩ **banaa-ne**]=**ko** saddaf=ko
Anjum.F=Nom necklace.M=Nom not make-Inf.Obl=Acc Saddaf.F=Dat
kah-egii
say-Fut.F.Sg
'Anjum will tell Saddaf not to make a necklace.'

c. anjum saddaf=ko [nahĩĩ **banaa-ne=ko** **kah-egii**]
Anjum.F=Nom Saddaf.F=Dat not make-Inf.Obl=Acc say-Fut.F.Sg
haar
necklace.M=Nom
'Anjum will tell Saddaf not to make a necklace.'
'Anjum will not tell Saddaf to make a necklace.'

The permissive as a complex predicate should have a different pattern of behavior. Since the two predicates in the permissive appear to function as a unit, the prediction with regard to negation is that the negative should only be able to appear to the left of **both** the infinitive and the finite verb, as in (27c). However, the negative can appear between the two verbs and negate only the finite verb *degii* 'will let'. This is shown in (27a). In (27b), the negative can negate the infinitive separately when the infinitive and its argument form a constituent.[10] And, as with the instructive, when the negative precedes both the predicates, an ambiguous interpretation as to the scope of the negative is possible.[11]

(27) a. anjum saddaf=ko [haar **banaa-ne**] nahĩĩ
Anjum.F=Nom Saddaf.F=Dat necklace.M=Nom make-Inf.Obl not
de-gii
give-Fut.F.Sg
'Anjum will not let Saddaf make a necklace.'

b. anjum [haar nahĩĩ **banaa-ne**] saddaf=ko
Anjum.F=Nom necklace.M=Nom not make-Inf.Obl Saddaf.F=Dat
de-gii
give-Fut.F.Sg
'Anjum will let Saddaf not make a necklace.'

c. anjum saddaf=ko [nahĩĩ **banaa-ne** **de-gii**]
Anjum.F=Nom Saddaf.F=Dat not make-Inf.Obl give-Fut.F.Sg
haar
necklace.M=Nom
'Anjum will not let Saddaf make a necklace.'
'Anjum will let Saddaf not make a necklace.'

The fact that the two predicates in the permissive function as a single unit with respect to agreement, anaphora and control is again not reflected in the phrase structure. With regard to negation, as well as scrambling, the infinitive and the finite verb may form a constituent, but they need not.

[10] It is furthermore possible to negate both the infinitive and the finite verb simultaneously by placing a *nahĩĩ* 'not' in front of each predicate separately. That is, it is possible for both the instructive and the permissive in (26) and (27) to contain two negatives, where each negative takes scope over one of the predicates.

[11] Dayal (p.c.) points out that an analysis in terms of Neg raising may be able to account for the ambiguous readings in (26c). This then would allow an alternative analysis to the one I propose in terms of two differing c-structure realizations. I do not pursue this possibility here.

3.3.3 Coordination

Anything that is a constituent can be coordinated in Urdu. If something is not conjoinable, then it is definitely not a constituent. However, it is not necessarily true that if something can be coordinated it must form a constituent. The examples in (28) illustrate a simple case. In (28a), two NPs are coordinated and the result is well-formed. One NP consists of an adjective and a noun, *garm anḍe* 'hot eggs', while the other consists of only the noun *roṭii* 'bread'. In (28b), on the other hand, items from different constituents have been conjoined and the sentence is ill-formed. Here the object *haar* 'necklace' and a part of the predicate, the main verb *banaa* 'make', are conjoined with the object *xat* 'letter' and the verb *lik*^{h} 'write'. Each main verb forms a constituent with the auxiliaries *rahii hai*. As the auxiliaries are not included in the coordination, the result is ill-formed.

(28) a. anjum [[garm anḍe] aur [roṭii]] xariid-tii
Anjum.F=Nom hot eggs.M.Obl=Nom and bread.F=Nom buy-Impf.F.Sg
hai
be.Pres.3.Sg
'Anjum buys hot eggs and bread.'

b. *anjum [haar banaa] aur [xat lik^{h}]
Anjum.F=Nom necklace.M=Nom make and letter.M=Nom write
rah-ii
Stat-Perf.F.Sg
hai
be.Pres.3.Sg
'Anjum is making a necklace and writing a letter.'

The instructive and the permissive both contrast with the data in (28b). As (29) shows, the instructive allows two possibilities for coordination. In (29a) the infinitive complements (*haar banaane* 'make necklace' and *xat lik*^{h}*ne* 'write letter') can be conjoined. This is as expected, given that the instructive is not a complex predicate. However, in (29b) two predicates, an infinitive and a finite matrix verb, are coordinated with another two predicates. It would appear that the infinitive and the finite verb form a $\overline{\text{V}}$, and that in (29b) two $\overline{\text{V}}$s are coordinated.

(29) a. anjum=ne saddaf=ko [[haar banaa-ne] aur
Anjum.F=Erg Saddaf.F=Dat necklace.M=Nom make-Inf.Obl and
[xat lik^{h}-ne=ko]]
letter.M=Nom write-Inf.Obl=Acc

kah-aa
say-Perf.M.Sg
'Anjum told Saddaf to make a necklace and write a letter.'

b. anjum=ne saddaf=ko roṭii [[xariid-ne=ko kah-aa] aur
Anjum.F=Erg Saddaf.F=ko bread.F=Nom buy-Inf.Obl=Acc say-Perf.M.Sg and
[k^{h}aa-ne=ko
eat-Inf.Obl=Acc
kah-aa]]
say-Perf.M.Sg
'Anjum told Saddaf to buy and eat bread.'

The permissive again displays exactly the same kind of pattern as the instructive. The well-formedness of (30b) is not surprising because the infinitive and the finite verb are expected to form a constituent. However, it is not clear why (30a) should be possible, when (28b) is not. If the permissive is really functioning as a single predicate, its behavior with respect to coordination should be like that of the single predicate in (28b), and not like that of the two predicate Tell Construction in (29a).[12]

[12]In light of the argument incorporation data discussed later on with respect to infinitives, the issue arises whether the permissive example in (30a) might not be an example of such incorporation. If it were indeed so, that only infinitives with incorporated arguments could conjoin, then the data in (30) could easily be accounted for under that view, and the postulation of two differing phrase structure realizations might seem like a less attractive solution (thanks to Dayal (p.c.) for pointing this out to me). However, sentences which cannot involve argument incorporation show the same pattern. Consider the parallel behavior of the instructive in (i) and the permissive in (ii).

i. mer-ii cacii=ne mujhe aaj subaah urdu-bazaar
my-F.Sg aunt.F=Erg I.Dat today morning Urdu-bazaar
jaa-ne=ko aur udir pupp^{h}ii nurjaahaan=se mil-ne=ko
go-Inf.Obl=Acc and there aunt.F Nurjahan=Inst meet-Inf.Obl=Acc
ka-haa
say-Perf.M.Sg
'My aunty told me this morning to go the Urdu Bazaar and meet Aunt Nurjahan there.'

ii. mer-ii cacii=ne mujhe aaj subaah urdu-bazaar
my-F.Sg aunt.F=Erg I.Dat today morning Urdu-bazaar
jaa-ne aur udir pupp^{h}ii nurjaahaan=se mil-ne
go-Inf.Obl and there aunt.F Nurjahan=Inst meet-Inf.Obl
di-yaa
give-Perf.M.Sg
'My aunty let me go to the Urdu Bazaar this morning and meet Aunt Nurjahan there.'

(30) a. anjum=ne saddaf=ko [[haar banaa-ne] aur
Anjum.F=Erg Saddaf.F=Dat necklace.M=Nom make-Inf.Obl and
[xat lik^{h}-ne]]
letter.M=Nom write-Inf.Obl
di-yaa
give-Perf.M.Sg
'Anjum let Saddaf make a necklace and write a letter.'

b. anjum=ne saddaf=ko roṭii [[xariid-ne d-ii] aur
Anjum.F=Erg Saddaf.F=Dat bread.F=Nom buy-Inf.Obl give-Perf.F.Sg and
[k^{h}aa-ne
eat-Inf.Obl
d-ii]]
give-Perf.F.Sg
'Anjum let Saddaf buy and eat bread.'

The coordination data thus provide more evidence for two differing possible phrase structure realizations of both the permissive and the instructive. Along with the scrambling and negation data, it shows that the instructive and the permissive cannot be differentiated in terms of phrase structure. In each case, the behavior of the permissive is exactly that of the Tell Construction. And yet the instructive is not a complex predicate, while the permissive clearly is.

3.3.4 Complex Predicate Analysis not affected by Scrambling

This section shows that the agreement, anaphora, and control facts which support a complex predicate analysis of the permissive are not crucially affected by scrambling. Under an LFG approach, which allows two or more differing c-structure realizations to correspond to the same f-structure, this is as expected. However, if complex predicate formation is defined in terms of movement at phrase structure (head-to-head movement in Baker (1988), for example), then the agreement, anaphora and control data would be predicted to be significantly affected by scrambling. The fact that this is not the case again argues against a pure phrase structure analysis of complex predicate formation.

3.3.4.1 Agreement

Recall that with respect to agreement, there was no evidence for an embedded object in the permissive. This contrasted with data from the instructive. The pertinent facts are

repeated in (31). For the permissive in (31a), the finite verb agrees with the object *ciṭṭʰii* 'note', indicating that it is a matrix argument. In (31b) the verb does not agree with *ciṭṭʰii* 'note'. Here the object must be embedded.

(31) a. anjum=ne saddaf=ko **ciṭṭʰii** likʰ-ne **d-ii**
Anjum.F=Erg Saddaf.F=Dat note.F=Nom write-Inf.Obl give-Perf.F.Sg
'Anjum let Saddaf write a note.'

b. anjum=ne saddaf=ko [**ciṭṭʰii** likʰ-ne=ko] **kah-aa**
Anjum.F=Erg Saddaf.F=Dat note.F=Nom write-Inf.Obl=Acc say-Perf.M.Sg
'Anjum told Saddaf to write the note.'

Now consider an alternative realization of the permissive, shown in (32a). Here the verb still agrees with the object *ciṭṭʰii* 'note'. Furthermore, in the parallel example of the instructive in (32b), the verb *kahaa* 'said' still cannot agree with the object *ciṭṭʰii* 'note'.

(32) a. anjum=ne [**ciṭṭʰii** likʰ-ne] saddaf=ko **d-ii**
Anjum.F=Erg note.F=Nom write-Inf.Obl Saddaf.F=Dat give-Perf.F.Sg
'Anjum let Saddaf write a note.'

b. anjum=ne [**ciṭṭʰii** likʰ-ne=ko] saddaf=ko **kah-aa**
Anjum.F=Erg note.F=Nom write-Inf.Obl=Acc Saddaf.F=Dat say-Perf.M.Sg
'Anjum told Saddaf to write the note.'

Scrambling thus has no effect on the verb agreement facts. The permissive still behaves as if it is a simple predicate.

3.3.4.2 Control

The example in (33) repeats the crucial control data for the permissive. Only *Anjum*, the single available subject in the sentence, is a possible controller of the participial adverbial *darvaaza kʰol kar* in (33). In (34), where the infinitive *rakʰne* 'put' is scrambled with its arguments (*samaan* 'luggage' and *kamre* 'room') the control facts do not change.

(33) **anjum=ne**$_i$ [____$_{i,*j}$ darvaazaa kʰol kar] saddaf=ko$_j$ [rakʰ-ne
Anjum.F=Erg door.M=Nom open having Saddaf.F=Dat put-Inf.Obl
di-yaa] samaan=ko
give-Perf.M.Sg luggage.M=Acc
kamre=mẽ
room.Obl=in
'Anjum, having opened the door, let Saddaf put the luggage in the room.'

(34) **anjum=ne**$_i$ [____$_{i,*j}$ darvaazaa k^{h}ol kar] [samaan=ko kamre=mẽ
Anjum.F=Erg door.M=Nom open having luggage.M=Acc room.Obl=in
rak^{h}-ne] saddaf=ko$_j$
put-Inf.Obl Saddaf.F=Dat
di-yaa
give-Perf.M.Sg
'Anjum, having opened the door, let Saddaf put the luggage in the room.'

Again, the behavior of the permissive with regard to control does not change because of a difference in c-structure.

3.3.4.3 Anaphora

The permissive examples in (35) restate the antecedency facts for the reflexive *apn-aa* 'self', and the pronominal *us=kaa*. As (35a) shows, the reflexive *apn-aa* can only refer to the subject *Anjum*, and not to *Saddaf*. In (35b), the pronominal *us=kaa* cannot refer to the subject, but can refer to the indirect object *Saddaf*, or another discourse referent.

(35) a. **anjum=ne**$_i$ saddaf=ko$_j$ [calaa-ne d-ii] apn-ii$_{i,*j}$ gaaṛii
Anjum.F=Erg Saddaf.F=Dat drive-Inf.Obl give-Perf.F.Sg self-F car.F=Nom
'Anjum let Saddaf drive her (Anjum's) car.'

b. anjum=ne$_i$ **saddaf=ko**$_j$ [calaa-ne d-ii] us=kii$_{*i,j,k}$ gaaṛii
Anjum.F=Erg Saddaf.F=Dat drive-Inf.Obl give-Perf.F.Sg pron=Gen.F car.F=Nom
'Anjum let Saddaf drive her (Saddaf's or somebody else's) car.'

The anaphora facts are exactly the same when the two verbs of the permissive do not form a constituent. This is demonstrated in (36).

(36) a. [apn-ii$_{i,*j}$ gaaṛii calaa-ne] **anjum=ne**$_i$ saddaf=ko$_j$ d-ii
self-F car.F=Nom drive-Inf.Obl Anjum.F=Erg Saddaf.F=Dat give-Perf.F.Sg
'Anjum let Saddaf drive her (Anjum's) car.'

b. [us=kii$_{*i,j,k}$ gaaṛii calaa-ne] anjum=ne$_i$ **saddaf=ko**$_j$ d-ii
pron=Gen.F car.F=Nom drive-Inf.Obl Anjum.F=Erg Saddaf.F=Dat give-Perf.F.Sg
'Anjum let Saddaf drive her (Saddaf's or somebody else's) car.'

The accumulated evidence in this section has shown that the particular c-structure represention posited for the permissive is completely independent of its behavior with respect to verb agreement, control and anaphora.

3.3.5 Summary

The data in this section present a strong argument in favor of defining a complex predicate at a level distinct from phrase structure. In my view, two predicates can only form a complex predicate when the combination of their argument structures results in a simple f-structure. A complex predicate like the permissive is thus exactly like a simple predicate in that it has only one subject, one object, etc. Complement constructions like the instructive, on the other hand, must a have a complex f-structure. As LFG defines grammatical functions at a separate level from phrase structure, the differences and similarities between the permissive and the instructive can easily be accounted for in terms of f-structure and c-structure properties. The fact that the permissive, but not the instructive, is a complex predicate is expressed at the level of f-structure. An abbreviated f-structure representation for the permissive in (37) is shown in (38).[13]

(37) anjum=ne saddaf=ko ciṭṭʰii likʰ-ne d-ii
Anjum.F=Erg Saddaf.F=Dat note.F=Nom write-Inf.Obl give-Perf.F.Sg
'Anjum let Saddaf write a note.'

(38)

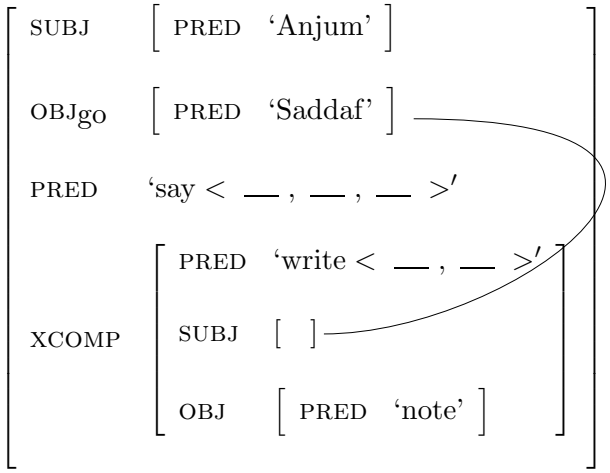
$$\begin{bmatrix} \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'Anjum'} \end{bmatrix} \\ \text{OBJ}_{\text{GO}} & \begin{bmatrix} \text{PRED} & \text{'Saddaf'} \end{bmatrix} \\ \text{PRED} & \text{'let-write} < \_\_ , \_\_ , \_\_ > \text{'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'note'} \end{bmatrix} \end{bmatrix}$$

The f-structure of the permissive has a single PRED 'let-write'. There is only one subject, one object and one indirect object. The permissive thus has a simple, or flat, f-structure. The f-structure in (40) for the instructive, on the other hand, contains two PREDs. The PRED 'say' takes an XCOMP as one of its arguments. This XCOMP in turn contains another argument taking PRED ('write'). The instructive therefore has a complex f-structure. The data from agreement, anaphora and control, which indicated that the permissive and the instructive differ from one another, are thus accounted for at the level of f-structure.

[13]The f-structure is abbreviated in the sense that I have only included the basic necessities. A complete f-structure would list such attributes as number, case, gender, tense, etc.

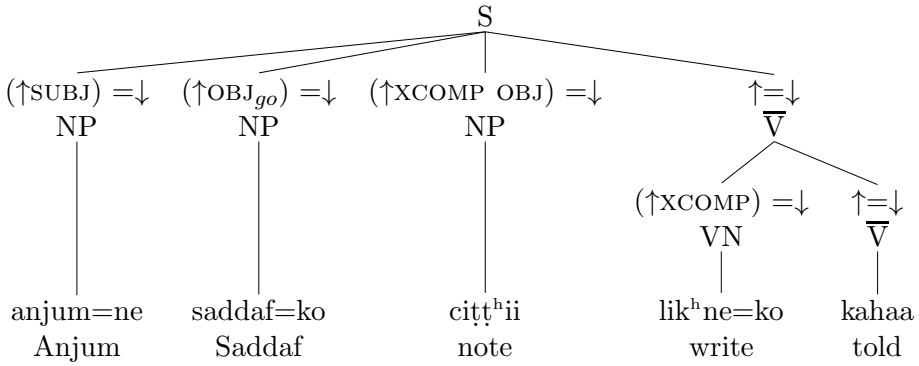
(39) anjum=ne saddaf=ko [ciṭṭʰii likʰ-ne=ko] kah-aa
Anjum.F=Erg Saddaf.F=Dat note.F=Nom write-Inf.Obl=Acc say-Perf.M.Sg
'Anjum told Saddaf to write a note.'

(40)

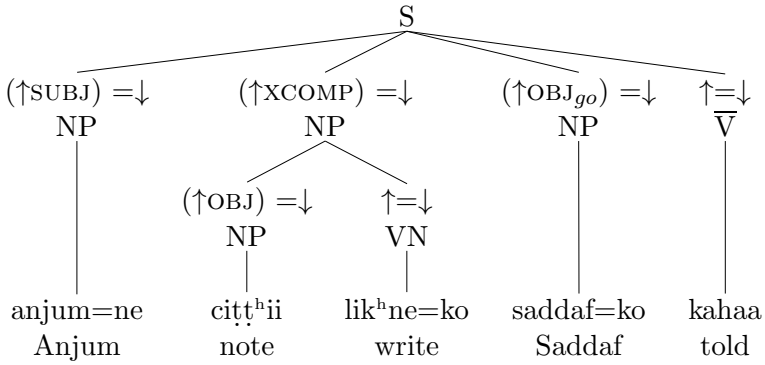


Although the permissive and the instructive behaved exactly the same with regard to scrambling, negation, and coordination, these phenomena also suggested two differing phrase structures for both of the constructions. The separation of grammatical function information from phrase structure makes it possible in LFG for a given sentence to have more than one c-structure realization, as long as the requirements of completeness and coherence are met at f-structure. I therefore propose that the permissive and the instructive be viewed as having two possible c-structure realizations. Both (41a) and (41b) are possible representations of the instructive, and both (42a) and (42b) are possible realizations of the permissive. If it is granted that the permissive and the instructive can be realized as either of the c-structures below, then their behavior with respect to scrambling, coordination and negation is completely accounted for.

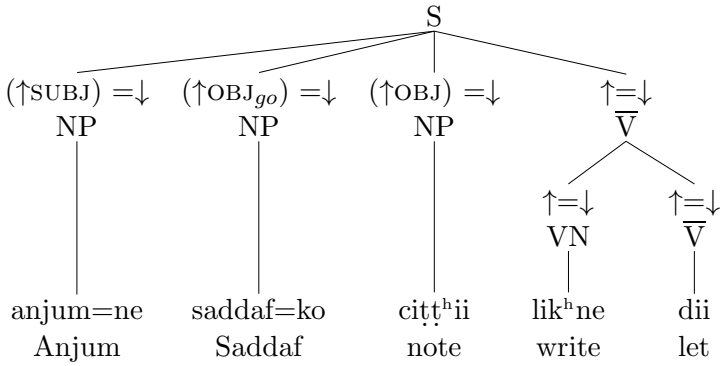
(41) a.


S
(↑SUBJ) =↓
NP
(↑OBJ$_{go}$) =↓
NP
(↑XCOMP OBJ) =↓
NP
↑=↓
V̄
(↑XCOMP) =↓
VN
↑=↓
V̄
anjum=ne
Anjum
saddaf=ko
Saddaf
cit̤t̤^{h}ii
note
lik^{h}ne=ko
write
kahaa
told

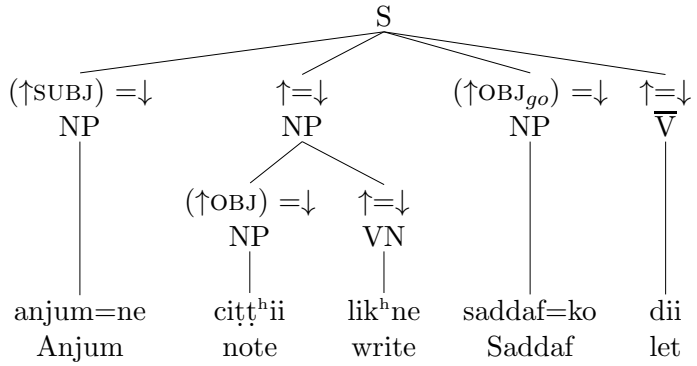
b.


S
(↑SUBJ) =↓
NP
(↑XCOMP) =↓
NP
(↑OBJ$_{go}$) =↓
NP
↑=↓
V̄
(↑OBJ) =↓
NP
↑=↓
VN
anjum=ne
Anjum
cit̤t̤^{h}ii
note
lik^{h}ne=ko
write
saddaf=ko
Saddaf
kahaa
told

(42) a.


S
(↑SUBJ) =↓
NP
(↑OBJ$_{go}$) =↓
NP
(↑OBJ) =↓
NP
↑=↓
V̄
↑=↓
VN
↑=↓
V̄
anjum=ne
Anjum
saddaf=ko
Saddaf
cit̤t̤^{h}ii
note
lik^{h}ne
write
dii
let

b.



The separation of information into f-structure and c-structure described here provides a nice account of complex predicates in general, since they can be defined as having a flat f-structure, independent of their particular phrase structure realization. However, the architecture of LFG, as formulated originally, does not allow for the possibility of discontinuous heads. As a consequence the annotations on the c-structure for the permissive in (42b) actually cannot be solved for a corresponding f-structure. Both the predicates *likʰ* 'write' and *de* 'let' must combine in some way to form the single PRED 'let-write' at f-structure. Several proposals have recently been formulated within LFG to account for discontinuous heads at c-structure. The analysis of complex predicates I present in Chapters 5 and 6 most closely resembles that of Alsina (1993). Alsina advocates an argument structure approach to Romance and Bantu causatives, and formulates a notion of *Predicate Composition*, by which two or more c-structure heads are composed into a single f-structure PRED. For other approaches and discussion of the problem see Andrews and Manning (1993), Kaplan and Wedekind (1993), and Dalrymple, Lamping, and Saraswat (1993).

In Chapter 4, I present evidence from Aspectual complex predicates for an *elaborated argument structure* approach to complex predicates, and then provide an analysis for the permissive and the Aspectual complex predicates in Chapters 5 and 6. However, before I go on to the next chapters, I first investigate the structure of the infinitival constituent for both the permissive and the instructive. Notice that I have represented infinitive phrases as NPs, and infinitive predicates with the category label VN in the c-structures in (39) and (40). This is meant to indicate that infinitive clauses are actually NPs headed by verbal nouns. In the next section, I show that infinitives in Urdu have the external distribution of an NP, while displaying the internal characteristics of a V. I also provide evidence for the verbal noun status of Urdu infinitives, and take a more detailed look at the structure of verbal nouns.

3.4 The Structure of Infinitives

3.4.1 Infinitive Clauses are NPs

In this section, I argue that the category of the infinitive constituent in the permissive complex predicate and the Tell Construction is an NP. The constituent headed by an infinitive has the distribution of an NP, can take case markers and undergo some further morphological processes that only apply to NPs.

Kachru (1980:40) refers to Hindi infinitive constituents as 'infinitival complements'. This analysis is essentially correct, as the sentences in (43) show. In (43a) and (43c) the infinitive *haar banaane* 'make necklace' functions as an argument of the finite verb. In (43b), the *haar banaane* is an adjunct of *bʰejaa* 'send'.

(43) a. anjum=ne saddaf=ko [haar banaa-ne=ko] kah-aa
Anjum.F=Erg Saddaf.F=Dat necklace.M=Nom make-Inf.Obl=Acc say-Perf.M.Sg
'Anjum told Saddaf to make a necklace.'

b. anjum=ne saddaf=ko [haar banaa-ne] bʰej-aa
Anjum.F=Erg Saddaf.F=Dat necklace.M=Nom make-Inf.Obl send-Perf.M.Sg
'Anjum sent Saddaf to make a necklace.'

c. anjum=ne saddaf=ko [haar banaa-ne] di-yaa
Anjum.F=Erg Saddaf.F=Dat necklace.M=Nom make-Inf.Obl give-Perf.M.Sg
'Anjum let Saddaf make a necklace.'

T. Mohanan (1992:21–25) discusses infinitives in the context of Noun Incorporation and represents the infinitive as a $\overline{\text{V}}$. Infinitives are also examined in Davison's (1988, 1990, 1991a, 1991b) and Mahajan's (1990) analyses of 'long distance agreement' in Hindi/Urdu. Davison consistently assumes that infinitive constituents such as *haar banaane* in (43) must be analyzed as CPs, while Mahajan treats them as VPs.

On the other hand, both Davison (1990:10, 1991a:10) and T. Mohanan (1990:99) briefly allude to the fact that Hindi/Urdu infinitives are nominal in character as well. Furthermore, Srivastav (1991c) examines infinitives in the context of wh-movement and concludes that they must be gerunds and should be treated as an IP whose head is a +N category, as proposed by Baker (1985) and Milsark (1988). I follow up on these observations by demonstrating in detail that the constituent headed by the infinitive has the distribution of an NP, can take case markers, and undergo some further morphological processes that only apply to NPs. The entire infinitive 'clause' must therefore be analyzed as an NP.

While infinitives behave as if they are NPs 'clause' externally, the infinitive itself behaves like a verb 'clause' internally most of the time. Infinitives in Urdu thus display the dual behavior exhibited by verbal nouns in languages like Japanese, Korean (Manning 1993) and, to a certain extent, English (Milsark 1988). However, under no circumstances can the infinitive constituent be analyzed as a CP.

3.4.1.1 Morphology

Glassman (1977:87) asserts that an infinitive in Urdu like *banaane* is "...in reality a verbal noun. As such, it inflects according to the rules for masculine nouns." An infinitive is formed by affixing the morpheme *-naa* to the bare stem of a verb. The bare stem for 'make' in the examples of the permissive and the instructive we have seen is *banaa-*. The morpheme *-naa* is actually the masculine form of the infinitive, and also functions as the default.

The examples in (44) show that the morphology on the infinitive can vary according to the gender of its object. In fact, the matrix verb also agrees with the embedded object of the infinitive, thus creating an effect of agreement across clause boundaries. Recall that the generalization for verb agreement is that a verb agrees with its highest nominative argument, and that agreement is usually clause-bounded. I return to the question of "long distance agreement" later. For now, I confine myself to the morphology on the infinitive.

(44) a. mujʰe [**gaaṛii** **calaa-nii**] **aa-tii** hai
I.Dat car.F=Nom drive-Inf.F come-Impf.F.Sg be.Pres.3.Sg
'I know how to drive a car.'

b. mujʰe [**ṭ̃āgaa** **calaa-naa**] **aa-taa** hai
I.Dat tonga.M=Nom drive-Inf.M come-Impf.M.Sg be.Pres.3.Sg
'I know how to drive a tonga.'

c. mujhe [**kitaab-ẽ** **paṛʰ-ne**] **aa-te** hãĩ
I.Dat book-Pl=Nom read-Inf.Pl know-Impf.Pl be.Pres.3.Pl
'I know how to read books.'

In (44a) the infinitive agrees with the feminine nominative *gaaṛii* 'car', and the morpheme *-nii* indicates feminine gender on the infinitive. In (44b), on the other hand, the embedded nominative object *tonga* is masculine and the infinitive is in the correspondingly masculine form *-naa*.[14] When the infinitive agrees with a plural entity, as in (44c) the appropriate morpheme is *-ne*.

[14] A tonga is two-wheeled horse-drawn carriage. It is used for transporting goods or functions as a taxi.

This *-ne* would appear to be same *-ne* as on the *banaane* 'make' in the instructive, the purposive, and the permissive in (43a), (43b), and (43c) respectively. However, although the two forms are homophonous, they do not perform the same function. The *-ne* in the permissive is not a marker of agreement, rather, it is a reflection of the non-nominative status of the infinitive. As (45) and (46) show, the *-ne* on *banaane* is invariable: it does not vary in order to agree with the nominative argument in the sentence. In (45b) a feminine infinitive to correspond to the feminine *gaaṛii* 'car' is ungrammatical. Similarly, in (46b), a masculine infinitive to correspond to the masculine *tonga* is bad.

(45) a. anjum=ne saddaf=ko [**gaaṛii calaa-ne**] d-ii
Anjum.F=Erg Saddaf.F=Dat car.F=Nom drive-Inf.Obl give-Perf.F.Sg
'Anjum let Saddaf drive the car.'

b. *anjum=ne saddaf=ko [**gaaṛii calaa-nii**] d-ii
Anjum.F=Erg Saddaf.F=Dat car.F=Nom drive-Inf.F give-Perf.F.Sg
*'Anjum let Saddaf drive the car.'

(46) a. anjum=ne saddaf=ko [**ṭ̃agaa calaa-ne**] di-yaa
Anjum.F=Erg Saddaf.F=Dat tonga.M=Nom drive-Inf.Obl give-Perf.M.Sg
'Anjum let Saddaf drive the tonga.'

b. *anjum=ne saddaf=ko [**ṭ̃agaa calaa-naa**] di-yaa
Anjum.F=Erg Saddaf.F=Dat tonga.M=Nom drive-Inf.M give-Perf.M.Sg
*'Anjum let Saddaf drive the tonga.'

The paradigm of inflection for the infinitive is essentially parallel to the paradigm for masculine nouns ending in *-aa*, such as *laṛkaa* 'boy'. These masculine nouns are inflected either when they are plural, or when they appear in non-nominative case. This is illustrated below with the accusative/dative marker *ko* as an example of non-nominative case marking.

	Nom. Singular	Nom. Plural	Dative/Accusative	Feminine
Masc. Noun	laṛk**aa**	laṛk**e**	laṛk**e**=ko	laṛk**ii** (girl)
Infinitive	banaa**naa**	banaa**ne**	banaa**ne**=ko	banaa**nii**

The inflected infinitive in the permissive *banaane* 'make', thus seems to pattern with masculine nouns ending in *-aa*. There is one addition in that the infinitive can also carry feminine inflection, as indicated above.

Notice that there is no overt case marker on the infinitive *calaane* 'drive' in the permissive constructions in (45) and (46). However, the infinitive of the instructive is marked with

a *ko*, which looks exactly like the accusative/dative marker *ko*. In addition, the 'oblique' morpheme *-e* can also appear on masculine nouns when they are not followed by an overt case marker. In some locatives, as in (47), an overt locative case marker may not be present, but the masculine noun, in this case *ḍaakxaanaa* 'post office', is inflected with the *-e*.

(47) anjum ḍaakxaane ga-yii
Anjum.F=Nom post office.M.Obl go-Perf.F.Sg
'Anjum went to the post office.'

The morphology on infinitive predicates thus suggests that the infinitive is functioning as a verbal noun. The morpheme *-n-* acts as a nominalizer in Urdu, while the morphemes *-aa/-ii/-e* function as indicators of gender and number. The permissive and the instructive again differ in that the infinitive in a permissive construction is inflected with the oblique *-e*, but does not show overt case, while the infinitive in a Tell Construction bears the overt marker *ko*. I suggest that this is a direct consequence of the fact that the infinitive in a permissive forms part of a complex predicate.

3.4.1.2 Case

This section demonstrates that the *ko* marker on the infinitive in an instructive such as (48) is indeed a case marker, and not a kind of complementizer.

(48) anjum=ne saddaf=ko [haar banaa-ne=**ko**] kah-aa
Anjum.F=Erg Saddaf.F=Dat necklace.M=Nom make-Inf.Obl=Acc say-Perf.M.Sg
'Anjum told Saddaf to make a necklace.'

The sentence in (49) is version of (48), but utilizes a finite complement clause. The embedded clause in (49) is introduced by the complementizer *ke* 'that'. Although Urdu is a head final language, complementizers canonically appear at the beginning of a clause. Thus *ke* in (49) behaves like a canonical complementizer, while the *ko* in (48) does not. Furthermore, the *ko* only appears on a few types of nonfinite clauses. Therefore, if it were analyzed as a complementizer, it would have to be analyzed as a very atypical complementizer, especially as it never appears with a finite embedded clause like (49).

(49) anjum=ne saddaf=se$_i$ kah-aa [**ke** vo$_i$ haar
Anjum.F=Erg Saddaf.F=Inst say-Perf.M.Sg that Pron.3.Sg necklace.M=Nom
banaa-ye]
make-Subj
'Anjum told Saddaf that she (Saddaf) should make a necklace.'

The examples in (50) and (51), taken from Davison (1991b), show that *ko* is not the only postposition which can appear after an inflected infinitive. In (50), the infinitive *aa-ne* 'come' is followed by the genitive *kii*. The infinitives in (51) are followed either by *ko*, the locative *par* 'on', or the postposition *ke liye* 'for'.

(50) [un-ke aa-ne**=kii** baat] mahatvapuurn hai
their=Gen.Obl come-Inf.Obl=Gen.F matter.F=Nom important be.Pres.3.Sg
'The fact of their coming is important.'

(51) a. raad^{h}aa=ne mohan=ko [kitaab paṛ^{h}-ne=**ko**] majbuur
Radha.F=Erg Mohan.M=Dat book.F=Nom read-Inf.Obl=Acc force
ki-yaa
do-Perf.M.Sg
'Radha forced Mohan to read a book.'

b. raad^{h}aa=ne mohan=ko [kitaab paṛ^{h}-ne=**par**] majbuur
Radha.F=Erg Mohan.M=Dat book.F=Nom read-Inf.Obl=on(Loc) force
ki-yaa
do-Perf.M.Sg
'Radha forced Mohan to read a book.'

c. raad^{h}aa=ne mohan=ko [kitaab paṛ^{h}-ne=**ke liye**] majbuur
Radha.F=Erg Mohan.M=Dat book.F=Nom read-Inf.Obl=for force
ki-yaa
do-Perf.M.Sg
'Radha forced Mohan to read a book.'

In light of the examples in (50) and (51), it is likely that the *ko* in the instructive in (48) is a case marker and not a complementizer. Case markers in Urdu do not otherwise appear on non-nominals.[15] Also note the special signficance of the genitive marker *-kii* in (50). The fact an infinitive can act as a genitive argument of a noun, *baat* 'matter' in (50), is a very good indication that the infinitive is acting like a noun here. It is furthermore clear that the entire infinitive constituent in the instructive in (48) is acting as the direct object of the verb *kah* 'say' (also see Davison (1991a) for an analysis of infinitive constituents as satisfying argument positions in a theta-grid). The *ko* case marker in the instructive in (48) therefore marks the infinitive constituent as the direct object of the verb *kah* 'say'.[16]

[15]They may appear on adjectives. However, I would argue that the adjectives in these cases are really functioning as nouns.

[16]T. Mohanan (1993b) has pointed out that there is a constraint on double occurences of a single case-marker like *ko* on arguments in Hindi. The instructive I use is in fact not very good for T. Mohanan. In

Case marking thus provides the first piece of tangible evidence that the infinitive constituent *haar banaane* 'make necklace' in (48) must be an NP. Case markers in Urdu are restricted to appearing on NPs. They can never appear on an embedded finite clause such as the *ke* 'that' CP in (52).

(52) *anjum=ne saddaf=se$_i$ kah-aa [ke vo$_i$ haar
Anjum.F=Erg Saddaf.F=Inst say-Perf.M.Sg that Pron.3.Sg necklace.M=Nom
banaa-ye=**ko**]
make-Subj=Acc
*'Anjum told Saddaf that she (Saddaf) should make a necklace.'

Evidence from coordination also shows that, with regard to case, infinitive clauses like *haar banaane* 'make necklace' behave like NPs. Example (53) coordinates two simple NPs. Notice that the case marker *ko* can appear either on both NPs, as in (53b), or only the rightmost NP in the coordinated structure, as in (53a).[17]

(53) a. [[laṛkiyõ] aur [laṛkõ=ko]] ṭofii d-o
girl.F.Pl and boy.M.Pl-Dat toffee.F=Nom give-Imp
'Give the boys and girls some candy.'

b. [[laṛkiyõ=ko] aur [laṛkõ=ko]] ṭofii d-o
girl.F.Pl=Dat and boy.M.Pl-Dat toffee.F=Nom give-Imp
'Give the boys and girls some candy.'

The instructive in (54) parallels the coordination facts in (53). The case marker *ko* can appear only on the rightmost infinitive, as in (54a), but it can also appear on both infinitives, as in (54b).

---

her dialect the comitative (instrumental) *se* is preferred for the indirect object *Saddaf* in (50). There is, however, a further issue of politeness which must also be taken into account. The use of *ko* to mark the indirect object is somehow more "direct", and therefore more impolite, than the use of the comitative *se* in constructions like (49). In discussions with T. Mohanan it emerged that besides the "double-ko" constraint, which must be violated sometimes in order to allow both dative and accusative case-marking in sentences like (i.), she was also reluctant to use the instructive with a *ko* because it was very impolite.

(i.) mãĩ=ne aapn-ii laṛkii=ko us daktar=ko dek^h-a-yaa
I=Erg self-F.Sg girl=Acc that doctor=Dat see-Caus-Perf.M.Sg
'I took my girl to that doctor.'

[17]Recall that T. Mohanan (1992) shows that case markers in Hindi/Urdu must be clitics which attach phrasally, not lexically, so the data in (53a) is expected. However, a good analysis of the overall behavior of clitics in Urdu remains a subject for investigation.

(54) a. anjum=ne saddaf=ko [[haar banaa-ne] aur
Anjum.F=Erg Saddaf.F=Dat necklace.M=Nom make-Inf.Obl and
[xat lik^{h}-ne=**ko**]]
letter.M=Nom write-Inf.Obl=Acc
kah-aa
say-Perf.M.Sg
'Anjum told Saddaf to make a necklace and write a letter.'

b. anjum=ne saddaf=ko [[haar banaa-ne=**ko**] aur
Anjum.F=Erg Saddaf.F=Dat necklace.M=Nom make-Inf.Obl=Acc and
[xat lik^{h}-ne=**ko**]]
letter.M=Nom write-Inf.Obl=Acc
kah-aa
say-Perf.M.Sg
'Anjum told Saddaf to make a necklace and write a letter.'

Furthermore, true complementizers like *ke* do not display the same behavior as the *ko* under coordination. This is demonstrated in (55). In particular, the attempts at coordination in (55a) and (55b) are ill-formed, while the parallel infinitive constructions in (54a) and (54b) above are good. Only when the two conjuncts are contained within the same finite complement, as in (55c), is coordination possible (see Srivastav (1991c) for details on wh-dependencies in Hindi and the structure of such complements).

(55) a. *anjum=ne saddaf=se$_i$ kah-aa [ke vo$_i$
Anjum.F=Erg Saddaf.F=Inst say-Perf.M.Sg Pron.3.Sg necklace.M=Nom
haar banaa-ye] aur [vo$_{*i}$ xat
make-Subj and that Pron.3.Sg letter.M=Nom
lik^{h}e]
write-Subj
'Anjum told Saddaf that she (Saddaf) should make a necklace and that she should write a letter.'

b. *anjum=ne saddaf=se$_i$ kah-aa [ke vo$_i$ haar
Anjum.F=Erg Saddaf.F=Inst say-Perf.M.Sg that Pron.3.Sg necklace.M=Nom
banaa-ye] aur [ke vo$_i$ xat
make-Subj and that Pron.3.Sg letter.M=Nom
lik^{h}e]
write-Subj
'Anjum told Saddaf that she (Saddaf) should make a necklace and that she should write a letter.'

c. anjum=ne saddaf=se$_i$ kah-aa [ke vo$_i$ [haar
Anjum.F=Erg Saddaf.F=Inst say-Perf.M.Sg that Pron.3.Sg necklace.M=Nom
banaa-ye] aur [xat
make-Subj and letter.M=Nom
lik^{h}e]]
write-Subj
'Anjum told Saddaf that she (Saddaf) should make a necklace and write a letter.'

In summary, evidence from case marking and coordination strongly suggests that infinitive constituents should be analyzed as NPs, and not CPs. The subsequent sections further substantiate the argument.

3.4.1.3 Finiteness and Case

Srivastav (1991c) and Davison (1991b) show that nonfinite and finite clauses in Hindi/Urdu have differing distributions. Finite clauses can only appear "clause externally" while nonfinite clauses appear in "clause internal argument positions" (Davison 1991b:1). This is demonstrated by the examples in (56). The instructive in (56a) contains an infinitive which can appear clause internally. The sentences in (56b) and (56c) contain corresponding finite embedded clauses. When the finite clause appears "outside" the main clause, as in (56b), the sentence is good. However, when the finite clause appears "inside" the matrix clause in (56c), the result is ungrammatical.

(56) a. anjum=ne saddaf=ko [kitaab paṛ^h-ne=ko] kah-aa
Anjum.F=Erg Saddaf.F=Dat book.F=Nom read-Inf.Obl=Acc say-Perf.M.Sg
'Anjum told Saddaf to read a book.'

b. anjum=ne saddaf=se$_i$ kah-aa [ke vo$_i$ kitaab
Anjum.F=Erg Saddaf.F=Inst say-Perf.M.Sg that Pron.3rd.Sg book.F=Nom
paṛ^h-e]
read-Subj
'Anjum told Saddaf that she (Saddaf) should read a book.'

c. *anjum=ne saddaf=se$_i$ [ke vo$_i$ kitaab paṛ^h-e]
Anjum.F=Erg Saddaf.F=Inst that Pron.3rd.Sg book.F=Nom read-Subj
kah-aa
say-Perf.M.Sg
'Anjum told Saddaf that she (Saddaf) should read a book.'

Davison explains this contrast between nonfinite and finite clauses in Urdu with a modification of Stowell's (1981) Case Resistance Principle: "Finite inflection may not be directly combined with case." (Davison 1991b:3). According to Davison, nonfinite clauses can only appear in positions which are governed and are assigned Case. However, the facts cited by Davison can be accounted for just as well under the theory that nonfinite clauses like the one in (56a) are NPs. In fact, this is essentially very close to Srivastav's (1991a, 1991c) approach to the problem. Srivastav also invokes the Case Resistance Principle, but for her it interacts with the fact that infinitive clauses must be analyzed as projections of IP, and are thus not subject to the Case Resistance Principle. Finite embedded clauses as in (56b), on the other hand, are analyzed as CPs adjoined to IP and are thus subject to the Case Resistance Principle. Srivastav thus posits a clear category difference between finite and nonfinite embedded complements.

As Bresnan (1991) presents evidence from Bantu against the Case Resistance Principle, an analysis of the distribution facts which only draws on the NP status of infinitive constituents seems preferable. The phrase structure rules in (57) represent the difference in distribution between the infinitives and embedded finite clauses such as the "that" clause.

(57) $S \rightarrow NP^* \ \overline{V} \ NP^*$
$S \rightarrow S \ , \ CP$

Infinitives can appear anywhere NPs usually can, i.e., anywhere within an S. The finite clause introduced by *ke* 'that', on the other hand, is not an NP, it is a CP. As such, it does not pattern with NPs, rather, it appears at the edge of a matrix clause because it is adjoined to S (or IP).

3.4.1.4 Correlatives

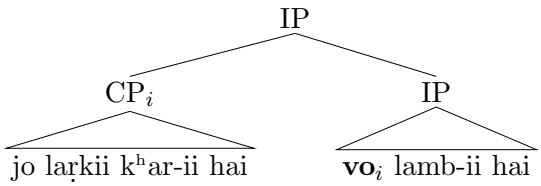
As already briefly mentioned above, Srivastav's (1991b) work on the syntax and semantics of correlatives in Hindi helps to illustrate more clearly that the infinitives are indeed functioning as NPs. Although the issue of correlatives is only tangentially relevant here, I believe her analysis helps to illuminate the structure of nonfinite versus finite clauses.

Specifically, Srivastav (1991b:682) observes that "... it is shown that CPs in Hindi cannot appear in case-marked positions. Thus finite complements of verbs must appear postverbally, thereby accounting for the non-rigid SOV pattern of the language." She identifies two types of correlatives in Hindi. An example, taken from Srivastav (1991b:652), of the type relevant to this paper is given in (58).

(58) [**jo** laṛkii kʰaṛ-ii hai] **vo** lamb-ii hai
which girl.F=Nom standing-F be.Pres.3.Sg Pron.3.Sg tall-F be.Pres.3.Sg
'The girl who is standing is tall.'

Srivastav (1991b:653) analyzes the correlative in (58) as "a quantificational structure in which the relative clause binds the main clause nominal." The structure she posits for the correlative in (58) is shown in (59). The CP, the relative clause, is coindexed with the pronoun *vo* in the main clause. This coindexation gives rise to the correlative reading of (58).

(59)



Srivastav (1991b:655) furthermore suggests that "only NPs with demonstratives qualify as variables that can be bound in such configurations." An example, taken from Srivastav (1991b:648), is shown in (60). Here the NP *laṛkii* 'girl' does not function as a demonstrative, so the relative clause cannot be coindexed with it, and the resulting sentence is ungrammatical. The sentence in (60) contrasts with the sentence in (58).

(60) *[**jo** laṛkii khaṛ-ii hai] **laṛkii** lamb-ii hai
which girl.F=Nom standing-F be.Pres.3.Sg girl.F=Nom tall-F be.Pres.3.Sg
'The girl who is standing, the girl is tall.'

If one takes a closer look at the finite "that" clauses, they seem to pattern with the correlative construction in (58). That is, they behave as if they are subordinate clauses which must coindex with a demonstrative in the main clause. The pattern is illustrated in (61), taken from Davison (1991b:3). The sentence in (61a) is good because the demonstrative *is* 'this' is available for coindexing with the finite "that" clause. But if there is no demonstrative in the main clause, as in (61b), the resulting example is bad.

(61) a. raadhaa=ne mohan=ko$_i$ [**is**$_j$ baat=par] majbuur ki-yaa [**ke**$_j$
Radha.F=Erg Mohan.M=Dat this matter.F=on force do-Perf.M.Sg that
vo$_i$ kitaab
Pron.3.Sg book.F=Nom
paṛʰ-e]
read-Subj
'Radha forced Mohan to read a book.'

b. *raadhaa=ne mohan=ko$_i$ [**baat=par**] majbuur ki-yaa [**ke**$_?$
Radha.F=Erg Mohan.M=Dat matter.M=on force do-Perf.M.Sg that
vo$_i$ kitaab
Pron.3.Sg book.F=Nom
paṛ^h-e]
read-Subj
'Radha forced Mohan to read a book.'

Infinitives do not show the contrast in (61). In example (62a) an attempt is made to coindex the infinitive constituent with a demonstrative in the matrix clause. This sentence is just as bad as (62b), in which no coindexation is attempted. The sentences are out because *baat=par* 'on this matter' and *kitaab paṛ^{h}ne=par* 'on reading the book' are both trying to fill the same argument position of the predicate *majbuur ki-yaa* 'force'. Coindexation is possible with the "that" clause in (61a) because the "that" clause is a CP and not an NP like the infinitive constituent in (62).

(62) a. *raadhaa=ne mohan=ko [**is**$_j$ baat=par] [kitaab
Radha.F=Erg Mohan.M=Dat this matter.F=on book.F=Nom
paṛ^h-ne=par$_j$] majbuur
read-Inf.Obl=on force
ki-yaa
do-Perf.M.Sg
'Radha forced Mohan to read a book.'

b. *raadhaa=ne mohan=ko [**baat=par**] [kitaab paṛ^h-ne=par]
Radha.F=Erg Mohan.M=Dat matter.M=on book.F=Nom read-Inf.Obl=on
majbuur
force
ki-yaa
do-Perf.M.Sg
'Radha forced Mohan to read a book.'

Finite "that" clauses thus pattern with the correlatives examined by Srivastav, and could be analyzed along the same lines as suggested by Srivastav for correlatives. Infinitives, on the other hand, do not pattern with correlatives or "that" clauses, indicating that they are not CPs.

3.4.1.5 Participials vs. Infinitives

The nonfinite embedded clauses presented so far are, of course, not the only kind of nonfinite clause in Urdu. Another kind, called a "participial construction" by Kachru (1980:34), is illustrated by the sentence in (63).

(63) [ro-tii huii] baccii=ko bula-o
cry-Impf.F.Sg being.F child.F=Acc call-Imp
'Call the crying child!'

Notice that the morphology on the verbs *ro-tii huii* 'be crying' differs from that of the infinitives. There is no morphology to suggest that the embedded verb here may be functioning as a verbal noun. The defining characteristic of the participial in (63) is that it is formed with a participial form of the verb *ho* 'be'. In (63) the *huii* is feminine because it agrees with the feminine noun *baccii* 'child'.

Kachru (1980) and Davison (1991b) note that the participial in (63) can be used either as an adjective or an adverbial. In (64a) the participial is functioning as an adjective. In (64b) it is used adverbially.

(64) a. [[ro-taa huaa] laṛkaa] a-yaa
cry-Impf.M.Sg being.M boy.M=Nom come-Perf.M.Sg
'The crying boy came.'

b. laṛkaa [ro-taa huaa] a-yaa
boy.M=Nom cry-Impf.M.Sg being-M come-Perf.M.Sg
'The boy came crying.'

The significance of these participials is that they do not pattern with the infinitives. It is not just the case that finite clauses must be differentiated from nonfinite clauses with regard to their phrase structure positions, distinctions must also be drawn within the category of nonfinite clauses. Infinitives have the distribution of NPs, while the participials in (64) pattern with adjectives and adverbials. The example in (65a) contrasts with (64a). In (64a) an adjectival participial yields a perfectly good sentence. However, in (65a), where a nominative infinitive has been substituted into the adjectival position, the result is bad.

(65) a. *[[ro-naa] laṛkaa] a-yaa
cry-Inf.M boy.M=Nom come-Perf.M.Sg
'The crying boy came.'

b. *laṛkaa [ro-naa] a-yaa
boy.M=Nom cry-Inf.M come-Perf.M.Sg
'The boy came crying.'

Similarly, (65b) contrasts with (64b). In (64b) the participial is good. The infinitive *ro-naa* 'to cry' in the same position in (64b), however, results in an ill-formed sentence.

The accumulated evidence from distribution has thus so far consistently pointed to the conclusion that infinitives behave as if they are NPs. And although I do not demonstrate it here exhaustively, it is quite clear that infinitives can appear wherever NPs usually can: in subject, object, indirect object, or adjunct position.

3.4.1.6 Wh Questions

Further evidence for the NP status of infinitives comes from Srivastav's (1991c) examination of the scope of Hindi/Urdu Wh-questions. Infinitives and finite complements contrast very clearly with respect to question formation in Hindi/Urdu. Example (66a) contains a wh-word in an embedded infinitive. This sentence must be interpreted as a direct question. The sentence in (66b), on the other hand, is formed with a finite complement and can only be interpreted as an indirect question.[18]

(66) a. tum [kyaa kar-naa] jaan-te ho
you=Nom what do-Inf.M know-Impf.Obl be.Pres.2.Sg
'What do you know to do?'

b. tum jaan-te ho [ke us=ne kyaa ki-yaa]
you=Nom know-Impf.Obl be.Pres.2.Sg that Pron.3.Sg-Erg what do-Perf.M.Sg
'You know what he did?'

Since (66a) must be interpreted as a direct question, the wh-word *kyaa* 'what' is taken to have wide scope. In the indirect question in (66b), on the other hand, the wh-word has narrow scope. Srivastav explains the fact that a finite complement acts a scope island in Hindi/Urdu by analyzing the finite complement as a CP in which the wh-word moves to the local SPEC of CP. Since finite clauses are scope islands in Hindi, they cannot move any further. If infinitives are analyzed as a gerundive of either category IP or NP (Srivastav's analysis leaves the possibilities open), then the contrast between (66a) and (66b) is easily explained. In an infinitive constituent there is no local SPEC of CP for the wh-word to move to, so it moves to the matrix SPEC of CP, and causes a wide scope reading.

These facts make sense under my approach as well. If the infinitive is an NP, then the wh-word should be able to move[19] and thus give rise to a wide scope reading. The constraint

[18]These examples are taken from Srivastav (1991c:178).

[19]The scrambling data shown for the permissive is duplicated by all constructions which contain infinitives.

that only direct daughters of S may scramble accounts for the fact that the *kyaa* 'what' in (66b) may not scramble out of the CP and must therefore have a narrow scope reading.[20] Note that in a theory like LFG, a treatment of wh-words in terms of c-structure movement is not required. Constraints on wh-scope are stated in terms of f-structure, rather than in terms of c-structure categories. However, an exploration of the above facts in LFG-particular terms would take me too far afield here, so I instead move on to other evidence for the NP status of Urdu infinitives.

3.4.1.7 Morphology: -valaa

Finally, the NP status of infinitives is further confirmed by data involving the suffix *-valaa* 'one'. Glassman (1977:304) characterizes *-valaa* 'one' as "a suffix (which) may imply the possessor, seller, agent, or distributor of something, have the sense of 'the one with' or be used to convey value or price. It inflects to agree with whatever it is used with." T. Mohanan (1992) identifies two types of *-valaa*, one which attaches lexically, as in (67), and one which attaches to a phrase, but the difference is not directly relevant here.

(67) sabzii-valii a-yii hai
vegetable.F-one.F=Nom come-Perf.F.Sg be.Pres.3.Sg
'The vegetable seller has arrived.'

When *-valaa* attaches lexically, it can attach either to a noun, as in (67), or to an adjective, as in (68). The output may either be an adjective or a noun, regardless of whether *-valaa* was suffixed to a noun or an adjective. In other words, N+valaa could function as a noun, as in (67), or it could function as an adjective, as in (69), adapted from Glassman (1977:304).

(68) laal-valii ṭopii mujhe dek^{h}a-o
red-one.F hat.F=Nom I.Dat show-Imp
'Show me the red hat!'

(69) do rupae-valaa ṭikaṭ le a-o
two rupee-one.M stamp.M=Nom take come-Imp
'Bring a two-rupee stamp!'

[20] It is actually possible for the *kyaa* 'what' to appear at the very front of the sentence. However, as Srivastav (1991c) also shows, this is an instance of topicalization, which is subject to different constraints. Furthermore, finite clauses do not allow movement at LF, so wide scope readings are always ruled out. See Srivastav (1991c), Mahajan (1990), and Gurtu (1985) for a more detailed discussion of the constraints involved. Note that the simple constraint on scrambling I state above does not cover instances of topicalization. For some suggestive work focusing on Russian see King (1993).

Similarly, a combination of Adj+valaa can result either in an adjective, as in (68), or in a noun, as in (70). Note that the Adj+valaa, *cʰoṭii-valii* 'small one', is case marked in (70). So when X+valaa is functioning as a noun, it is possible to mark it with case.

(70) cʰoṭii-valii=ko zaraa bula-o
small.F-one.F=Acc just call-Imp
'Just call that small one (girl).'

The examples in (67)–(70) exhaustively illustrate the uses of *-valaa* when it is attached lexically. It cannot appear on verbs. The sentences in (71b–c) are ungrammatical because the suffix *-valaa* appears on a verb. As *-valaa* can appear on infinitives, this further motivates the analysis of the infinitive constituent as an NP.

(71) a. laṛkii haar banaa-tii hai
girl.F=Nom necklace.M=Nom make-Impf.F.Sg be.Pres.3.Sg
'The girl is making a necklace.'

b. *laṛkii haar **banaa-tii-valii** hai
girl.F=Nom necklace.M=Nom make-Impf.F.Sg-one.F be.Pres.3.Sg
'The girl is one making a necklace.'

c. *laṛkii sundar **hai-valii**
girl.F=Nom beautiful be.Pres.3.Sg-one.F
'The girl is a beautiful one.'

Verma (1971:104) describes two uses of phrasal *-valaa*. One use is an "adjectivization transformation" in which *-valaa* appears on an infinitive constituent, which then functions as an adjective. It can also be used as a marker of immediate future, and in these instances can also only appear on infinitives. Both uses of *-valaa* are demonstrated by (72), taken from Verma (1971:104).

(72) laṛkaa paṛʰ-ne=valaa hai
boy.M=Nom read-Inf.Obl=one.M be.Pres.3.Sg
'The boy is the studious type.'
'The boy is about to start studying.'

An infinitive with *-valaa* can function as an NP, as well as an AP. This is demonstrated by (73). Here the phrase *gʰar jane-vale=ko* 'the one going home' is case marked with the accusative *ko*.

(73) [gʰar jaa-ne]=vale=ko bula-o
house go-Inf.Obl=one.Obl=Acc call-Imp
'Call the one going home.'

Recall that *-valaa* can be affixed either to a noun or an adjective, and result in either a noun or an adjective. Since I have already demonstrated that infinitives do not have the distribution of an adjective, or AP, the evidence from *-valaa* once again suggests that infinitive phrases should be analyzed as NPs.

For the sake of completeness, I show in (74) and (75) that *-valaa* cannot appear on a finite embedded clause or a nonfinite participial.

(74) *anjum=ne saddaf=se$_i$ kah-aa [ke vo$_i$ kitaab
Anjum.F=Erg Saddaf.F=Inst say-Perf.M.Sg that Pron.3.Sg=Nom book.F=Nom
paṛ^h-e]=**vale**
read-Subj-one.Obl
'Anjum told Saddaf that she (Saddaf) should read a book.'

(75) *[[ro-tii huii]=**valii** baccii=ko] bula-o
cry-Impf.F.Sg being.F=one.F child.F=Acc call-Imp
'Call the crying child.'

The *ke* 'that' clause in (74) and the participial in (75) are both headed by verbs. This shows that *-valaa* cannot be suffixed to a clause headed by a finite verb. It can appear on infinitives exactly because these are nominalized forms.

3.4.2 Verbal Nouns

The accumulated evidence of the previous sections provides a convincing argument that infinitives in Urdu are NPs. However, the infinitive predicates themselves are ambiguous as to whether they are an N or a V. The case marking and adverbial modification properties of infinitives indicate that the infinitive must really be analyzed as a gerundive or verbal noun. It cannot simply be accorded the status of other Ns which head an NPs in the language.

Recall that Verma (1971) describes the suffix *-valaa* as a marker of immediate future. The relevant example is repeated here in (76). Now, if the *paṛ^{h}ne* 'read' here is indeed an N, it should pattern like a simple noun with regard to *-valaa* 'one'. However, this is not the case. Simple nouns can never take on aspectual properties through the attachment of *-valaa*. So, while (76) has a sense of imminent action, no such interpretation can be given for (77).

(76) laṛkaa paṛ^h-ne=valaa hai
boy.M=Nom read-Inf.Obl=one.M be.Pres.3.Sg
'The boy is about to start studying.'

(77) laṛkii sabzii-valii hai
girl.F=Nom vegetable.F-one.F be.Pres.3.Sg
'The girl is a vegetable seller.'
*'The girl is about to be a vegetable seller.'

The infinitive cannot be modified by adjectives, as one would expect if it were an N. The sentence in (78a), for example, is ungrammatical because the adjective *acc*^{h}*aa* 'good' modifies the infinitive *banaane* 'make'. Infinitives can only successfully be modified by adverbials, as in (78b).

(78) a. *anjum=ne saddaf=ko haar acc^{h}aa banaa-ne
Anjum.F=Erg Saddaf.F=Dat necklace.M=Nom good.M make-Inf.Obl
di-yaa
give-Perf.M.Sg
'Anjum let Saddaf make a necklace good.'

b. anjum=ne saddaf=ko haar jaldii=se banaa-ne
Anjum.F=Erg Saddaf.F=Dat necklace.M=Nom hurry=Inst make-Inf.Obl
di-yaa
give-Perf.M.Sg
'Anjum let Saddaf make a necklace quickly.'

Furthermore, the case marking on the arguments of infinitives can be both nominal and verbal in nature. In most of the examples involving infinitives presented so far, the arguments of the infinitive were marked with verbal (nominative, accusative) rather than nominal (genitive) case. This suggests that the infinitive must be analyzed as a verbal noun. Also recall from the previous section that arguments can freely move out of infinitive constituents. Infinitives differ from ordinary NPs in this respect, since elements contained within an ordinary NP do not have the same freedom of movement. The phenomenon of 'long distance' agreement (Davison 1991a, Mahajan 1990) provides further evidence for the verbal noun status of infinitives, and also leads to a better understanding of the structure of infinitive clauses.

3.4.3 "Long Distance" Agreement

The basic pattern of "long-distance" agreement is illustrated in (79). Here the object embedded in the infinitival complement (*gaaṛii* 'car' and *ṭãgaa* 'tonga' in (79a) and (79b) respectively) determines the agreement morphology on the infinitive predicate (*calaa-nii* vs.

calaa-naa 'to drive'), and is ultimately responsible for the agreement morphology on the matrix verb. Because agreement in Urdu/Hindi is clause-bounded in all other constructions, the pattern in (79) is exceptional and in need of an explanation.

(79) a. naadyaa=ko [**gaaṛii calaa-nii**] **aa-tii** hai
Nadya.F=Dat car.F=Nom drive-Inf.F.Sg come-Impf.F.Sg be.Pres.3.Sg
'Nadya knows how to drive a car.'

b. naadyaa=ko [**ṭ̃agaa calaa-naa**] **aa-taa** hai
Nadya.F=Dat tonga.M=Nom drive-Inf.M.Sg come-Impf.M.Sg be.Pres.3.Sg
'Nadya knows how to drive a tonga.'

Mahajan (1989, 1990) and Davison (1985, 1988, 1990, 1991a, 1991b) represent two differing lines of research which have been concerned with long distance agreement in Hindi. Both approaches attempt to bring long distance agreement in line with agreement in simple clauses. Neither approach, however, can explain quite the range of data presented here. If the infinitive constituents in (79) are analyzed as NPs, as suggested in this chapter, and if agreement is taken to be with nominative argument NPs, long distance agreement follows from the same principles as agreement in simple clauses.

The percolation of agreement observed in (79) is optional. Previous approaches to agreement in Urdu/Hindi have not been able to satisfactorily explain this optionality, or why infinitives are able to take either verbal or nominal arguments. I present evidence that there are three different kinds of infinitive NPs. In one kind the embedded object is incorporated by the infinitive, and therefore does not trigger agreement. In examples like (79), on the other hand, the lower object is not incorporated and does trigger "long distance" agreement. The third type of infinitive is unlike the other two in that it takes genitive (nominal rather than verbal) arguments and does not give rise to long distance agreement. I argue that this infinitive displays only nominal properties because it enters the syntax as a noun. In contrast, the other two infinitives are nominalized in the syntax, and as such display both verbal and nominal properties.

3.4.3.1 Previous Analyses

Most of the data presented in this section have been noted previously either by Mahajan (1989, 1990) or Davison (1985, 1988, 1990, 1991a, 1991b), or both. As a language with object agreement, Urdu/Hindi poses problems for theories of syntax which assume that agreement is a characteristic of subjects. The challenge, then, is not only to bring long

distance in line with cases of local agreement, but also to successfully formulate an analysis of local agreement.

Mahajan (1990) does this by making use of both a Spec of Agr_s and a Spec of Agr_o within AgrP, and by formulating the basic pattern of agreement for simple clauses as follows. As perfect participles and psych verbs have ergative and dative (overtly case marked) subjects, and it is primarily in these constructions that agreement with a nominative object is possible, he takes perfect participles and psych verbs to be non Case assigning verbs. Objects of non Case assigning verbs must move to Spec of Agr_o in order to receive structural Case. In Spec of Agr_o, the objects trigger verb agreement. Mahajan thus employs AgrP and movement for Case reasons to arrive at the essential generalization that verbs can only agree with NPs not overtly marked with a case clitic. Infinitives are taken to be somewhat like perfectives and psych predicates in that they are optionally non Case assigning.

There are several disadvantages inherent to this approach. For example, the assumption that psych predicates and perfective participles are non Case assigning, and infinitives only optionally non Case assigning, is stipulative (also see Butt (1993b) for arguments against the postulation of AGR for an analysis of agreement and object specificity in Hindi/Urdu). Mahajan's analysis, however, does have the advantage that agreement is taken to be a purely local phenomenon.

Davison (1991a) analyzes infinitives as CP arguments of the matrix verb. Agreement is taken to be a case of ϕ feature percolation. Arguments carry ϕ feature specifications, which are percolated upward along with a theta-grid (Speas 1990). Case clitics block ϕ features. If more than one ϕ feature is percolated upward, only the leftmost one results in agreement on the verb. Because infinitive constituents satisfy an argument position in the theta-grid of the matrix verb, the ϕ features get percolated upwards in these CPs. Percolation of ϕ features cannot take place out of non-argument CPs.

Although Davison's approach accounts for a wider range of data than Mahajan's theory of agreement, there are some issues which do not receive a satisfactory explanation. For example, there are finite CPs, such as "that" clauses, which parallel the function of some infinitives. If the infinitive CP can be analyzed as an argument of the matrix verb, finite CP complements could be analyzed as satisfying an argument position as well. However, "long distance" agreement never takes place out of finite CPs. As with Mahajan's analysis, the primary advantage of Davison's approach is that the long distance percolation of features does not differ from local agreement. Furthermore, the possible argument status of infinitive

constituents is recognized.

In the next few sections I present the facts previously noted about infinitives and briefly show how the data can be accounted for under my basic approach to infinitives and agreement.

3.4.3.2 Blocking of Agreement by Case

Long distance agreement is blocked when there is a case clitic on the infinitive, as in (80). The embedded feminine object *gaaṛii* 'car' is nominative, but does not trigger agreement on either the infinitive or the matrix verb.

(80) anjum=ne saddaf=ko [**gaaṛii** calaa-ne]-**ko**
Anjum.F=Erg Saddaf.F=Dat car.F=Nom drive-Inf.Obl-**Acc**
kah-aa
say-Perf.M.Sg
'Anjum told Saddaf to drive a car.'

Under my approach, the pattern of agreement in (80) is expected. The matrix verb has default masculine singular *-aa* morphology because there is no nominative argument NP in the matrix clause it can agree with: the subject and indirect object NPs as well as the infinitive constituent *gaaṛii calaa-ne=ko* 'to drive a car' are all non-nominative. Additionally, the infinitive predicate *calaa-ne* 'to drive' cannot show agreement with its nominative object because the presence of the case clitic *ko* induces the oblique inflection *-ne*.

3.4.3.3 Matrix plus Embedded Agreeement

The matrix verb and the infinitive predicate can agree with different arguments. In (81) the infinitive predicate agrees with its nominative object *roṭii* 'bread' and the matrix verb agrees with the subject *Ram*. The grammaticality of (81) is again expected under my approach. The infinitive predicate agrees with its only nominative argument, while the matrix verb agrees with its highest nominative argument, the subject.

(81) raam [roṭii kʰaa-nii] caah-taa tʰaa
Ram.M=Nom bread.F=Nom eat-Inf.F.Sg want-Impf.M.Sg be.Past.3.Sg
'Ram wanted to eat the bread.'

For Mahajan (1990) the possibility of simultaneous matrix and embedded agreement is problematic. As the trace of the matrix nominative subject in these sentences already

occupies a position in AgrP, the lower object cannot move to a matrix Spec of Agr position and trigger agreement with the infinitive on the way.[21] Therefore, the sentence in (81) is predicted to be ungrammatical under Mahajan's approach.[22]

3.4.3.4 Genitive Arguments

The infinitive in (82a) takes nominal arguments. The matrix verb *lag-taa* 'seem' here agrees with the infinitive, but the infinitive *kaṛak-naa* 'crackling' does not show agreement with its feminine argument *bijlii* 'lightning'. Rather, the genitive clitic *-kaa*, which behaves like an adjective in Urdu in that it always agrees with the head noun, agrees with the infinitive. Example (82b) contrasts minimally with (82a) (both these examples are adapted from Davison (1990)). Here the embedded argument is nominative and the effect of 'long distance' agreement is observed.

(82) a. adnaan=ko [**bijlii=kaa** **kaṛaknaa**] acc^{h}aa nahĩĩ
Adnan.M=Dat lightning.F=Gen.M crackle.Inf.M good.M not
lag-taa
seem-Impf.M.Sg
'Adnan does not like the crackling of lightning.'

b. adnaan=ko [**bijlii** **kaṛak-nii**] acc^{h}ii nahĩĩ
Adnan.M=Dat lightning.F=Nom crackle-Inf.F good.F not
lag-tii
seem-Impf.F.Sg
'Adnan does not like lightning crackling.'

I propose that the crucial difference between (82a) and (82b) is that in (82a) the infinitive is a "true" noun. It is formed in the lexicon and enters the syntax as a masculine noun which takes a genitive argument.

[21] Spec of Agr_s and and Spec of Agr_o cannot be filled at the same time because it is impossible to have both subject and object agreement simultaneously in a clause. Mahajan postulates that Spec of Agr_s may simply be missing in these cases, as specifiers can be optional (Fukui and Speas 1986).

[22] Mahajan proposes an explanation by which imperfective participles govern the lower Spec of Agr_o. When the infinitive does not assign Case to the lower object, the object can move to the lower Spec of Agr_o to receive Case. It therefore shows agreement with the infinitive, but not with the matrix verb. This explanation does not follow independently from any further data in Hindi.

3.4.3.5 Optionality of Agreement

A more puzzling phenomenon is illustrated in (83) and (84). The sentence in (83) is an instance of "long distance" agreement. However, as (84) shows, the agreement between the infinitive predicate and its nominative object *gaaṛii* 'car' is optional.

(83) naadyaa=ko [**gaaṛii** **calaa-nii**] **aa-tii** hai
Nadya.F=Dat car.F=Nom drive-Inf.F.Sg come-Impf.F.Sg is
'Nadya knows how to drive a car.'

(84) naadyaa=ko [gaaṛii **calaa-naa**] **aa-taa** hai
Nadya.F=Dat car.F=Nom drive-Inf.M.Sg come-Impf.M.Sg is
'Nadya knows how to drive a car.'

Both Mahajan (1989) and Davison (1988) (attributed to Hook (1979:29–30)) observe that in (83) the object *gaaṛii* 'car' is more specific than the object *gaaṛii* 'car' in (84). I argue that this difference in specificity is directly attributable to the fact that (84) represents an incorporated structure while (83) does not. Example (84) denotes abstract 'car-driving', while (83) refers to 'driving a car'.

3.4.3.6 Incorporating Infinitives

Two differing types of data lead to the conclusion that the infinitive constituent *gaaṛii calaa-ne* 'car-driving' in (84) must be analyzed as a form of incorporation: the object cannot be scrambled away from the infinitive, and it cannot be modified (also see T. Mohanan (1992) for a discussion of such constructions within her examination of Noun Incorporation in Hindi).

The sentences in (85) and (86) illustrate the difference in scrambling possibilities between the agreeing and nonagreeing infinitives in (83) and (84). In the case of long distance agreement, it is possible to scramble either the entire infinitive constituent *gaaṛii calaa-nii* 'driving a car', or just the embedded object *gaaṛii* 'car' to the front of the sentence. This is illustrated in (85a) and (85b).

(85) a. [**gaaṛii** **calaa-nii**] [naadyaa=ko] [aa-tii hai]
car.F=Nom drive-Inf.F.Sg Nadya.F=Dat come-Impf.F.Sg be.Pres.3.Sg
'Nadya knows how to drive a car.'

b. [**gaaṛii**] [naadyaa=ko] [**calaa-nii** aa-tii hai]
car.F=Nom Nadya.F=Dat drive-Inf.F.Sg come-Impf.F.Sg be.Pres.3.Sg
'Nadya knows how to drive a car.'

When the infinitive does not agree with its object, the scrambling possibilities differ. Although it is still possible to scramble the entire infinitive constituent to the front of the sentence in (86a), example (86b) shows that it is not possible to scramble the embedded object by itself. This suggests that in the nonagreeing example in (86), the embedded object is incorporated.

(86) a. [**gaaṛii calaa-naa**] [naadyaa=ko] [aa-taa hai]
car.F=Nom drive-Inf.M.Sg Nadya.F=Dat come-Impf.M.Sg be.Pres.3.Sg.
'Nadya knows car-driving.'

b. *[**gaaṛii**] [naadyaa=ko] [**calaa-naa** aa-taa hai]
car.F=Nom Nadya.F=Dat drive-Inf.M.Sg come-Impf.M.Sg be.Pres.3.Sg
'Nadya knows car-driving.'

Examples (87) and (88) illustrate the modification facts. While it is possible to modify the embedded object with a modifier expressing specificity when it agrees with the infinitive, this is not possible when the object and the infinitive show no agreement. In (87) the infinitive agrees with its object. As (87a) and (87b) show, here modification with either a genitive NP or a determiner is possible.

(87) a. naadyaa=ko [adnaan=kii **gaaṛii calaa-nii**] **aa-tii**
Nadya.F=Dat Adnan.M=Gen.F.Sg car.F=Nom drive-Inf.F.Sg come-Impf.F.Sg
hai
be.Pres.3.Sg
'Nadya knows how to drive Adnan's car.'

b. naadyaa=ko [kaii **miṭʰ̃ayãã banaa-nĩĩ**] **aa-tĩĩ**
Nadya.F=Dat several sweets.F.Pl=Nom make-Inf.F.Pl come-Impf.F.Pl
hãĩ̃
be.Pres.3.Pl
'Nadya knows how to make several sweets.'

On the other hand, when the infinitive does not agree with its object in (88), it is not possible to modify that object.

(88) a. *naadyaa=ko [adnaan=kii **gaaṛii calaa-naa**] **aa-taa**
Nadya.F=Dat Adnan.M=Gen.F.Sg car.F=Nom drive-Inf.M.Sg come-Impf.M.Sg
hai
be.Pres.3.Sg
'Nadya knows how to drive Adnan's car.'

b. *naadyaa=ko [kaii **miṭʰayãã** **banaa-naa**] **aa-taa**
Nadya.F=Dat several sweets.F.Pl=Nom make-Inf.M.Pl come-Impf.M.Pl
hai
be.Pres.3.Sg
'Nadya knows how to make several sweets.'

This again indicates that the nominative object is incorporated by the infinitive predicate in (88).

3.4.3.7 Analysis

The analysis of infinitive agreement presented here has the advantages of previous approaches in that "long distance" agreement is reduced to a chain of local agreement relations. If infinitives are NPs, the possibility that they can function as arguments of a predicate follows immediately. Furthermore, if agreement is with nominative arguments, the "long distance" agreement facts can be accounted for in exactly the same manner as local agreement.

In LFG agreement is stated at the level of f-structure. Recall that this level contains the representation of the grammatical relations of a given clause and also encodes such information as tense, aspect, gender, number and case. The agreement facts for Urdu can easily be accounted for within LFG. In Urdu, a given expression can only be well-formed if the predicate agrees with a nominative argument. This is 'checked' at f-structure. If there is more than one nominative argument in an expression, the predicate must agree with the higher one. The notion of 'higher' is ultimately derived from a thematic hierarchy (Bresnan and Kanerva 1989) from which theta roles are mapped on to grammatical functions at f-structure. Thus, a given PRED at f-structure must agree with a nominative argument (SUBJ, OBJ).

The apparent optionality of agreement, repeated here in (89) and (90), follows from the fact that there are two infinitive constituents which differ structurally. The embedded object in (90) is incorporated, while the embedded object in (89) is not.

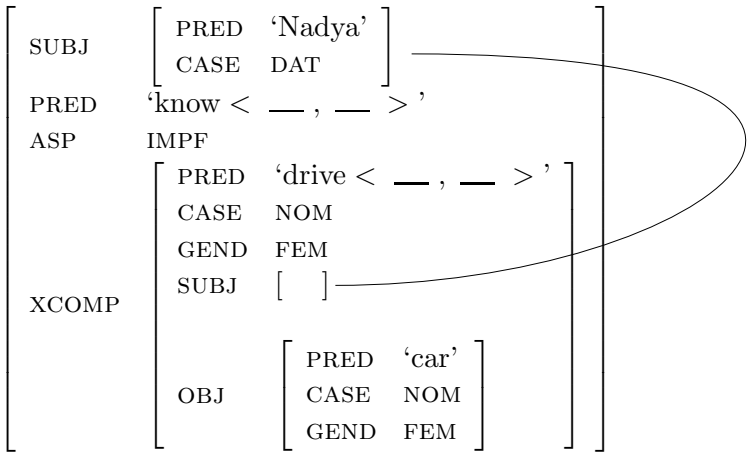
(89) naadyaa=ko [**gaaṛii** **calaa-nii**] **aa-tii** hai
Nadya.F=Dat car.F=Nom drive-Inf.F.Sg come-Impf.F.Sg be.Pres.3.Sg
'Nadya knows how to drive a car.'

(90) naadyaa=ko [gaaṛii **calaa-naa**] **aa-taa** hai
Nadya.F=Dat car.F=Nom drive-Inf.M.Sg come-Impf.M.Sg be.Pres.3.Sg
'Nadya knows car-driving.'

In her examination of Hindi Noun Incorporation, T. Mohanan (1992) discusses several differing constructions in Hindi, among them examples similar to (90), and argues for a lexical analysis of Noun Incorporation. However, it is possible for some material, such as emphatic particles, to intervene between the *gaaṛii* 'car' and the infinitive *calaa-naa* 'drive' in (90). I therefore represent the incorporation at the level of c-structure, as well as at f-structure. However, I believe that ultimately a deeper semantic analysis, in terms of the notion of *predicate-modification* (Ramchand 1993, De Hoop 1992) will prove to be the key to a more complete understanding of the construction. I do not pursue this analysis here, as it would lead me too far afield.

The f-structure corresponding to the non-incorporated infinitive in (89) is shown in (91). The infinitive constituent *gaaṛii calaa-nii* 'to drive a car' is represented as an XCOMP (complement) of the matrix predicate *aa-tii* 'know'. The XCOMP PRED 'drive' must agree with the embedded XCOMP OBJ 'car' in (91) because the object is nominative and an argument of the XCOMP PRED. In turn, the matrix PRED 'know' agrees with its nominative XCOMP argument.

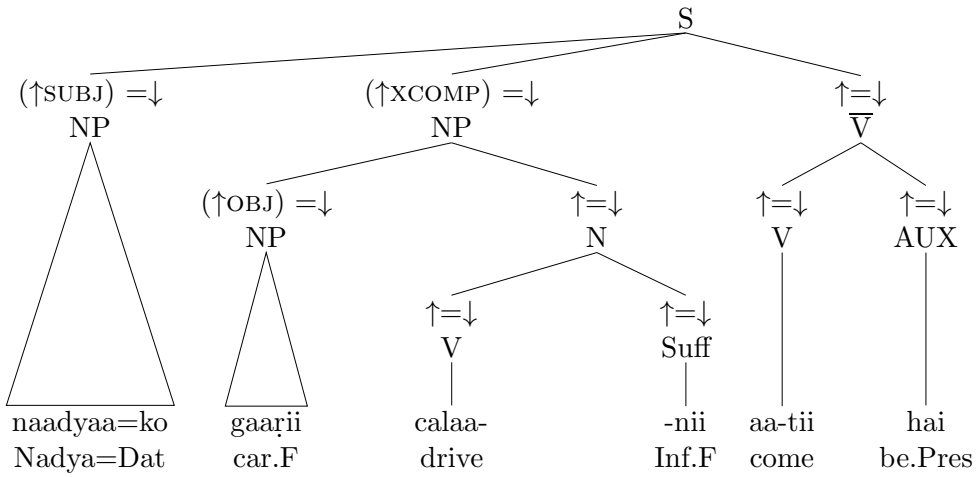
(91)



The c-structure corresponding to the f-structure in (91) is shown in (92). Within LFG the concordance of agreement shown in (91) must actually be derived from *constraint* equations listed in the lexical entries of the agreement morphology, *-ii* in (92). These equations articulate the constraint that the highest nominative argument in the clause must have the same gender and number as the agreement morpheme on the verb. If this constraint is

not met, then the f-structure corresponding to the c-structure is not well-formed, and the sentence is ruled out as ungrammatical.

(92)



The infinitive in (92) is represented as a verb *calaa* 'drive', which has been nominalized by the morpheme *-nii*. The infinitive predicate *calaa-nii* 'to drive' is thus a verbal noun. Although it heads a constituent which has nominal properties, the arguments of the infinitive predicate can appear in direct (nominative, accusative) case, rather than being marked by the genitive, as is generally the case for arguments of nominals.

The f-structure for the incorporated infinitive is shown in (93). The crucial difference between this f-structure and the f-structure for the non-incorporated infinitive is that here the infinitive predicate *calaa-naa* 'to drive' does not have an object argument. Rather, the object *car*, forms an XCOMP PRED with the infinitive 'drive'.

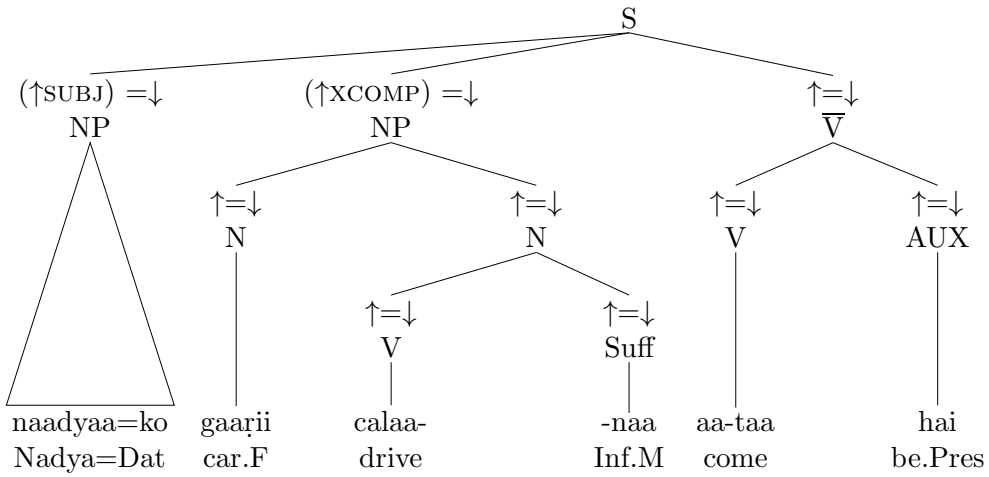
(93)

$$\begin{bmatrix} \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'Nadya'} \\ \text{CASE} & \text{DAT} \end{bmatrix} \\ \text{PRED} & \text{'know} < \_\_ , \_\_ > \text{'} \\ \text{ASP} & \text{IMPF} \\ \text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{'car-drive} < \_\_ > \text{'} \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{MASC} \\ \text{SUBJ} & [\quad] \end{bmatrix} \end{bmatrix}$$

Since the XCOMP SUBJ in (93) is controlled by the matrix subject, and is therefore "empty", there is no nominative argument that the XCOMP PRED *car-drive* can agree with. The

agreement feature of the XCOMP must therefore be the default masculine. The matrix PRED 'know', however, does agree with the nominative masculine XCOMP. The impossibility of "long distance" agreement in this construction is thus directly attributable to incorporation. The corresponding c-structure, which expresses the incorporation is shown in (94). The incorporation in (94) is indicated by the fact that the *gaaṛii* 'car' in (94) is not an NP, and is not annotated as being the object of the infinitive *calaa-naa* 'drive'. Incorporation in (94) is thus treated as a kind of compounding, which results in the 'car-driving' PRED in (93). As mentioned previously, I believe that a deeper semantic analysis of this construction in terms of predicate-modification (Ramchand 1993) is needed, but, pending further research, (93) and (94) accurately characterize the structure indicated by the data.

(94)



At this point, only the infinitive in (95), which takes genitive arguments, remains to be accounted for. As the f-structure in (97) and the rough sketch of constituency in (96) illustrate, this infinitive is not nominalized in the syntax. It is a noun which is formed in the lexicon.

(95) adnaan=ko [**bijlii=kaa** **kaṛak-naa**] acc^{h}aa nahĩĩ
Adnan.M=Dat lightning.F=Gen.M crackle-Inf.M good.M not
lag-taa
seem-Impf.M.Sg
'Adnan does not like the crackling of lightning.'

(96) [$_{NP}$ [$_{NP}$ bijlii-kaa] [$_{N}$ kaṛaaknaa]]

(97)

$$\begin{bmatrix} \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'Adnan'} \\ \text{CASE} & \text{DAT} \end{bmatrix} \\ \text{PRED} & \text{'like} < \underline{\quad}, \underline{\quad} > \text{'} \\ \text{ASP} & \text{IMPF} \\ \text{NEG} & + \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'crackling} < \underline{\quad} > \text{'} \\ \text{CASE} & \text{NOM} \\ \text{GEND} & \text{MASC} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'lightning'} \\ \text{CASE} & \text{GEN} \\ \text{GEND} & \text{MASC} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The genitive clitic *kaa* agrees with the masculine head noun *kaṛaknaa* 'crackling'. Since the infinitive constituent is an argument of the matrix verb *lag-taa* 'seem', the matrix verb must agree with the infintive constituent.

The data from "long distance" agreement, which I have here reanalyzed as a case of successive local agreement, in combination with the infinitive data presented previously, shows not only that infinitives must be treated as NPs, but several different types of infinitive NPs must also be distinguished. This section has identified and provided an account for three differing infinitives: 1) a verbal noun formed in the syntax; 2) a verbal noun similarly formed in the syntax, but which incorporates its argument; 3) a verb which is nominalized in the lexicon and enters the syntax as a noun.

3.5 Conclusion

This chapter has examined the structure of the Urdu permissive in some detail. In particular, I have demonstrated that it must be analyzed as a complex predicate, but that its status as a complex predicate cannot be represented straightforwardly at phrase structure. Evidence for this comes from a comparison with the instructive, which superficially resembles the permissive, but cannot be analyzed as a complex predicate. It was further shown that the infinitive constituents of both the permissive and the instructive are NPs, and not CPs or VPs. These infinitival NPs are headed by verbal nouns, which are either nominalized in the syntax, or formed in the lexicon. Finally, an examination of "long distance" agreement led to a detailed analysis of the internal structure of the differing types of infinitive NPs

that can be identified. The data and analyses presented in this chapter provide not only the motivation, but also a crucial backdrop, for an analysis of complex predicate formation in terms of argument structure. This analysis is formulated in Chapters 5 and 6, and further motivated by the examination of another type of Urdu complex predicate, the Aspectual complex predicate, in the next chapter.

Chapter 4

Aspectual Complex Predicates

4.1 Introduction

This chapter examines a class of constructions I refer to as *Aspectual* complex predicates. These V-V complex predicates do not exhibit the structural mismatch in evidence for the Urdu permissive, but they display intriguing semantically based restrictions on complex predicate formation and case assignment. I argue that a systematic account of these alternations can be achieved most successfully through the postulation of an argument structure which contains more fine-grained semantic information than thematic role labels. Recall that the permissive complex predicate in contrast to the complement instructive construction, argued for a process of complex predicate formation at argument structure. The Aspectual complex predicates in this chapter demonstrate that, not only must complex predicate formation be expressed at argument structure, but the type of information needed at argument structure pertinent for linking, case-marking, and the syntactic well-formedness of an expression must also be more fine-grained. After laying out the evidence from Aspectual complex predicates in this chapter, I go on to formulate a level of elaborated argument structure based on Jackendoff's (1990) theory of Conceptual Semantics in the next chapter.

In present day Urdu, Aspectual complex predicates are preferred to simple predicates. Native speakers will insist that the action seems *incomplete* or unsituated when only a simple verb is used. This intuition is in accord with the existing literature, which characterizes the second verb in the predicate as an aspectual marker. According to Masica (1976:143), this second verb is used to contribute "completion, suddenness, directionality, benefaction, intensity, violence, stubbornness, reluctance, regret, forethought, thoroughness, etc." Hook

(1991) defines a complex predicate as follows:

> With the exception of some dialect of Shina every contemporary Indo-Aryan language or dialect has a set of auxiliary (or "vector") verbs homophonous with members of its inventory of basic lexical verbs. As full lexical verbs these express a change in location or posture, or an action that entails such a change: GO, GIVE, TAKE, THROW, LET GO, GET UP, COME, STRIKE, SIT, FALL, etc. A *compound verb* (CV) comprises the finite form of one of these following a non-finite or stem form of a main or primary verb.

An example of such a complex predicate is given in (1).

(1) anjum=ne xat **lik^{h} li-yaa**
Anjum.F=Erg letter.M=Nom **write take-Perf.M.Sg**
'Anjum wrote a letter (completely).'

Hook, like others who have written on South Asian languages, uses the term *auxiliary* to describe the finite component of compound verbs because he assumes that it only contributes aspectual information, and does not contain any information pertinent to the syntax. That is, he assumes that the second verb in the compound has no bearing on issues like case-marking and argument structure.

I show that constructions such as the one in (1) must be analyzed as complex predicates, rather than as cases in which the second verb, the *light verb*[1] is merely functioning as an aspectual auxiliary. The light verb *li-yaa* 'took' in (1) does contribute a sense of completion to the action, but I show that it also contributes to the argument structure of the entire, complex, predicate.

The next few sections first introduce the various types of light verbs commonly found in the Lahori dialect, then go on to examine the grammatical function and phrase structure of Aspectual complex predicates. In contrast to the permissive complex predicate, there is no problematic mismatch between the c-structure and the f-structure. Consequently, these sections are rather short and expository. The truly interesting argument structure properties of the Aspectual complex predicates are presented once the pertinent structural information, which allows a comparison with the permissive in the previous chapter, has been discussed.

[1]This terminology is used by Jespersen (1954) and has, for example, been more recently used for Japanese complex predicates by Grimshaw and Mester (1988) and Matsumoto (1992), and for Hindi by T. Mohanan (1990).

As mentioned, Aspectual complex predicates provide evidence for an elaborated level of argument structure. I show that the light verbs in Aspectual complex predicates not only contribute aspectual information in terms of inception and completion, but also express whether or not a given action was performed volitionally. These semantic contributions are reflected in the syntax through case-marking on the subject. I discuss the aspectual and volitionality components of the light verbs in some detail and then briefly sketch an argument structure approach to complex predicates. The details of the analysis in terms of the necessary elaborated argument structure and its relationship to the phrase (constituent) and grammatical function (functional) structure of a given expression are left for Chapters 5 and 6.

4.2 Types of Light Verbs

Hook (1974) provides an exhaustive list of verbs which can function as light verbs in Hindi/Urdu. Of the various light verbs discussed in the literature, Hook finds about twenty-four in fairly common use, while the others are obscure. Table (2) below lists the light verbs most commonly used in the dialect of Urdu I examine. For each of the light verbs in Table (2) there is a corresponding "heavy" verb which is identical in form to the light verb. The meaning of the heavy verb is what is given as a gloss in (2).[2]

(2)

Common Light Verbs	
Based on (di)transitives (Ergative Subject)	Based on Intransitives (Nominative Subject)
le 'take'	aa 'come'
de 'give'	jaa 'go'
ḍaal 'put'	paṛ 'fall'
maar 'hit'	mar 'die'
nikaal 'pry out'	nikal 'emerge'
	cuk 'finish'
	baiṭʰ 'sit'
	uṭʰ 'rise

[2]It is striking that complex predicates crosslinguistically seem to utilize the same sets of core light verbs (*do, take, come, go, give, hit*, etc.), and that these light verbs are generally based on main verbs still in active use in the language. A broad, crosslinguistic comparison of light verbs and the kinds of complex predicate constructions they occur in would clearly be of interest.

As can be seen, the light verbs in Table (2) fall into two basic classes. Verbs like *le* 'take', which are based on transitive or ditransitive main verbs, require that the subject of the complex predicate must be ergative in the perfective. Light verbs like *aa* 'come' are based on intransitives and require that the subject of the complex predicate be nominative. The second half of this chapter demonstrates how the case marking on the subject of a complex predicate is a direct consequence of semantic information, such as conscious choice (volitionality) and inception/completion, which is contributed by the light verb.

Of the light verbs in Table (2), I concentrate on the light verbs *le* 'take', *de* 'give', *ḍaal* 'put', *paṛ* 'fall', *uṭʰ* 'rise' and *jaa* 'go'. These light verbs are broadly representative. The verbs *le, de*, and *ḍaal* are the kind which require ergative subjects in the perfective, signal conscious choice and the completion of an action. The light verb *paṛ* 'fall', on the other hand, denotes the inception of a not consciously controlled action and requires a nominative subject in the perfective. The syntactic and semantic properties of these light verbs are presented in the next few sections.

4.3 Grammatical Function Structure

This section establishes that Aspectual complex predicates are exactly like simple predicates in terms of f-structure phenomena. That is, Aspectual complex predicates and simple predicates do not differ with respect to verb agreement, anaphora, and control.

4.3.1 Agreement

As already discussed for the permissive in the previous chapter, a finite verb in Urdu always agrees with one of its nominative arguments. If the subject is non-nominative, the verb agrees with the nominative object. If both the subject and object are non=Nominative, the verb has the default masculine singular agreement morphology *-aa*. Furthermore, agreement is clause-bounded.

The sentences in (3) illustrate the agreement pattern for a simple predicate in which the subject is ergative. In (3a) the verb *likʰ-aa* 'wrote' agrees with the masculine nominative object *xat* 'letter'. Similarly, in (3b) the verb *likʰ-ii* 'wrote' agrees with the feminine nominative object *cit̪ṭʰii* 'note'.

(3) a. anjum-ne **xat** **lik^{h}-aa**
Anjum.F=Erg letter.M=Nom write-Perf.M.Sg
'Anjum wrote a letter.'

b. anjum-ne **ciṭṭ^{h}ii** **lik^{h}-ii**
Anjum.F=Erg note.F=Nom write-Perf.F.Sg
'Anjum wrote a note.'

The above pattern of agreement for a simple predicate is mirrored by the complex predicate in (4). In (4a) the finite part of the complex predicate, the light verb *liy-aa* 'took' agrees with the masculine nominative *xat* 'letter', and in (4b) the finite light verb *l-ii* 'took' agrees with feminine *ciṭṭ^{h}ii* 'note'.

(4) a. anjum=ne **xat** **lik^{h}** **li-yaa**
Anjum.F=Erg letter.M=Nom write take-Perf.M.Sg
'Anjum wrote a letter (completely).'

b. anjum=ne **ciṭṭ^{h}ii** **lik^{h}** **l-ii**
Anjum.F=Erg note.F=Nom write take-Perf.F.Sg
'Anjum wrote a note (completely).'

The fact that the light verb in (4) is able to agree with the nominative object shows that Aspectual complex predicates cannot consist of a matrix light verb which subcategorizes for an embedded complement. The bare stem form used for main verbs in Aspectual complex predicates also appears in other constructions in which the verb in the stem form does head an embedded clause. This is illustrated in (5) with the participial adverbial encountered in the previous chapter.

(5) anjum=ne saddaf=ko [ciṭṭ^{h}ii **lik^{h}** **kar**] dek^{h}-aa
Anjum.F=Erg Saddaf.F=Acc note.M=Nom write having see-Perf.M.Sg
'Anjum saw Saddaf after having written a note.'

Recall that the matrix finite verb in constructions containing a participial adverbial will never agree with the embedded object (*ciṭṭ^{h}ii* 'note' in this case). As object agreement is possible in the sentences in (4), the form *lik^{h} l-ii* 'write took' must be analyzed as a complex predicate, while the verbal complex *lik^{h} kar dek^{h}-aa* 'having written, saw' cannot be a complex predicate.

The examples in (6) and (7) show that intransitive complex predicates also pattern the same as intransitive simple predicates with respect to agreement. The sentences in (6) contain the simple predicate *gir* 'fall'. In (6a), the verb agrees with the nominative masculine subject *Adnan*, and in (6b), it agrees with its nominative feminine subject *Nadya*.

(6) a. adnaan **gir-aa**
Adnan.M=Nom fall-Perf.M.Sg
'Adnan fell.'

b. naadyaa **gir-ii**
Nadya.F=Nom fall-Perf.F.Sg
'Nadya fell.'

The sentences in (7) illustrate the corresponding complex predicate pattern. The finite light verb agrees with the nominative subject: in (7a) *ga-yaa* 'went' agrees with the masculine *Adnan*, while in (7b) it agrees with the feminine *Nadya*.

(7) a. adnaan **gir gay-aa**
Adnan.M=Nom fall go-Perf.M.Sg
'Adnan fell (completely).'

b. naadyaa **gir gay-ii**
Nadya.F=Nom fall go-Perf.F.Sg
'Nadya fell (completely).'

Aspectual complex predicates thus pattern like simple predicates with respect to verb agreement. Data from anaphora and control confirm this pattern.

4.3.2 Anaphora

The generalization for Urdu *apn-ii* 'self' and *us=kii* 'her' was presented in the previous chapter for the permissive. Recall that *apn-ii* 'self' takes a subject as an antecedent while *us=kii* 'her' may corefer with a non-subject. This section illustrates very briefly that with regard to anaphora, Aspectual complex predicates pattern exactly like simple predicates. The sentences in (8) illustrate the pattern for the simple ditransitive predicate *de* 'give'.

(8) a. mumtaaz=ne$_i$ naadyaa=ko$_j$ apn-ii$_{i,*j}$ gaaṛii d-ii
Mumtaz.F=Erg Nadya.F-Dat self-F.Sg car.F=Nom give-Perf.F.Sg
'Mumtaz gave Nadya self's (Mumtaz's) car.'

b. mumtaaz=ne$_i$ naadyaa=ko$_j$ us=kii$_{*i,j}$ gaaṛii d-ii
Mumtaz.F=Erg Nadya.F-Dat pron=Gen.F.Sg car.F=Nom give-Perf.F.Sg
'Mumtaz gave Nadya her (Nadya's) car.'

In (8a), the reflexive *ap-nii* 'self' can only take the subject *Mumtaz* as an antecedent. The *us=kii* in (8b), on the other hand, can only refer to the non-subject *Nadya*. Exactly the same pattern is found for the Aspectual complex predicate *de d-ii* 'gave (completely)' in (9).

(9) a. mumtaaz=ne$_i$ naadyaa=ko$_j$ apn-ii$_{i,*j}$ gaaṛii de d-ii
Mumtaz.F=Erg Nadya.F=Dat self-F.Sg car.F=Nom give give-Perf.F.Sg
'Mumtaz gave Nadya self's (Mumtaz's) car (completely).'

b. mumtaaz=ne$_i$ naadyaa=ko$_j$ us=kii$_{*i,j}$ gaaṛii de d-ii
Mumtaz.F=Erg Nadya.F=Dat pron=Gen.F.Sg car.F=Nom give give-Perf.F.Sg
'Mumtaz gave Nadya her (Nadya's) car (completely).'

With regard to anaphora, then, the Aspectual complex predicates again pattern like simple predicates.

4.3.3 Control

Recall that the participial adverbial headed by *kar* 'having' must always be controlled by the subject. This is illustrated in (10) for the simple ditransitive predicate *de* 'give'. Here only the subject *Mumtaz*, and not the non-subject *Nadya*, can control the participial adverbial *gʰar jaa kar* 'having gone home'. Example (11) illustrates the fact for a corresponding complex predicate.

(10) mumtaaz=ne$_i$ naadyaa=ko$_j$ [____$_{i,*j}$ gʰar jaa kar] kitaab
Mumtaz.F=Erg Nadya.F=Dat house.M=Loc go having book.F=Nom
d-ii
give-Perf.F.Sg
'Mumtaz, having gone home, gave Nadya a book.'

(11) mumtaaz=ne$_i$ naadyaa=ko$_j$ [____$_{i,*j}$ gʰar jaa kar] kitaab
Mumtaz.F=Erg Nadya.F=Dat house.M=Loc go having book.F=Nom
de
give
d-ii
give-Perf.F.Sg
'Mumtaz, having gone home, gave Nadya a book (completely).'

The Aspectual complex predicate in (11) is thus again exactly parallel to the simple predicate in (10).

4.3.4 Summary

This section has illustrated that Aspectual complex predicates pattern exactly like simple predicates with regard to verb agreement, anaphora, and control. Recall that this was true

for the permissive complex predicate in the previous chapter as well. The Aspectual complex predicates thus meet one of the criteria formulated for complex predicates: they must be analyzed as having a simple f-structure. That is, they can be analyzed as containing only one nucleus PRED, and therefore only one subject, one object, etc.

The next section explores the phrase structure properties of Aspectual complex predicates. Unlike the permissive, where the actual phrase structure realization did not necessarily reflect its status as a complex predicate, Aspectual complex predicates do form a tight unit at phrase structure.

4.4 Phrase Structure

The light verb in Aspectual complex predicates contributes to the syntax and semantics of the complex predicate so subtly, it is not immediately obvious to the native speaker that the two verbs are each lexical items in their own right. Because the combination of a main verb with a light verb forms a very tight unit intuitively, Aspectual complex predicates are often analyzed as being a single lexical item (Glassman 1977), or as a main verb with an auxiliary (Hook 1991). This section shows that the two verbs are each lexical items in their own right and combine into a complex predicate in the syntax.

4.4.1 Syntactic Composition

The two verbs in an Aspectual complex predicate do form a tight constituent at phrase structure as they cannot be scrambled away from one another, a modifier cannot appear between the two verbs, and the coordination facts are exactly parallel to those of simple predicates containing auxiliary markers. In fact, with the exception of negation, the phrase structure properties of the Aspectual complex predicates are exactly parallel to those of simple predicates with auxiliaries.[3] That the light verbs in Aspectual complex predicates can nevertheless not be analyzed as simple auxiliaries is shown in the section discussing the semantic constraints on complex predicate formation. This section shows that neither the Urdu auxiliaries nor the light verb in an Aspectual complex predicate can be analyzed as forming a lexical item with the main verb of a sentence.

[3]The Aspectual complex predicates are quite unlike other simple and complex predicates in that it is impossible to negate an Aspectual complex predicate. I present a more detailed discussion of this in a later section.

The emphatic particles *hii* and *bʰii* can appear after any lexical item in an expression.[4] The particles cannot ever appear in the middle of a lexical item. This is illustrated in (12). The emphatic particle *bʰii* can appear in various positions in the sentence, but, as (12c) shows, it cannot appear in the middle of a morphologically complex lexical item such as *banaa-tii* 'make-Impf.F.Sg'.

(12) a. anjum haar **bʰii** [banaa-tii hai]
Anjum.F=Nom necklace.M=Nom too make-Impf.F.Sg be.Pres.3.Sg
'Anjum makes necklaces also.'

b. anjum haar [banaa-tii **bʰii** hai]
Anjum.F=Nom necklace.M=Nom make-Impf.F.Sg too be.Pres.3.Sg
'Anjum also makes necklaces.'

c. *anjum haar [banaa-**bʰii**-tii hai]
Anjum.F=Nom necklace.M=Nom make-too-Impf.F.Sg be.Pres.3.Sg
'Anjum also makes necklaces.'

The examples in (13) illustrate the same pattern for an Aspectual complex predicate. The emphatic particle *bʰii* can appear outside of the complex predicate, or inside of the verbal complex. It cannot, however, intrude in a lexical item. Both (13c) and (13d) are bad because the *bʰii* appears inside the morphologically complex lexical items *banaa* 'make' and *liyaa* 'took' respectively.[5]

(13) a. anjum=ne haar **bʰii** [banaa liy-aa]
Anjum.F=Erg necklace.M=Nom too make take-Perf.M.Sg
'Anjum made a necklace as well.'

b. anjum=ne haar [banaa **bʰii** liy-aa]
Anjum.F=Erg necklace.M=Nom make too take-Perf.M.Sg
'Anjum indeed made a necklace.'

c. *anjum=ne haar [ban-**bʰii**-aa liy-aa]
Anjum.F=Erg necklace.M=Nom be made-too-Caus take-Perf.M.Sg
'Anjum indeed made a necklace.'

[4]Glassman (1977) describes the differences between the two as follows: "The particle *bʰii* is 'inclusive' because it carries a meaning of 'as well' or 'too'. The particle *hii*, on the other hand, is 'exclusive' in the sense that conveys a meaning of 'just this'."

[5]Although I do not generally gloss it as such, *banaa* 'make' is actually a causativized form of *ban* 'be made'. One of the causative morphemes in Urdu is *-aa-* and there is a general, regular correspondence between intransitives such as *ban* 'be made' and transitives such *ban-aa* 'make' in Urdu. However, I do not gloss each transitive verb containing the morpheme *-aa-* as a causativized intransitive because the causativization in these cases seems to have become lexicalized.

d. *anjum=ne haar [banaa li-**b**^{h}**ii**-aa]
Anjum.F=Erg necklace.M=Nom make take-too-Perf.M.Sg
'Anjum indeed made a necklace.'

Each member of an Aspectual complex predicate must thus be considered to be a lexical item in its own right. Light verbs in this respect are parallel to auxiliaries.

4.4.2 Scrambling

The constituents of a $\overline{\text{V}}$ cannot be scrambled among one another in Hindi/Urdu. The sentence in (14) provides an example of a representative predicate.[6] It consists of a finite form of the verb *lik*^{h} 'write' and the auxiliary *ho* 'be'. The predicate can scramble only as a unit, as in (14b), but not in any other order.

(14) a. anjum=ne xat [**lik**^{h}**-aa** **hai**]
Anjum.F=Erg letter.M=Nom write-Perf.M.Sg be.Pres.3.Sg
'Anjum has written a letter.'

b. anjum=ne [**lik**^{h}**-aa** **hai**] xat
Anjum.F=Erg write-Perf.M.Sg be.Pres.3.Sg letter.M=Nom
'Anjum has written a letter.'

c. *anjum=ne **lik**^{h}**-aa** xat **hai**
Anjum.F=Erg write-Perf.M.Sg letter.M=Nom be.Pres.3.Sg
'Anjum has written a letter.'

d. *anjum=ne **hai** xat **lik**^{h}**-aa**
Anjum.F=Erg be.Pres.3.Sg letter.M=Nom write-Perf.M.Sg
'Anjum has written a letter.'

e. *anjum=ne xat **hai** **lik**^{h}**-aa**
Anjum.F=Erg letter.M=Nom be.Pres.3.Sg write-Perf.M.Sg
'Anjum wrote a letter.'

Similarly, the two verbs of the complex predicate cannot be scrambled away from one another. The examples in (15) illustrate the facts for a representative Aspectual complex predicate. The complex predicate can only be scrambled as a unit, as in (15b). Examples (15c–e) show that when the verbs in a complex predicate are scrambled away from one another, the result is ill-formed. The other possible word order combinations not shown below in which the two verbs are separated from one another are ill-formed as well.

[6] Kachru (1980:13–14) describes a typical predicate in Hindi as consisting of the main verb and aspect, tense and mood markers.

(15) a. anjum=ne xat [**lik**^{h} **liy-aa**]
Anjum.F=Erg letter.M=Nom write take-Perf.M.Sg
'Anjum wrote a letter.'

b. anjum=ne [**lik**^{h} **liy-aa**] xat
Anjum.F=Erg write take-Perf.M.Sg letter.M=Nom
'Anjum wrote a letter.'

c. *anjum=ne **liy-aa** xat **lik**^{h}
Anjum.F=Erg take-Perf.M.Sg letter.M=Nom write
'Anjum wrote a letter.'

d. *anjum=ne **lik**^{h} xat **liy-aa**
Anjum.F=Erg write letter.M=Nom take-Perf.M.Sg
'Anjum wrote a letter.'

e. *xat **lik**^{h} anjum=ne **liy-aa**
letter.M=Nom write Anjum.F=Erg take-Perf.M.Sg
'Anjum wrote a letter.'

The Aspectual complex predicates thus pattern like simple predicates with auxiliaries and unlike the permissive complex predicate with respect to scrambling.

4.4.3 Modification

In this section, I illustrate the adverbial modification pattern with regard to the adverb *kal* 'yesterday/tomorrow'. In (16) the adverb *kal* 'yesterday/tomorrow' appears in canonical position, to the left and outside of the $\overline{\text{V}}$ *lik*^{h}*-aa t*^{h}*aa* 'had written'.

(16) anjum=ne **kal** xat [lik^{h}-aa t^{h}aa]
Anjum.F=Erg yesterday letter.M=Nom write-Perf.M.Sg be.Past.3.Sg
'Anjum had written a letter yesterday.'

Although *kal* can appear in various positions, as (17) shows, it cannot intrude between the two members of the simple predicate. This is demonstrated in (18).

(17) a. anjum=ne xat **kal** [lik^{h}-aa t^{h}aa]
Anjum.F=Erg letter.M=Nom yesterday write-Perf.M.Sg be.Past.3.Sg
'Anjum had written a letter yesterday.'

b. **kal** anjum=ne xat [lik^{h}-aa t^{h}aa]
yesterday Anjum.F=Erg letter.M=Nom write-Perf.M.Sg be.Past.3.Sg
'Anjum had written a letter yesterday.'

(18) *anjum=ne xat [lik^{h}-aa **kal** t^{h}aa]
Anjum.F=Erg letter.M=Nom write-Perf.M.Sg yesterday be.Past.3.Sg
'Anjum had written a letter yesterday.'

The examples in (19) show that the adverb *kal* 'yesterday' cannot occur between the main and the light verb in an Aspectual complex predicate either.

(19) a. anjum=ne xat **kal** [lik^{h} liy-aa]
Anjum.F=Erg letter.M=Nom yesterday write take-Perf.M.Sg
'Anjum wrote the letter yesterday.'

b. *anjum=ne xat [lik^{h} **kal** liy-aa]
Anjum.F=Erg letter.M=Nom write yesterday take-Perf.M.Sg
'Anjum wrote the letter yesterday.'

With respect to adverbial modification, Aspectual complex predicates thus behave like a simple predicate with an auxiliary. The coordination data confirm this pattern.

4.4.4 Coordination

The sentence in (20a) illustrates an instance in which two predicates containing auxiliaries are coordinated. Examples (20b) and (20c) show that any attempts at coordination which separate one of the main verbs from its auxiliary are ill-formed.

(20) a. nadyaa haar [[banaa rah-ii t^{h}ii] aur
Nadya.F=Nom necklace.M=Nom make Stat-Perf.F.Sg be.Past.F.Sg and
[us=hii vakt pahin rah-ii
that=Emph time wear Stat-Perf.F.Sg
t^{h}ii]]
be.Past.F.Sg
'Nadya was making a necklace and wearing it at the same time.'

b. *nadyaa haar [[banaa aur pahin] rah-ii t^{h}ii]
Nadya.F=Nom necklace.M=Nom make and wear Stat-Perf.F.Sg be.Past.F.Sg
'Nadya was making a necklace and wearing it at the same time.'

c. *nadyaa [[haar banaa aur haar pahin]
Nadya.F=Nom necklace.M=Nom make and necklace.M=Nom wear
rah-ii
Stat-Perf.F.Sg
t^{h}ii]
be.Past.F.Sg
'Nadya was making a necklace and wearing it at the same time.'

Similarly, the two verbs in an Aspectual complex predicate cannot be coordinated separately. This is shown in (21). Example (21a) shows a successful coordination of two complex predicates. The examples in (21b) and (21c), on the other hand, are ill-formed because one of the main verbs has been separated from its light verb.

(21) a. nadyaa=ne gaaṛii [[xariid l-ii] aur [calaa l-ii]]
Nadya.F=Erg car.F=Nom buy take-Perf.F.Sg and drive take-Perf.F.Sg
'Nadya bought a car and drove it.'

b. *nadyaa=ne gaaṛii [[xariid aur calaa] l-ii]
Nadya.F=Erg car.F=Nom buy and drive take-Perf.F.Sg
'Nadya bought a car and drove it.'

c. *nadyaa=ne [[gaaṛii xariid] aur [gaaṛii calaa]] l-ii
Nadya.F=Erg car.F=Nom buy and car.F=Nom drive take-Perf.F.Sg
'Nadya bought a car and drove it.'

The data presented so far has shown that the Aspectual complex predicates behave like simple predicates in Urdu, and therefore should be analyzed as being contained under a $\overline{\text{V}}$, like simple predicates and their auxiliaries (see T. Mohanan 1990). The next section is concerned with the internal structure of Aspectual complex predicates.

4.4.5 Internal Structure of Aspectual Complex Predicates

Having established that Aspectual complex predicates cannot be analyzed as a single lexical item, the question of their internal syntactic structure must be resolved. At this point the two structures shown in (23) would seem to be equally possible and plausible.[7]

(22) a. $[\text{V} \quad \overline{\text{V}}]_{\overline{\text{V}}}$

b. $[\text{V} \quad \text{V}]_{\overline{\text{V}}}$

The next two sections examine the likelihood of these two possibilities and provide evidence that the internal structure of Aspectual complex predicates is best represented by (22b).

[7]The structure shown in (i) is a further logical possibility.

(i) $[\overline{\text{V}} \ \text{V}]_{\overline{\text{V}}}$

However, since the main verb in Aspectual complex predicates always appears in a stem form and never expands to include auxiliaries or modifiers, (i) is not a plausible structure.

4.4.5.1 Auxiliaries

The sentence in (23) provides an example of a simple predicate with the maximal number of possible auxiliaries. The *rah* 'stay' is an aspectual marker and signals a state.[8] The auxiliary *ho* 'be' acts as a tense marker and is optional in the present tense, but obligatory in the past tense and when following the auxiliary *rah*.

(23) naadyaa xat lik^{h} rah-ii hai
Nadya.F=Nom letter.M=Nom write stay-F.Sg be.Pres.3.Sg
'Nadya is writing a letter.'

Recall that the auxiliaries cannot be scrambled away from one another or the main verb. Following T. Mohanan, I take this to indicate that the auxiliaries are not contained under a separate phrasal constituent, but are lexical items, which form a $\overline{\text{V}}$ with the main verb. I therefore propose that a simple $\overline{\text{V}}$ is generated by the rule in (24), where 'Stat' allows for stative auxiliaries like *rah* 'stay' in (23).

(24) $\overline{\text{V}} \longrightarrow$V (Stat) (Aux)

As shown in (25), the light verb in an Aspectual complex predicate is not only always inflected, it is also always possible for an Aspectual complex predicate to take an auxiliary.[9]

(25) a. anjum=ne xat lik^{h} **li-yaa** **hai**
Anjum.F=Erg letter.M=Nom write take-Perf.M.Sg be.Pres.3.Sg
'Anjum has written a letter (completely).'

b. anjum gir **paṛ-ii** **hai**
Anjum.F fall fall-Perf.F.Sg be.Pres.3.Sg
'Anjum has fallen (completely, suddenly).

It is also possible to construct a sentence with the stative (progressive) marker *rah* 'stay' in a restricted set of contexts. As discussed in a later section, Aspectual complex predicates generally signal either the inception or completion of a given action. The combination of the stative marker with an Aspectual complex predicate is thus pragmatically odd in most contexts. However, sentences like (26) do occur.

[8]The *rah* 'stay' also functions as a main verb in the language. I treat the stative *rah* as an auxiliary rather than a light verb because it makes no semantic contribution other than aspect, i.e., it does not contribute to the argument structure in any way. Its only function is the expression of stativity, and there are no restrictions as to which main verbs it may combine with.

[9]Most of the sentences presented in this chapter actually sound more natural with an auxiliary.

(26) log aa jaa-te **rah-te** hãĩ
people.M.Pl come go-Impf.M.Pl stay-Impf.M.Pl be.Pres.3.Pl
'People keep on arriving.'

The fact that Aspectual complex predicates can occur with auxiliaries, as well as the stative marker *rah* 'stay', indicates that Aspectual complex predicates interact with auxiliaries just as simple predicates do.

4.4.5.2 N-V Complex Predicates

The structure of N-V complex predicates as analyzed by T. Mohanan (1990), and their interaction with Aspectual complex predicates, suggests that the light verb in Aspectual complex predicates must be analyzed as a simple V, and not a $\overline{\text{V}}$. The sentence in (27c) is an example of a combination of the V-V Aspectual complex predicate in (27a) and the N-V complex predicate in (27b).

(27) a. anjum **aa ga-yii**
Anjum.F=Nom come go-Perf.F.Sg
'Anjum has arrived.'

b. anjum=ko **ɣussaa a-yaa**
Anjum.F=Dat anger.M come-Perf.M.Sg
'Anjum became angry (lit. Anger came to Anjum).'

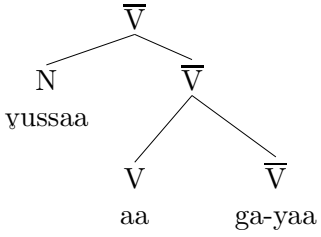
c. anjum=ko **ɣussaa aa ga-yaa**
Anjum.F=Dat anger.M come go-Perf.M.Sg
'Anjum became angry.'

T. Mohanan (1990) proposes the structure in (28) for N-V complex predicates.

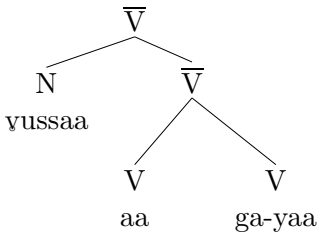
(28) $[\text{N } \overline{\text{V}}]_{\overline{\text{V}}}$

This structure, given the two possibilities for the internal structure of Aspectual complex predicates, allows two possible analyses of the N-V-V complex predicate in (28). The two possibilities are shown in (29) and (30).

(29)



(30)



Evidence for the structure in (30) and against the structure in (29) comes from topicalization. T. Mohanan (1990) shows that the light verb in a N-V complex predicate can be topicalized. This is illustrated by (31). The entire complex predicate can be topicalized as well, but not the noun by itself (T. Mohanan 1990:277). While I will not provide an analysis of topicalization in Urdu here, but instead refer the reader to Dwivedi (1993), what does emerge from T. Mohanan's work is that only phrasal constituents, and not lexical items, can be topicalized. Since the light verb in (31) is analyzed as a $\overline{\text{V}}$ and the noun as an N, only the light verb can be be topicalized.

(31) **a-yaa** anjum=ko **ɣussaa**
come-Perf.M.Sg Anjum.F=Dat anger.M
'Anjum became angry.'

As demonstrated in (32), light verbs in Aspectual complex predicates cannot be similarly topicalized. This indicates that the light verb in Aspectual complex predicates is not a $\overline{\text{V}}$, but should instead be analyzed as a V.

(32) *__ga-yii__ anjum **aa**
go-Perf.F.Sg Anjum.F=Nom come
'Anjum arrived.'

Furthermore, in a combination of an N-V and an Aspectual complex predicate, the expression is only grammatical if both the verbs in the construction are topicalized as a constituent, as in (33a), or if the entire N-V-V complex predicate is topicalized. All other attempts at topicalization in (33) are ill-formed.

(33) a. [**aa** **ga-yaa**] anjum=ko **ɣussaa**
come go-Perf.M.Sg Anjum.F=Dat anger.M
'Became Anjum angry.'

b. [**ɣussaa** **aa** **ga-yaa**] anjum=ko
anger.M come go-Perf.M.Sg Anjum.F=Dat
'Became Anjum angry.'

c. *ga-yaa anjum=ko ɣussaa aa
go-Perf.M.Sg Anjum.F=Dat anger.M come
'Became Anjum angry.'

d. *aa anjum=ko ɣussaa ga-yaa
come anjum=Dat anger.M go-Perf.M.Sg
'Became Anjum angry.'

e. *ɣussaa aa anjum=ko ga-yaa
anger.M come Anjum.F=Dat go-Perf.M.Sg
'Became Anjum angry.'

The topicalization data thus argue for the structure in (30) as representative of a N-V-V complex predicate. This in turn indicates that the only plausible structure for an Aspectual complex predicate is one in which both verbs are simple Vs contained under a $\overline{\text{V}}$.

4.4.6 Summary

This section has shown that although an Aspectual complex predicate cannot be analyzed forming a single lexical item, the two verbs do form a tighter syntactic unit than the constituents of a permissive complex predicate. The coordination and scrambling facts showed that the main verb and the light verbs in Aspectual complex predicates could be analyzed as paralleling the structure of a simple predicate with auxiliaries. However, as light verbs can themselves appear with the full range of auxiliaries in Urdu, they cannot be completely like auxiliaries. In the next section, I argue explicitly that light verbs must be distinguished from auxiliaries since they contribute to the argument structure of a clause.

Evidence from the interaction of N-V complex predicates with Aspectual complex predicates indicated that the internal structure of an Aspectual complex predicate must be as shown in (34).

(34) $[\text{V V}]_{\overline{\text{V}}}$

Furthermore, since Aspectual complex predicates occur with auxiliaries and statives, I propose (35) as a general phrase structure rule for the generation of predicates in Urdu.

(35) $\overline{\text{V}} \rightarrow$V (V) (STAT) (AUX)

There are two noteworthy implications of (35). One is that the entire $\overline{\text{V}}$ should be able to topicalize. This is indeed the case. The other is that it rules out V-V complex predicates

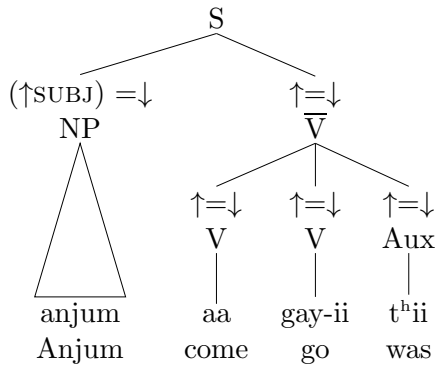
consisting of more than two verbs. This is desirable, as there seem to be no occurences of V-V-V complex predicates. As (36) shows, Aspectual complex predicates do interact with the permissive, but since the infinitive predicate in the permissive was shown to be an N, the sentence in (36) is again an instance of a N-V-V complex predicate.

(36) naadyaa baccõ=ko ṭofii [$_{\overline{V}}$ [$_{N}$xariid-ne] [$_{\overline{V}}$ de
Nadya.F=Nom child.Pl=Dat toffee.F=Nom buy-Inf.Obl give
paṛ-ii]]
fall-Perf.F.Sg
'Nadya suddenly allowed the children to buy candy.'

Finally, I provide a c-structure and an f-structure representation for the typical Aspectual complex predicate in (37). The c-structure is shown in (38), and the corresponding f-structure in (39).

(37) anjum **aa** **ga-yii** **t^{h}ii**
Anjum.F=Nom come go-Perf.F.Sg be.Past.F.Sg
'Anjum had arrived.'

(38)



(39)

$$\begin{bmatrix} \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'Anjum'} \end{bmatrix} \\ \text{PRED} & \text{'come-complete} < \underline{\quad} > \text{'} \\ \text{TENSE} & \text{Past} \\ \text{ASP} & \text{Perf} \end{bmatrix}$$

The PRED value *come-complete* in (39) is a combination of the argument structures of the main verb *come* and the light verb *go*. I leave a full-fledged discussion of exactly how this PRED value is arrived at to Chapter 5, where I present an argument structure analysis of

complex predicates. The account I propose in Chapter 5 is crucially motivated by the facts presented in the next section.

4.5 Argument Structure

Aspectual complex predicates display some rather intriguing restrictions on the combination of light verbs with main verbs. The fact that not simply any combination of main and light verb is well-formed has been remarked on (Hook 1974), but an analysis which successfully predicts possible combinations has not been formulated. I show in this section that the light verb plays a role in determining the case of the subject, and that the case marking on the subject is correlated with the semantic notions of conscious choice (volitionality) and inception/completion. I propose that if these semantic notions are represented at the level of argument structure, then not only do the case-marking facts follow, but a solid basis for a theory which articulates well-formedness criteria for Aspectual complex predicates can also be established.

To begin with, I argue that a light verb cannot be treated as an auxiliary, as has been suggested in the literature (Hook 1991). Auxiliaries in Urdu function as tense and aspect markers, and place no restrictions on the kind of main verb they may be combined with. In addition, the case on the subject of Aspectual complex predicates is affected by the semantic information the light verb contributes. Auxiliaries do not make a similar, regular, semantic contribution and thus do not similarly affect the argument structure of a clause.

Light verbs are thus sometimes described as having more 'semantic content' than an auxiliary. I propose that the rather vague notion of semantic content must be expressed in terms of argument structure. Light verbs do not have the fully articulated argument structure of their full counterparts, but they do retain at least one argument. When this argument fuses with the argument structure of the main verb to produce a complex predicate, the fused arguments must agree in terms of the semantic information encoded in them. The next sections illustrate how the case marking on subjects is systematically conditioned by the light verb.

4.5.1 Conscious Choice and Ergativity

I have identified two kinds of information which determine whether a main and a light verb are compatible. A light verb selects not only for conscious choice, but also for the aspectual

factor of completion/inception. This section focuses on the interaction of conscious choice and the ergative case.[10]

4.5.1.1 The Basic Pattern

Light verbs like *ḍaal* 'put' and *paṛ* 'fall' play a part in determining whether or not the actor is interpreted to have conscious control over the action. The light verb *ḍaal* 'put' attributes conscious choice to the actor while *paṛ* 'fall' does not. In (40), for example, where *ḍaal* 'put' combines with the main verb *pii* 'drink', the actor intended to drink the medicine. In (41), on the other hand, where *paṛ* 'fall' forms a complex predicate with *gir* 'fall', the subject *Sheila* is seen as having no control over the action.

(40) us=ne davaii pii ḍaal-ii
Pron=Erg medicine.F=Nom drink put-Perf.F.Sg
'He downed the medicine.'

(41) šiila gir paṛ-ii
Sheila.F=Nom fall fall-Perf.F.Sg
'Sheila fell.'

The next set of examples illustrates in some more detail the effect the light verbs *paṛ* 'fall' and *ḍaal* 'put' have on the semantics of an expression, and the role they play in the selection of case on the subject. Example (42) shows that *ro* 'cry' is an intransitive verb with a nominative subject in the perfective.[11] In (43b) the light verb *ḍaal* 'put' gives the act of crying a deliberate, completed aspect. In (43a), on the other hand, the light verb *paṛ* 'fall' contributes a meaning of suddenness and surprise (lack of control). Notice that while the subject of the main verb *ro* 'cry' in (42) can only be nominative in the perfective, the subject of a complex predicate formed with *ro* 'cry' can be either nominative as in (43a), or ergative, as in (43b).

(42) a. vo ro-yaa
Pron=Nom cry-Perf.M.Sg
'He wept.'

b. *us=ne ro-yaa
Pron=Erg cry-Perf.M.Sg
'He wept.'

[10] I use the term 'conscious choice' in preference to 'volitionality' as it would seem to express the semantics involved more clearly.

[11] Recall that ergativity is correlated largely with perfectivity in Urdu.

(43) a. **vo** ro **paṛ-aa**
Pron=Nom cry fall-Perf.M.Sg
'He fell to weeping (burst into tears).'

b. **us=ne** ro **ḍaal-aa**
Pron=Erg cry put-Perf.M.Sg
'He wept heavily (on purpose).'

In fact, as will be demonstrated, the light verb *paṛ* 'fall' can only appear in a complex predicate with a nominative subject, while *ḍaal* 'put' requires that the subject of the complex predicate be ergative in the perfective. Examples (44) and (45) illustrate complex predicates in which *ḍaal* 'put' and *paṛ* 'fall' are combined with a transitive main verb, *gaa* 'sing' which must take an ergative subject in the perfective. The light verb *ḍaal* 'put' again appears with an ergative subject and a reading of conscious choice, while the complex predicate with *paṛ* 'fall' must have a nominative subject and describes a sudden, unplanned action.

(44) a. us=ne gaanaa gaa-yaa
Pron=Erg song.M=Nom sing-Perf.M.Sg
'He sang a song.'

b. *vo gaanaa gaa-yaa
Pron=Nom song.M=Nom sing-Perf.M.Sg
'He sang a song.'

(45) a. **us=ne** gaanaa gaa **ḍaal-aa**
Pron=Erg song.M=Nom sing put-Perf.M.Sg
'He sang a song (completely, forcefully).'

b. **vo** gaanaa gaa **paṛ-aa**
Pron=Nom song.M=Nom sing fall-Perf.M.Sg
'He fell to singing (burst out into song).'

The intransitive verb *ciikʰ* 'scream' can take either an ergative or a nominative subject in the perfective. When the subject is ergative, as in (46b), the action must be interpreted as more intentional than when the subject is nominative as in (46a). Example (47) shows that again both *paṛ* 'fall' and *ḍaal* 'put' can form a complex predicate with *ciikʰ*. In (47a), the light verb *paṛ* 'fall' again cooccurs with a nominative subject and an involuntary action. In (47b), on the other hand, the light verb *ḍaal* 'put' has an ergative subject and conveys a sense of conscious control over the action.

(46) a. vo ciikʰ-aa
Pron=Nom scream-Perf.M.Sg
'He screamed (despite himself).'

b. us=ne ciik^{h}-aa
Pron=Erg scream-Perf.M.Sg
'He screamed (on purpose).'

(47) a. **vo** ciik^{h} **paṛ-aa**
Pron=Nom scream fall-Perf.M.Sg
'He began screaming suddenly (despite himself).'

b. **us=ne** ciik^{h} **ḍaal-aa**
Pron=Erg scream put-Perf.M.Sg
'He screamed violently (on purpose).'

The correlations that have emerged so far are clear. The light verb *ḍaal* 'put' contributes a reading of conscious choice to an action and assigns the ergative case to the subject of the complex predicate. The light verb *paṛ* 'fall', on the other hand, implies a lack of conscious choice and assigns the nominative, unmarked case to the subject. The transitivity/intransitivity of a given verb, or its inherent case marking requirements on the subject were seen to play no role.

4.5.1.2 Similar Light Verbs

Some examples of other light verbs which work like *ḍaal* 'put' and *paṛ* 'fall' are *maar* 'hit', *baiṭ*^{h} 'sit', and *uṭ*^{h} 'rise'. The light verb *maar* 'hit' is a *ḍaal* 'put' type verb. An example of it as a main verb is given in (48); its light verb use is illustrated in (49). As can be seen, the light verb *mar* 'hit' in (49) also assigns the ergative case and attributes conscious choice to the subject.

(48) šiila=ne aadmii=ko maar-aa
Sheila.F=Erg man.M-Acc hit-Perf.M.Sg
'Sheila hit the man.'

(49) **us=ne** xat lik^{h} **maar-aa**
Pron=Erg letter.M=Nom write hit-Perf.M.Sg
'He dashed off a letter.'

The example in (50) demonstrates how *bait*^{h} 'sit', a *paṛ* 'fall' type verb, works. The verb *uṭ*^{h} 'rise' is shown in (51) and functions similarly. Both of these light verbs assign the nominative case and denote a lack of conscious choice.

(50) **vo** ro **baiṭ**^{h}**-ii**
Pron=Nom weep sit-Perf.F.Sg
'She wept (caused by an external event).'

(51) **vo** bol **uṭʰ-ii**
Pron=Nom speak rise-Perf.F.Sg
'She burst out talking.'

There are thus two kinds of verbs that relate to conscious choice: those that function like *ḍaal* 'put' and those that function like *paṛ* 'fall'. The *ḍaal* 'put' type verbs mark the subject of a complex predicate with the ergative. The *paṛ* 'fall' type verbs require a nominative subject.

4.5.1.3 The Expanded Pattern

The interaction between conscious choice and ergativity is further confirmed by some data from modification. However, ultimately, the picture sketched above is in need of some refinement, as not every instance of an ergative case marker in the language can necessarily be interpreted as a marker of conscious choice.

The correlation of *ḍaal* 'put' type light verbs with conscious choice predicts that they should be incompatible with any additions to the expression that suggest otherwise. So, the modifier *ɐaltii=se* 'by mistake' in combination with a *ḍaal* 'put' type light verb should at the very least be infelicitous. Similarly, a complex predicate formed with a *paṛ* 'fall' type light verb, should not be able to be modified by the phrase *dʰiyaan=se* 'with care'. The examples in (52) and (53) show that *ḍaal* 'put' type light verbs do indeed work the way they are expected to. In (52b) the song can be sung by mistake; however, in (53b) where *ḍaal* 'put' attributes conscious choice to the agent, the action cannot be taken to be accidental.

(52) a. us=ne gaanaa ga-yaa
Pron=Erg song.M=Nom sing-Perf.M.Sg
'He sang a song.'

b. us=ne ɐaltii=se gaanaa ga-yaa
Pron=Erg mistake.F=Inst song.M=Nom sing-Perf.M.Sg
'He sang the song by mistake.'

(53) a. us=ne gaanaa gaa ḍaal-aa
Pron=Erg song.M=Nom sing put-Perf.M.Sg
'He sang the song (completely, forcefully).'

b. #us=ne ɐaltii=se gaanaa gaa ḍaal-aa
Pron=Erg mistake.F=Inst song.M=Nom sing put-Perf.M.Sg
'He sang the song (completely, forcefully) by mistake.'

Complex predicates formed with *paṛ* 'fall' type light verbs in combination with the modifier *ɗʰiyaan=se* 'with care' are also semantically infelicitous. This is shown in (54) and (55). Example (54b) is felicitous in an appropriate context, such as a situation in which a child is upset, but is attempting to avoid incurring the further wrath of a parent: the child needs to cry, but tries to keep it within bounds and do it carefully. However, in (55b), the light verb *paṛ* is incompatible with the conscious choice reading made explicit by the modifier *ɗʰiyaan=se* 'with care'.

(54) a. baccii ro-yii
child.F=Nom cry-Perf.F.Sg
'The child cried.'

b. baccii dʰiyaan=se ro-yii
child.F=Nom attention.M=Inst cry-Perf.F.Sg
'The child cried carefully.'

(55) a. baccii ro paṛ-ii
child.F=Nom cry fall-Perf.F.Sg
'The child cried (burst into tears).'

b. #baccii dʰiyaan=se ro paṛ-ii
child.F=Nom attention.M=Inst cry fall-Perf.F.Sg
#'The child burst into tears carefully.'

Although the interaction of the light verbs with the modifiers is as predicted, the above set of examples does also bring out an issue which I have glossed over so far. Up to now an ergative case marker on the subject has been automatically equated to something like a feature [+conscious]. However, (52b) shows that the presence of an ergative subject alone does not necessarily induce a conscious choice reading. I would argue that the appearance of the ergative here is not so much attributable to conscious choice, but to the requirement of the transitive-perfective paradigm for ergativity. As discussed in Chapter 2, the ergative case functions both as a semantic and a structural case. In an unmarked transitive, perfective sentence like (52b), the ergative case is fulfilling a structural requirement, and does not necessarily indicate a meaning of conscious choice.

I propose that most verbs are underspecified with regard to the notion conscious choice. A specification of [±conscious] may be expressed through adverbs, light verb constructions and the ergative case when it is not necessarily required by the transitive-perfective paradigm, as in (46), repeated here as (56).

(56) a. vo ciik^{h}-aa
Pron=Nom scream-Perf.M.Sg
'He screamed (despite himself).'

b. us=ne ciik^{h}-aa
Pron=Erg scream-Perf.M.Sg
'He screamed (on purpose).'

The indication of conscious choice through a *ḍaal* 'put' type light verb also correlates with the appearance of the ergative, as was demonstrated in the previous two sections. A conscious choice reading is therefore not directly tied to the appearance of ergative versus nominative case on the subject, but depends on a complex interaction between the semantic contribution of the light verb, the specification of conscious choice on main verbs, and the appearance of the ergative case marker. For a closer look at how ergativity, voltionality and verbal paradigms can interact in a complicated way see Inman (1993) on Sinhala.

4.5.1.4 Restrictions on Complex Predicate Formation

I proposed above that most verbs are underspecified with regard to conscious choice: one can sing intentionally, or one could have been so happy, one just burst into song unintentionally. Main verbs which are underspecified are thus compatible with both the *ḍaal* 'put' and the *paṛ* 'fall' type verbs. This section takes a quick look at examples of complex predicates formed with a main verb which is inherently either negatively or positively specified for conscious choice.

In (57), the main verb is *dek*^{h} 'see'. It can form a complex predicate with the light verb *le* 'take' in (57a), but not with the light verb *paṛ* 'fall' in (57b). The ill-formedness of (57b) is expected if the light verb *paṛ* 'fall' is taken to encode a lack of conscious choice, while the main verb *dek*^{h} 'see' is specificed positively for conscious choice: when the argument structures of the two predicates combine to form a complex predicate, there is a clash in semantic information.

(57) a. naadyaa=ne aaj ek hiran **dek**^{h} **li-yaa**
Nadya.F=Erg today one deer.M=Nom see take-Perf.M.Sg
'Nadya saw a deer today.'

b. *naadyaa aaj ek hiran **dek**^{h} **paṛ-ii**
Nadya.F=Nom today one deer.M=Nom see fall-Perf.F.Sg
'Nadya fell to seeing a deer today.'

In (58a) the main verb *gir* 'fall' is negatively specified for conscious choice (one does not usually fall on purpose), while the light verb *le* 'take' requires a conscious choice reading. Again there is a mismatch in features and the argument structures of the two verbs cannot combine into a well-formed complex predicate. On the other hand, as expected under this view, the verb *gir* 'fall' is compatible with the [−conscious] light verb *paḍ* 'fall' in (58b).

(58) a. *naadyaa=ne **gir li-yaa**
Nadya.F=Erg fall take-Perf.M.Sg
'Nadya fell (on purpose).'

b. naadyaa **gir paṛ-ii**
Nadya.F=Nom fall fall-Perf.F.Sg
'Nadya fell (suddenly, involuntarily).'

While I have only discussed two examples here, a broad examination of Aspectual complex predicates reveals that the respective well- or ill-formedness of a whole range of main and light verb combinations can indeed be accurately predicted under this view. Main verbs underspecified for conscious choice can combine with either type of light verb, while main verbs which are inherently either negatively or positively specified can only combine with a matching light verb. There are, or course, exceptions to this paradigm, but once the notions of inception and completion are taken into account as well, the set of criteria for the well-formedness of an Aspectual complex predicate is almost complete.

4.5.1.5 Summary

Light verbs in Urdu fall into two basic classes: one which requires an ergative subject in the perfective, and one which always takes a nominative subject. It was pointed out in a previous chapter that the ergative case in Urdu is associated with the semantic notion of conscious choice (see also T. Mohanan (1990) for Hindi and Pandharipande (1990) for Marathi). A closer look at the interaction of the two types of light verbs with main verbs shows that light verbs actually determine the case marking on the subject and also require that the main verb be semantically compatible in terms of conscious choice. A light verb like *paṛ* 'fall', which is specified negatively for conscious choice, requires a nominative subject and cannot combine with a main verb that is specified positively for conscious choice. On the other hand, a light verb like *ḍaal* 'put', which is specified positively for conscious choice, requires an ergative subject in the perfective and cannot combine with a main verb specified negatively for conscious choice.

Light verbs thus clearly contribute to the determination of case on the subject of a complex predicate. Because the semantics of conscious choice are crucial for determination of the well-formedness of a complex predicate, and for the case marking on subjects, I propose that this semantic information be included at the level of argument structure. I again leave the precise representation of conscious choice at argument structure and a detailed analysis of complex predicate formation in general for Chapter 5.

4.5.2 Unaccusativity

The light verb *paṛ* 'fall' discussed in the previous section is often analyzed as unaccusative (Grimshaw 1987, Belletti 1988). In the previous section, I identified this light verb as denoting a lack of conscious choice on the part of the actor, and showed that light verbs of the *paṛ* 'fall' type could not combine with main verbs specified for conscious choice. In the elaborated argument structure approach I present in Chapter 5, I treat the light verb *paṛ* 'fall' as taking a single argument: an actor which is negatively specified for conscious choice.

Before the presentation of that analysis, however, the question of unaccusativity must be considered. If *paṛ* 'fall' is analyzed as an unaccusative verb with a single theme argument, then the complex predicate formation facts would seem to follow as well. A theme argument would be negatively specified for conscious choice inherently because a theme never has control over an action. A light verb with a theme argument therefore could not unify with the agent argument of a main verb.

An unaccusative analysis, however, does not allow an immediate explanation of the contrast in (59). The main verbs *gaa* 'sing' and *pii* 'drink' in (59a) and (59b), respectively, are both transitive. They might be considered to have the same type of agent/actor arguments. However, *gaa* 'sing' can combine with a *paṛ* 'fall' type light verb, but *pii* 'drink' cannot.[12]

(59) a. anjum gaanaa **gaa paṛ-ii**
Anjum.F=Nom song.M=Nom sing fall-Perf.F.Sg
'Anjum burst out into song.'

b. *anjum paanii **pii paṛ-ii**
Anjum.F=Nom water.M=Nom drink fall-Perf.F.Sg
'Anjum fell to drinking water.'

[12]Note that (59a) is actually much better with the light verb *uṭʰ* 'rise'. I have used *paṛ* 'fall' for greater ease of exposition. The light verb *uṭʰ* 'rise' does not form a complex predicate with *pii* 'drink', so the contrast remains constant.

Furthermore, the sentences in (60), which contrast the light verb *paṛ* 'fall' with the light verb *jaa* 'go', demonstrate that the *paṛ* 'fall' type light verbs cannot be analyzed as unaccusative.

(60) a. vo ciikʰ **paṛ-ii**
Pron=Nom scream fall-Perf.F.Sg
'She screamed involuntarily, suddenly.'

b. *vo ciikʰ **gay-ii**
Pron=Nom scream go-Perf.F.Sg
'She completed screaming.'

Recall that the main verb *ciikʰ* 'scream' is one of the verbs in Urdu which can appear either with an ergative or a nominative subject in the perfective. Recall that main verbs of this type are able to form complex predicates with both the *paṛ* 'fall' and the *ḍaal* 'put' type light verbs. I propose that the verb *ciikʰ* 'scream' therefore should be analyzed as having an agentive argument unspecified for conscious choice.

An analysis which treats both the light verb *paṛ* 'fall' in (60a) and the light verb *jaa* 'go' in (60b) as unaccusatives cannot immediately provide an account for the fact that the complex predicate *ciikʰ paṛ* 'scream involuntarily' in (60) is well-formed, while the complex predicate *ciikʰ jaa* 'complete screaming' in (60b) is not.

A close look at the patterns of complex predicate formation show that the light verb *paṛ* 'fall', but not *jaa* 'go', can combine with the main verbs *bol* 'speak', *ro* 'cry', *kah* 'say', *cal* 'walk', *sun* 'listen, hear', and *has* 'laugh'. Each of these verbs in fact may be analyzed as unergative (Grimshaw 1987).

I propose that *jaa* 'go' is indeed an unaccusative light verb which cannot combine with unergatives. The light verb *paṛ* 'fall', on the other hand, displays a differing pattern of complex predicate formation, and cannot be similarly analyzed as unaccusative. Rather, it must be analyzed as taking an actor argument which is negatively specified for conscious choice.[13]

Van Valin (1987) argues that a syntactic, monolithic approach to verb classification, such as the Unaccusative Hypothesis (Perlmutter 1978, Burzio 1986) is not able to account for the various fine semantic distinctions found between verb classes, and argues for an approach to verb classification based on Lexical Semantics. The data from Urdu Aspectual

[13]In fact, the distinction recently proposed by B. Levin and Rappaport Hovav (1993) between intransitives describing *internally caused* vs. *externally caused* events may be what is needed here. This remains to be explored.

complex predicates serves to confirm Van Valin's (1987) argument, and crucially motivates the elaborated level of argument structure I introduce in Chapter 5.

4.5.3 Inception and Completion

4.5.3.1 The Basic Pattern

Aspectual information in terms of inception and completion must also be recognized as an integral part of complex predicate formation and the determination of case marking on the subject. Although the semantic notion of conscious choice discussed in the last section is able to account for the grammaticality or ill-formedness of large number of complex predicates, the notion of conscious choice by itself is not sufficient to account for the full range of complex predicate formation. The sentences in (61) and (62), for example, illustrate four complex predicates which ought to be possible. The main verb *bʰuul* 'forget' in Urdu, as in English, denotes a non-volitional act in most contexts. Why, then, should both (61a) and (61b) be ill-formed? Similar reasoning applies to (62). The verbs *so* 'sleep' and *baiṭʰ* 'sit' can express either a volitional or an unintentional action. But as (62a) shows, these main verbs are incompatible with the light verb *paṛ* 'fall'.

(61) a. *anjum kahaanii bʰuul paṛ-ii
Anjum.F=Nom story.F=Nom forget fall-Perf.F.Sg
'Anjum forgot the story (started forgetting).'

b. *anjum kahaanii bʰuul uṭʰ-ii
Anjum.F=Nom story.F=Nom forget rise-Perf.F.Sg
'Anjum forgot the story (started forgetting).'

(62) a. *anjum so paṛ-ii
Anjum.F=Nom sleep fall-Perf.F.Sg
'Anjum fell asleep (started sleeping).'

b. *anjum baiṭʰ paṛ-ii
Anjum.F=Nom sit fall
'Anjum sat down suddenly.'

Besides denoting involuntariness, however, the *paṛ* 'fall' type light verbs also contribute a sense of suddenness and inception (Hook 1974). I argue that the aspectual notion *inception* accounts for the unacceptability of the sentences in (61) and (62).

Similarly, it has been long noted that the *ḍaal* 'put' type light verbs denote completion (Hook 1974). This accounts for the ill-formedness of the complex predicate in (63), which consists of the stative main verb *jaan* 'know' and the completive light verb *le* 'take'.[14]

(63) *anjum=ne ye jaan li-yaa
Anjum.F=Erg this=Nom know take-Perf.M.Sg
'Anjum has known this (completely).'

While the idea that aspect in terms of completion interacts with the argument structure of predicates in some interesting ways has been discussed in the literature, the notion of inception has not generally been considered. After providing some background on the literature, I go on to show that both the notions of completion and inception are needed in order to correctly account for the facts of complex predicate fomation.

4.5.3.2 Background on Aspect and Argument Structure

Tenny (1987) has shown that aspectual information, in particular the notion of delimitation, must be visible to syntactic processes. Tenny's solution is to encode the appropriate delimitation of an action in the lexical entry of each verb. Even though Tenny uses the neutral term "delimiter", she actually focuses almost exclusively on the properties of completion. Similarly, Moens and Steedman (1988) critique a purely linear model of time and, based on data from English, propose a more complicated picture in terms of a contingency-based event structure and aspect coercion, but do not situate inception in their event structure.

With regard to South Asian languages, several different approaches have been taken. While it is clear from work done by Bashir (1989), DeLancey (1986), and Pandharipande (1989) that inception is central in phenomena encountered with regard to complex predicates and causatives, the insights from South Asian languages have not been explicitly integrated with more general theories of verb classification and aspect (Dowty 1979, Vendler 1967).

[14]Dayal (p.c.) points out that (63) has a close correlate, shown in (i), in which the combination of the main verb *jaan* 'know' and the light verb *le* 'take' is good. In this case, though, as Dayal also points out, the *jaan* 'know' must have an activity rather than a state reading, and as such a point of completion can be specified by the light verb *le* 'take'. In its activity reading, the verb *jaan* carries the sensse of 'realize' rather than 'know'.

(i.) aur tum ye **jaan** **l-o** ke mãĩ tumhaarii baat kab
and you=Nom this=Nom know take-Imp that I=Nom your matter ever
hii nahĩ̃ĩ maanũ-gii
Emph not accept
'And you realize this, that I will never accept what you are saying.'

Ramchand (1990) and Singh (1990) examine complex predicates in Bangla and Hindi, respectively. Ramchand's central proposal is that complex predicate formation in Bangla can be accounted for in terms of three central "roles": cause, change, and completion. These roles both represent the core arguments of the complex predicate, and encode aspectual information. She notes that complex predicates express completion even when the host verb in its simple form is not necessarily completive. The completive aspect of the complex predicate comes from an inherent property of the light verb, which is taken to augment the lexical aspectual structure of the main verb. Although Ramchand is also primarily concerned with the aspectual notion completion, she points out that in the case of the Bangla light verb *pɔɽa* 'fall', inception must also be taken into account for an exhaustive explanation of Bangla complex predicate formation. However, as she uses a relatively abstract notion of completion, which simply names the new state that is arrived at by the change argument, a 'completion' could actually occur at the temporal beginning of an event, rather than at the end. In other words, Ramchand's notion of completion contains both inception and completion.

Singh (1990) adopts a 'two component theory of aspect' proposed by Smith (1991) and separates out *situation type* (states, events) from *viewpoint* (perfective, imperfective). She argues that light verbs are markers of telicity, and that the particular role of light verbs is to focus on the different stages of a telic event. She takes a telic event to have three diferrent stages: Initial, Final, and Result. Her definition of Initial and Final corresponds exactly to the notions of inception and completion discussed in this chapter. The light verb *paɽ* 'fall' is analyzed as focusing on the Initial stage of a telic event, while light verbs like *ɖaal* 'put' or *le* 'take' focus on the Final stage.[15] Singh shows in some detail that the *le* 'take' type light verbs do indeed encode a notion of completeness, so I will not duplicate her findings here. However, while she does recognize the aspectual importance of inception, she does not spend very much time on a treatment of the inceptive light verbs. The next section therefore examines the evidence for inception in Urdu in some detail.

[15]The Result stage is represented by the light verb *cuk* 'lift', which carries a meaning of 'the action has really been done completely and here is the result'. I do not discuss this particular light verb, but its precise semantic connotations in comparison with the other completive light verbs is clearly of interest.

4.5.3.3 Evidence for Inception

As already discussed, there are some main verbs which cannot combine with the light verb *paṛ* 'fall' to form a complex predicate. This incompatibility cannot be explained by conscious choice because the main verbs do not require a volitional agent.[16] The relevant examples are repeated here in (64) and (65). If the light verb *paṛ* 'fall' indeed focuses on the inception of an event, it is possible that the reason it cannot form a complex predicate with main verbs like *so* 'sleep' or *bʰuul* 'forget' is that these verbs do not describe an action with an identifiable starting point.

(64) *anjum **so** **paṛ-ii**
Anjum.F=Nom sleep fall-Perf.F.Sg
'Anjum fell asleep.'

(65) *anjum kahaanii **bʰuul paṛ-ii**
Anjum.F=Nom story.F=Nom forget fall-Perf.F.Sg
'Anjum started forgetting the story (suddenly).'

Some other verbs which do not require conscious agency, but which can nevertheless not combine with the light verb *paṛ* 'fall' as a light verb are: *jaan* 'know', *samaj* 'understand', *duub* 'sink', and *baiṭʰ* 'sit'.

The set of examples in (66) and (67), in which I have tried to render as accurate an English gloss as possible, provide a rather striking illustration of the phenomenon. The light verb *paṛ* 'fall' can combine with the main verb *uṭʰ* 'rise', but not with the main verb *baiṭʰ* 'sit'. Similarly, in (67) a combination with the main verb *jaag* 'wake', but not with the main verb *so* 'sleep', is possible.

(66) a. anjum uṭʰ paṛ-ii
Anjum.F=Nom rise fall-Perf.F.Sg
'Anjum rose (suddenly) at that very instant.'

b. *anjum baiṭʰ paṛ-ii
Anjum.F=Nom sit fall-Perf.F.Sg
'Anjum sat (suddenly) at that very instant.'

(67) a. anjum jaag paṛ-ii
Anjum.F=Nom wake fall-Perf.F.Sg
'Anjum woke (suddenly) at that very instant.'

[16]The reader may have noticed that although I use the term *conscious choice* for the most part, I do also use the term *volitionality*. The two notions may be argued to be semantically distinct, but are close enough for the purposes of this dissertation and I therefore do not address the issue here.

b. *anjum so paṛ-ii
Anjum.F=Nom sleep fall-Perf.F.Sg
'Anjum slept (suddenly) at that very instant.'

I propose that the light verb *paṛ* 'fall', in addition to indicating non-volitionality, also serves to "pick out" or *focus* on the point of inception of an action. When a verb describes an action which does not have a point of inception that can be readily identified, independent of point of perspective, then the light verb *paṛ* 'fall' cannot combine with it. So, in (66a), both the performer of the action and any observers in the room can readily agree on when the action of getting up commenced. However, the action of sitting down is a different matter. It is very easy to identify when the action of sitting down has been completed, but it is not clear when exactly it commenced. *Anjum* may argue that she actually began the action of sitting down when she began moving towards the chair, observers in the room may insist that it was when she already by the chair, or perhaps hovering above it, etc. A similar, albeit intuitive, chain of reasoning applies to (67). When I wake up, I can immediately identify the instant at which I commenced being awake, and so can any observers watching me. The action of sleeping is a different matter. Although observers may be able to identify the exact instant at which I fell asleep, I myself cannot ever identify the point at which I began to be asleep.

Note that it is of course possible to say *Anjum started sleeping* in Urdu. However, the verb *lag* 'be attached to', which is used to denote the beginnings of an action differs substantially from the *paṛ* 'fall' type light verbs in both syntactic and semantic terms. The differences between the types of semantics and their interaction with the event structure of a main verb is discussed in the next section.

The kind of sentence illustrated in (68), in which a given action is distributed over two clauses, shows that the *paṛ* 'fall' type light verbs do indeed pick out the point of inception of an event. If, as in (68), the action is not explicitly defined as having an endpoint, it is presumed to continue in the second half of the sentence. In other words, both of the verbs *sang* in (68) refer to one and the same action of singing by *Anjum*.

(68) Anjum sang for a little while and then she sang some more.

If *paṛ* 'fall' type light verbs do focus on the inception of an action, then a complex predicate with *paṛ* 'fall' should not be able to appear in the latter half of a sentence like (68), and be able refer to the same action described by the predicate in the first half. In other words, an

inceptive predicate should denote the beginning of a new action, and therefore should not be able to continue a previous event.

This is illustrated in the set of examples in (69). In each of these sentences, *Anjum* sings for a little while in the first clause. Only the examples in (69a) and (69b) are good under the reading that *Anjum*'s same action of singing was continued in the second clause. In (69a), *Anjum* sang a little, sang some more in the second clause, but has not necessarily finished singing. When a completive light verb like *le* 'take' appears in the second clause, the complex predicate refers to the same action as in the first clause, and the action of singing must be interpreted as having been completed.

(69) a. anjum=ne t^{h}orii d^{h}eer keliye gaa-yaa aur p^{h}ir us=ne aur b^{h}ii
Anjum.F=Erg few time for sing-Perf.M.Sg and then Pron=Erg more too
ga-yaa
sing-Perf.M.Sg
'Anjum sang for a little while and then she sang some more.'

b. anjum=ne t^{h}orii d^{h}eer keliye ga-yaa aur p^{h}ir us=ne aur b^{h}ii
Anjum.F=Erg few time for sing-Perf.M.Sg and then Pron=Erg more too
gaa
sing
li-yaa
take-Perf.M.Sg
'Anjum sang for a little while and then she sang some more (completed singing).'

c. anjum=ne t^{h}orii d^{h}eer keliye ga-yaa aur p^{h}ir vo aur b^{h}ii
Anjum.F=Erg few time for sing-Perf.M.Sg and then Pron=Nom more too
gaa
sing
paṛii
fall-Perf.F.Sg
∗'Anjum sang for a little while and then she sang some more.'
'Anjum sang for a little while and then she fell to singing something else.'

In (69c) the use of the inceptive light verb *paṛ* 'fall' in the second clause prohibits a reading in which the action of singing in the second clause is a continuation of the singing in the first clause. Rather, a whole new event of singing, separate and distinguishable from the singing event in the first clause, is begun in the second clause.

Finally, just as a completive light verb like *le* 'take' in (70a) is incompatible with a stative main verb like *jaan* 'know', the inceptive light verb *paṛ* 'fall' also cannot form a

complex predicate with a stative. This is shown in (70b).

(70) a. *anjum=ne ye jaan li-yaa
Anjum.F=Erg this=Nom know take-Perf.M.Sg
'Anjum has known this (completely).'

b. *anjum ye jaan paṛ-ii
Anjum.F=Nom this=Nom know fall-Perf.F.Sg
'Anjum fell to knowing this.'

While it is quite clear at this point that the *le* 'take' type light verbs focus on the completion of a telic event, while the *paṛ* 'fall' type light verbs pick out the inception of an action, the role of the verb *lag* 'be attached to' in describing the beginning of an event must be explored. The use of this verb was mentioned very briefly, and is discussed in a little more detail in the next section.

4.5.3.4 Point of Inception versus Beginning an Event

As already mentioned, it is of course possible to say *Anjum began to sleep* or *Anjum began to sit* in Urdu. The verb *lag* 'be attached to' when combined with a verbal noun (infinitive) means *begin* or *start*. The sentences in (71) show how the beginning of an event is expressed with *lag* 'be attached to'. The examples in (72) repeat the corresponding ill-formed inceptive complex predicates for comparison.

(71) a. anjum baiṭʰ-ne lag-ii
Anjum.F=Nom sit-Inf.Obl be attached-Perf.F.Sg
'Anjum began to sit down.'

b. anjum so-ne lag-ii
Anjum.F=Nom sleep-Inf.Obl be attached-Perf.F.Sg
'Anjum began to sleep.'

(72) a. *anjum baiṭʰ paṛ-ii
Anjum.F=Nom sit fall-Perf.F.Sg
'Anjum sat (suddenly) at that very instant.'

b. *anjum so paṛ-ii
Anjum.F=Nom sleep fall-Perf.F.Sg
'Anjum slept (suddenly) at that very instant.'

The verb *lag* 'be attached to', unlike the light verb *paṛ* 'fall', is compatible with any main verb. Although both *lag* 'be attached to' and *paṛ* 'fall' clearly express inception in some way,

I propose that the two verbs are actually doing two very different things. The verb *lag* 'be attached to' embeds the entire event structure of the infinitive predicate it combines with under its own event structure. The light verb *paṛ* 'fall', on the other hand, combines with another predicate in such a way that there is no embedding of one event within another. Rather, it picks out the main verb's inherent point of inception. If the main verb has no identifiable point of inception, a complex predicate cannot be formed. For the verb *lag* 'be attached to' it is immaterial what the internal event structure of the infinitive predicate is: *lag* 'be attached to' simply embeds the entire event structure within its own.

Some evidence for this proposal comes from adverbial modification. In the sentences with *lag* 'be attached to', there are two identifiable events, the event of beginning x, and the event x itself. The sentences with *lag* 'be attached to' in (73), given an appropriate context, allow a separate modification of these two events. This is not possible for the complex predicates formed with *paṛ* 'fall' in (74).[17] So, in (73), *Anjum* may have been engaged in the action of laughing or getting up, but may not have been able to complete the action due to an atmosphere of general hilarity in the room. Thus she in fact began the action of laughing or getting up some two minutes ago, and is finally managing to engage in the action now. The *do miṇṭ* 'two minutes' thus modifies the event of beginning (*lag* 'begin/be attached to'), while the *ab hii* 'now' modifies the event of getting up or laughing.

(73) a. anjum ab hii uṭʰ-ne do miṇṭ pahle lag-ii
Anjum.F=Nom now Emph rise-Inf.Obl two minute before be attached-Perf.F.Sg
'Anjum began getting up now two minutes ago.'

b. anjum ab hii has-ne do miṇṭ pahle lag-ii
Anjum.F=Nom now Emph laugh-Inf.Obl two minute before be attached-Perf.F.Sg
'Anjum began laughing now two minutes ago.'

(74) a. *anjum ab hii uṭʰ paṛ-ii do miṇṭ pahle
Anjum.F=Nom now Emph rise fall-Perf.F.Sg two minute before
'Anjum now got up (suddenly) two minutes ago.'

b. *anjum ab hii has paṛ-ii do miṇṭ pahle
Anjum.F=Nom now Emph laugh fall-Perf.F.Sg two minute before
'Anjum now suddenly fell to laughing two minutes ago.'

[17] Recall that an adverbial cannot appear between a main verb and a light verb in Aspectual complex predicates – hence the difference in modifier placement.

In the next chapter, I formulate the difference between the inceptive light verb and the *lag* 'begin/be attached to' in terms of different types of Events contained in their respective argument structures. A verb such as *lag* 'begin/be attached to' subcategorizes for an Event argument, and thus is able to embed another predicate, and the event that predicate describes, under its own event structure. A light verb, on the other hand, does not subcategorize for an Event. Rather, its aspectual structure is "deficient" in the sense that it is only specified for either completion or inception. This particular aspectual specification must then be unified with the event structure of the main verb as part of the process of complex predicate formation.

4.6 Conclusion

This chapter has shown that Aspectual complex predicates form a tight unit at c-structure under a $\overline{\text{V}}$ since they could not be scrambled away from one another or be separated by adverbial modifiers. However, a main verb in combination with a light verb cannot be analyzed as a compounded lexical item. Light verbs also must be differentiated from auxiliaries because they can themselves combine with the full range of auxiliaries possible in Urdu. More importantly, though, light verbs were shown to have more semantic content than auxiliaries. Light verbs impose restrictions on which main verbs they can combine with to form a complex predicate, and determine the case marking on the subject. In particular, the notions of conscious choice and inception/completion were seen to play a role in complex predicate formation and the determination of the subject's case.

I maintain that since light verbs do contribute to the argument structure of a complex predicate, at least in terms of the subject of the complex predicate, an argument structure representation of a light verb cannot be empty, as proposed by S. Rosen (1989) for Romance, for example. In particular, I propose that the semantic notions which interact with the case-marking of the subject be represented at argument structure. In the next chapter, I adopt some of Jackendoff's theory of Conceptual Semantics in order to be able to represent the semantic information pertinent to case marking in Urdu at the level of an elaborated argument structure.

Beyond the phenomena already presented in this chapter, which can be accounted for straight-forwardly under the view of complex predicate formation formulated in the next

chapter, there are some further interesting issues with regard to Aspectual complex predicates which remain to be resolved, but are beyond the scope of this dissertation.[18] I mention these issues here because they are extremely intriguing and should be the subject of further research.

It has long been noted that it is not possible to negate an Aspectual complex predicate (Bhatia 1973, Hook 1974, Kachru 1980). This is illustrated in (75). While the simple predicate *gaa* 'sing' in (75a) can be negated with *nahĩĩ* 'not', the Aspectual complex predicates in (75b) and (75c) cannot be similarly negated. Interestingly enough, the analogs of Aspectual complex predicates in Bangla cannot be negated either (Ramchand, p.c.).

(75) a. us=ne gaanaa nahĩĩ ga-yaa
Pron=Erg song.M=Nom not sing-Perf.M.Sg
'He didn't sing a song.'

b. *us=ne gaanaa nahĩĩ gaa li-yaa
Pron=Erg song.M=Nom not sing take-Perf.M.Sg
'He didn't sing a song (completely, forcefully).'

c. *vo gaanaa nahĩĩ gaa paṛ-aa
Pron=Nom song.M=Nom not sing fall-Perf.M.Sg
'He didn't fall to singing (burst out into song).'

Singh (1990) represents the only work I know of which offers a coherent account of these facts. Her suggestion is that, as light verbs focus on the point of inception or completion of a particular telic event, the use of negation is incompatible with the focusing on a particular stage of a telic event. This account strikes me intuitively as proceeding in the right direction: although one can negate an event as a whole, one cannot negate the particular point of inception or completion of an event. However, it is not quite clear how this can be expressed semantically in a more concrete and logical manner. A purely syntactic approach to this phenomenon, on the other hand, does not seem immediately possible. It is not clear to me how Aspectual light verbs would differ structurally from simple predicates with auxiliaries in such a way as to preclude negation. In conclusion, then, while Singh (1990) offers some insights into the phenomena that seem preferable to a purely syntactic account, there clearly remains more work to be done.

[18] An issue I have not made explicit here for lack of space is the much noticed alternation between the light verb *le* 'take' and *de* 'give', where the latter, but not the former, usually induces a beneficiary reading. This alternation can be accounted for straight-forwardly under the analysis in Chapter 5: both light verbs are postulated as having two arguments, an actor and a receiver/goal, but in the case of *le* 'take', the second argument is coindexed with the first, so a beneficiary reading does not arise.

A close look at the literature on complex predicates, the possibilities of combinations and usage (see particularly Hook 1974 for a good compilation and description) reveals any number of further interesting issues. One of these is the possibility of the reversal of main and light verb. An example of such a reversal, taken from Hook (1974:9) is shown in (76).

(76) a. us=ne j^{h}aṭke=se lagaam **k^{h}ĩ̃ic d-ii**
Pron=Erg force=with reins.F=Nom pull give-Perf.F.Sg
'He jerked back on the reins.'

b. us=ne j^{h}aṭke=se lagaam **de k^{h}ĩ̃ic-ii**
Pron=Erg force=with reins.F=Nom give pull-Perf.F.Sg
'He jerked back on the reins.'

The possibility of this type of reversal would actually be predicted by the argument structure analysis formulated in Chapter 5, in combination with the phrase structure rule posited for the generation of complex predicates, repeated here in (77).

(77) $\overline{\text{V}} \longrightarrow$V (V) (Stat) (Aux)

However, it is not clear to me what these kinds of 'light verb reversals' signify, or what the difference in meaning between (77a) and (77b) might be. In the long run, a theory of complex predicate formation should be able to provide a precise account for the kind of data in (77).

Another interesting question is raised by the examples in (78) and (79). The complex predicates *aa ga-yii* 'come went' and *uṭ^{h} baiṭ^{h}-ii* 'rise sat' in (78a) and (79a) respectively are well-formed. Notice that each of the main verbs in (78a) and (79a) can also function as light verbs.

(78) a. naadyaa aa ga-yii
Nadya.F=Nom come go-Perf.F.Sg
'Nadya arrived.'

b. *naadyaa jaa a-yii
Nadya.F=Nom go come-Perf.F.Sg
'Nadya went.'

(79) a. naadyaa uṭ^{h} baiṭ^{h}-ii
Nadya.F=Nom rise sit-Perf.F.Sg
'Nadya got up.'

b. *naadyaa baiṭʰ uṭʰ-ii
Nadya.F=Nom sit rise-Perf.F.Sg
'Nadya got up.'

However, the other logical (reversed) possibilities for forming complex predicates with these verbs are ill-formed. This is shown in (78b) and (79b).[19] Intuitively the restriction would appear to be of a directional type: coming before going in (78a) is okay, but not going before coming in (78b). This again remains an issue for future research.

As already mentioned, the next chapter presents an argument structure analysis which can account for most of the data discussed in this chapter. While there remain some further issues to be explored, I believe the analysis of complex predicate formation I propose will provide the concrete basis needed for subsequent research.

[19]The example in (79b) is actually well-formed under a non-complex predicate reading. If it is interpreted as being an instance of a *kar* 'having' participial adverbial in which the *kar* 'having' is optional, the sentence is good. The appropriate reading in this case would be: *Anjum got up after having been sitting.*

Chapter 5

Complex Predicate Formation

5.1 Introduction

This chapter describes an argument structure approach to complex predicate formation within LFG. Borrowing heavily from Jackendoff's (1990) theory of Conceptual Semantics, I postulate an elaborated level of a(rgument)-structure and formulate a linking procedure in Chapter 7 which not only relates a-structure to f-structure, but also a-structure to c-structure. The relationship between c-structure and f-structure remains essentially as formulated by Kaplan and Bresnan (1982), but is modified somewhat to allow for discontinuous heads, as in the Urdu permissive, and to allow for a greater reliance on the information provided by morphological case marking in a free word order language like Urdu.

Chapter 3 demonstrated that the differences between the permissive and the instructive are most clearly expressed at the level of f-structure. That is, the two predicates of the permissive form a complex predicate, in the sense that it functions like a single PRED at f-structure, while the instructive consists of a matrix PRED, which takes a complement (an XCOMP). The problem posed by the permissive for LFG, and theories of syntax in general, is how two or more discontinuous syntactic heads may be represented as the head of a monoclausal sentence. Within LFG, it was seen that two or more syntactic heads must somehow combine to form a single PRED at f-structure.[1] However, the early LFG

[1]Within a structural approach like Government-Binding, one solution to the problem of complex predicates would seem to lie in working out how a process like head-to-head movement could apply to the permissive, but not to the instructive, and still be able to account for the fact that the constructions do not differ with regard to scrambling, negation, and coordination. An argument structure approach along the lines of Grimshaw (1990) and S. Rosen (1989), or an analysis relying on θ-grids as in Li (1990) for Chinese

formalism developed by Kaplan and Bresnan (1982) does not allow two or more predicates at c-structure to unify into a single f-structure PRED, as required by the Urdu permissive.

Recent work in LFG has sought to tackle the issues raised by complex predicates like the Urdu permissive, or Romance causatives and restructuring verbs. Butt, Isoda, and Sells (1990) proposed that complex predicate formation take place at a-structure, and that a-structure be related to c-structure through a projection called α.[2] The α projection was designed to work in parallel with the function (projection) ϕ, which formally relates c-structure to f-structure representations. In addition, a projection ϕ', the inverse of α composed with the function ϕ, was defined. This projection formally related the level of a-structure to f-structure, and allowed a combination of the argument structures of discontinuous syntactic heads into a single PRED at f-structure. However, this approach was found to be ultimately unworkable. For one, the problem of complex predicate formation was ultimately only restated at the level of a-structure, but the issue of how two discontinuous syntactic heads could be related to a single PRED at f-structure was not actually solved.

Subsequently, several different proposals have been put forward. Andrews and Manning (1993) point out that while the classic LFG architecture was able to deal well with certain phenomena like discontinuous constituents because of the fact that information from separate c-structure nodes could be unified into a single f-structure component, this collapsing of structure also caused information to be lost that needed to remain distinct. In particular, they addressed problems from stacked adjectival modification and Romance complex predicates, and proposed an architecture of LFG which distinguished between semantic (O) and syntactic heads (H), and held all levels of representation to be *copresent*, so that the relevant information can always *spread*, and not become inaccessible, as in classic LFG. While the proposal put forward by Andrews and Manning (1993) is indeed able to solve many of the problems associated with complex predicates successfully, I do not adopt it here. The representation of functional, phrasal, and semantic information within the same structure would seem to go against the spirit of clearly factoring out distinct levels of representation, and then investigating how the syntactic processes stated at these separate levels

resultatives, would be able to successfully differentiate between the permissive and the instructive in terms of complex predicate formation at argument structure. However, as a difference in argument structure is generally taken to correspond to a phrase structural difference, the identical structural properties of the two constructions would still present a challenge.

[2]Kaplan (1987) described the relationship between c-structure and f-structure in terms of a *projection* architecture. The formal algorithm relating the levels of representation is defined as a *projection* ϕ, which 'projects' information from c-structure to f-structure.

of representation interact with one another.

Kaplan and Wedekind (1993) see the problem of complex predicates as consisting of the framework's inability to relate a complex semantic (or argument) structure to a simple f-structure. A *restriction operator* is therefore defined, which is able to pick out a part of an f-structure. The simple f-structure representation of the sentence in (1) can thus be decomposed into parts and related to the corresponding parts of the complex argument structure: the arguments *Saddaf* and *letter* of *write*, for example, can be related to the partial f-structure consisting of only the PRED *let-write*, the indirect object (OBJ2) *Saddaf*, and the object (OBJ) *letter*. The subject is factored out through the restriction operator.

(1) anjum=ne saddaf=ko xat lik^{h}-ne di-yaa
Anjum.F=Erg Saddaf.F=Dat letter.M=Nom write-Inf.Obl give-Perf.M.Sg
'Anjum let Saddaf write a letter.'

However, this approach requires that the lexical entry of each main verb in the language specifies how the grammatical functions of the verb are affected when it combines with the permissive *let*. Presumably, then, since many different types of complex predicates exist in Urdu, and since they also interact with one another, the lexical entry for each main verb in the language would need to specify what exactly happens in terms of complex predicate formation for each of the different light verbs it might combine with. As light verbs can also "stack" (a permissive can be combined with an Aspectual light verb), the range of possible combinations would also need to be stipulated lexically. This approach, then, is clearly undesirable as it is not able to provide any kinds of generalizations about complex predicate formation (see Butt (1994) for a more detailed discussion) and must therefore rely on unneccsary lexical stipulation.

Dalrymple, Lamping, and Saraswat (1993) introduce a method of semantic composition for complex predicates, which relies on linear logic representations for assembling and composing semantic information specified in the lexical entry of an item. The results of semantic composition can then be interpreted in a way that is consistent with the principles of LFG. This approach is essentially compatible with my analysis of Urdu complex predicate and in fact, tackles a problem not explicitly treated in this dissertation. I am mostly concerned with the mapping between a-structure, f-structure, and c-structure, while Dalrymple, Lamping, and Saraswat assume the existence of principles for the mapping between a-structure and f-structure, in order to tackle the problem of semantic interpretation.

My approach most closely resembles that formulated in Alsina (1993). Alsina examines

Romance causatives and argues that they, like the Urdu permissive, must be formed in the syntax, and not in the lexicon. He formulates a process of *Predicate Composition*, which allows the combination of distinct argument structures in the syntax. A complex argument structure thus composed in the syntax is related to the level of f-structure by the *Functional Mapping Theory* formulated by Alsina. While the analysis of Urdu complex predicates is very close to Alsina (1993), Romance differs from Urdu in that the c-structure constrains complex predicate formation rather strictly in Romance. These constraints cannot be applied straightforwardly to Urdu, but must be loosened and reformulated.

Most of the approaches described above utilize a level of argument structure in some form or another.[3] Indeed, the idea that complex predicates involve processes at the level of argument structure is not new, nor is it confined to work within LFG. I will therefore not spend much time here justifying an argument structure approach to complex predicates. The previous two chapters provide clear evidence for an argument structure approach. For one, both the permissive and the Aspectual complex predicates show that more than one predicate is contributing to the overall argument structure of the complex predicate. In the permissive, at least one argument is always clearly contributed by one predicate, but not the other. In the Aspectual complex predicates, it was seen that the light verb plays a role in determining the case-marking on the subject and would therefore seem to be contributing to the argument structure of the complex predicate in some way.

I do, however, diverge from the current trend to represent as little information at the level of argument structure as possible. Within T. Mohanan's (1990) framework of multidemensional representation, argument structure is argued to consist merely of slots, which are linked to a more elaborate semantic representation. The same idea is found in Grimshaw (1990), S. Rosen (1989), and Ritter and S. Rosen (1993), who represent argument structure as a hierarchical organization of variables like 'x' and 'y'. These variables are linked to a full-blown LCS (Lexical Conceptual Structure). Detailed information is not taken to play a significant role at argument structure; it is merely the hierarchical organization of variables which is significant. A remarkable feature of this approach is that it allows empty argument structures, particularly for Romance restructuring verbs (S. Rosen 1989) and the Japanese light verb *suru* 'do' (Grimshaw and Mester 1988). These constructions are discussed in

[3]Matsumoto (1992) provides an account of Japanese complex predicates within LFG and untilizes a level of argument structure as well. However, he is not explicitly concerned with the relationship of one level of representation to another.

some detail in Chapter 7, where I show that the approach to complex predicates formulated in this chapter can be extended to provide an account for Romance restructuring verbs and Japanese light *suru* 'do'. However, an argument structure which abstracts away entirely from semantic information cannot easily account for the facts presented by Urdu Aspectual complex predicates, where volitionality and inception/completion influence the case marking of the subject, and were seen to preclude certain possibilities for complex predicate formation.

Alsina (1993) also argues for a more abstract representation, but does represent some semantic information directly at a-structure. In particular, he bases the representation of arguments on Dowty's (1991) system of proto-roles. Alsina thus achieves a medium between the traditional use of θ-role labels, and Grimshaw's (1990) and T. Mohanan's (1990) proposals that no semantic information whatsoever be encoded at argument structure. He classifies arguments either as *proto-agents* or *proto-patients* and derives principles of causative formation and mapping from a-structure to f-structure from the relatively abstract semantic information provided by a-structure representations.

As already mentioned, my approach to complex predicate formation in the syntax is very close to Alsina's; however, I differ on the matter of argument structure representation. Rather than arguing for more abstraction at argument structure, I propose to adopt LCS representations (Jackendoff 1990) directly for argument structure representation. The details of this *elaborated a-structure* approach are described in this chapter. Once the representation of argument structure has been established, I formulate a process of complex predicate formation, and then go on to describe principles, which link a-structure, f-structure, and c-structure to one another in Chapter 7.

5.2 Representation of Argument Structure

5.2.1 General

As already mentioned, the elaborated level of a-structure I propose borrows heavily from Jackendoff's (1990) theory of Conceptual Semantics. In Jackendoff (1990), the organization of information in Lexical Conceptual Structures (LCS) is reformulated, partly with the goal of developing a theory of linking, which determines how a given LCS may be linked up to the syntax. In his discussion of linking procedures, Jackendoff explicitly mentions that the theory of Conceptual Semantics should be easily compatible with any theory of syntax,

including LFG.

The work presented in this chapter thus represents an experiment which attempts to combine a version of Conceptual Semantics with the formal architecture of LFG. Furthermore, although Jackendoff intends Conceptual Semantics to apply crosslinguistically, most of the discussion in Jackendoff (1990), and earlier work, is limited to data from English. This chapter therefore also examines the viability of the application of LCSs to Urdu. In addition, complex predicates present an entirely new range of constructions to be examined within the framework of Conceptual Semantics.

Many of the ideas and formalisms developed in Jackendoff (1990) are particularly attractive for a treatment of Urdu complex predicates. For example, the semantic factors of conscious choice (volitionality) and inception/completion, which were found to play a role in Aspectual complex predicate formation, are easily expressed with the help of LCSs. Jackendoff also defines a process of *Argument Fusion*, which I redefine and extended to account for complex predicate formation. The basics of the linking process, which Jackendoff designs to be compatible with Government-Binding, can similarly be redefined and extended to be compatible with LFG's mapping theory from a-structure to f-structure. Finally, the detailed semantic information contained in an LCS allows a formulation of semantic case marking in Urdu, which has been argued for in T. Mohanan (1990) and Butt and King (1991). The formulation of case marking principles based on semantic information in turn allows a cleaner account of the principles relating functional annotations to c-structure representations in a free word order language like Urdu.

It should be noted, however, that while I adopt structures which look very much like the LCSs in Jackendoff (1990), I do not here adopt the entire theory of Conceptual Semantics. Rather, I only make use of those pieces of the formalism which are needed for a representation of argument structure. I do not propose to represent any information not directly required for linking or case marking purposes. In effect, then, I am merely adopting a subset of the theory of Conceptual Semantics: the fact that the structures I use look similar to Jackendoff's LCSs does not imply that I have adopted other parts of the theory. My adaptation of Jackendoff (1990) is therefore exactly as described in the next few sections. I furthermore refer to the level of information represented as *a-structure*, not Lexical Conceptual Structure. This is in part to make clear that I do not base myself entirely on Jackendoff (1990), but only adapt a subset of the ideas for my purposes, and in part to emphasize that the information I represent is exactly the information pertinent for

argument structure processes.

5.2.2 Levels of Representation

Jackendoff argues that his conception of LCSs renders a separate level of argument structure superfluous. An LCS can do all the work expected from an argument structure. Another advantage, in his view, is that he does not need to posit an 'arbitrary' thematic role hierarchy. Rather, a hierarchy falls out from the way an LCS is organized. In fact, he argues, Conceptual Semantics gives a meaning to the rather arbitrary thematic role labels generally used in theories of syntax.

While the idea that LCSs can finally provide us with some graspable semantic reality for the arguments of a predicate is without question appealing,[4] Jackendoff's claim that a separate level of argument structure is not needed is not as clear cut. For example, take the abbreviated LCS for one version of the simple verb 'give' in (2). It shows that the verb 'give' is of category V and describes an *Event*, as opposed to a *State*, for example. The Event described is one in which an actor α causes (CS) something to go into the possession of a beneficiary β. The first line, headed by the function CS represents the *Thematic Tier*, which essentially encodes the meaning of the verb. The second line, the *Action Tier*, which is headed by the function AFF indicates actor/patient/beneficiary relations. In this case, the use of the "+" on AFF indicates that there is an actor and a beneficiary (a "-" would indicate an actor and a patient). The arguments of AFF are generally coindexed with a slot on the Thematic Tier. The first argument of AFF, the actor, is thus also identified as the causer of the action by means of coindexation with the Greek letter α. The beneficiary of the action is coindexed through the Greek letter β with a locational goal, the argument of TO, on the Thematic Tier. Notice that only one of the coindexed slots may be empty. There are thus three empty square brackets ([ ]) in (2). These represent unfilled argument slots and indicate that *give* has three arguments.

(2)

$$\left[\begin{array}{l} \text{give} \\ \text{V} \\ \left[\begin{array}{l} \text{CS}([\alpha], \text{GO}_{Poss}([\ ]_A, \text{TO}[\beta])) \\ \text{AFF}^{+}([\ ]^{\alpha}_{A}, [\ ]^{\beta}_{A}) \end{array}\right]_{EVENT} \end{array}\right]$$

[4] See Ramchand (1993) for a theory which rigorously defines θ-role labels semantically within a theory of aspect and argument structure.

These three arguments are clearly marked with an **A** in order to distinguish them from empty adjunct slots in a given LCS. Jackendoff thus actually has an argument structure embedded within the LCS in (2). Essentially, the issue is one of representation. Instead of positing a semantic structure and a separate argument structure, Jackendoff represents the pertinent information within a single picture, within a single level of representation.

Recall that the common thread in a variety of literature (Grimshaw (1990), S. Rosen (1989), Ritter and S. Rosen (1993), T. Mohanan (1990), Alsina (1993)) is to represent no semantic information, or as little as possible, at the level of argument structure. In these approaches, such underspecified argument structure slots are linked to an LCS, or something akin to an LCS, which provides the detailed semantic information for the NPs associated with a slot at the argument structure. These approaches thus all explicitly separate out the semantics from a level of argument structure.

In contrast to these approaches, I believe that the semantic information directly relevant to argument structure processes, such as causativization, passivization, complex predicate formation, and also the determination of semantic case, should be represented exactly at the argument structure, and not have to be retrieved from another level whenever necessary. In fact, although the approaches formulated in Grimshaw (1990), S. Rosen (1989), or Ritter and S. Rosen (1993) purport to make use of no semantic information, a close examination of the analyses reveal that semantic information is actually relied on implicitly.[5]

As this chapter progresses, it will become clear that there is in fact an abstract level of argument structure contained within the elaborated a-structures I propose. This is the list of arguments available for linking once argument fusion or control has taken place. But, as these arguments stand in a one-to-one correspondence to the grammatical functions at f-structure, a representation of the skeletal/abstract list of hierarchically organized arguments at a separate level appears to be redundant.

[5]S. Rosen's account of complex predicates, for example, relies on the fact that only argument structures containing an *Event* can merge with another argument structure. However, this is never explicitly made clear. Furthermore, the distinction between external and internal arguments, which she indicates by means of different bracketing conventions at argument structure, must also rely on semantic information in noncanonical cases.

5.2.3 Adaptations of Functions

For my purposes, I assume that any empty slot in an a-structure represents an argument. I thus dispense with the rather stipulative **A** subscript employed by Jackendoff to distinguish arguments from adjuncts. As I do not treat adjuncts in this dissertation, this particular revision is not a problem within the current scope of investigation. Furthermore, although I do not address adjuncts, I believe that once a viable theory of linking from a-structure to c-structure is articulated, both arguments and adjuncts can be distinguished at a-structure without stipulations as to which is which. An LCS for the English verb *give* without the **A** subscripts is shown in (3).

(3)

$$\left[\begin{array}{l} \text{give} \\ \text{V} \\ \left[\begin{array}{l} \text{CS}([\alpha], \text{GO}_{Poss}([\ \ ], \text{TO}[\beta])) \\ \text{AFF}^{+}([\ ]^{\alpha}, [\ ]^{\beta}) \end{array}\right]_{EVENT} \end{array}\right]$$

A feature of the LCS in (3) I will not adopt is the category information specified on the second line. Jackendoff makes use of the category information for linking purposes, but within an LFG approach, the a-structure information is just one part of a given item's lexical entry. I therefore take the category information to be specified separately in the lexical entry.

As mentioned, the two lines contained within the square brackets labeled EVENT represent the *Thematic Tier* and the *Action Tier*. The Thematic Tier provides the basic meaning of the lexical item in question. The Action Tier contains some of the same information as the Thematic Tier, but it is formulated to highlight actor/patient/beneficiary relationships. Information about motion and location, for example, are never represented on the Action Tier, only on the Thematic Tier.

The Action Tier's main function within Jackendoff's system is to represent the relationship of the actor/patient/beneficiary roles, so that these particular arguments will be primary in terms of linking. That is, when it comes to determining the ranking of arguments for linking to syntax, the arguments represented at the Action Tier are always ranked higher than the arguments represented at the Thematic Tier. The Action Tier thus ensures that those types of roles, which canonically appear as subjects and direct objects (i.e., those

which are assigned Structural Case), are indeed canonically linked to subjects and direct objects. Another way of viewing the Action Tier is that it essentially encodes the analog to Dowty's (1991) idea of proto-roles. Predictably, Jackendoff encounters some difficulty in the examination of psych predicates. Here the argument linked to the subject is generally an experiencer, not an actor/agent. Jackendoff therefore invents a function REACT, which essentially inverts the order of roles found in the function AFF (Affect) in (3) above. Once the order of roles is 'reversed', linking can proceed straightforwardly.

For my purposes, the Action Tier is useful in that it allows an expression of volitionality/conscious choice in addition to facilitating the formulation of linking principles. However, the argumentation in favor of the function REACT is not convincing, and the account of pysch predicates it allows is not complete. In Urdu, the subjects of psych predicates are always marked with the dative case. The dative case always appears on goals, both abstract and locational (T. Mohanan 1990). An account of psych predicates can thus be given through a combination of semantic case assignment and already existing linking principles, which makes the existence of the function REACT at the Action Tier superfluous. I believe that the Urdu pattern also provides the basis for an account for psych predicates in languages like English and Romance; however, I do not investigate this topic here.

Notice that AFF has been referred to as a *function* in the discussion above. Jackendoff defines all the capitalized components of an LCS as functions which take certain arguments and map them onto a region or place. While I do not here make use of the actual mapping mechanism, I adopt the terminology. Functions like AFF, CS, or GO often have different instantiations. The function AFF, for example, has four manifestations: AFF^+, AFF^-, AFF^0, and AFF. The function AFF^+ takes two arguments, an actor and a beneficiary (for verbs of helping), AFF^- denotes an action involving opposition (an actor and a patient), and AFF^0 is meant to denote the actor/patient roles in permissive (letting) events. Finally, AFF is the underspecified version.

In addition, the function AFF must not always take exactly two arguments. In the representation of an unergative verb like *dance*, for example, the Action Tier may only contain one argument, as shown in (4) below. The Action Tier for an unaccusative verb like *come*, on the other hand, would be as in (5). The difference between the two representations is that in (4) the argument α is the *first* argument of AFF, while in (5), it is the *second* argument. In (4) an actor argument is represented, while (5) depicts a theme.

(4) AFF([α],)

(5) AFF(, [α])

The Action Tier thus allows for quite a variety of representations. This is crucial not only for a successful analysis of Urdu complex predicates, but also for a treatment of the Romance and Japanese complex predicates examined in Chapter 7.

Like the function AFF, there are several instantiations of the function CS (cause): CS$^+$, CS$^-$ and CS^u. CS$^+$ denotes the application of force with a successful outcome,[6] CS$^-$ denotes an unsuccessful outcome, and CS^u an undetermined result. I will simply adopt these, but represent CS^u as the unmarked CS.

Jackendoff uses CS to describe both the application of force within lexical items like 'make' (cause to exist) and 'kill' (cause to die), and the application of force denoted by overtly realized causatives, either morphological or syntactic ('make cook'). It is not quite clear to me how, or if, the causation implicit in lexical items like 'kill' differs from overt causatives like 'make die'. It seems desirable that lexical decomposition be constrained (Fodor 1970), and that perhaps there should be an explicit difference formulated between overt causatives like the English *make repair*, or the Italian *fare riparare* 'make repair', and the meaning of causation implicit in a verb like *give*. Clear cut evidence for such a distinction cannot be found in Urdu, as a large number of transitive verbs are formed from intransitives with the causative morpheme *-aa*. Some examples are *ban* 'be made' and *banaa* 'make', *jal* 'be burning' and *jalaa* 'burn', *mar* 'be dead' and *maaraa* 'hit/kill'. However, there is also another causative morpheme *-vaa*, which requires a causer as well as a causee. While there may arguably be both a semantic and syntactic difference in English between 'kill' and the periphrastic causative 'make die', the two causative morphemes in Urdu seem to be quite closely related. The distinction between the two Urdu causatives can in fact be analysed within Alsina and Joshi's (1991) argument structure approach to causatives. Two types of causatives are identified crosslinguistically: a two place causative (Causer Event), and a three place causative (Causer Causee Event). Some languages like Romance may have one or the other, while some languages like Malayalam may have both. Urdu would appear to have both. Within this dissertation, I will simply use the function CS for both the causatives overly marked by the morphemes *-aa* or *-vaa*, and the causation implicit in verbs like *de* 'give'. If the need for a distinction between the two types should arise, the variant

[6]This notation is equivalent to Jackendoff's (1983) former formulation of CAUSE. The differing versions of CS represent a finer grained analysis.

CAUSE might be used to denote overtly realized causation. However, the analysis of Berber causatives presented in Alalou and Farrell (to appear) in terms of Jackendovian LCSs shows that the unified treatment of causation implicit in Jackendoff's approach provides the basis for a successful analysis of Berber causatives.

One other issue to be mentioned in connection with the LCS in (3) is the differing semantic fields a given function may range over. The function GO, for example, is specified as pertaining to *possession* in (5). Other typical semantic fields a function might be interpreted over are spatial, temporal, or identificational (Jackendoff 1990:26). I will make use of the notion of differing semantic fields that a function may range over. In particular, I will introduce GO_{Info} for the a-structure representation of the instructive.

Finally, recall that the Action Tier takes precedence in terms of linking. This is expressed by the manner in which arguments are bound to one another. In (3) the first argument of CS is filled with α, and bound to the first argument of AFF^+ by coindexation, and not vice versa. Thus, the argument of AFF^+ takes precedence. The argument of TO is similarly bound to the second argument of AFF^+ by the β index. Furthermore, if the argument of TO were not bound by anything on the Action Tier, the verb *give* would have the equivalent of a goal argument, rather than that of a beneficiary, as in (1). Jackendoff uses these differing possibilities of coindexation to formulate an account of the English Dative Alternation. As there is no equivalent to the English Dative Alternation in Urdu, I do not utilize the distinction between a goal and a beneficiary, but simply always treat all argument of TO uniformly as goals. However, I do adopt the coindexation of arguments and the primacy of the Action Tier in terms of linking.

5.2.4 Argument Structure Hierarchy

The primacy of the Action Tier, coupled with the ordering of arguments from left to right in the order of their appearance, determines a hierarchy of arguments. The theory of linking developed by Jackendoff relies on this implicit argument structure hierarchy. While I do not adopt Jackendoff's formulation of linking principles, I do adopt the implicit hierarchy of arguments as determined by the structure of an LCS, or a-structure. Figure (6) lists the arguments of functions in the order of most prominent to least prominent, and also shows the traditional θ-role label which most closely corresponds to the argument of a particular function. For example, the first argument of GO_{Poss} in an LCS like (3) is always a theme,

the first argument of AFF$^+$ is an actor, and the second argument of AFF$^+$ is a beneficiary.[7]

(6)

Position at LCS	**Corresponding Role**
First argument of AFF	actor
Second argument of AFF$^-$	patient
Second argument of AFF$^+$	beneficiary
First argument of CS	agent
First argument of Location and Motion functions	theme
Argument of TO	goal
Argument of FROM	source

The ranking of arguments shown in (6) essentially corresponds to the thematic hierarchy in (7).[8] Note that the labels *patient* and *beneficiary* are listed together. This, however, does not imply that the hierarchy is indeterminate with regard to these roles, rather, it indicates that either a patient or a beneficiary will be realized, but never both at the same time.[9]

(7) Actor > Patient, Beneficiary > Theme > Location, Source, Goal

Jackendoff's thematic hierarchy differs somewhat from the hierarchy in (8), which is argued for within LFG (Bresnan and Kanerva 1989).

(8) Agent > Beneficiary > Recipient/Experiencer > Instrumental > Theme/Patient > Location

Within the scope of this dissertation, the differences between the hierarchies do not appear crucial. That is, there seems to be sufficient overlap in the two hierarchies, so that the linking of an a-structure based on Jackendoff's framework, and therefore an implicit usage of (7), by means of a Mapping Theory based on (8) does not cause any problems within my analysis of Urdu complex predicates.

[7]The first argument of CS, an agent, in practice is always coindexed with the first argument of AFF, an actor.

[8]While Jackendoff does not make explicit use of an argument structure hierarchy, the system of organization implicitly clearly defines a thematic hierarchy.

[9]A reviewer notes that sentences like *Peel me a grape.* would appear to be counterexamples to Jackendoff's claim, but that he might argue that *grape* here is actually a theme. Given the slippery distinction between patients and themes, all apparent counterexamples to his claim might then arguably involve themes, rather than patients.

5.2.5 The Aspect Tier

Jackendoff (1990) discusses a theory of Events, States, and Inchoatives at some length, and develops the system further in Jackendoff (1991). In particular, he combines a notion of boundedness with a consideration of the internal structure of nouns (mass, aggregate, etc.) to arrive at a characterization of verb types. He is not only able to represent the verb types identified by Vendler (1967), but is also able to identify some additional types.

However, while Jackendoff uses the notion of boundedness to identify both endpoints and *inceptions* of events, the formalism cannot be readily applied to the inception/completion facts of the Aspectual complex predicates examined in Chapter 4. A crucial distinction exists between a verb like *commence* (see Jackendoff (1991:40) for a treatment of this verb), and an Aspectual light verb like *paṛ* 'fall'. The former indicates that an action is beginning by embedding an event structure within its own event structure. The light verb *paṛ* 'fall', on the other hand, picks out the precise point of inception of the event, but does not embed that event within an event structure of its own.

I therefore do not adopt the details of Jackendoff's analysis of event types and instead propose the presence of an *Aspect Tier*, in addition to the Thematic and Action Tiers, at a-structure. The Aspect Tier contains only one function ASP, which has three slots. Each of these slots can be specified either positively, with a '1', or negatively with a '0'. They can also be unspecified and be left empty. Some possibilities are shown in (9) and (10). The first slot represents the starting point of an event, the second slot the duration, and the third slot the end point.

(9) ASP (1 _ _)

(10) ASP (0 _ _)

The Aspect Tier in (9) indicates a verb positively specified for inception and underspecified for both duration and completion. This is thus an appropriate representation for the *paṛ* 'fall' type light verbs. Recall that the verb *so* 'sleep' could not combine with the light verb *paṛ* 'fall' and I concluded that this particular main verb must be negatively specified for inception. The Aspect Tier in (10) is therefore representative of a verb like *so* 'sleep'.

This system allows twenty-seven possible aspectual specifications, which is a great deal more than usually postulated in the literature on aspectual classes (Vendler 1967, Dowty 1979). I do not, however, propose that each of the possible specifications be representative of a lexical class. Take, for example, the Aspect Tier in (11). It indicates a predicate which

is negatively specified for inception, positively specified for duration, and underspecified for completion. A predicate which might correspond to this Aspect Tier is shown in (12).

(11) ASP (0 1 _)

(12) naadyaa so rah-ii t^{h}ii
Nadya.F=Nom sleep Stat-F.Sg be.Past.F.Sg
'Nadya was sleeping.'

Here the predicate is composed of a base form *so* 'sleep', a stative marker *rah* 'stay' and the past tense auxiliary. The Aspect Tier in (11) represents a combination of the various lexical specifications of the predicate's components: the negative specification on inception is contributed by the main verb *so* 'sleep', and the positive specification of duration by the stative marker. In the next section, I propose that argument structure information be unified in the process of complex predicate formation. I take the same notion of unification to hold for the combination of aspectual information in simple predicates.

The Aspect Tier thus allows a ready expression of the patterns discussed in Chapter 4. Some main verbs, like *so* 'sleep', were seen to be negatively specified for inception. These verbs could not combine with a light verb like *paṛ* 'fall', which are positively specified for inception. A full account of the role of the Aspect Tier in complex predicate formation is shown in the next section. Note that the formulation of an Aspect Tier does not by any means constitute a deep analysis of aspectual phenomena in Urdu. This remains to be undertaken. In the meantime, I regard the Aspect Tier as a tool which allows a first understanding of the aspectual phenomena at hand.

The figure in (13) illustrates my proposal for the representation of a-structure, using the already familiar main verb *de* 'give' as an example. There are now three tiers: the Thematic Tier, the Action Tier and the Aspect Tier. Arguments are no longer marked explicitly by the subscript **A**, and the recipient is not taken to be a beneficiary, but a simple goal.

(13)

$$\begin{bmatrix} \text{de 'give'} \\ \begin{bmatrix} \text{CS}([\alpha], \text{GO}_{Poss}([\ ], \text{TO}[\ ])) \\ \text{AFF}([\ ]^{\alpha},) \\ \text{ASP}(\_\ \_\ \_) \end{bmatrix}_{E} \end{bmatrix}$$

As a final remark on basic issues of representation, I would like to point out that while I have kept my representation of a-structure very close to Jackendoff's proposals for Lexical

Conceptual Structures, the format in (13) is not incompatible with the general utilization of attribute value matrices (AVM) for the representation of information in LFG. The a-structure in (13) could easily be expressed as an AVM.

5.3 Complex Predicate Formation

5.3.1 Representation of Light Verbs

A common thread in the literature on complex predicates has been that one of the components of the predicate is judged to be *light* or *incomplete* in some sense. That is, one of the verbs in the complex is classified as being something inbetween an auxiliary and a main verb (Jespersen 1954, Masica 1976, Cattell 1984, Grimshaw and Mester 1988, S. Rosen 1989, Alsina 1993).[10] It is also striking that in language after language (Japanese, Korean, Chinese, Hindi/Urdu, Marathi, Bangla, Malayalam, Tamil, Tepehua, Sranan, Yoruba) the same basic set of lexical items tend to participate in complex predicate or serial verb formation. Typical examples are *give*, *take*, *go*, *come*, *put*, *hit*, and *fall*.

In particular, the set of light verbs crosslinguistically also tend to have "full" or main verb analogs which are phonetically identical to the light verbs. Urdu is a typical example of this. The full verb *de* 'give' has not only one, but two light verb analogs: the permissive *de* 'let', and the light verb *de* 'completion/volitionality/forcefulness' participating in Aspectual complex predicate formation.

The intuition has been that the light verb is "lighter" in some sense than the corresponding full verb. That is, the light verb is a *semantically bleached* version of a corresponding full verb. While this intuition has not been formalized in the literature on South Asian languages, Grimshaw and Mester (1988) and S. Rosen (1989) express the equivalent of this intuition by proposing that light verbs are light because they have either a completely empty, or merely an incomplete, argument structure. Alsina (1993) treats light verbs as *Incomplete Predicates* which must combine with another argument taking predicate in order to be syntactically well-formed.

While I also treat light verbs as Incomplete Predicates, I do not believe that an empty argument structure, as proposed in Grimshaw and Mester (1988), or S. Rosen (1989), provides the basis for a successful account of complex predicate formation. The evidence

[10] In one tradition, which Hook (1974) follows in his examination of Hindi V-V complexes, the 'light' verb is refered to as a *vector* verb, a distinct type of auxiliary.

from Urdu has shown that neither the Aspectual light verbs, nor the permissive, can be analyzed as having a completely empty argument structure. Furthermore, in Chapter 7 I show that the analysis of both Japanese *suru* 'do', which is examined in Grimshaw and Mester (1988), and the Italian restructuring verbs examined in S. Rosen (1989), follows more naturally from an account based on the elaborated a-structure described in this chapter.

In particular, I believe that the distinction between the Thematic Tier and the Action Tier at a-structure provides exactly the right mechanism for the intuition prevalent in the literature that a light verb is somehow a semantically bleached version of a corresponding full verb. Since the Thematic Tier encodes the purely semantic information, while the Action Tier encodes more abstract actor/patient relationships, a natural way to express semantic bleaching is to have the Thematic Tier contain less semantic information than a corresponding full verb. Some information on the Action Tier may also be lost, but I would expect that the semantic information encoded at the Thematic Tier is "bleached" away in the process of historical change before the information at the Action Tier. A given intransitive verb like *paṛ* 'fall' thus might lose the information on the Thematic Tier, so that it no longer meant 'fall', but merely conveyed involuntariness, inception, etc. Hence it would retain an argument at the Action Tier, which would continue to play a role in case assignment in Urdu, or auxiliary selection in Italian (see Chapter 7).

The above discussion is, of course, merely suggestive. However, I believe that elaborated a-structure representations could readily provide the key to understanding the historical processes which give rise to light verbs crosslinguistically. While some work on related phenomena has been done within historical linguistics (see Givón (1979) and Lord (1976)) on related questions, this area remains to be investigated.

5.3.2 Aspectual Complex Predicates

5.3.3 Transparent Events and Fusion

As proposed in Alsina (1993), I treat light verbs as Incomplete Predicates. However, I adopt the idea formulated in Butt, Isoda and Sells (1990) that the pertinent characteristic of light verbs is the presence of a *transparent Event* at a-structure. Again, the idea of a bleaching process is suggestive for the intuition behind a *transparent Event*. A transparent Event (E_T) is an argument which corresponds to an Event in the full version of the verb. But, the Event has been whittled away at over time and now has become transparent, in

the sense that it cannot stand on its own any more: complex predicate formation must take place.

However, the notion of a transparent Event as described so far could also be taken as being essentially identical to that of a diacritic, which could simply be expressed as [+ complex predicate formation] and be contained in a light verb's lexical entry. I would like to emphasize here, however, that while it is difficult to characterize the semantic nature of a transparent Event in contrast with a "true" Event, I do believe that the notion of a transparent Event can be semantically motivated, given an explicit examination of event structure and its interaction with complex predicates. Needless to say, such an in-depth investigation goes far beyond the scope of this dissertation, so I can only attempt to describe the intuition behind the notion of a transparent Event.

A transparent Event in contrast to a simple Event has something of a deficient nature, it cannot stand on its own and must either unify with another event structure, or lean on it in some way. While causatives arguably describe two separable events (the causing event and the caused event), the connection between these two events is somehow tighter than between the respective events described by an embedded predicate and the matrix verb. The exact nature of the difference is difficult to characterize, but I believe that the notion of a transparent Event, when explored more fully can contribute to an understanding of this problem.

A graspable difference between a transparent and a simple Event, besides the triggering of complex predicate formation, in terms of the material treated in this dissertation, is that while both transparent and simple Events may function as arguments of a predicate, only simple Events may be case marked. A transparent Event argument can never receive case, and again, I would argue that this is due to its somewhat deficient nature.

In conclusion, then, I take there to be an underlying semantically characterizable difference between a transparent Event and a simple Event, but the formulation of this characterization is not something I can accomplish here. For the present, I simply take a transparent Event to be a defining characteristic of a light verb. At argument structure, a transparent Event is indicated by an "$\{\ \}_{E_T}$". Only transparent Events may trigger complex predicate formation. This is stated explicitly in (14) and (15) below.

(14) **Definition of Light Verbs:** Every light verb contains a transparent Event argument.

(15) **Transparent Event:** A transparent Event requires combination with the a-structure of another predicate and triggers *Event* or *Argument Fusion*.

As illustrated in detail in the following sections, I represent the Thematic, Action and Aspect Tiers in Aspectual light verbs as contained within an E_T, while the permissive light verb *de* 'give' contains an E_T argument at the Thematic Tier. This is representative of their differing properties. While both types of constructions are complex predicates in that two predicates combine to form a single PRED, two differing types of Fusion must be defined. Argument Fusion coindexes two arguments so that only one argument remains visible for linking purposes.[11] Event Fusion unifies the information contained within two Events. If there is a clash of information, for example, if there are contradictory specifications for conscious choice or inception/completion, then Event Fusion fails.

(16) **Event Fusion:**
An a-structure containing a transparent Event must be fused with another a-structure. This is accomplished through unification: the highest arguments of each a-structure are unified with one another, then the next highest, etc. The information at the respective Aspect Tiers of the a-structures must also be unified into a single Aspect Tier. Two arguments or Aspect Tiers with incompatible specifications may not be fused.

(17) **Argument Fusion:**
If one a-structure is embedded within another as a transparent Event, then the highest argument of the embedded a-structure is fused with lowest argument of the matrix a-structure. This is expressed by filling the embedded argument slot with the Greek letter index of the matrix argument slot.

As shown in the next section, in the case of Aspectual light verbs the transparent Event contains the Action and Aspect Tiers, so the relevant mechanism for complex predicate formation is Event Fusion. Since the a-structure of the permissive light verb contains a transparent Event argument and thus embeds the a-structure it must combine with, Argument Fusion is the relevant process here. Note that the *highest* or *lowest* status of arguments is determined by the hierarchy implicit in an a-structure: the Action Tier always takes precedence, and arguments ordered before other arguments from left to right are 'higher'. The next section proposes a-structure representation of Aspectual light verbs

[11] Note that Jackendoff also defines a process of Argument Fusion. While Argument Fusion usually applies within a single, simple LCS in Jackendoff's system, the process of Argument Fusion defined here only applies to a complex a-structure, consisting of one or more a-structures embedded within one another. However, since a transparent Event must always trigger the process of Argument Fusion defined here, I do not believe that there is a question of incompatibility. Furthermore, since Argument Fusion in Jackendoff's system and my system perform essentially the same function of rendering one argument invisible for linking purposes, I do not propose to make a distinction in terms of terminology.

and demonstrates how Event Fusion yields the right results in terms of complex predicate formation.

5.3.4 Aspectual Complex Predicates

5.3.4.1 The *paṛ* 'fall' (inceptive) type light verbs

The light verbs which form a component of the Aspectual complex predicates examined in Chapter 4 do not express much of the semantics of the corresponding full verbs. I therefore represent the Aspectual light verbs with a defective Thematic Tier. As shown in (18), for the light verb *paṛ* 'fall', the Thematic Tier contains nothing. The transparent Event, the defining characteristic of a light verb, encloses the entire a-structure. Note that in (18), the transparent Event is represented as '{ }$_{E_T}$' (an 'Event' (E) will be represented as '[ ]$_E$'). This notation is used throughout. The representation of the transparent Event in (18) expresses the fact that the event structure of Aspectual light verbs is defective in some sense as well. That is, as discussed in Chapter 4, they do not denote an event by themselves, but rather serve to focus on a particular sub-part of the event denoted by the main verb.

(18)

$$\left[\begin{array}{l} \text{paṛ 'fall'} \\ \left\{\begin{array}{l} \text{AFF}_{-\text{cc}}([\ ],) \\ \text{ASP}(1\ \_\ \_) \end{array}\right\}_{E_T} \end{array}\right]$$

The semantics equivalent to that of the full verb *paṛ* 'fall' are thus not expressed in the representation of the light verb *paṛ*. However, (18) does have an argument at the Action Tier. This argument is negatively specified for conscious choice by means of the [−cc] notation. Note that the subscript on the AFF function only ever pertains to the first argument of AFF: the second argument of AFF (patient/beneficiary/theme) may not be similarly specified for conscious choice. This part of the formalism is directly adopted from Jackendoff (1990), who argues for the need of a ±volitionality feature at the Action Tier.

Another important feature of the representation in (18) is the Aspect Tier. As can be seen, the first slot of the Aspect Tier is positively specified with a '1'. This indicates that the light verb *paṛ* is positively specified for inception, but underspecified for duration

and completion. The elaborated a-structure in (18) is thus able to accurately represent the basic properties which were found to play a role in case assignment and, therefore, argument structure. It has been observed (Hook 1974, Masica 1976, etc.) that while light verbs do represent the semantics of the corresponding full verb in some weak sense, the use of light verbs also gives rise to interpretations not immediately derivable from the coresponding full verbs. The implication of suddenness with the use of the light verb *paṛ* 'fall' is one example. The implication of intensity or violence associated with an action is another (see Masica 1976). The question of how particular light verbs can acquire such semantic interpretations, or an inquiry into exactly which parts of the semantics of the full verb are retained is an interesting one. However, it is not one I address here. I have not been able to determine that other meanings of light verbs, such as suddenness, intensity, violence, thoroughness, etc., play a role at the level of argument structure. I therefore only represent the notions of inception/completion and conscious choice in elaborated a-structure shown in (19).

Recall that the light verb *paṛ* 'fall' is only compatible with verbs which do not necessarily require conscious control over the action by the subject. For example, the light verb *paṛ* 'fall' is compatible with the main verb *gir* 'fall', but not with *banaa* 'make'. The contrast is again illustrated in (19).

(19) a. *anjum haar banaa paṛ-ii
Anjum.F=Nom necklace.M=Nom make fall-Perf.F.Sg
'Anjum fell to making the necklace impulsively.'

b. anjum gir paṛ-ii
Anjum.F=Nom fall fall-Perf.F.Sg
'Anjum fell suddenly, accidentally.'

The a-structures for the main verbs *banaa* 'make' and *gir* 'fall' are shown in (20) and (21) respectively. The main verb *banaa* 'make' is depicted as a typical transitive main verb with two arguments, an actor and a theme. Note that the argument of the function BE is not coindexed with an argument slot at the Action Tier. If I represented the thing being made on the Action Tier as well, it would be interpreted as a patient, rather than as a theme. An example of an a-structure with a theme argument is the a-structure for the main verb *gir* 'fall'. As shown in (21), the only argument of *gir* 'fall' is a second argument of AFF at the Action Tier. The verb *gir* 'fall' is thus represented as an unaccusative verb with a single

theme argument.[12]

(20)

$$\left[\begin{array}{l} \text{banaa 'make'} \\ \left[\begin{array}{l} \text{CS}([\alpha], \text{BE}[\ ]) \\ \textbf{AFF}_{+\textbf{cc}}([\ ]^{\alpha},\ ) \\ \text{ASP}(\_\ \_\ \_) \end{array}\right]_{E} \end{array}\right]$$

(21)

$$\left[\begin{array}{l} \text{gir 'fall'} \\ \left[\begin{array}{l} \text{Go}_{Down}([\alpha]) \\ \textbf{AFF}(\ \ ,[\ ]^{\alpha}) \\ \text{ASP}(\_\ \_\ \_) \end{array}\right]_{E} \end{array}\right]$$

The combination of the a-structures corresponding to the complex predicate formed by *banaa* 'make' and *paṛ* 'fall' in (19a) is shown in (22). The two a-structures in (18) and (20) are combined into the single a-structure shown in (22). Event Fusion, as defined in the previous section, in this case fuses each of the arguments of each of the a-structures with one another, from highest to lowest. It also combines the two Aspect Tiers into a single one. As the light verb *paṛ* 'fall' contains only a single argument, only this argument is fused with the highest argument of *banaa* 'make'. The a-structure in (22) shows the result of Event Fusion, and actually also that of the linking process discussed in the next chapter. I have filled in the argument slots in (22) to allow greater readability of the structures, but in fact, the actual result of Event Fusion would have empty slots where *Anjum* and *necklace* are specified in (22). Only when a-structure, c-structure, and f-structure representations are linked up with one another, can the empty argument slots at a-structure be filled.

(22)

$$\left[\begin{array}{l} \text{*banaa paṛii 'made impulsively'} \\ \left[\begin{array}{l} \text{CS}([\alpha], \text{BE}[\textit{necklace}]) \\ \textbf{AFF}_{+\textbf{cc}/-\textbf{cc}}([\textit{Anjum}]^{\alpha},\ ) \\ \text{ASP}(1\ \_\ \_) \end{array}\right]_{E} \end{array}\right]$$

[12]The verb *gir* 'fall', unlike the light verb *paṛ* 'fall' must be unaccusative as it can combine with the light verb *jaa* 'go', and never allows an ergative subject.

The argument which is the result of fusion is marked with the letter α in (22). The contribution of the light verb *paṛ* 'fall' is reflected by the positive specification for inception at the Aspect Tier, and the negative specification of conscious choice on the Action Tier. Notice that while the Aspect Tier is well-formed, the a-structure in (22) does not represent an acceptable combination because there is a clash of features at the Action Tier. The fusion of the two highest arguments in each of the a-structures of *banaa* 'make' and *paṛ* 'fall' therefore cannot be carried out successfully in (22).

Contrast this with the successful combination of a-structures in (24). Here the a-structures of the main verb *gir* 'fall' and the light verb *paṛ* 'fall' have been combined to form the complex predicate in (23). This time, there there is no clash of features as the highest argument of *gir* 'fall' is a theme and therefore is inherently negatively specified for conscious choice. Event Fusion can be applied successfully, and the complex predicate in (23) is well-formed. Note that the *Anjum* in (24) is a first argument of AFF, thus when there is a difference in argument position (first argument of AFF in the light verb *paṛ* 'fall' vs. second argument of AFF in the main verb *gir* 'fall'), the argument position indicated in the a-structure of the light verb is the one which determines the position of the argument in the complex a-structure. As will be seen in the section on the permissive, this is consistent with the process of Argument Fusion.

(23) anjum gir paṛ-ii
Anjum.F=Nom fall fall-Perf.F.Sg
'Anjum fell suddenly, accidentally.'

(24)

$$\left[\begin{array}{l} \text{gir paṛii 'fall suddenly, accidentally'} \\ \left[\begin{array}{l} \text{Go}_{Down}([\alpha]) \\ \textbf{AFF}_{\textbf{-cc}}([Anjum]^{\alpha}, \quad) \\ \text{ASP}(1 \_ \_) \end{array}\right]_{E} \end{array}\right]$$

Recall that since the light verb *paṛ* 'fall' is positively specified for inception, it cannot combine with a main verb like *bʰuul* 'forget', which is negatively (0) specified for inception. This is shown in (25). The a-structure in (26) shows that a combination of the a-structure for *bʰuul* 'forget' in (25) with the a-structure of *paṛ* 'fall' is not possible.

(25) *anjum kahaanii **bʰuul paṛ-ii**
Anjum.F=Nom story.F=Nom forget fall-Perf.F.Sg
'Anjum forgot the story.'

(26)

$$\left[\begin{array}{l} \text{b}^{\text{h}}\text{uul 'forget'} \\ \left[\begin{array}{l} \text{CS}([\alpha], \text{GO}_{Info}([\ ], \text{FROM}[\alpha])) \\ \text{AFF}([\ ]^{\alpha},) \\ \textbf{ASP}(0\ \_\ \_) \end{array}\right]_{E} \end{array}\right]$$

(27)

$$\left[\begin{array}{l} *\text{b}^{\text{h}}\text{uul paṛii 'forget inceptively'} \\ \left[\begin{array}{l} \text{CS}([\alpha], \text{GO}_{Info}([story], \text{FROM}[\alpha])) \\ \text{AFF}_{-\text{cc}}([Anjum]^{\alpha},) \\ \textbf{ASP}(\textbf{0/1}\ \_\ \_) \end{array}\right]_{E} \end{array}\right]$$

The contribution of the light verb *paṛ* 'fall' in the a-structure above is represented by the [−cc] specification on the Action Tier, and by the positive specification for inception on the Aspect Tier. While there is no clash of conscious choice features in the a-structure in (27), this time there is a clash of specifications on the Aspect Tier. The complex predicate *b*$^{\text{h}}$*uul paṛii* is thus not realizable.

I should repeat here again that I do not pretend to claim that the introduction of an Aspect Tier would be able to capture and solve the various issues to do with aspect and its interaction with argument structure. In particular, see Ramchand (1993) for a discussion of the broader issues, and a recent theory on the interaction between aspect and argument structure.

5.3.4.2 The *le* 'take' (completive) type light verbs

The *le* 'take' type of Aspectual light verbs were seen to require an ergative subject in the perfective, and to be specified positively for conscious choice and completion. As *paṛ* 'fall' was used as a representative of the inceptive type Aspectual light verbs, so I use *le* 'take' as a representative of the completive light verbs.

The a-structure representation for *le-* 'take' is shown in (28). The light verb *le* 'take' is positively specified for completion, and underspecified for inception and duration at the Aspect Tier. Like the light verb *paṛ* 'fall', it has no information on the Thematic Tier, but has one argument at the Action Tier. This argument is specified positively for conscious choice. As with the light verb *paṛ* 'fall', the light verb *le* 'take' is itself a transparent Event.

(28)

$$\left[\begin{array}{l} \text{le 'take'} \\ \left\{\begin{array}{l} \\ \text{AFF}_{+\text{cc}}([\ ],) \\ \text{ASP}(\_\ \_\ 1) \end{array}\right\}_{E_T} \end{array}\right]$$

The process of complex predicate formation and the application of Event Fusion is essentially identical to the *paṛ* 'fall' type light verbs. Recall that the light verb *le* 'take' cannot combine with a main verb that is specified negatively for either conscious choice or completion. It can, of course, combine with main verbs with are underspecified for conscious choice or completion: Event Fusion only results in an ill-formed structure when there is a clash of features. The a-structure below shows a combination of the a-structures of *banaa* 'make' and *le* 'take'. An example of a corresponding complex predicate is shown in (29).

(29) anjum=ne haar banaa li-yaa
Anjum.F=Erg necklace.M=Nom make take-Perf.M.Sg
'Anjum completed making the necklace.'

(30)

$$\left[\begin{array}{l} \text{banaa liyaa 'made completely'} \\ \left[\begin{array}{l} \text{CS}([\alpha], \text{BE}[necklace]) \\ \mathbf{AFF}_{+\mathbf{cc}}([Anjum]^{\alpha},\ ) \\ \text{ASP}(\_\ \_\ 1) \end{array}\right]_{E} \end{array}\right]$$

The a-structure in (30) is well-formed because Event Fusion was able to apply without resulting in a clash of features. The a-structure corresponding to the complex predicate in (31), however, is not possible. The one argument of the main verb *gir* 'fall' is inherently negatively specified for conscious choice, as it is a second argument, not a first argument of AFF, but the argument of the light verb *le* 'take' carries a positive specification. A clash of features, therefore, prevents the successful fusion of the two highest arguments of each a-structure: the complex predicate in (31) is ill-formed.

(31) *anjum=ne gir li-yaa
Anjum.F=Erg fall take-Perf.M.Sg
'Anjum fell completely on purpose.'

(32)

$$\left[\begin{array}{l} \text{*gir liyaa 'fall on purpose'} \\ \left[\begin{array}{l} \text{Go}_{Down}([\alpha]) \\ \textbf{AFF}_{\textbf{+cc/-cc}}([Anjum]^{\alpha}, \ \ ) \\ \text{ASP}(1 \_ \_) \end{array}\right]_{E} \end{array}\right]$$

As the mechanics for a successful combination of specifications on the Aspect Tier are essentially identical to the examples with *paṛ* 'fall' given in the previous section, I do not provide examples of a-structure combinations with feature clashes on the Aspect Tier for *le* 'take'.

5.3.4.3 The light verb *jaa* 'go'

I argued previously that the *paṛ* 'fall' type light verbs could not be analyzed as unaccusative. Crucial data came from the light verb *jaa* 'go'. The behavioral distribution of *paṛ* 'fall' and *jaa* 'go' shows clear differences in that the light verb *paṛ* 'fall' can combine with unergative main verbs, while the light verb *jaa* 'go' cannot. I therefore proposed that *jaa* 'go' should be analyzed as unaccusative. This is reflected by the a-structure for *jaa* 'go' in (33).

(33)

$$\left[\begin{array}{l} \text{jaa 'go'} \\ \left\{\begin{array}{l} \\ \text{AFF}(\quad, [\ ]) \\ \text{ASP}(\_ \_ 1) \end{array}\right\}_{E_T} \end{array}\right]$$

The a-structure in (33) shows that *jaa* 'go' contains only one argument at the Action Tier, the second argument of AFF. Recall that this is essentially equivalent to a theme argument. Since only the first argument of AFF may ever be specified positively or negative for conscious choice, the second argument of AFF is necessarily [−cc]. Furthermore, note that the light verb *jaa* 'go' is completive, as indicated by the Aspect Tier. Although this feature was not discussed previously, a quick survey of complex predicates formed with *jaa* 'go' shows that it contributes completive aspect.[13]

[13]The *jaa* 'go' is the only light verb derived from an intransitive which signals completion. It would appear to be used with all the main verbs which cannot combine with a *le* 'take' type completive main verb due to clashing conscious choice specifications.

The light verb *jaa* 'go' is only compatible with verbs whose subjects cannot be volitional. This implies that *jaa* 'go' cannot form a complex predicate with standard transitive verbs. For example, *jaa* 'go' is compatible with the main verb *ban* 'be made', but not with *banaa* 'make'. The contrast is illustrated in (34).

(34) a. *anjum haar banaa gay-ii
Anjum.F=Nom necklace.M=Nom make go-Perf.F.Sg
'Anjum completed making the necklace.'

b. kursii ban gay-ii
chair.F=Nom be made go-Perf.F.Sg
'The chair is made (is finished).'

The a-structures for the main verbs *banaa* 'make' and *ban* 'be made' are shown in (35) and (36) respectively. As is shown below, only the combination of (36) with the a-structure for the light verb *jaa* 'go' in (33) is possible. This follows directly from the requirements of Event Fusion.

(35)

$$\left[\begin{array}{l} \text{banaa 'make'} \\ \left[\begin{array}{l} \text{CS}([\alpha], \text{BE}[\ ]) \\ \textbf{AFF}_{+\textbf{cc}}([\ ]^{\alpha}, \quad) \\ \text{ASP}(\_\ \_\ \_) \end{array}\right]_E \end{array}\right]$$

(36)

$$\left[\begin{array}{l} \text{ban 'be made'} \\ \left[\begin{array}{l} \text{BE}([\alpha]) \\ \textbf{AFF}(\quad, [\ ]^{\alpha}) \\ \text{ASP}(\_\ \_\ \_) \end{array}\right]_E \end{array}\right]$$

The combination of the a-structures corresponding to the complex predicate formed by *banaa* 'make' and *jaa* 'go' in (34a) is shown in (37). The two a-structures in (33) and (35) are combined into the single a-structure shown in (37).

(37)

$$\left[\begin{array}{l} \text{*banaa gay-ii 'made completely'} \\ \left[\begin{array}{l} \text{CS}([\alpha], \text{BE}[\textit{necklace}]) \\ \textbf{AFF}_{+\textbf{cc}}(\quad, [\textit{Anjum}]^{\alpha}) \\ \text{ASP}(\_\ \_\ \_1) \end{array}\right]_E \end{array}\right]$$

Again, while the Aspect Tier in (37) is well-formed, the a-structure in (37) does not represent an acceptable combination because there is actually no first argument of AFF, which could be positively specified for conscious choice. The fusion of the two highest arguments in each of the a-structures of *banaa* 'make' and *paṛ* 'fall' therefore cannot be carried out successfully in (37).

Contrast this with the successful combination of a-structures in (39). Here the a-structures of the main verb *gir* 'fall' and the light verb *jaa* 'go' have been combined to form the complex predicate in (38). This time each of the highest arguments in the respective a-structures are themes, and thus [−cc], so there is no clash of features. Event Fusion can be applied successfully, and the complex predicate in (38) is well-formed.

(38) kursii ban gay-ii
chair.F=Nom be made go-Perf.F.Sg
'The chair is made (is finished).'

(39)

$$\left[\begin{array}{l} \text{ban gay-ii 'made completely'} \\ \left[\begin{array}{l} \text{BE}([\alpha]) \\ \textbf{AFF}(\quad, [chair]^{\alpha}) \\ \text{ASP}(\_\ \_\ \_1) \end{array}\right]_{E} \end{array}\right]$$

Again, I do not show differing combinatory possibilities involving the Aspect Tier, as they are quite straight-forward and were already demonstrated for the light verb *paṛ* 'fall'. Instead, I move on to illustrate the process of *Argument Fusion* for the formation of the permissive complex predicate.

5.3.5 The Permissive

The meaning of the permissive *de* 'let' is quite close to the meaning of the main verb *de* 'give'. The permissive has not undergone as much semantic bleaching as evident in the Aspectual light verbs. This is evident from the fact that it contributes more than one argument to the complex predicate and, in fact, determines the case marking of the subject and the indirect object in (40).

(40) anjum=ne saddaf=ko haar **banaa-ne** **di-yaa**
Anjum.F=Erg Saddaf.F=Dat necklace.M=Nom make-Inf.Obl give-Perf.M.Sg
'Anjum let Saddaf make a necklace.'

I propose that the act of permission be visualized as a metaphorical giving of an Event to somebody. I therefore take the a-structure of the full verb *de* 'give' in (41) as the basis for the a-structure of permissive *de-* 'let'.

(41)

$$\left[\begin{array}{l} \text{de- 'give'} \\ \left[\begin{array}{l} \text{CS}([\alpha], \text{GO}_{Poss}([\quad], \text{TO}[\quad])) \\ \text{AFF}_{+cc}([\quad]^{\alpha},) \\ \text{ASP}(\_\ \_\ \_) \end{array}\right]_{E} \end{array}\right]$$

The a-structure for the permissive *de-* 'let' is shown in (42). The transparent Event is in (42) is represented as an argument of the permissive. In particular, it is an argument of the function GO_{Poss} at the Thematic Tier. In this case, the a-structure of another predicate is substituted in for the transparent Event argument. Notice that this difference in Event types is the only difference between the a-structure for the full verb *de* 'give', and the permissive *de* 'let'. The idea that the Urdu permissive is semantically less bleached than the Aspectual light verb, and very close in meaning and syntax to the full verb *de* 'give' can thus be expressed easily and systematically at the level of elaborated a-structure.

(42)

$$\left[\begin{array}{l} \text{de 'let'} \\ \left[\begin{array}{l} \text{CS}([\alpha], \text{GO}_{Poss}(\{\quad\}_{E_T}, \text{TO}[\quad])) \\ \text{AFF}_{+cc}([\quad]^{\alpha},) \\ \text{ASP}(\_\ \_\ \_) \end{array}\right]_{E} \end{array}\right]$$

Recall that since the transparent Event is an argument at the Thematic Tier and does not contain the various Tiers of the permissive, as is the case for Aspectual light verbs, the process of Argument Fusion rather than Event Fusion is relevant here. The definition of Argument Fusion is repeated here in (43).

(43) **Argument Fusion:**
If one a-structure is embedded within another as a transparent Event, then the highest

argument of the embedded a-structure is fused with the lowest argument of the matrix a-structure. This is expressed by filling the embedded argument slot with the Greek letter index of the matrix argument slot.

The complex a-structure in (45) illustrates how the a-structures of the main verb *banaa* 'make' and the permissive *de* 'let' are combined to form a permissive complex predicate. The a-structure of the main verb *banaa* 'make' repeated in (44) is substituted into the E_T slot on the Thematic Tier of the permissive *de* 'let'.

(44)

$$\left[\begin{array}{l}\text{banaa 'make'}\\ \left[\begin{array}{l}\text{CS}([\alpha],\text{BE}[\ ])\\ \textbf{AFF}_{+\textbf{cc}}([\ ]^{\alpha},\ )\\ \text{ASP}(\_\ \_\ \_)\end{array}\right]_{E}\end{array}\right]$$

(45)

$$\left[\begin{array}{l}\text{banaane diyaa 'let make'}\\ \left[\begin{array}{l}\text{CS}([\alpha],\text{GO}_{Poss}(\left\{\begin{array}{l}\text{CS}([\gamma],\text{BE}[necklace])\\ \textbf{AFF}_{+cc}([\beta]^{\gamma},\ )\\ \text{ASP}(\_\ \_\ \_)\end{array}\right\}_{E_T},\textbf{TO}[Saddaf]^{\beta}))\\ \text{AFF}_{+cc}([Anjum]^{\alpha},)\\ \text{ASP}(\_\ \_\ \_)\end{array}\right]_{E}\end{array}\right]$$

Except for the highest argument of *banaa* 'make', the β and γ marked argument of AFF$_{+\text{CC}}$, nothing else in the embedded a-structure of *banaa* 'make' is affected. One a-structure is thus really embedded within another one. The Aspect Tier of *banaa* 'make' is therefore also simply embedded, and not unified, with that of the light verb *de* 'let'. Note that the $[\ ]_E$, in which the a-structure of *banaa* 'make' in (44) is enclosed, is no longer indicated in (45). This indicates that, while the event structure and Aspect Tier of *banaa* 'make' are independently definable, they are also connected to the event structure of the light verb *de* 'let'. As mentioned before, this is an area which requires further research.

The transparent Event argument of the permissive *de* 'let', in the complex a-structure in (45) triggers Argument Fusion. Under Argument Fusion, the lowest matrix argument is fused with the highest embedded argument. This is clearly indicated in (45). The permittee *Saddaf* is marked with an β. Because the Action Tier is always primary, the argument

Anjum on the Action Tier is higher than *Saddaf*. The lowest matrix argument is therefore *Saddaf*. Similarly, the highest embedded argument is the argument of the function AFF+CC embedded within the E_T. This argument slot has been filled with the index β, is fused to the argument *Saddaf*, and therefore invisible for linking to syntax.

Note again that I have only used the completely filled in a-structure in (45) for purposes of elucidation: it is easier to read the a-structure with the argument slots already filled in. The result of the actual combination of the a-structure of *banaa* 'make' in (44) with the a-structure of *de* 'let' in (43) is shown in (46).

(46)

$$\left[\begin{array}{l}\text{banaane diyaa 'let make'}\\ \left[\begin{array}{l}\text{CS}([\alpha], \text{GO}_{Poss}(\left\{\begin{array}{l}\text{CS}([\gamma], \text{BE}[\ ])\\ \textbf{AFF}_{+cc}([\beta]^{\gamma}, \ \ )\\ \text{ASP}(\_\ \_\ \_)\end{array}\right\}_{E_T}, \textbf{TO}[\ ]^{\beta}))\\ \text{AFF}_{+cc}([\ ]^{\alpha}, \ \ )\\ \text{ASP}(\_\ \_\ \_)\end{array}\right]_{E}\end{array}\right]$$

In this a-structure none of the argument slots have been filled in. But notice that Argument Fusion has coindexed the argument of the embedded AFF with the argument of TO in the matrix a-structure. When the a-structure in (46) is linked to the syntax, there are only three argument slots which must be filled: AFF_{+cc} ([ ]$^{\alpha}$,), TO [ ], and BE [ ]. The issues which arise in linking to both f-structure and c-structure are discussed in the next chapter. In particular, I show that the idea of an E_T which triggers Argument Fusion is instrumental in accounting for the essential characteristic of a complex predicate: how a complex a-structure forms a single f-structure PRED (or syntactic head).

5.4 Conclusion

This chapter has introduced a level of elaborated a-structure which is an adaptation of Jackendoff's (1990) theory of Conceptual Semantics. Note that while I have stayed very close to Jackendoff's original notation, the a-structures introduced in this chapter are actually standard attribute-value matrices (AVMS) through which paths can be specified, just as for the f-structures. So while the notation I use for a-structures may appear to be substantially different from the standard f-structure AVMs, the a-structures here formally do not deviate

from standard LFG.

An elaborated a-structure goes against the current trend in work on argument structure processes, which either make use of traditional θ-role labels, or argue that only the hierarchical organization, but none of the semantic information, of arguments should be considered. I showed that the fine-grained semantic information needed for a simple analysis of Aspectual complex predicates is easily and systematically expressed in an elaborated a-structure. While immediate evidence for an elaborated a-structure approach cannot be culled from the properties of the Urdu permissive, the elaborated a-structure approach allows a simple representation and analysis of the permissive light verb *de* 'let' as well.

The notion of a transparent Event further allows the formulation of a theory of complex predicate formation at argument structure. A transparent Event is taken to be the characteristic property of light verbs and the presence of a transparent Event at a-structure triggers either Event Fusion or Argument Fusion, depending on what role the transparent Event is playing. In the case of Aspectual light verbs, where the light verb is itself a transparent Event, Event Fusion must apply. Under Event Fusion, each argument, starting with the highest, in the a-structure of a light verb is successively fused, or unified, with an argument in the a-structure of a main verb. The Aspect Tiers contained in each of the a-structures must be similarly fused or unified. If there is a clash in features on either the Aspect Tier or in the fusion of one of the arguments, the complex predicate is not well-formed.

In the case of the permissive light verb, the transparent Event is an argument of the verb at the Thematic Tier. Here, the a-structure of a main verb is embedded in the a-structure of the light verb and Argument Fusion applies. Argument Fusion serves to fuse the highest embedded argument with the lowest matrix argument, so that only one argument remains for purposes of linking.

The introduction of an elaborated level of a-structure, in combination with the notion of a transparent Event, thus allows a simple treatment of complex predicate formation at a-structure. The fact that there are two differing processes of Fusion accurately reflects the differing nature of the two light verb constructions involved. While both constructions form complex predicates, the kind of complex predicates which are formed differ somewhat at the level of a-structure. At the level of f-structure, however, these differences are no longer in evidence. The next chapter first discusses how a-strucure and f-structure are related to one another, and then presents a formulation of the relationship of these levels of representation to c-structure. I propose that these levels of representation are constrained

by one another, and that only an interplay of information represented at each level allows a complete analysis of complex predicates.

Chapter 6

Linking

6.1 Introduction

The preceding chapters have shown that complex predicates, especially complex predicates like the permissive, pose a challenge for theories of syntax because of the inherent mismatch of syntactic and semantic information they display: two or more semantic heads correspond to a single syntactic head. As shown in the previous chapters, within LFG this mismatch of information can be easily represented at differing levels of representation. The distinct semantic heads are represented at a-structure, where complex predicate formation also takes place. The fact that two or more distinct semantic heads behave like a single predicate in terms of agreement, anaphora and control phenomena is represented through the postulation of a single PRED at f-structure. The c-structure representation, on the other hand, may be either complex or simple.

In this chapter I now proceed to show how these differing levels of representation are related to one another. Recall that the particular challenge posed by the permissive is that two discontinous heads are able to combine in the syntax to form a single complex predicate. The preceding chapter articulated a theory of complex predicate formation in terms of an elaborated argument structure approach. This theory of complex predicate formation applies to the permissive, as well as the Aspectual complex predicates, where the two members of the complex predicate form a tighter unit than the two predicates which form the permissive. Although I do not include a treatment of the Urdu causative here, I also intend it to apply to causatives, which are formed morphologically in Urdu, i.e., in the lexicon. The formation of a complex predicate at a-structure is thus truly

independent of the actual surface appearance of the complex predicate. Whether it consists of discontinous heads in the syntax, or whether it is formed morphologically, the processes of Argument or Event Fusion discussed in the preceding chapter always apply (see Alsina (1993) for detailed argumentation). Furthermore, as already mentioned above, all types of complex predicates have a complex a-structure resulting from Argument or Event Fusion, which always corresponds to a simple f-structure, where "simple" refers to the fact that the f-structure contains only one nuclear PRED. I propose this as a defining characteristic of complex predicates. This is articulated in (1).

(1) **Definition of a Complex Predicate:**

- The argument structure is complex (two or more semantic heads contribute arguments).
- The grammatical functional structure (f-structure) is that of a simple predicate. It is *flat*: there is only a single predicate (a nuclear PRED) and a single subject.
- The phrase structure (c-structure) may be either simple or complex. It does not necessarily determine the status of a complex predicate.

Alsina (1993), in his examination of Romance and Bantu causatives provides a detailed analysis of the relationship between a-structure and f-structure, and presents a proposal for a more developed theory of linking within LFG. I therefore do not spend much time on the relationship between a-structure and f-structure, but instead refer the reader to Alsina (1993). As will become evident, the mapping between a-structure and f-structure is quite straight forward, given the elaborated a-structure approach to complex predicate formation I presented in Chapter 5. The majority of the chapter thus concentrates on the relationship of both a-structure and f-structure to c-structure. In particular, I propose to make use of the notion of semantic case, in addition to the requirements of "structural", or grammatical, case assignment in order to facilitate the linking between the three distinct levels of representation.

6.2 Linking a-structure and f-structure

6.2.1 Mapping Theory

Lexical Mapping Theory (LMT) was developed within LFG to account for such diverse phenomena as Locative Inversion (Bresnan and Kanerva 1989), Object Asymmetries and

Applicatives (Bresnan and Moshi 1990, Alsina and Mchombo 1989), passivization, causitivization, etc. Essentially, LMT provides a theory which relates a-structure to f-structure. That is, it relates a hierarchically organized argument structure to grammatical relations.

As already discussed in Chapter 2, Alsina (1993) proposes a version of mapping theory called *Functional Mapping Theory* (FMT). This theory is specifically designed to account for Romance causatives, which are formed in the syntax like the Urdu permissive. Despite the fact that FMT is designed to account for syntactically formed complex predicates, I do not adopt it here. While the concepts utilized by FMT are more intuitive and better justified than the concepts underlying LMT, the basic principles of mapping are essentially the same. Furthermore, Alsina applies FMT to a version of argument structure which is not immediately compatible with the elaborated a-structures used here: Alsina's a-structures provide as little semantic information in as abstract a way as possible, while elaborated a-structures allow a fairly detailed representation of the relationships involved.

I therefore utilize a version of LMT for the mapping from a-structures to f-structures. As already mentioned in Chapter 2, LMT was designed to apply only in the lexicon, but I define it to apply in the syntax as well and will therefore dub it Mapping Theory (MT), rather than Lexical Mapping Theory.

Recall that grammatical functions are classified according to the features [±r] (thematically restricted) and [±o] (objective). These features are applied to thematic roles according to various principles. The features assigned to a thematic role determine which grammatical function a particular thematic role is mapped to. Figure (2) shows the basic correspondences.

(2)

Grammatical Function	**Features**
SUBJ	[−r, −o]
OBJ	[−r, +o]
OBJ_θ	[+r, +o]
OBL_θ	[+r, −o]

The basic principles of MT needed for the purposes of this chapter are the *Intrinsic Role Classifications* and the *Default Role Classifications*. Intrinsically, patientlike roles (themes or patients) roles carry the feature [−r]. Secondary patientlike roles (i.e., beneficiary) are intrinsically classified as [+o] (Bresnan and Zaenen 1990). All other roles are [−o]. This ensures that agents usually wind up as subjects (nonobjective), and themes and patients as objects.

The [+o] intrinsic classification of secondary patientlike roles is intended to account for languages in which verbs may have multiple patientlike roles that result in double object constructions, for example. Since the recipient/goal of the Urdu ditransitive verb *de* 'give' is marked with dative case, and never with a postposition, I assume that it is not an oblique, but an indirect object (OBJ_θ). I therefore assume that goals in Urdu have the intrinsic feature [+o].

The Default Role Classifications assign [−r] to the highest thematic role. All other roles receive [+r] by default. A feature [+r] can, of course, not be assigned by Default to a thematic role that is already specified for [−r]: such clash of features is not allowed.[1]

These are the basic principles needed for an account of the Aspectual complex predicates, the permissive, as well as the non-complex predicate instructive. The next sections detail the application of MT to the a-structures of each of these constructions.

6.2.2 The Permissive

The complex a-structure corresponding to the permissive complex predicate in (3) is repeated in (4) for easy reference.

(3) anjum=ne saddaf=ko haar **banaa-ne** **di-yaa**
Anjum.F=Erg Saddaf.F=Dat necklace.M=Nom make-Inf.Obl give-Perf.M.Sg
'Anjum let Saddaf make a necklace.'

(4)

$$\left[\begin{array}{l}\text{banaane diyaa 'let make'}\\ \left[\begin{array}{l}\text{CS}([\alpha], \text{GO}_{Poss}(\left\{\begin{array}{l}\text{CS}([\gamma], \text{BE}[\ ])\\ \textbf{AFF}_{+cc}([\beta]^{\gamma}, \ \ )\\ \text{ASP}(\_\ \_\ \_)\end{array}\right\}_{E_T}, \textbf{TO}[\ ]^{\beta}))\\ \text{AFF}_{+cc}([\ ]^{\alpha}, \ \ )\\ \text{ASP}(\_\ \_\ \_)\end{array}\right]_{E}\end{array}\right]$$

Since Argument Fusion has applied to the a-structure in (4) and rendered one argument invisible for linking, there are exactly three empty argument slots which need to be linked up to an f-structure representation. The three arguments to be linked are listed from highest

[1] Bresnan and Zaenen (1990) describes a more recent version of LMT, but the essential components are the same as given here.

to lowest in Figure (5), as determined by the organization of the a-structure in (4). This figure displays the application of MT to the unfilled argument slots in (4). These arguments are the only ones available for linking. Note that the list of arguments in (5) represents a covert argument structure in some sense, since it is a 'flat' list of arguments derived from the composition and embedding of two distinct argument structures. While it can be argued that the list in (5) should be encoded at a separate level of representation, I see little motivation to invest the derived list in (5) with an independent status.

(5)

	AFF$_{+cc}$([ ]$^{\alpha}$,)	TO[ ]$^{\beta}$	BE[ ]
<	(ag)	(go)	(th)>
	|	|	|
intrinsic	[−o]	[+o]	[−r]
default	[−r]	[+r]	
GF	SUBJ	OBJ$_{go}$	OBJ

The argument slot AFF$_{+cc}$([ ]$^{\alpha}$,) essentially corresponds to an agent (actor), the slot TO[ ]$^{\beta}$ to a goal, and BE[ ] to a theme (this follows Jackendoff's (1990) system, as discussed in the section on the representation of argument structure). As shown in (5), the 'agent' argument is assigned the intrinsic feature [−o], the 'goal' the feature [+o], and the 'theme', the feature [−r]. The Default role classifications then assign [−r] to the highest argument, the 'agent', and [+r] to all the other roles. The theme cannot be assigned [+r] because it already has a [−r] specification.

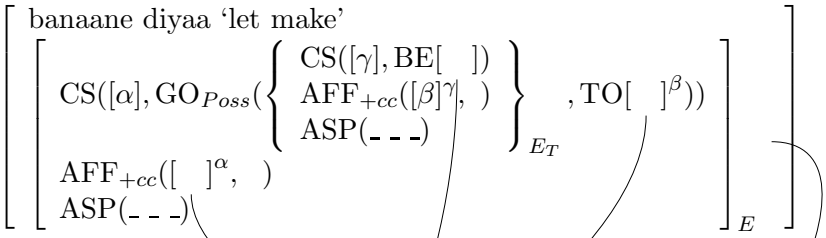
Although the transparent Event serves as an argument on the Thematic Tier in (4), it is invisible for linking: it is transparent in the sense that it does not itself participate in the linking process. This is in contrast to simple Events, which are visible for linking purposes, and which are generally linked to complements. This is demonstrated in the next section for the simple Event argument of the predicate *kah* 'tell'.

According to the correspondences of features to grammatical functions previously shown in (2), the list of arguments is mapped on to grammatical functions as shown in (5). The 'agent' corresponds to a SUBJ, the goal to an OBJ$_{go}$, and the 'theme' to an OBJ. Notice that because the 'theme' only carries a [−r] feature, it could also be possible to map it onto a SUBJ. However, in this case a well-formedness condition applies, which states that there can only be one subject in a clause. The 'agent' argument has already been mapped to SUBJ, so the only grammatical function possible for the 'theme' is the OBJ.

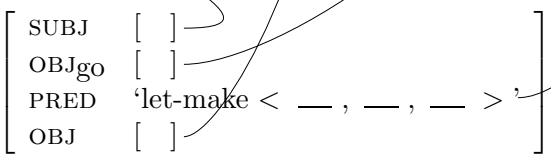
The application of MT thus relates the complex a-structure in (6) to the flat f-structure in (7). Recall that this is exactly the desired result for the permissive. The data examined

in Chapter 3 showed that while the permissive involved the combination of two distinct heads (and argument structures), it must be considered to be monoclausal, and therefore have a simple f-structure, as in (7).

(6)



(7)



The f-structure PRED is represented as *let-make*, with three arguments. Within LFG, the value of a PRED provides the meaning of the expression. For example, imagine the attribute-value pair [PRED 'necklace'] were represented in the OBJ sub-part of the f-structure above. This would indicate that the lexical item corresponding to the grammatical function object had the meaning 'necklace'. However, the meaning of the lexical item is not merely 'necklace'. Rather, the label 'necklace' is a shorthand notation which stands for a more detailed semantic representation. This detailed semantic representation will specify information such as 'worn around neck', etc. I follow Alsina (1993) in assuming that the notation *let-make* at f-structure is similarly an abbreviated representation of the semantics represented in more detail at a-structure. The value of the PRED in (7), then, is actually the entire a-structure in (6). This is indicated by the line connecting the outer Event of the complex a-structure in (6) with the PRED value in (7).

6.2.3 The Instructive

An example of the instructive is repeated in (8). Recall that the instructive was shown not to form a complex predicate. This is reflected in the a-structure for *kah* 'say' shown

in (9): there is no transparent Event argument. The semantics of *kah* 'say' are expressed as consisting of a causer who causes some piece of information to go to somebody. In the representation below, the piece of information is represented as an Event.

(8) anjum=ne saddaf=ko haar banaa-ne=ko kah-aa
Anjum.F=Erg Saddaf.F=Dat necklace.M=Nom make-Inf.Obl=Acc say-Perf.M.Sg
'Anjum told Saddaf to make a necklace.'

(9)

$$\left[\begin{array}{l} \text{kah 'say'} \\ \left[\begin{array}{l} \text{CS}([\alpha], \text{GO}_{Info}([\ ]_E, \text{TO}[\ ])) \\ \text{AFF}([\ ]^{\alpha},\ ) \\ \text{ASP}(\_\ \_\ \_) \end{array}\right]_E \end{array}\right]$$

Because the Event argument of the function GO at the Thematic Tier in (9) is not a transparent Event, as it was in the a-structure of the permissive *de* 'let', complex predicate formation is not triggered when the a-structure for *kah* 'say' combines with the a-structure for *banaa* 'make' in (10). The a-structure of *banaa* 'make' here is embedded within the a-structure of *kah* 'say', but instead of Argument Fusion, the complex a-structure in (10) illustrates an instance of *Argument Control*.

(10)

$$\left[\begin{array}{l} \text{banaane kah 'tell to make'} \\ \left[\begin{array}{l} \text{CS}([\alpha], \text{GO}_{Info}(\left[\begin{array}{l} \text{CS}([\gamma], \text{BE}[necklace]) \\ \text{AFF}_{+cc}([\ ]^{\gamma,\beta},\ ) \\ \text{ASP}(\_\ \_\ \_) \end{array}\right]_E, \text{TO}[Saddaf]^{\beta})) \\ \text{AFF}([Anjum]^{\alpha},\ ) \\ \text{ASP}(\_\ \_\ \_) \end{array}\right]_E \end{array}\right]$$

In (10), I have represented the combination of a-structures with all the argument slots already filled in so as to facilitate the interpretation of (10). The complex a-structure corresponding to the instructive *banaane-ko kahaa* 'told to make' which actually results from a combination of the two distinct a-structures is shown in (11).

(11)

$$\left[\begin{array}{l} \text{banaane kah 'tell to make'} \\ \left[\begin{array}{l} \text{CS}([\alpha], \text{GO}_{Info}(\left[\begin{array}{l}\text{CS}([\gamma], \text{BE}[\ ]) \\ \text{AFF}_{+cc}([\ ]^{\gamma,\beta},\ ) \\ \text{ASP}(\_\ \_\ \_)\end{array}\right]_E, \text{TO}[\ ]^{\beta})) \\ \text{AFF}([\ ]^{\alpha},\ ) \\ \text{ASP}(\_\ \_\ \_)\end{array}\right]_E \end{array}\right]$$

As with Argument Fusion, the highest embedded argument is coindexed with the lowest matrix argument. The crucial difference between Argument Fusion and Argument Control, however, is that the highest embedded is not *filled*, i.e., fused, with the lowest matrix argument. The lowest matrix argument is the argument of TO, which is indexed by β. The highest embedded argument is the argument of embedded AFF (marked by both β and γ). These two arguments are coindexed, but not completely identified with one another. Since there is no transparent Event in (11) and neither Argument Fusion nor Event Fusion are triggered, the arguments of *kah* 'say' and *banaa* 'make' are linked up separately by Mapping Theory. The predicate *kah* 'say' has three arguments: the sayer (argument of matrix AFF, indicated by α), the person something is said to (argument of TO, indicated by β), and the Event argument. These three arguments are mapped to SUBJ, OBJ_{go}, and XCOMP, respectively. The application of MT shown in (12) is essentially the same as with the permissive. The "agents" have the intrinsic feature [−o], the "themes" the feature [−r], and the goal argument is [+o]. The Default rule assigns [−r] or [+r] where possible. The correspondences between a-structure and f-structure are shown in (13) and (14).

(12)

	$\text{AFF}([\ ]^{\alpha},\ )$	$\text{TO}[\ ]^{\beta}$	$\text{GO}_{Info}[\ ]$
<	(ag)	(go)	(th/Ev)>
	|	|	|
intrinsic	[−o]	[+o]	[−r]
default	[−r]	[+r]	
GF	SUBJ	OBJ_{go}	XCOMP

(13)

$$\left[\begin{array}{l} \text{kah 'say'} \\ \left[\begin{array}{l} \text{CS}([\alpha], \text{GO}_{Info}([\ ]_E, \text{TO}[\ ])) \\ \text{AFF}([\ ]^{\alpha},\ ) \\ \text{ASP}(\_\ \_\ \_)\end{array}\right]_E \end{array}\right]$$

(14)

$$\begin{bmatrix} \text{SUBJ} & [\ ] \\ \text{OBJ}_{\text{GO}} & [\ ] \\ \text{PRED} & \text{`say} < \_\_ , \_\_ , \_\_ > \text{'} \\ \text{XCOMP} & [\ ] \end{bmatrix}$$

Note that although "that" and "wh" complements are examined in Bresnan (1982b) for the relationship between predicate-argument structure and grammatical functions, the treatment of Event arguments has not been made explicit in subsequent work on the theory of mapping between a-structure and f-structure. The thematic hierarchy used in the development of MT does not feature an Event argument, and there are no principles for linking of Events.

According to Jackendoff's (1990) system of relating traditional θ-role labels to the argument slots of functions, the Event argument of *kah* 'say' essentially corresponds to a theme because it is the argument of a GO function. As already illustrated above, I therefore propose to treat the Event argument like a theme for the purposes of MT. The intrinsic feature assigned to the Event argument in (14) is thus [−r]. However, as this "theme" is also an Event, the argument is not mapped onto an OBJ at f-structure, as required by the mapping principles, but is realized as an XCOMP. MT as it stands must therefore be augmented by the Mapping Principle in (15). This principle states that an argument which is an Event must be mapped on to a complement.[2]

(15) θ_{Ev} is mapped onto (X)COMP.

The principle leaves open whether an Event argument is mapped onto an XCOMP or a COMP. Control and complementation has been stated at the level of f-structure in LFG (Bresnan 1982a), and a difference is made between functional control, which involves an XCOMP, and anaphoric control, which involves a COMP. In this dissertation, I have only discussed instances of functional control, so a more detailed discussion on control and complemention is beyond the present scope of investigation. However, I believe that the elaborated level of a-structure formulated here allows for a semantic theory of control, such as the one

[2]Note that I do not intend the term "Event" to encompass such NPs as *dinner* or *movie*.

advocated by Sag and Pollard (1991). Thus, more detailed semantic information present at a-structure would allow MT to determine whether a given Event should be mapped into an XCOMP or a COMP. Also recall from Chapter 3 that the case marking on infinitives ranges over a variety of case markers: accusative, comitative, locative, genitive, etc. As shown in the section on linking to phrase structure, elaborated a-structure representations provide information relevant for semantic case marking. The case marking for a given infinitival complement can thus also be determined by means of the information represented at an elaborated a-structure.

The predicate *banaa* 'make' has two arguments: the maker (argument of embedded AFF, indicated by β and γ), and the thing made (argument of BE). The application of MT, illustrated in (16), links these two arguments SUBJ and OBJ. The correspondences are again shown in (17) and (18).

(16)

	AFF($[\ ]^{\gamma,\beta}_{+cc}$,)	BE[ ]
<	(ag)	(th)>
	|	|
intrinsic	[−o]	[−r]
default	[−r]	
GF	SUBJ	OBJ

(17)

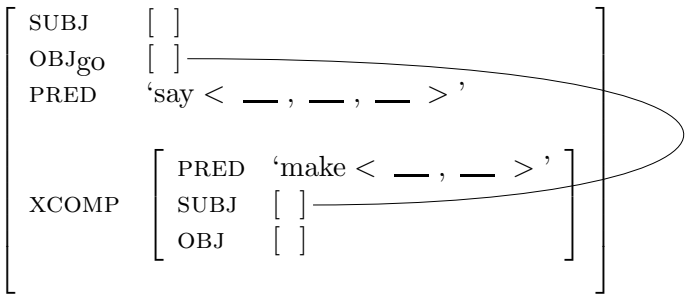
$$\left[\begin{array}{l} \text{banaa 'make'} \\ \left[\begin{array}{l} \text{CS}([\alpha], \text{BE}[\ ]) \\ \text{AFF}_{+\text{cc}}([\ ]^{\alpha},\ ) \\ \text{ASP}(\_\ \_\ \_) \end{array}\right]_{E} \end{array}\right]$$

(18)

$$\left[\begin{array}{ll} \text{SUBJ} & [\ ] \\ \text{PRED} & \text{'make} < \underline{\quad}, \underline{\quad}, \underline{\quad} > \text{'} \\ \text{OBJ} & [\ ] \end{array}\right]$$

The application of MT to the instructive is now complete. The complex, skeletal f-structure which the arguments of the complex a-structure in (11) map on to is shown in (19).

(19)



I refer to (19) as a *skeletal* f-structure because none of the subparts of the f-structure have been filled in, i.e., linked to a c-structure representation, as yet. That part of linking is discussed in the remainder of the chapter, after the mapping between a-structure and f-structure has been illustrated for Aspectual complex predicates. For now, note that the skeletal f-structure again corresponds exactly to the desired result. Data from agreement, control, and anaphora showed that the instructive could not be a complex predicate, but consisted of a matrix predicate with an embedded complement. This is exactly what the f-structure in (19) represents. Thus, the elaborated level of a-structure I posit, in combination with a theory of complex predicate formation allows a clear differentiation of complex predicates from complement constructions, and may also be used as the base for a semantic theory of control, such as the one advocated in Sag and Pollard (1991).

6.2.4 Aspectual complex predicates

The application of MT to a-structures of Aspectual complex predicates is straightforward. An example of a representative Aspectual complex predicate is given in (20). The a-structure corresponding to the fusion of the a-structures of the light verb *le* 'take' and the main verb *banaa* 'take' is repeated in (21).

(20) anjum=ne haar banaa li-yaa
Anjum.F=Erg necklace.M=Nom make take-Perf.M.Sg
'Anjum made the necklace completely, on purpose.'

(21)

$$\begin{bmatrix} \text{banaa liyaa 'made completely'} \\ \begin{bmatrix} \text{CS}([\alpha], \text{BE}[\ ]) \\ \textbf{AFF}_{+\text{cc}}([\ ]^{\alpha},\ ) \\ \text{ASP}(\_\ \_\ 1) \end{bmatrix}_{E} \end{bmatrix}$$

There are two arguments available for linking in (20): the first argument of AFF_{+cc} (agent) and the argument of BE (theme). As illustrated in (22), the "agent" is intrinsically [−o], while the "theme" is [−r]. The default specifications then assign [−r] to the agent, the agent is mapped to the f-structure SUBJ, and the theme is mapped to the OBJ. The corresponding skeletal f-structure is shown in (23).

(22)

	$\text{AFF}_{+cc}([\ ]^{\alpha},\ )$	BE [ ]
<	(ag)	(th)>
	|	|
intrinsic	[−o]	[−r]
default	[−r]	
GF	SUBJ	OBJ

(23)

$$\begin{bmatrix} \text{SUBJ} & [\ ] \\ \text{PRED} & \text{'complete-make} < \underline{\quad}, \underline{\quad} > \text{'} \\ \text{OBJ} & [\ ] \end{bmatrix}$$

In summary, this section has shown that MT can be applied successfully to the elaborated, complex a-structures of permissive and Aspectual complex predicates, as well the complement instructive. While each of the three constructions is represented by a complex a-structure, the theory of complex predicate formation articulated in Chapter 5 in combination with Mapping Theory places each of the complex a-structures in exactly the right correspondence with the various differing f-structures required by the data examined in the previous two chapters.

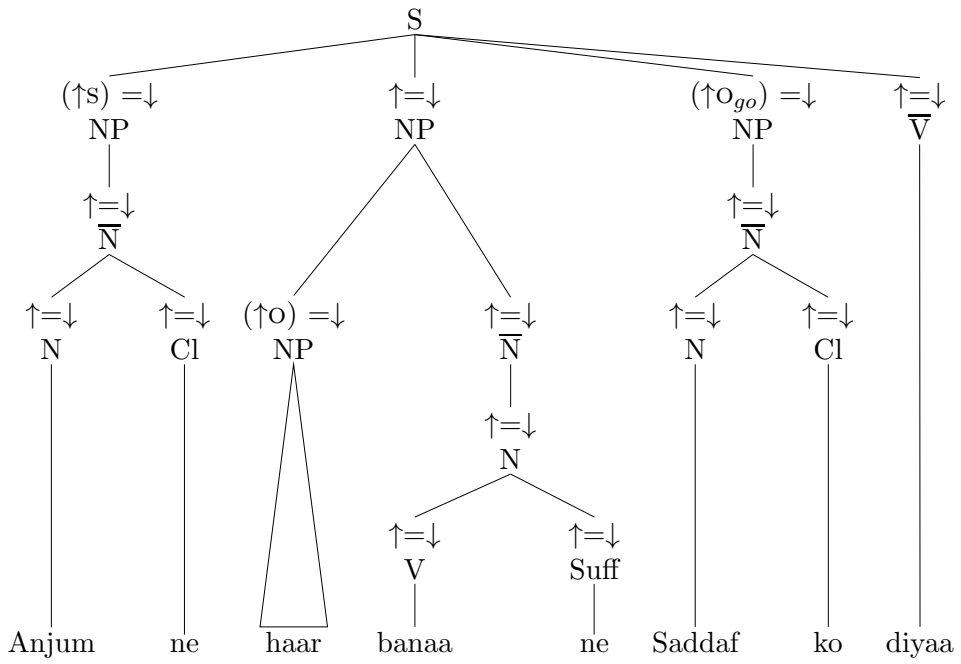
6.3 Linking to Phrase Structure

6.3.1 Predicate Composition

In Bresnan and Kaplan (1982), an algorithm relating c-structure to f-structure was proposed, by which annotations on c-structures were 'solved' to arrive at an f-structure corresponding to a given c-structure representation. An example of a possible c-structure representation for the Urdu permissive in (24) is shown in (25) (in the interests of space, SUBJ has been abbreviated to S, and OBJ to O). Because this section is concerned with the relationship of c-structure to a-structure and f-structure, details of the c-structure representation are of greater importance than in previous chapters. The c-structure in (24) is therefore more detailed than any of the previous representations.

(24) anjum=ne haar banaa-ne saddaf=ko di-yaa
Anjum.F=Erg necklace.M=Nom make-Inf.Obl Saddaf.F=Dat give-Perf.M.Sg
'Anjum let Saddaf make a necklace.'

(25)



In addition to the functional information expressed by the annotations on the nodes, functional information is also contributed by the leaves of the tree. In the interests of space, I have not shown the functional information contained in the lexical entries of terminal nodes in (25). The lexical entry for *di-yaa* 'let', a morphologically complex form, which is

put together in the lexicon, for example, contains the functional information shown in (26). This information "flows" upward in the c-structure in (25) by means of the annotation ↑=↓, and plays a role in the determination of the f-structure corresponding to the c-structure in (25).

(26) diyaa (↑PRED =) 'let< __ , __ >'
(↑ASPECT) = PERF
(↑PERS) = 3
(↑NUM) = SG
(↑GEND) = MASC

I will not go into further details on the algorithm which 'solves' and unifies the functional information annotated at c-structure to produce a well-formed f-structure. Recall, however, that the algorithm does not provide a way to unify PREDs. This is a problem for a treatment of the permissive, as the two syntactic heads *di-yaa* 'let' and *banaa-ne* 'make' in (25) cannot be unified into a single PRED at f-structure.

I follow Alsina (1993) in proposing that the PRED value of the verb *di-yaa* 'let' in (26), for example, actually stands for the a-structure representation of *de* 'let'. Alsina further proposes that predicates be allowed to *compose* in the syntax through a combination of their respective a-structures. The notion of *Predicate Composition* formulated by Alsina is applicable to the Urdu permissive as well. The c-structure in (25) is annotated as being double-headed. The ↑=↓ notation allows the PRED information, i.e., the a-structures associated with each of the predicates, to flow up towards the top of the tree, where the information from each predicate must be combined in order for the structure to be well-formed. Alsina formalizes the notion of Predicate Composition by redefining the interpretation of '↑=↓' as shown in (27).

(27) ↑=↓ ⟶ ↑PRED = function of composition of ↓PRED
elsewhere ↑ATTRIBUTE =↓ATTRIBUTE

The first line of (27) allows the composition of two ↑=↓ specifications, i.e., the composition of two PREDs. The original 'meaning' of the notation, which passes information from the annotated node up to the mother, is the *elsewhere case* under this system. However, the precise formulation proposed for Predicate Composition is not quite right for Urdu. The syntactically formed causatives which Alsina examines are from Romance languages. As the c-structure in Romance languages is more constrained than in Urdu, a Romance light verb always is sister to a VP complement, where the main verb the light verb combines with

is located. The relatively free word order of Urdu does not allow a similar strict reliance on precedence or dominance at c-structure. I return to the issue of how to identify the heads relevant for Predicate Composition after a discussion of the determination of c-structure annotations.

6.3.2 Determination of c-structure annotations

Recall that annotations on c-structure are not stipulated, but are considered to follow from independent principles of the language. In Romance and English, for example, the annotations (↑SUBJ) =↓ and (↑OBJ) =↓ are determined by principles of phrase structure position. Roughly: the subject NP is sister to VP, while the object NP is sister to the V. In a language where position does not strictly determine the grammatical relation of an NP, the determination of c-structure annotations must follow from other principles.

K.P. Mohanan (1982) proposes *Principles of Case Interpretation* for a treatment of Malayalam within LFG. A sample principle is given in (28). This principle is formally encoded within LFG as the functional equations in (29), which an NP node at c-structure is annotated with.

(28) Interpret accusative case as the direct object OBJ.

(29) (↑OBJ) =↓
(↓CASE) = ACC

When an NP node is annotated with the equations in (29), and thus identified as the direct object of the sentence, the noun contained within the NP must bear accusative case. If it does not, the f-structure corresponding to the c-structure is ill-formed because the CASE features of the f-structure corresponding to that particular NP node fail to unify.

The choice as to which of the Case Interpretation principles annotates which NP node at c-structure is free. In other words, a given c-structure is generated with underdetermined annotations on its NP nodes. Thus, in a free word order language like Malayalam, the phrase structure rule in (30) generates a basic sentence. The notation (↑GF) = ↓, where GF is short for "grammatical function", in combination with $\overline{\mathrm{N}}^*$ permits any number of grammatical functions in the sentence. This allows the positions of subject, object, etc. to vary.[3]

[3]Note that the particular notation ↑GF was not available to K.P. Mohanan (1982). However, it correctly expresses the intent of his analysis.

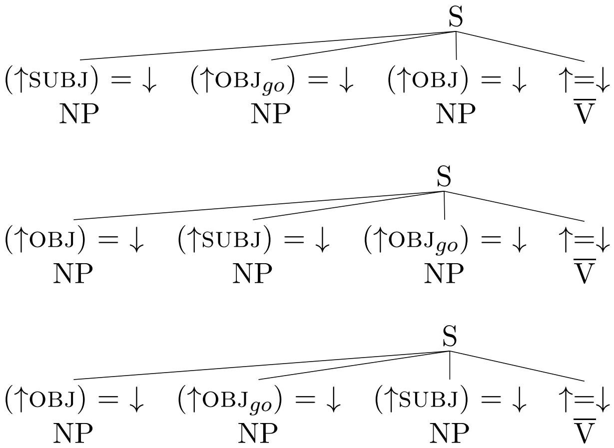
(30) S ⟶ $\overline{\text{N}}$* $\overline{\text{V}}$
(↑GF) =↓

In the c-structure realization of a clause in Malayalam, the underdetermined (↑GF) =↓ may be instantiated by any one of the Case Interpretation principles. The Completeness Principle (see Chapter 2) then rules out all the combinations of annotations which do not result in a well-formed f-structure.

Under this approach, there are fifteen differing c-structure possibilities for the Urdu sentence in (31).[4] Three abbreviated c-structure possibilities are shown in (32).

(31) haar saddaf=ko anjum=ne banaa-ne di-yaa
necklace.M=Nom Saddaf.F=Dat Anjum.F=Erg make-Inf.Obl give-Perf.M.Sg
'Anjum let Saddaf make a necklace.'

(32) a. b. c.



While the Completeness Principle eventually rules out all of the c-structure possibilities which do not yield a well-formed f-structure for (29), the various differing c-structure possibilities are posited by the theory and must be sifted through. A claim to maximal efficiency can therefore not be made. I propose to improve on this state of affairs by utilizing the level of a-structure, which was not fully developed within LFG at the time of K.P. Mohanan's (1982) treatment of Malayalam. In particular, the combination of a-structures in the syntax, and the subsequent mapping of a-structure to f-structure constrain the possibilities of annotations at c-structure. I take the information encoded by case clitics seriously and use it to determine grammatical function annotations at the c-structure. The next section

[4] The annotations on the c-structures nodes need not necessarily be different: one could instantiate the annotation SUBJ more than once.

describes the basic machinery needed. The subsequent sections then demonstrate how it provides an account for the permissive and Aspectual complex predicates, as well as the instructive, and a complex predicate consisting of more than two predicates.

6.3.3 Phrase Structure and Case Clitics

Essentially following K.P. Mohanan (1982) for Malayalam, I propose that a basic clause in Urdu is generated by the immediate dominance rule in (33a).[5] The use of the Kleene star on the GF* allows for functional uncertainty (see Kaplan and Zaenen 1989). This notation allows the expression of equations such as '(↑XCOMP XCOMP OBJ) =↓', which enable reference to an arbitrarily deep embedded item. The fragment of Urdu phrase structure rules needed to account for Urdu complex predicates is shown in (33b–f). By convention, heads such as 'V' are not annotated explicitly with '↑=↓'.[6]

(33) a. $S \longrightarrow \underset{(\uparrow GF^*)=\downarrow}{NP^*} \quad \overline{V} \quad \underset{(\uparrow GF^*)=\downarrow}{NP^*}$

b. $\overline{V} \longrightarrow V \quad (V) \quad (Stat) \quad (Aux)$

c. $\overline{V} \longrightarrow \underset{(\uparrow GF^*)=\downarrow}{\overline{N}} \quad \overline{V}$

d. $\overline{N} \longrightarrow N \quad (Cl)$

e. $NP \longrightarrow \underset{\downarrow\in(\uparrow Adj)}{AdjP^*} \quad \overline{N}$

f. $NP \longrightarrow \underset{(\uparrow GF^*)=\downarrow}{NP} \quad \overline{N}$

Evidence from Aspectual complex predicates showed that their structure was that of two Vs contained under a $\overline{V}$. Aspectual complex predicates are thus realized by the rule in (33b). As the rule also indicates, the stative *rah* 'stay' along with the auxiliary are optionally realized. The rule in (33c) generates N-V complex predicates of the type examined by T. Mohanan (1990, 1993c) and also allows a representation of Infinitive-V contained under

[5]Note that while this rule generates all the possible basic word orders in Urdu, it does not take into account focus and topicalization. In order to give a full account of word order and its interaction with focus and topicalization, an approach like King's (1993) needs to be worked out for Urdu.

[6]The annotation on the AdjP in (33e) reflects that within LFG adjectives are treated as belonging to a set. See Bresnan (1982a) for an early formulation and Andrews and Manning (1993) for a more recent discussion.

a $\overline{\text{V}}$. Note that this rule may apply recursively to produce complex predicates consisting of more than two predicates. The $\overline{\text{N}}$ in (33c) allows for the fact that infinitives may bear case. According to the arguments presented in T. Mohanan (1992), case clitics in Hindi/Urdu attach phrasally, not lexically, so an infinitive bearing a case clitic must be a $\overline{\text{N}}$. This is further reflected by the rule in (33d). The rule in (33e) allows the generation of standard NPs, while (33f) represents the structure of infinitival NPs, as argued for in Chapter 3. Finally, note that the $\overline{\text{N}}$ in (33c) is annotated with $(\uparrow \text{GF}^*) = \downarrow$. As is discussed in the next sections on the permissive and the instructive, this permits the $\overline{\text{N}}$ to be realized either as an XCOMP in the case of the instructive, or as a head, with an $\uparrow = \downarrow$ annotation, in the case of complex predicates such as the permissive or the N-V complex predicates examined by T. Mohanan.

Recall that K.P. Mohanan (1989b), T. Mohanan (1990), and Butt and King (1991) argue for a semantic basis to case in languages such as Hindi/Urdu and Malayalam. However, case cannot be purely semantic; it needs to make reference to grammatical functions as well.[7] In my approach to Urdu, this is reflected in the lexical entries of the case clitics.

(34) a. -ne $(\uparrow \text{CASE}) = \text{ERG}$
$(\text{SUBJ}\uparrow)$

b. -ko $(\uparrow \text{PRED ARG FN} =_{\text{C}} \text{TO})$
$(\uparrow \text{CASE}) = \text{DAT}$
$(\text{OBJ}_{go}\uparrow) \vee (\text{SUBJ}\uparrow)$
$\vee$
$(\text{PRED ARG FN} \neq \text{TO})$
$(\uparrow \text{CASE}) = \text{ACC}$
$(\text{OBJ}\uparrow)$
$(\uparrow \text{SPEC}) = +$

c. -se $(\uparrow \text{CASE}) = \text{INST}$
$(\text{OBL}_{\theta}\uparrow)$

d. -k- $(\uparrow \text{CASE}) = \text{GEN}$

e. -mẽ $(\uparrow \text{CASE}) = \text{LOC}$
$(\text{OBL}\uparrow)$

Under this system, the case clitics assign case to a given NP. The ergative and accusative cases even exclusively determine the grammatical function of an NP, since they require that

[7] See Chapter 2 for a discussion of this issue and the notions of inherent vs. structural Case.

the NP marked by the clitic be a subject or an object. This is expressed by the notation (SUBJ↑) or (OBJ↑).[8] The ergative and accusative case clitics can thus be thought of as assigning structural Case.

The entry for *ko* allows it to be realized either as dative or accusative. It can be realized as a dative only if there is an argument slot at a-structure which is an argument of a TO function. Here the a-structure directly constrains the interpretation of grammatical functions: a dative argument in Urdu can only be realized as either an indirect object, or a subject. If this condition is not met, then *ko* must be interpreted as an accusative, and the NP marked by it must be the direct object of the sentence.

The genitive in (34d) is simply represented as *-k-*. This is because it agrees in gender and number with its head noun. The agreement morphology is assumed to combine with the genitive in the lexicon, and is not represented here. Also note that the nominative case is not included among the entries above. Rather than providing an entry for a phonologically null case, I formulate a principle which requires that all NPs must receive case. This principle operates on f-structure. With regard to the nominative, f-structures with the grammatical function attributes SUBJ and OBJ are checked for a case value. If there is none, then nominative case is assigned. Note that it will not do to simply assign nominative case to any argument without a CASE value at f-structure. Locatives in Urdu may appear without an overt case clitic or post-position, but they cannot be analyzed as nominative (T. Mohanan 1990).

The basic machinery needed for a complete analysis of the permissive and Aspectual complex predicates is now in place. The next sections demonstrate its application and the interaction of the differing levels of representation in some detail.

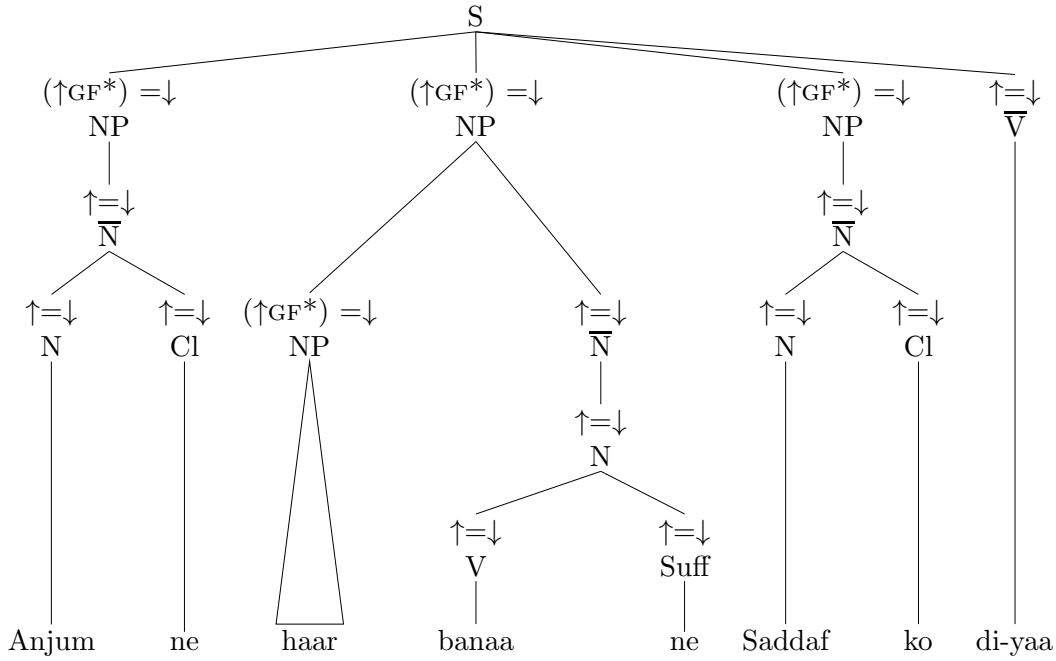
6.3.4 The Permissive

The c-structure corresponding to the Urdu permissive in (35), as generated by the phrase structure rules given in the previous section, is repeated again in (36).

(35) anjum=ne haar banaa-ne saddaf=ko di-yaa
Anjum.F=Erg necklace.M=Nom make-Inf.Obl Saddaf.F=Dat give-Perf.M.Sg
'Anjum let Saddaf make a necklace.'

[8] While this notation is part of standard inside-out functional uncertainty (see Kaplan and Zaenen 1989), King (1993) first proposed applying it in this form. It simply indicates that the mother node must fulfill the specified gramamtical function. So the annotation (SUBJ↑) on the ergative clitic requires that the mother node, an NP, be the subject of the sentence.

(36)



Predicate Composition as defined in Alsina (1993) cannot be applied to the c-structure in (36). There is only one node which is marked ↑=↓ at the top-most level. In this case the light verb is not sister to a VP complement also annotated with a ↑=↓, as is the case in Romance.

One possible solution to this problem would be to posit two phrase structure rules: one which produces the c-structure in (36), and one which allows NPs to be marked as a head with the notation ↑=↓. This NP could then straightforwardly be subject to Predicate Composition. So, for example, in (36) the NP dominating *haar banaa-ne* 'make necklace' could be annotated with ↑=↓ and Predicate Composition as formulated by Alsina could apply.[9] However, since the definition of the Kleene star allows any number of instantiations, including zero, the postulation of two differing phrase structure rules is actually not needed. A zero instantiation of (↑GF*)=↓ is ↑=↓. The particular realization of a given (↑GF*)=↓ annotation is primarily determined by the entries for case clitics given above, which draw on a-structure as well as f-structure information.

I propose an approach which builds up a complete representation of a sentence like (35) through an interaction of the different levels of representation. This is best demonstrated by

[9]This is in fact the solution Alsina (1993) proposes. He allows either a ↑=↓ or (↑XCOMP) =↓ annotation on VPs. The Principles of Coherence and Completeness then rule out the ill-formed possibilities.

example. The information immediately accessible from (36) is that the predicate *di-yaa* 'let' is a head. However, its a-structure (which is represented in the lexical entry, and accessible at c-structure because the information has been passed up through the ↑=↓ notation) contains a transparent Event. In order for the sentence to be grammatical, the a-structure of *di-yaa* 'let' must therefore be combined with the a-structure of another predicate. Since there is no predicate available within the $\overline{\text{V}}$ constituent containing *di-yaa* 'let', the predicate must be located elsewhere in the c-structure if the sentence is well-formed. The combination of a-structures is subject to the as yet informally stated constraint in (37).

(37) **Constraint on Predicate Composition:** Only a-structures of predicates dominated by all the same S nodes may combine.

The constraint in (37) ensures that a light verb cannot combine with a predicate in another clause. S (or IP) can thus be thought of as a barrier for complex predicate formation. Because of the demonstrated variability in word order with permissive complex predicates, no further constraints on a-structure composition are statable. Nor are they needed. Since the information from the lexical entries of all of the lexical items in (36) is available, it is a simple matter to determine that the only lexical entry which contains an a-structure is the entry for the V *banaa* 'make'. The a-structure of *di-yaa* 'let' can therefore combine with the a-structure of *banaa* 'make' to form the complex, complete, a-structure in (38).

(38)

$$\left[\begin{array}{l} \text{banaane diyaa 'let make'} \\ \left[\begin{array}{l} \text{CS}([\alpha], \text{GO}_{Poss}(\left\{\begin{array}{l} \text{CS}([\gamma], \text{BE}[\ \ ]) \\ \textbf{AFF}_{+cc}([\beta]^{\gamma}, \ \ ) \\ \text{ASP}(\_\ \_\ \_) \end{array}\right\}_{E_T}, \textbf{TO}[\ \ ]^{\beta})) \\ \text{AFF}_{+cc}([\ \ ]^{\alpha}, \ \ ) \\ \text{ASP}(\_\ \_\ \_) \end{array}\right]_{E} \end{array}\right]$$

As demonstrated previously, the application of MT to this a-structure yields the skeletal f-structure in (39).

(39)

$$\left[\begin{array}{ll} \text{SUBJ} & [\ \ ] \\ \text{OBJ}_{\text{GO}} & [\ \ ] \\ \text{PRED} & \text{'let-make} < \underline{\quad}, \underline{\quad}, \underline{\quad} > \text{'} \\ \text{OBJ} & [\ \ ] \end{array}\right]$$

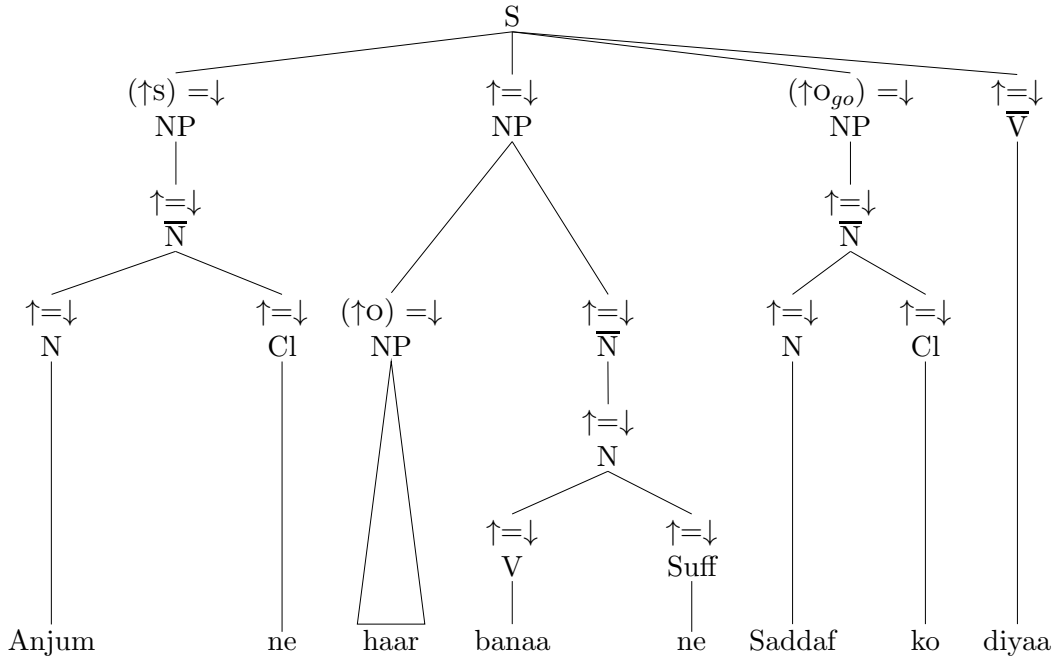
At this point, the lexical entries of the case clitics provide the key to linking up the skeletal f-structure in (39) with the underdetermined c-structure in (36) and the unfilled a-structure slots in (38). If these three levels of representation correspond to one another in a well-formed way, then the sentence in (35) is well-formed. The lexical entry for the case clitic *ne* specifies that its mother node must be assigned ergative case, and be the subject of the sentence. The annotation on the NP containing *Anjum=ne* can therefore straightforwardly be determined as (↑SUBJ) =↓.

Similarly, the lexical entry for the case clitic *ko*, repeated here in (40), determines that the NP *Saddaf=ko* must be dative. The first part of the "or" statement in (40) is satisfied because there is a slot at a-structure which is an argument of a TO function. The NP *Saddaf=ko* could either be realized as an indirect object, or a subject. However, as a subject has already been identified, the NP must be annotated as (↑OBJ$_{go}$) =↓.

(40) -ko (↑PRED ARG FN $=_c$ TO)
(↑CASE) = DAT
(OBJ$_{go}$ ↑) ∨ (SUBJ↑)
∨
(PRED ARG FN ≠ TO)
(↑ CASE) = ACC
(OBJ↑)
(↑SPEC) = +

At this point, there is only one grammatical function at f-structure left unlinked. At c-structure, there are two NPs which carry underdetermined annotations: the NP immediately dominating *haar*, and its mother node, the NP dominating *haar banaa-ne* 'make necklace'. The functional annotation on either of these two NPs may now be instantiated as (↑OBJ)=↓ or as ↑=↓. The Principle of Coherence serves to rule out the possibility in which *haar banaa-ne* 'make necklace' is annotated as the object, so the fully annotated c-structure in (41) represents the only well-formed possibility. This c-structure corresponds to the f-structure in (42) and the a-structure in (43).

(41)



(42)

$$\begin{bmatrix}
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`Anjum'} \\ \text{CASE} & \text{ERG} \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{FEM} \\ \text{PERS} & 3 \end{bmatrix} \\
\text{OBJ}_{\text{GO}} & \begin{bmatrix} \text{PRED} & \text{`Saddaf'} \\ \text{CASE} & \text{DAT} \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{FEM} \\ \text{PERS} & 3 \end{bmatrix} \\
\text{PRED} & \text{`let-make} < __ , __ , __ > \text{'} \\
\text{ASPECT} & \text{PERF} \\
\text{GEND} & \text{MASC} \\
\text{NUM} & \text{SG} \\
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`necklace'} \\ \text{CASE} & \text{NOM} \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{MASC} \\ \text{PERS} & 3 \end{bmatrix}
\end{bmatrix}$$

(43)

$$\left[\begin{array}{l}\text{banaane diyaa 'let make'}\\ \left[\begin{array}{l}\text{CS}([\alpha],\text{GO}_{Poss}(\left\{\begin{array}{l}\text{CS}([\gamma],\text{BE}[necklace])\\ \textbf{AFF}_{+cc}([\beta]^{\gamma},\ \ )\\ \text{ASP}(\_\ \_\ \_)\end{array}\right\}_{E_T},\textbf{TO}[Saddaf]^{\beta}))\\ \text{AFF}_{+cc}([Anjum]^{\alpha},\ \ )\\ \text{ASP}(\_\ \_\ \_)\end{array}\right]_{E}\end{array}\right]$$

The aspect, gender and number specifications of the various PREDs are represented at the f-structure in (42). This fully specified f-structure is determined by the algorithm relating c-structures to f-structures, which essentially solves the annotations on the c-structure nodes. This algorithm serves to unify the functional information specified by the annotations and checks for consistency. For example, verb agreement, agreement between an adjective and its head, or the unification of information from discontinuous NPs (such as in Warlpiri (Simpson and Bresnan 1982)) are achieved through the processes of information unification defined by the algorithm.

Notice that in the f-structure in (42), the case of the object *haar* 'necklace' is given as nominative. This is due to the the well-formedness condition on case stated in (44), which assigns nominative case to a subject or object not specified for case by information from a case clitic. The well-formedness condition in (44) is preferable to the postulation of lexical entry for a phonologically null nominative case clitic because locative NPs do not always bear an overt locative marker. A phonologically null case clitic opens the door to c-structure representations in which locative NPs are annotated as nominative. By the Principles of Coherence and Completeness, these c-structures could not correspond to well-formed f-structures, but the possibilities would need to be examined before they were rejected. The assignment of default nominative to subjects and objects has the advantage that it does not give rise to possibilities which need to be examined and rejected.

(44) **Default Nominative Case**
If an f-structure f is the value of the attribute SUBJ or OBJ at the level of f-structure, and f does not contain a value for CASE (i.e., is not specified for case), then (f CASE) = NOM.

So, the OBJ 'necklace' in (43) was not assigned case by a clitic, and it received nominative case by default. The assignment of default nominative case is consistent with the idea that

all NPs bear case, whether it be overtly realized or not. It also corresponds to the notion that the nominative is a purely structural case. In Government-Binding, the nominative is assigned on the basis of the position of the NP in Spec of IP. The statement in (44) is consistent with the case-theoretic view expressed by Government-Binding theory, but incorporates the additional advantage that it does not necessarily tie the nominative case to a particular phrase structure position, but rather to a direct argument: either the subject or object.

The three levels of representation above thus are all linked to one another. The a-structure is linked to heads at c-structure, and in turn is linked to the f-structure in (43) by the principles of MT. The determination of specific grammatical function annotations on c-structure, which is based on information contributed by the lexical entries of the case clitics at c-structure, then sets up further correspondences between the empty subparts of the skeletal f-structure and NP nodes at c-structure. In addition to this, the f-structure is also linked tightly to the c-structure by the mapping algorithm which solves the annotations at c-structure for the full f-structure in (43).

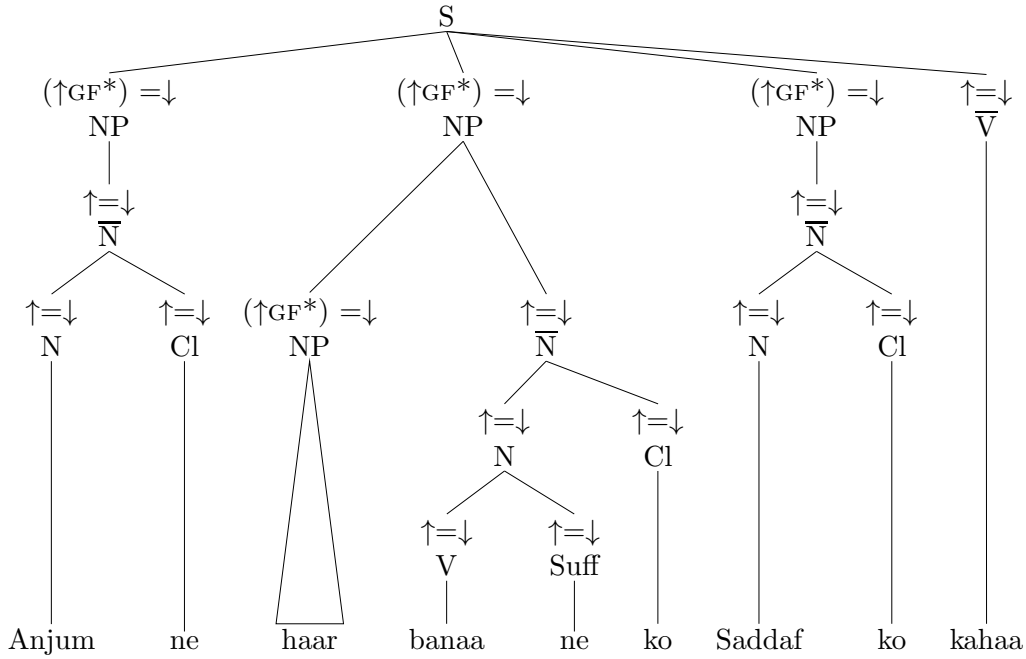
Although I have not demonstrated it explicitly, it should be clear that the other possible c-structure realization of the permissive, in which the two predicates are contained under the same $\overline{\text{V}}$, presents no challenge to the approach presented here. In the alternative realization, there are four nodes with underdetermined annotations: *Anjum=ne* 'Anjum=Erg', *Saddaf=ko* 'Saddaf=Dat/Acc', *haar* 'necklace', and the infinitive *banaa-ne* 'make'. The case clitics straightforwardly identify Anjum as the subject, and Saddaf as the indirect object. The finite verb *di-yaa* 'let' has already been identified as containing a transparent Event argument, which was filled by the a-structure of the infinitive. The infinitive *banaa-ne* is thus a complex predicate forming head, and its annotation must be realized as $\uparrow=\downarrow$. The remaining NP *haar* 'necklace' is instantiated as the only grammatical function remaining unlinked at the skeletal f-structure – the OBJ. So, while the c-structure realization differs from the c-structure in (41), the f-structure and a-structure representations corresponding to the alternative c-structure realization are exactly as in (42) and (43). The approach to complex predicate formation I propose thus again yields exactly the results needed to account for the collection of data in Chapter 3.

6.3.5 The Instructive

An example of the instructive is repeated here in (45). The phrase structure rules given previously generate the c-structure in (46) as one of the two possible representations for (45).

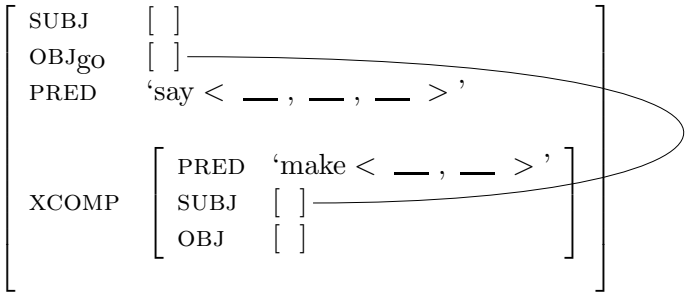
(45) anjum=ne haar banaa-ne=ko saddaf=ko kah-aa
Anjum.F=Erg necklace.M=Nom Saddaf.F=Dat make-Inf.Obl=Acc say-Perf.M.Sg
'Anjum told Saddaf to make a necklace.'

(46)



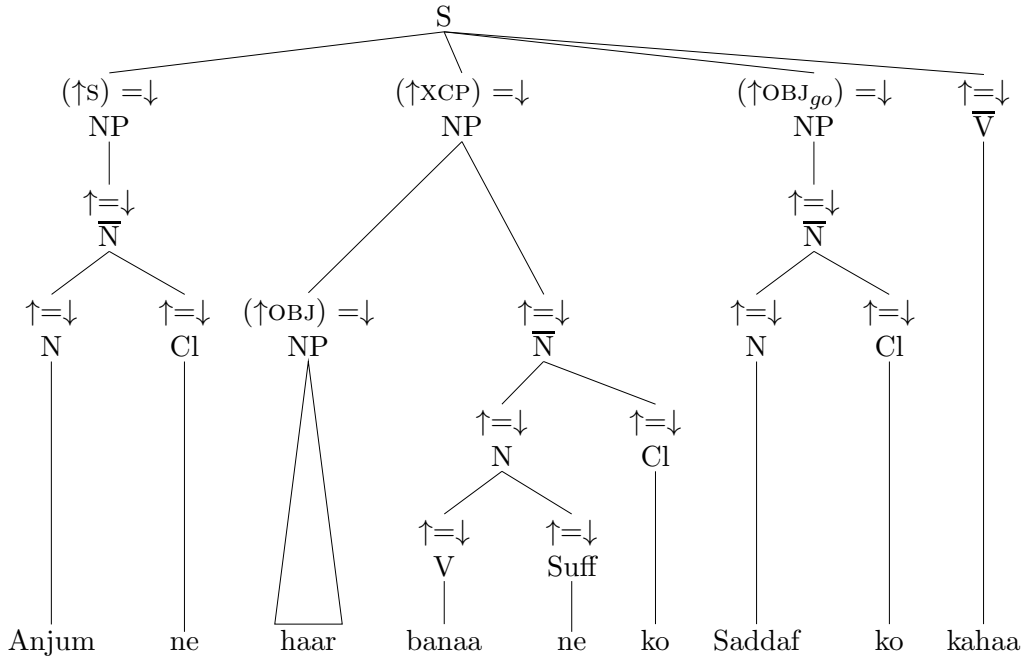
Predicate Composition, as defined for the permissive does not apply to the instructive. The immediately identifiable head in (46) is *kahaa* 'said'. There is no transparent Event at a-structure, so its argument structure does not *compose* with another predicate. The a-structures of *kahaa* 'say' and *banaane* 'make' are therefore linked up separately by MT. However, as the *kahaa* 'said' requires an Event argument to be satisfied, an a-structure of another predicate must be embedded within it. Mapping from a-structure to f-structure through MT, along with an embedding of a-structures, yields the skeletal f-structure in (47).

(47)



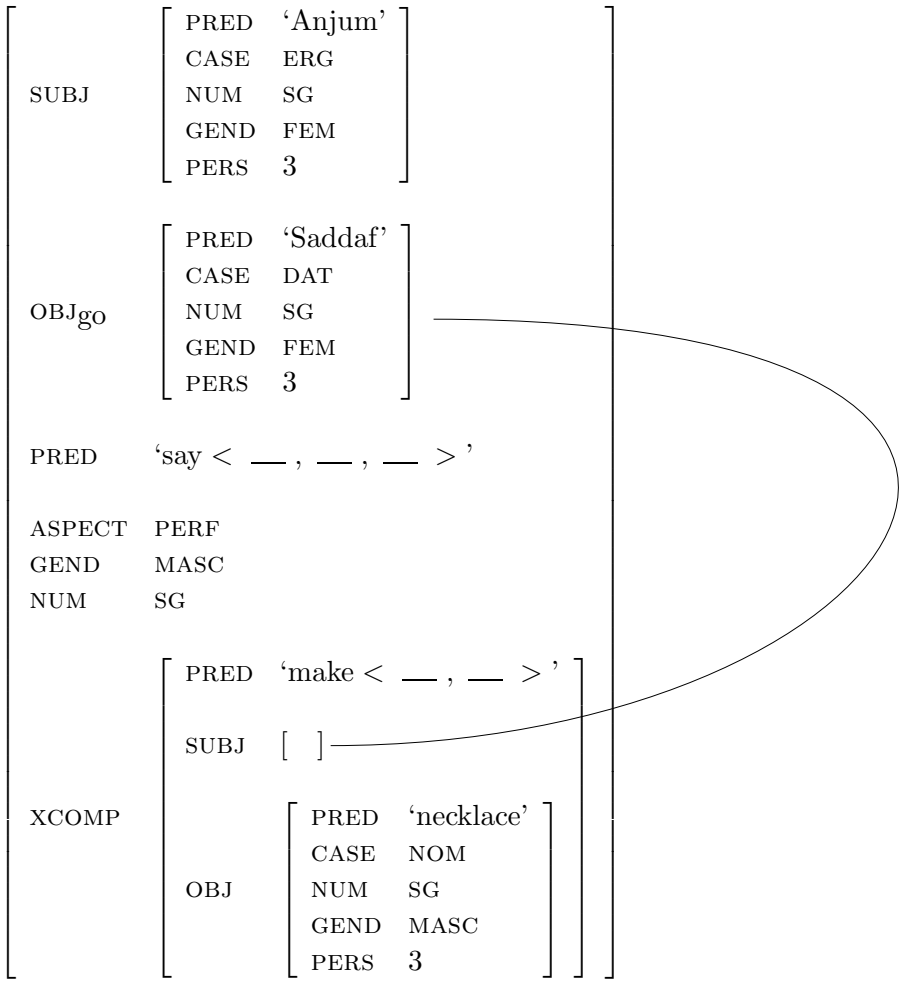
Again, the entry for the ergative case clitic *ne* requires that the NP node dominating *Anjum* be annotated as the subject. This time, however, there are two *ko* case markers. There is an argument of a TO function at a-structure, so the condition for dative case is satisfied for one of these *ko* clitics. When the *banaane* 'make' is annotated as the OBJ_{go} of the clause, the full f-structure arrived at ultimately is not complete. When the NP containing *Saddaf* is annotated as the indirect object, on the other hand, the levels of representation can be linked up successfully, and the f-structure is complete. The *ko* on the *banaane* 'make' must now be identified as accusative, and the fact that it is a predicate causes it to be linked to the XCOMP. This leaves the NP containing *haar*, which is linked to the XCOMP OBJ by default. The determination of c-structure annotations based on case information, and guided by the principles formulated above, is shown in (48) (XCOMP has been abbreviated as XCP in the interests of space).

(48)



The fully specified c-structure annotations in (48) can now be solved to yield the complete f-structure in (49). Since the grammatical functions represented at f-structure are linked to specific slots at a-structure through MT, the complex a-structure in (50), which represents the instructive given at the beginning of this section, is now complete as well.

(49)



(50)

$$\left[\begin{array}{l} \text{banaane kahaa 'told to make'} \\ \left[\begin{array}{l} \text{CS}([\alpha], \text{GO}_{Info}(\left[\begin{array}{l}\text{CS}([\gamma], \text{BE}[necklace]) \\ \text{AFF}_{+cc}([\ ]^{\gamma,\beta},\ ) \\ \text{ASP}(\_\ \_\ \_)\end{array}\right]_E, \text{TO}[Saddaf]^{\beta})) \\ \text{AFF}([Anjum]^{\alpha},\ ) \\ \text{ASP}(\_\ \_\ \_)\end{array}\right]_E \end{array}\right]$$

As was the case with the permissive, I again do not provide a detailed account of the alternative phrase structure realization of the instructive in which both the infinitive predicate *banaane* 'make' and the finite predicate *kahaa* 'said' are dominated by the same $\overline{\text{V}}$. Again, this is because the alternative phrase structure realization receives a straightforward account under this approach. There would be four nodes with underspecified (↑GF*)=↓ annotations: *Anjum=ne* 'Anjum=Erg', *Saddaf=ko* 'Saddaf=Dat/Acc', *haar* 'necklace', and *banaane=ko* 'make=Dat/Acc'. The ergative *ne* identifies *Anjum* as the subject, and the fact that *banaane=ko* is the only NP with an a-structure requires that it be linked to the Event argument of *kahaa* 'said'. The annotation on *banaane=ko* is thus instantiated as (↑XCOMP)=↓. The *Saddaf=ko* is now linked to the argument of TO and realized as the indirect object. The only remaining NP, *haar* 'necklace' is linked to the only grammatical function remaining available at f-structure: the XCOMP OBJ.

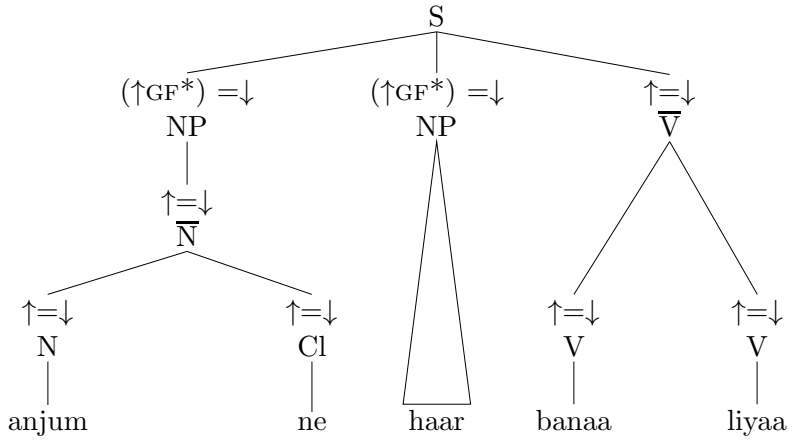
The utilization of information from case clitics to guide the determination of possible grammatical function annotations at c-structure thus demonstrably limits the number of possibilities that must be checked for well-formedness. Furthermore, as this section has shown, the determination of c-structure annotations does not have any adverse effects on the architecture of LFG, but rather, builds on recent advances in the explication of argument structure and mapping from a-structure to f-structure. The next section briefly demonstrates that the approach advocated here applies unproblematically to Aspectual complex predicates as well.

6.3.6 Aspectual Complex Predicates

Aspectual complex predicates show none of the disjunctive properties at c-structure which the permissive and the instructive display, so the linking procedure is very straightforward. Consider the representative Aspectual complex predicate in (51), and the corresponding c-structure realization in (52).

(51) anjum=ne haar banaa li-yaa
Anjum.F=Erg necklace.M=Nom make take-Perf.M.Sg
'Anjum made the necklace completely, on purpose.'

(52)



Since the a-structure of one of the V's contained under the $\overline{\text{V}}$, the a-structure of *le* 'take' contains a transparent Event, this predicate must compose with another one. In (52), there is another predicate immediately dominated by same $\overline{\text{V}}$ as *le* 'take', so this predicate is the first candidate for Predicate Composition (and indeed, it is the only candidate, but that need not be exhaustively determined). The complex a-structure which results from the fusion of the a-structures of *banaa* 'make' and *le* 'take', is shown in (53).

(53)

$$\left[\begin{array}{l} \text{banaa liyaa 'made completely'} \\ \left[\begin{array}{l} \text{CS}([\alpha], \text{BE}[\ ]) \\ \text{AFF}_{+\text{cc}}([\ ]^{\alpha},\ ) \\ \text{ASP}(\_\ \_\ 1) \end{array}\right]_{E} \end{array}\right]$$

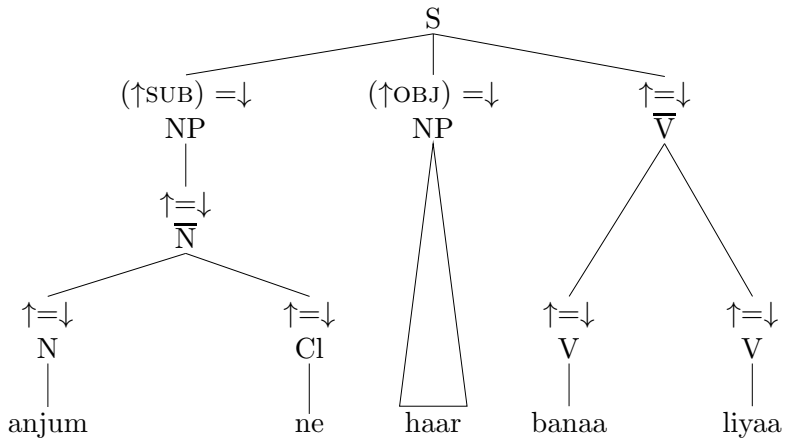
The application of MT to this a-structure results in the skeletal f-structure in (54).

(54)

$$\left[\begin{array}{ll} \textsc{pred} & \text{'complete-make} < \underline{\quad}, \underline{\quad} > \text{'} \\ \textsc{subj} & [\ ] \\ \textsc{obj} & [\ ] \end{array}\right]$$

The ergative *ne* again immediately identifies the NP containing *Anjum* as the subject. The NP dominating *haar* 'necklace' is linked to the one remaining unlinked OBJ grammatical function. The fully specified c-structure annotations are thus as shown in (55). The complete f-structure resulting from the annotations in (55) is shown in (56). Finally, the complete a-structure is shown in (57).

(55)



(56)

$$\begin{bmatrix} \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`Anjum'} \\ \text{CASE} & \text{ERG} \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{FEM} \\ \text{PERS} & 3 \end{bmatrix} \\ \text{PRED} & \text{`complete-make} < \_\_ , \_\_ > \text{'} \\ \text{ASPECT} & \text{PERF} \\ \text{GEND} & \text{MASC} \\ \text{NUM} & \text{SG} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`necklace'} \\ \text{CASE} & \text{NOM} \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{MASC} \\ \text{PERS} & 3 \end{bmatrix} \end{bmatrix}$$

(57)

$$\begin{bmatrix} \text{banaa liyaa `made completely'} \\ \begin{bmatrix} \text{CS}([\alpha], \text{BE}[necklace]) \\ \text{AFF}_{+\text{CC}}([Anjum]^{\alpha}, \ ) \\ \text{ASP}(\_\ \_\ 1) \end{bmatrix}_E \end{bmatrix}$$

The a-structure, f-structure, and c-structure representations shown above are thus clearly linked to one another and together determine the structure of a given sentence. In particular,

the method of relating the separate levels of representation allows a successful account of both the Aspectual and permissive complex predicates, and the complement constructions like the instructive.

6.4 Conclusion

Much of the analysis in this chapter has concentrated on how the differing levels of representation are related to, or linked up with one another. One could therefore be tempted to argue for a theory which represented the disparate types of information at a single level of syntactic representation. In my view, however, the separation of different types of information into separately identifiable levels of representation is exactly the component of this theory which allows a clear understanding of the structure of complex predicates. As suggested by the Urdu permissive and the Urdu Aspectual complex predicates, the defining characteristic of a complex predicate is a complex a-structure, which is linked to a simple f-structure. In other words, the crucial, defining characteristic of a complex predicate is that it consists of two or more semantic heads which correspond to a single syntactic head. This characteristic must be accounted for in any theory of syntax. The organization of LFG into separate, but interconnected, levels of representation allows a precise characterization of complex predicate formation.

Crucial to the success of the account of complex predicate formation at argument structure is the level of elaborated a-structure I propose. This level is based heavily on Jackendoff's (1990) formulation of Lexical Conceptual Structures (LCS) within a theory of Conceptual Semantics. While I do not adopt the theory of Conceptual Semantics, I adopt LCS-like representations for a level of elaborated a-structure. The kind of information represented in these structures, combined with the way the information is organized into differing tiers lends itself nicely to an account of complex predicate formation in terms of Argument/Event Fusion. In particular, the elaborated a-structures are ideal for an account of the restrictions on complex predicate formation observed for the Aspectual complex predicates.

Chapter 7

Extending the Approach

7.1 Introduction

The preceding chapters have presented two differing complex predicates in Urdu and a unifying analysis of complex predicate formation. In this chapter, I expand the domain of enquiry by surveying some well-known work on complex predicates in Romance languages and Japanese. Complex predicates of varying degrees of difference are found in languages all over the world and seem to be particularly popular in South Asia (to list just a few, see T. Mohanan (1993c) for Hindi N-V complex predicates, K.P. Mohanan (1989a) for Malayalam, Steever (1979) and Nagarajan (1990) for Tamil, Pandharipande (1990) for Marathi, and Krishnamurti (1989) for Telugu). However, I do not have the space here to present a discussion of all the languages in which complex predicates are either reported, or where some V-V sequences would appear to be complex predicates (Watters (1993) on Tepehua, Kural (1993) on Turkish). I therefore confine myself to a discussion of some of the better known cases and show how the analysis of complex predicates presented in this dissertation also accounts for complex predicates in Romance and Japanese.

In the last section of this chapter, I address the issue of serial verbs. The terms "serial verb" and "complex predicate" generally seem to be distributed according to language family: West African languages and creole languages have serial verbs, while other languages have complex predicates. Sometimes the terms are also used to denote one and the same entity. For example, Marathi complex predicates are referred to as compound verbs in Pandharipande (1989) and as serial verbs in Pandharipande (1990). Despite such overlap, there is a general feeling in the literature that complex predicates (or compound verbs)

must differ in some way from serial verbs. However, it has as yet not become clear exactly what characterizes a serial verb as opposed to a complex predicate. This overall confusion in the literature clearly runs counter to the interest of arriving at a unifying analysis of both complex predicates and serial verbs. The last section examines some of the work on serial verbs, and compares the structure of serial verbs to the structure of complex predicates, as formulated in this dissertation. I argue that serial verbs can indeed be distinguished from complex predicates once some clear criteria are established.

7.2 Romance Complex Predicates

There has been quite an impressive amount of work done on Romance complex predicates and related issues. I do not intend to address the entire body of literature,[1] but instead focus on the argument structure approach taken by S. Rosen (1989). Rosen provides an account of Romance causative, perception and restructuring verbs. She identifies a process of *argument merger* that she claims serves to unify all of the above complex predicates as well as other, crosslinguistic, complex predicate phenomena. Rosen's account is of particular interest here because her approach is very close to the Argument Fusion approach formulated for Urdu, and yet differs in some significant ways.

7.2.1 Restructuring Verbs

As mentioned, Rosen posits a process of *argument merger* to account for complex predicates in Romance. However, rather than arriving a unifying notion of argument merger, she formulates three different varieties: *light merger*, *complete merger*, and *partial merger*. Restructuring verbs trigger a *light merger*, which is analogous to the *argument transfer* analysis proposed by Grimshaw and Mester (1988) for the Japanese light verb *suru* 'do'. Essentially, Romance restructuring verbs like the Italian modal type *volere* 'want', or the verb of motion type like *andare* 'go', are described as *light* in the sense that they have a completely empty argument structure. The argument structure of the entire complex predicate is thus isomorphic with the argument structure of the main verb. This is illustrated in (1) for *volere* 'want'. The parantheses indicate the argument structure of the verbs. The verb

[1]For some more recent detailed discussions of the literature, see S. Rosen (1989), Alsina (1993), and Manning (1992). For original insights and discussions of the problem see Rizzi (1982), Aissen and Perlmutter (1983), and Davies and C. Rosen (1988).

volere 'want' has an empty argument structure, while the verb *leggere* 'read' has two arguments, x and y. Argument structure is hierarchically organized. The parentheses enclosing the argument y thus serve to set off y from x and indicate that x is the more prominent argument. The <e> in the argument structure representations stands for *event* and indicates that each of the verbs denotes an event. The events are identified with one another during argument merger. This is expressed by a bar which joins two events together and reflects the fact causatives and restructured clauses in Romance do not contain two separate IPs. Rosen proposes that embedded clauses in Romance complex predicates are VPs at phrase structure.[2]

(1) volere ( )<e>
'want'

leggere (x (y))<e>
'read'

⟶ volere leggere (x (y))<e> <e>
'want read'

Rosen is an advocate of the view, argued for in detail by Grimshaw (1990), that argument structure contains no semantic information whatsoever. The arguments of a predicate are merely indicated by place-holding variables such as "x" and "y" in (1). The semantic information usually associated with theta-roles is taken to be represented at the Lexical Conceptual Structure (LCS) of a given lexical item. Although Rosen argues that an argument structure analysis is the right approach to complex predicates, her proposals for the representation of argument structure and her analysis of light verbs differ substantially from the approach I present for Urdu. In what follows, I suggest that the elaborated level of a-structure allows a more insightful account of Romance restructuring verbs.

For one, it is not immediately clear how a restructuring verb like *volere* 'want' is to be distinguished from an auxiliary. Presumably, the argument structure of an auxiliary is just as empty and incomplete as the argument structure of *volere* 'want' in (1). Rosen (1989:174) posits that the semantic content of a light verb is represented at its LCS and that the difference between light and "heavy" versions of a predicate are not expressible in terms of their semantic contents, but are a direct result of whether or not actual arguments are mapped from an LCS to argument structure. In the case of heavy verbs, arguments

[2]But see Alsina (1993) for evidence that the embedded clauses in causatives cannot be treated as VP complements.

are mapped from an LCS to argument structure; in the case of light verbs they are not. Presumably, then, the difference between a restructuring verb and an auxiliary is defined at the semantic level of an LCS representation as well, although it is not quite clear how this is achieved.

The difference between Rosen's approach and the approach described in this dissertation may so far appear to be simply one of terminology. The idea of Lexical Conceptual Structures (LCS), which represent the semantics of an item, is based on Jackendoff (1990). The LCS of *volere* 'want' for Rosen, then, is roughly equivalent to an a-structure representation for *volere* 'want' under my approach. I argue, however, that the differences are not merely terminological. The fact that Rosen does not admit any semantic information in her argument structure representation does not allow her complete insight into auxiliary selection for Italian restructuring verbs.

Romance restructuring is characterized by three basic phenomena: clitic climbing, long object preposing, and auxiliary selection. The facts of auxiliary selection, however, can only be observed for Italian. As it is these facts which provide the data in favor of an elaborated a-structure approach, the discussion here is necessarily limited to Italian restructuring verbs.

The general facts of auxiliary selection for Italian are that unaccusatives and passives select *essere* 'be', while unergatives and transitives select *avere* 'have'. The light merger account illustrated in (1) predicts that auxiliary selection in complex predicates will always be determined by the main verb, because it is this verb's argument structure which completely determines the argument structure of the complex predicate.

This prediction initially holds up fairly well with the *volere* 'want' type light verbs, but breaks down completely for the *andare* 'go' type verbs. This is illustrated in (2) and (3). In (2a), the light verb *volere* 'want' is combined with the unaccusative predicate *andare* 'go'. The presence of the clitic *ci* in the matrix clause indicates that clitic climbing and, therefore, restructuring has occured. The sentence in (2a) is only good if the auxiliary is *essere* 'be' (*sarebbe*). Conversely, in (2b), only the auxiliary *avere* 'have' is good with the heavy verb *prendere* 'take'.

(2) a. Mario ci sarebbe proprio voluto andare.
*Mario ci avrebbe proprio voluto andare.
'Mario would really want to go there.'

b. *Giovanni lo avrebbe voluto prendere.
Giovanni lo avrebbe voluto prendere.
'Giovanni would have wanted to fetch it.'

Rosen argues that the contrast in (2) is easily explained under the view that *volere* 'want' has no argument structure of its own, so it is the main verb which solely determines the selection of the auxiliary. However, the data in (3) do not fit this generalization. Here the main verb is *prendere* 'take', so the auxiliary selected should be *avere* 'have' and not *essere* 'be'. But in fact, whenever an *andare* 'go' type light verb is involved, the selected auxiliary is always *essere* 'be' (Rosen 1989:194).[3]

(3) Giovanni lo è andato a prendere.
'Giovanni has (essere) gone to fetch it.'

The contrast between (2) and (3) is unexpected exactly because no argument structure information for *volere* 'want' and *andare* 'go' is considered to be relevant. A closer examination of the two types of verbs reveals that the former type consists of verbs which are all assumed to take an external argument, either an agent or an experiencer, in their heavy forms: *volere* 'want', *cominiciare* 'begin', *continuare* 'continue', and *dovere* 'must'. The latter type consists of verbs which are generally taken to be unaccusative: *andare* 'go', *venire* 'come', and *stare* 'stand' (see Rizzi 1982). Now, if one were to take the view that both the *volere* 'want' and *andare* 'go' type light verbs do make a contribution to the overall argument structure of the complex predicate, the data in (2) and (3) could be explained.

Rosen (1989:194) remarks that "... the andare-type verbs behave as if they have an argument structure of their own to contribute to the merged representation, ...". However, she cannot convince herself that the light verbs really should have a non-empty argument structure. The primary reason for her rejection of a non-empty argument structure approach is based on the *volere* 'want' type verbs. If *volere* 'want' were taken to contain an external argument, then the complex predicate *volere prendere* 'want to take' would contain two external arguments which were bound together by argument structure merger (Rosen 1989:188). In Rosen's framework, the binding of two external arguments must necessarily result in a PRO at the embedded subject position. Since there is evidence against an embedded PRO, Rosen opts for the empty argument merger analysis.

However, the distinction between Argument Fusion and Argument Control posited in the preceding chapter allows for the necessary differentiation between complex predicates

[3]These data are taken from Rosen (1989) and Rizzi (1982). However, C. Rosen (p.c.) reports that sentences with *andare* 'go' as a restructuring verb are extremely marginal in Italian. If this is the case, then these examples do not actually pose a problem for either the GB approach of S. Rosen (1989), or the RG approach in C. Rosen (1993). But see Rizzi (1982) for a discussion of these data.

and complement constructions. My approach thus allows a reconsideration of the idea that the restructuring light verbs might indeed have a non-empty argument structure. Consider first the a-structure for *andare* 'go' shown in (4). It indicates that *andare* 'go' has a theme and a transparent Event argument, and is thus unaccusative.[4]

(4)

$$\left[\begin{array}{l} \text{andare 'go'} \\ \left[\begin{array}{l} \text{GO}([\alpha], \text{TO}\{\ \ \}_{E_T}) \\ \text{AFF}(\ \ , [\ \ ]^{\alpha}) \\ \text{ASP}(\_\ \_\ \_) \end{array}\right]_E \end{array}\right]$$

The only difference between this restructuring light verb *andare* 'go' and its "heavy" counterpart is the fact that the light verb has a transparent Event at a-structure. The overall semantics of the light and heavy versions of a verb are otherwise exactly the same. This is in accordance with the fact that restructuring is always optional in Romance and that the semantics of the expression do not change when restructuring does occur.

The auxiliary selection facts for *andare* 'go' can now be accounted for as follows. When an *andare* type light verb forms part of a complex predicate, as in (6), its "theme" argument is fused with the highest embedded argument. I propose that whenever a theme argument fuses with any other type of argument in complex predicate formation, the auxiliary *essere* 'be' must be selected. The complex a-structure in (6) represents the complex predicate *andato a prendere* 'gone to fetch' in (5). The transparent Event argument of *andare* 'go' has triggered Argument Fusion: the "theme" argument *Giovanni* has been fused with the highest embedded, β marked, argument of *prendere* 'take'. This β marked argument is an external argument in the sense that it is the first argument of the function AFF in the embedded a-structure. However, since it has been fused with a theme, the auxiliary selected must be *essere*.

(5) Giovanni lo è andato a prendere.
'Giovanni has (essere) gone to fetch it.'

[4]Recall that the α marked argument of AFF in (4) is a theme rather than agent because it is the second, not the first, argument of AFF at the Action Tier. Furthermore, this argument must be a theme, and not a patient, because there is no actor specified at the Action Tier. This is my innovation and not necessarily consistent with Jackendoff (1990).

(6)

$$\left[\begin{array}{l} \text{andato a prendere 'go to fetch'} \\ \left[\begin{array}{l} \text{GO}([\alpha], \text{TO} \left\{\begin{array}{l} \text{CS}([\beta], \text{BE}([it], \text{AT}[\beta])) \\ \text{AFF}([\alpha]^{\beta}, \ \ ) \\ \text{ASP}(\_\ \_\ \_) \end{array}\right\}_{E_T} \ \ ) \\ \text{AFF}(\ \ , [Giovanni]^{\alpha}) \\ \text{ASP}(\_\ \_\ 1) \end{array}\right]_{E} \end{array}\right]$$

Now consider the case in (7), where the light verb *volere* 'want' combines with the unaccusative *venire* 'come' (taken from Rizzi 1982:20). Here the auxiliary *essere* 'be' is determined by *venire* 'come' rather than the light verb *volere* 'want'.[5]

(7) Maria è voluta venire con noi.
'Maria (essere) wanted to come with us.'

The a-structure in (8) shows the representation needed for the light verb *volere* 'want' under my approach. In (8), *volere* 'want' has both an external and a transparent Event argument.[6] The argument labeled α on the Action Tier is considered external because it is the first argument of AFF.

(8)

$$\left[\begin{array}{l} \text{volere 'want'} \\ \left[\begin{array}{l} \text{WANT}([\alpha], \{\ \ \}_{E_T}) \\ \text{AFF}_{+cc}([\ \ ]^{\alpha},) \\ \text{ASP}(\_\ \_\ \_) \end{array}\right]_{E} \end{array}\right]$$

The a-structure representation of the complex predicate *voluta venire* 'wanted to come' in (7) is shown in (9). Although the most prominent argument, *Maria*, here is shown as an external argument, notice that it fuses with a theme: the argument labeled both α and β on the embedded Action Tier (AFF). Recall that when a theme is fused with an argument

[5]Example (7) is actually also good with the auxiliary *avere* 'have'. However, when *avere* 'have' is selected, restructuring is argued not to have taken place, so that *volere* 'want' is the sole determiner of auxiliary selection. See Rizzi (1982) and Rosen (1989) for details.

[6]I use the term *external argument* here to avoid the labels agent or experiencer. In Urdu, the case marking on the wanter indicates that it must be an agent. In Romance, however, the wanter is usually considered to be an experiencer. The precise nature of the label is irrelevant, given the elaborated level of a-structure.

of any type, as in (9), the appropriate auxiliary is *essere* 'be'. In other words, all of the semantic information associated with the argument structure of unaccusatives like *venire* 'come' is not lost under argument fusion, but continues to play a role in the determination of the syntax of the entire complex predicate.

(9)

$$\left[\begin{array}{l} \text{voluta venire 'want to come'} \\ \left[\begin{array}{l} \text{WANT}([\alpha], \left\{\begin{array}{l} \text{COME}([\beta], \text{WITH}[us]) \\ \textbf{AFF}(\ \ , [\alpha]^{\beta}) \\ \text{ASP}(\_\ \_\ \_) \end{array}\right\}_{E_T}\ ) \\ \textbf{AFF}_{+cc}([Maria]^{\alpha}, \ \ ) \\ \text{ASP}(\_\ \_\ 1) \end{array}\right]_{E} \end{array}\right]$$

Manning (1992) argues that *volere* 'want' type light verbs are transparent in that they do not, ultimately, project an argument of their own into the syntax. Manning's position is thus very close to Rosen (1989), who proposed that the argument structure of Romance restructuring verbs is completely empty. The intent of both of these approaches is to account for the data in (10) and (11). In (10), the light verb *volere* 'want' combines with two verbs: an unaccusative (*andare* 'go') and a transitive verb (*prendere* 'take'). The auxiliary must always be *avere* 'have' and is considered to be determined by the transitive main verb *prendere* 'take'.

(10) Maria li avrebbe voluti andare a prendere lei stessa.
*Maria li sarebbe voluti andare a prendere lei stessa.
'Maria would have (have/*be) wanted to go fetch them herself.'

(11) Maria ci sarebbe dovuta cominciare ad andare.
'Maria would have (essere) had to begin to go there.'

Similarly, in (11), the auxiliary *essere* 'be' appears to be determined by the main verb *andare*. The *volere* 'want' type light verbs thus seem to be completely transparent for purposes of auxiliary selection, while the *andare* 'go' type light verbs clearly do project some type of information. One could therefore, propose that *volere* 'want' type light verbs must have an empty argument structure, while *andare* 'go' type light verbs project a theme. This approach not only fails to unify the restructuring phenomena in Italian, it also does not provide a completely satisfactory account for the data in (10). If *volere* 'want' is transparent

in (10), but *andare* 'go' does contribute a theme argument to the complex predicate, then why is the auxiliary *essere* 'be' not possible?

Rather than formulating an account where one type of restructuring verb is taken to be 'lighter' than another, I instead propose that the auxiliary selection facts be accounted for in terms of the actual process of complex predicate formation. Consider the a-structure representation in (12), which corresponds to the complex predicate *voluti andare a prendere* 'wanted to go to fetch' in (10).

(12)

$$\left[\begin{array}{l} \text{voluti andare a prendere 'wanted to go to fetch'} \\ \left[\begin{array}{l} \left[\begin{array}{l} \text{WANT}([\alpha], \left\{\begin{array}{l} \text{GO}([\beta], \text{TO} \left\{\begin{array}{l} \text{CS}([\gamma], \text{BE}([them], \text{AT}[\gamma])) \\ \textbf{AFF}([\alpha]^{\gamma}, \ \ ) \\ \text{ASP}(\_\ \_\ \_) \end{array}\right\}_{E_T}) \\ \text{AFF}(\ \ , [\alpha]^{\beta}) \\ \text{ASP}(\_\ \_\ \_) \end{array}\right\}_{E_T}) \\ \textbf{AFF}_{+cc}([Maria]^{\alpha},) \\ \text{ASP}(\_\ \_\ 1) \end{array}\right]_{E} \end{array}\right] \end{array}\right]$$

In (12), the light verb *volere* 'want' contributes the argument *Maria* to the complex predicate. This argument fuses with each of the highest arguments contributed by the other predicates. That is, all the argument slots on the Action (AFF) tier of the embedded predicates have been filled with the label α, indicating that they are fused with the argument *Maria*. The idea is that Argument Fusion, which is triggered by the presence of a transparent Event at a-structure, sequentially fuses each of the highest arguments of the embedded predicates to the matrix argument slot filled by *Maria*.[7] In addition to the semantic information associated with the matrix argument, only the semantic information of the last argument position to be fused is retained, and taken into consideration for the selection of the auxiliary. In other words, only information for two argument positions is accessible once Argument Fusion has taken place: the matrix argument and the argument position that has most recently undergone fusion. The information associated with intermediate argument positions is discarded. In (12), for example, the last argument fused with the argument slot

[7]Another possible way to conceive of Argument Fusion is to have each argument fuse, in turn, with the argument embedded immediately below it. This method, however, will not yield the right results for Romance.

filled by *Maria* is the argument of AFF labeled both α and γ. Since this argument and the matrix argument filled by *Maria* are both external arguments, the auxiliary selected must be *avere* 'have'.

In contrast, for the complex predicate in (11), the last argument fused is the equivalent to a theme (second argument of AFF). Although the matrix light verb *dovuta* 'must' provides an external argument, in this case the auxiliary *essere* 'be' must be selected because the last argument fused with is a theme, and whenever a theme is fused with any other type of argument, the selected auxiliary must be *essere* 'be'.

The utilization of an elaborated level of a-structure thus allows a more successful characterization of Romance restructuring verbs than was previously possible. Light verbs are distinguishable from auxiliaries by the virtue of the fact that light verbs have an argument structure while auxiliaries do not. The identical semantic content of light verbs and their heavy counterparts is also accounted for. The only difference between a light verb and its heavy counterpart is the fact that a light verb contains a transparent Event at a-structure. The presence of this transparent Event licenses restructuring. Since each light verb has a heavy counterpart which does not contain a transparent Event, the optionality of restructuring generally observed in Romance is also accounted for. Finally, and most importantly, the elaborated a-structure approach allows for a complete analysis of the auxiliary selection facts.

7.2.2 Romance Causatives

Rosen (1989) presents a treatment of Romance causatives much along the lines of Romance restructuring verbs. However, in order to be able to account for causatives, she needs to posit two further types of Argument Merger: Partial Merger and Complete Merger. The reason for the distinction between Partial and Complete Merger is the differing passivization facts in Italian, as compared to French and Spanish. The Italian causative can be passivized in its entirety, while the French and Spanish causatives do not allow passivization of the whole complex predicate. Rosen concludes that the French and Spanish causatives therefore only undergo a Partial Merger, in which the argument structures are combined, but the highest argument of the embedded argument structure retains its status as an external argument.

During Argument Merger, whether it be complete or partial, one of the arguments of the light verb is replaced with the argument structure of a full predicate. This is akin to the embedding of an argument structure within a transparent Event argument. And

indeed, the argument which is replaced by the argument structure of a heavy predicate must always be an Event within Rosen's system. However, as Rosen explicitly argues for a semantically underdetermined argument structure, it is not immediately clear how the generalization that Complete and Partial Merger can take place only in the presence of an Event argument is accounted for in a non-stipulative manner. Under the elaborated argument structure approach I formulate, this requirement is expected, and indeed crucial for the process of complex predicate formation.

In addition to the issues mentioned above, a great deal more can be said about Romance causatives, and indeed is said by Alsina (1993), who reviews the properties of Romance causatives and formulates an account within LFG in terms of argument structure composition. Recall that my approach is very close to Alsina's approach. I therefore do not engage with Rosen's approach to causatives in any more detail here, but instead refer the reader to Alsina's (1993) cross-linguistic account of causatives.

7.3 Japanese 'suru'

Like the Romance restructuring verbs and causatives, the Japanese light verb *suru* 'do' has received a considerable amount of attention in the literature (for a recent discussion, see Matsumoto (1992)). Grimshaw and Mester (1988) were the first to provide a complex predicate analysis for *suru* 'do' in terms of processes at argument structure. In particular, they proposed a mechanism of *Argument Transfer*, on which Rosen (1989) based her account of *Argument Merger* for Romance. Again, it is particularly interesting to examine the issues surrounding *suru* 'do' here because the approach originally presented in Grimshaw and Mester (1988) is close to the complex predicate account I develop for Urdu, and yet differs from it quite concretely. Furthermore, Japanese is a verb final language, the use of morphological case marking is extensive, and debate as to whether or not Japanese provides evidence for a VP is far from settled (see Yatabe 1993). As Japanese is thus very much like Urdu in these respects, a comparison of complex predicate formation across these two languages can only be instructive.

7.3.1 Differing types of 'suru'

At least three different types of *suru* 'do' are clearly identifiable. A detailed examination of the accumulated evidence may require finer distinctions between the differing uses of *suru*

'do' (see Matsumoto (1992) for some discussion). However, an investigation of such detail is beyond the scope of this work.

Grimshaw and Mester (1988) identify three types of *suru* 'do'. Though there is some disagreement as to which type of sentence is an illustration of which type of *suru* 'do', the bulk of the literature seems to accept these three basic categories. As might be expected, there is both a heavy and a light version of *suru* 'do'. In addition to these, there is a construction refered to as "incorporated" *suru* 'do'.

The sentences in (13) and (14), taken from Grimshaw and Mester, give examples of light and incorporating *suru* 'do', respectively. The bold-faced items in (13) and (14) are the nouns, referred to as θ-Transparent NPs in Grimshaw and Mester and θ-nouns in Sells (1989), which form a complex predicate with *suru* 'do'. The unusual feature of (13) is that while arguments of the sentence are contributed by the noun, the arguments in (13) have 'verbal' (nominative, accusative, topic) rather than 'nominal' (genitive) case marking. The peculiarity to be accounted for with light *suru* 'do' thus is that the arguments contributed by the noun do not have to appear in genitive case, but seem to be assigned Case by the light verb *suru* 'do'.

(13) John-wa Bill-to **aiseki-o** shita
John-Top Bill-with table-sharing-Acc do.Past
'John shared a table with Bill.'

(14) John-wa Bill-to **aiseki** shita
John-Top Bill-with table-sharing do.Past
'John shared a table with Bill.'

The only difference between (13), light *suru* 'do', and the "incorporating" *suru* 'do' in (14) is the fact that the θ-noun *aiseki* 'table-sharing' itself is not case marked in the incorporating construction. Although these two constructions have roughly the same meaning, there are some subtle semantic differences in terms of referentiality (Poser 1991) and in the requirement of agency. Incorporated *suru* 'do' can combine with both unaccusative and unergative nouns, while light *suru* 'do' can only combine with unergative nouns (Miyagawa 1989, Tsujimura 1990). Poser (1991) shows that the "incorporated" *suru* 'do' actually cannot be an instance of lexical incorporation. Instead, the *aiseki shita* 'table-sharing do' in (14) form an N-V complex.[8] It also does not show any of the intriguing "mixed" case

[8]It is tempting to analyze this N-V complex as a complex predicate, however, there is evidence that the "incorporating" *suru* 'do' actually does not contribute to a complex argument structure, but that it rather just functions as a placeholder for the aspectual and tense information (Sells p.c.).

marking properties of light *suru* 'do', which led Grimshaw and Mester (1988) towards an Argument Transfer analysis. The "incorporating" *suru* 'do' is thus more of an auxiliary than a light verb in my terms, and I leave this construction aside, instead concentrating on the properties of light *suru* 'do'.

7.3.2 Argument Transfer

7.3.2.1 Grimshaw and Mester 1988

Grimshaw and Mester (1988) argue that light *suru* 'do' is thematically incomplete because it assigns no theta-roles. However, they do describe light *suru* 'do' as a transitive verb in the sense that it assigns accusative Case to the θ-noun. They posit a process of Argument Transfer to account for the fact that the arguments of a sentence like (13) seem to be theta-marked by a verb, rather than a noun. If they were truly arguments of the noun, the arguments contributed by the noun *aiseki* 'table-sharing' in (13), for example, should be marked with the genitive case.

Grimshaw and Mester further argue that the behavior of light *suru* 'do' cannot be analyzed within usual conceptions of complex predicate formation, in which two predicates and their argument structures are merged into a single unit. They argue that the possibility of "mixed" structures, in which some of the θ-noun's arguments receive "verbal" case, while others are marked with the genitive, provide the crucial evidence for an Argument Transfer account. Within the system they propose, at least two of the θ-noun's arguments must be transferred to the empty argument structure of *suru* 'do', which then theta-marks them. When a noun is ditransitive, one of its arguments may be genitive while the other two receive 'verbal' case through *suru* 'do'. These facts are seemingly impossible to account for within a theory of complex predicate formation, such as the one I have formulated, in which two or more argument structures are combined into a single domain.

In what follows, I present some of the core data to be accounted for, along with a more detailed description of Grimshaw and Mester's approach. I then consider the data in light of Sells (1989) and show that light *suru* 'do' is actually amenable to the analysis of complex predicates I presented in the previous chapter, and that, in fact, complex predicate formation based on the elaborated argument structure I propose can account for the properties of light *suru* 'do' quite elegantly, and removes some of the mystery of why certain things have to be stipulated to be 'just so' within Grimshaw and Mester's proposals.

7.3.2.2 The Basic Pattern and Analysis

The basic pattern in need of an explanation is represented by (15) and (16), which are taken from Grimshaw and Mester. Note that while Grimshaw and Mester report (15b) to be questionable, Matsumoto (1992) finds this example perfectly acceptable. I will return to this point further on.

(15) a. John-wa Bill-to **aiseki-o** shita
John-Top Bill-with table-sharing-Acc do.Past
'John shared a table with Bill.'

b. ?John-wa [Bill-to-no **aiseki**]-o shita
John-Top Bill-with table-sharing-Acc do.Past
'John shared a table with Bill.'

(16) a. John-wa murabito-ni [ookami-ga kuru-to] **keikoku-o** shita
John-Top villagers-Dat wolf-Nom come-Comp warning-Acc do.Past
'John warned the villagers that the wolf was coming.'

b. John-wa murabito-ni [[ookami-ga kuru-to]-no **keikoku**]-o shita
John-Top villagers-Dat wolf-Nom come-Comp-Gen warning-Acc do.Past
'John warned the villagers that the wolf was coming.'

c. ?*John-wa [ookami-ga kuru-to] [murabito-e-no **keikoku**]-o shita
John-Top wolf-Nom come-Comp villagers-Dat-Gen warning-Acc do.Past
'John warned the villagers that the wolf was coming.'

d. *John-wa [murabito-e-no [ookami-ga kuru-to]-no **keikoku]-o**
John-Top villagers-Dat-Gen wolf-Nom come-Comp-Gen warning-Acc
shita
do.Past
'John warned the villagers that the wolf was coming.'

Grimshaw and Mester note that only the nouns which can also head a derived nominal are able to combine with light *suru* 'do' (Inoue 1976, but see Manning (1993) for arguments on the verbal noun status of those nominals). In Grimshaw's (1990) terms, such nouns are process nominals and are, crucially, the kinds of nouns which do have an argument structure of their own. The light verb *suru* 'do', on the other hand, is considered not to contribute any arguments of its own: it has the empty argument structure shown in (17). However, as the θ-nouns (the process nominals) *aiseki* 'table-sharing' and *keikoku* 'warning' in (15) and (16) are marked with the accusative *-o*, a Case which they can hardly be assigning to

themselves, Grimshaw and Mester must give light *suru* 'do' the ability to assign accusative Case. Light *suru* 'do' thus does not assign accusative Case to one of its own arguments, rather, it assigns Case to the NP which is its immediate sister.

(17) *suru*, V; ( )<acc>

I argue later that the Case assignment ability of *suru* 'do' is unnecessarily stipulative within Grimshaw and Mester's approach and show that if *suru* 'do' as a light verb is viewed as having an argument structure, albeit a semantically bleached one, as was the case with the Urdu light verbs, then the Case assignment facts, as well as the Argument Transfer data, fall out quite nicely. For the moment, however, I proceed with the discussion of the basic pattern and Argument Transfer analysis.

Because *suru* 'do' in (17) is thematically incomplete, it must combine with a noun which has an argument structure. The respective argument structures for the θ-nouns *aiseki* 'table-sharing' and *keikoku* 'warning' are given in (18a) and (18b).

(18) a. *aiseki* 'table-sharing' (Agent, Comitative)

b. *keikoku* 'warning' (Agent, Goal, Theme)

Rather than positing a process of complex predicate formation which merges the argument structures of the θ-noun and light *suru* 'do' as was proposed for Romance by Rosen (1989), Grimshaw and Mester posit a process of Argument Transfer. Under this system, some or all of the arguments of the noun are transferred to the empty argument structure of light *suru* 'do'. Light *suru* 'do' then theta-marks its newly acquired arguments and thus enables them to receive 'verbal' case. The process of Argument Transfer must, however, obey the rules in (19).

(19) i. At least one argument apart from the subject must be outside the NP (i.e., must have been transferred).

ii. The subject argument must always be outside the NP.

iii. For Nouns that take a Theme and a Goal, if the Theme argument is realized outside NP, the Goal must also be realized outside NP.

Crucially, this system allows for *Partial Transfer*, which appears to be needed for the data in (16). While (16a) is an example of *Complete Transfer*, in which all of the arguments of the noun have been transferred to *suru* 'do', the example in (16b) appears to show that the

possibility of Partial Transfer is needed. In (16b), the agent *John* and the goal *villagers* appear outside of the θ-NP. The theme *ookami-ga kuru-to-no* 'wolf come-that-Gen', on the other hand, bears genitive case and would appear to have remained inside the θ-NP. Because a theory of complex predicate formation, in which two argument structures are simply unified in some way (merger, fusion), would predict that an example like (16b) could not exist, examples like (16b) are crucial for the Argument Transfer analysis. In light of further evidence presented in Sells (1989), however, it becomes clear that it is indeed possible to account for data like (16b) within a theory of complex predicate formation by Argument Fusion.

Examples (15b) and (16d) are bad because they do not meet the requirements in (19i) that at least one argument besides the subject must be transferred. In (15b) the Comitative argument *Bill* is genitive and has thus not been transferred from the noun's argument structure. Similarly in (16d), both the Goal argument *villagers* and the Theme argument *that wolf coming* are marked with the genitive case, have thus not been transferred, and produce an ill-formed result. Example (16c), on the other hand, violates the constraint in (19iii). Here the Goal *villagers*, but not the Theme *that wolf coming* is marked with the genitive.

The various possibilities of Argument Transfer are thus as shown below, in the format proposed by Grimshaw and Mester.[9]

Complete Transfer:
aiseki 'table-sharing' (Agent, Comitative)
suru 'do' ( )<acc>
aiseki ( ) + *suru* (Agent, Comitative)<acc> (= (15a))

keikoku 'warning' (Agent, Goal, Theme)
suru 'do' ( )<acc>
keikoku ( ) + *suru* (Agent, Goal, Theme)<acc> (= (16a))

[9]Note that while Grimshaw and Mester are not explicit about the Comitative argument of *aiseki* 'table-sharing', I assume that it is treated like a Theme with regard to the noun.

Partial Transfer:
keikoku 'warning' (Agent, Goal, Theme)
suru 'do' ( )<acc>
keikoku (Theme) + *suru* (Agent, Goal)<acc> (= (16b))

Impossible:
aiseiki 'table-sharing' (Agent, Comitative)
suru 'do' ( )<acc>
aiseki (Comitative) + *suru* (Agent)<acc> (= (15b))

keikoku 'warning' (Agent, Goal, Theme)
suru 'do' ( )<acc>
keikoku (Goal, Theme) + *suru* (Agent)<acc> (= (16d))

The fact that only nouns like *keikoku* 'warning', which have a ditransitive argument structure, allow for Partial Argument Transfer follows from the fact that at least two arguments must always be transferred.

The constraints in (19) are stated descriptively in order to show how the data can be accounted for. Grimshaw and Mester motivate these constraints theoretically as follows. The constraint that the subject argument must always be outside of the NP follows from the fact that nouns in general can never realize an overt subject in the syntax. This is because the external argument of a noun is lexically suppressed. Furthermore, if Argument Transfer is viewed as only having taken place when at least one unsuppressed argument position has been transferred, the constraint in (19i) follows immediately. The last constraint in (19iii) falls out from the hierarchical organization of argument structure: Argument Transfer respects argument structure hierarchy and can only apply in an outside-in fashion. Thus, if a Theme has been transferred, the Goal must have been transferred as well in order to avoid a violation of the argument hierarchy.

7.3.2.3 Sells 1989

Sells (1989) takes up the notion of Argument Transfer as it pertains to Japanese *suru* 'do' and argues that the evidence Grimshaw and Mester use to justify the notion of Argument Transfer is actually not valid once the basic properties of the θ-nouns are considered in

their own right. He does go on to provide a more detailed examination of the data which would seem to motivate an Argument Transfer account after all, but also argues that the process of Transfer must be conceived of as pertaining to more than just theta-roles. In fact, under Sells's view, subcategorization information and Case-assigning properties must be transferred as well as the actual thematic roles. I will not detail Sells's proposal here, but will instead focus on the arguments which contribute to a better understanding of the nature of light *suru* 'do' and complex predicate formation in Japanese.[10]

Sells notes that the θ-nouns which can appear with light *suru* 'do' are predominantly of the Sino-Japanese class. In the literature, these nouns are also often referred to as verbal nouns because of some properties which set them apart from other Japanese nominals. For example, these θ-, or verbal, nouns may take verbal arguments (nominative, accusative, dative) in constructions other than with light *suru* 'do' as well. In fact, Iida (1987) identifies a *temporal affix construction* in which either an external temporal noun such as *ori* 'occasion', or a temporal affix licenses the θ-noun to assign verbal case to its arguments. An example is shown in (20). Here the arguments of *keikoku* 'warning' can either appear with the genitive (nominal) case in (20a), or can be marked with nominative and dative case, as in (20b).

(20) a. John-no Mary-e-no **keikoku**-no **ori**-ni Mary-wa kikanakatta
John-Gen Mary-Dat-Gen warning-Gen occasion-at Mary-Top listen.Past
'On the occasion of John's warning her, Mary did not listen.'

b. John-ga Mary-ni **keikoku**-no **ori**-ni Mary-wa kikanakatta
John-Nom Mary-Dat warning-Gen occasion-at Mary-Top listen.Past
'On the occasion of John's warning her, Mary did not listen.'

When a θ-noun is able to assign verbal case to its arguments, it also cannot be modified by an adjective: it must rather be modified by an adverb. Thus, in (20a), the noun *keikoku* 'warning' can only be modified by an adjective, while in (20b), it must be modified with an adverb. This pattern might suggest that *keikoku* 'warning' in (20a) is truly a noun, while in (20b) it is a verb. However, Sells presents some detailed evidence that *keikoku* 'warning' can never be a verb – it must always be analyzed as a noun.

[10]Matsumoto (1992) argues against a complex predicate analysis of light *suru* 'do'. He instead proposes that *suru* 'do' should be considered a simple control verb, which subcategorizes for an complement. This complement can either be manifested as an XCOMP, if there is a controlled embedded subject, or as an OBJ, if there is no embedded subject. However, the argumentation presented in favor of this view depends on a different factorization of levels from the one assumed here.

Furthermore, in light of the Double-o constraint discussed by Poser (1989), it would appear that there is no motivation for a Transfer account of light *suru* 'do'. It has long been noted that there is a restriction on having two adjacent NPs which bear the accusative case *-o*. Poser (1989) shows that the constraint is not merely a constraint on the surface appearance of *-o*, as might be supposed, but that it is also a "deep" constraint in the sense that it is a prohibition on two objects in the same clause. In Grimshaw and Mester's terms, this constraint requires that a verb cannot assign accusative Case twice, where the overt morphological realization of abstract accusative Case on an object can be either accusative *-o*, nominative *-ga*, or topic *-wa*. Within LFG, the constraint is articulated as a prohibition on two direct objects within the domain of a single PRED. Because the process of Argument Transfer in effect requires that *suru* in (21) assigns accusative Case both to the θ-noun *keikoku* 'warning' and one of the arguments of the θ-noun, in this case the complement *ookami-ga kuru-to* 'that wolf come', Sells argues that the Argument Transfer account is actually in violation of the Double-o constraint.[11]

(21) John-wa murabito-ni [ookami-ga kuru-to] **keikoku-o** shita
John-Top villagers-Dat wolf-Nom come-Comp warning-Acc do.Past
'John warned the villagers that the wolf was coming.'

Sells further argues that the only way the Double-o constraint can be satisfied for constructions like (21), is to allow the θ-noun to assign accusative Case to its object. Once it is granted that the θ-noun can, and indeed must, assign verbal case to its arguments, the original motivation cited by Grimshaw and Mester for an Argument Transfer analysis vanishes. However, Sells then goes on to provide evidence for the existence of Argument Transfer in constructions with light *suru* 'do'. First he notes, citing Matsumoto (1988), that when both the θ-noun and the Theme argument of the θ-noun are assigned accusative Case, special properties hold of the construction. These are listed in (22).

(22)
- Modification of the head N of the θ-NP is not possible.
- Scrambling of the θ-NP is not possible.

[11]Sells suggests there is evidence to show that although the complement *ookami-ga kuru-to* 'that wolf come' is not overtly case-marked, it does receive accusative Case. That is, it is functioning as a direct object. Matusomoto (1992) argues that the *keikoku-o suru* 'warning do' construction, which Sells and Grimshaw and Mester base much of their argumentation on, is actually an instance of heavy *suru* 'do'. He analyzes heavy *suru* as a ditransitive verb, which always requires a beneficiary argument. However, there appear to be no instances of simple sentences in which *suru* gives rise to a beneficiary construction. The fact that it can only do so in combination with a θ-noun would appear to argue against the existence of a heavy ditransitive *suru*.

- Passivization of the θ-NP is not possible.
- Anaphoric replacement of the θ-NP is not possible.

In particular, the evidence from passivization is crucial for the motivation of some kind of Transfer account. This is illustrated by (23)–(25). In (23), the Theme argument of the θ-noun, *100,000 yen*, is the subject of passivization. Example (24) further shows that a θ-noun, *keikoku* 'warning' in this case, cannot be the subject of passivization if its Theme, the complement *ookami-ga kuru-to* 'that wolf come', is accusative. However, if the Theme argument of the θ-noun is not accusative, but genitive as in (25), the θ-NP *kihu* 'donation' can be the subject of passivization.

(23) 100,000 en-ga sono toshokan-ni Taroo-kara **kihu**-o sareta
100,000 yen-Nom that library-Dat Taroo-from donation-Acc do.Pass.Past
'100,000 yen was given to that library by (from) Taroo'

(24) ?? **keikoku**-ga Taroo-niyotte [ookami-ga kuru-to] murabito-ni sareta
warning-Nom Taroo-by wolf-Nom come-Comp villagers-Dat do.Pass.Past
'The warning that the wolf was coming was done by Taroo to the villagers.'

(25) [100,000 en-no **kihu**-ga] sono toshokan-ni Taroo-kara sareta
100,000 yen-Gen donation-Nom that library-Dat Taroo-from do.Pass.Past
'100,00 yen was given to that library by Taroo.'

The fact that the Theme argument of the θ-noun can become the subject of the sentence under passivization in (23) indicates very strongly that it is functioning as a direct object of the active sentence and that therefore some type of Argument Transfer or complex predicate formation has taken place.

Sells accounts for these additional facts and the original data presented by Grimshaw and Mester by including both subcategorization and thematic role information in the process of Argument Transfer, and by allowing light *suru* 'do' to optionally assign the feature [+asp] under government. When light *suru* 'do' assigns [+asp] to a θ-noun under government, that θ-noun can assign verbal case to its arguments, but must also be governed by *suru* 'do', so that it cannot be subject to modification, scrambling, passivization, or anaphoric replacement. In addition, subcategorization information is taken to be crucial. Only when a θ-noun subcategorizes for an argument, can it assign verbal case. And when it does subcategorize for an argument, that argument is never transferred to the argument structure of *suru* 'do'. Sells is thus able to account both for examples like (23), where the Theme

argument has been transferred (and hence can be subject to passivization), and for the sentences in (26), where he takes the Theme argument of the θ-noun not to have been transferred, but to be Case marked by the θ-noun.

(26) Taroo-ga Hanako-no ronbun-o kibishiku **hihan**-o shimashita
Taroo-Nom Hanako-Gen paper-Acc severely criticism-Acc do.Past
'Taroo severely criticized Hanako's paper.'

When the θ-noun does not subcategorize for an argument, and that argument is not transferred to the argument structure of light *suru* 'do', the argument of the θ-noun bears genitive case. In examples like (27), therefore, both the agent and the goal of the θ-noun have been transferred. As the theme *ookami-ga kuru-to* 'that wolf come' is not subcategorized for by the θ-noun and also has not been transferred, it must bear the nominal genitive case *-no*.

(27) John-wa murabito-ni [[ookami-ga kuru-to]-no **keikoku**]-o shita
John-Top villagers-Dat wolf-Nom come-Comp-Gen warning-Acc do.Past
'John warned the villagers that the wolf was coming.'

However, it is not quite clear to me what motivates the subcategorization distinctions between nouns. It is also not clear how Grimshaw and Mester's generalization that at least one argument besides the subject must always be transferred is accounted for under the system proposed by Sells. Furthermore, note that although Sells makes a distinction between (24) and (26) in terms of the type of Transfer that has occurred (in (24) the Theme is subcategorized for, but in (26) it is not), the properties listed in (22) hold for both of these constructions. It would therefore appear that the two examples actually do not represent two differing types of Transfer, as Sells proposes, but that these examples should be given a unifying analysis.

7.3.3 Elaborated Argument Structure

7.3.3.1 Basic Approach

Grimshaw and Mester argue that light *suru* 'do' must have an empty argument structure, and, in particular, cannot have an 'external' argument because its subject must never satisfy any selectional restrictions. This, they claim, is quite unlike heavy *suru* 'do', which requires that its subject be agentive. However, the fact that Grimshaw and Mester show that at least the subject must always be transferred, and the fact that light *suru* 'do' is able to

assign accusative Case is not quite in tune with an analysis that requires an empty argument structure.

Indeed, the interaction of the Japanese causative with the empty argument structure of light *suru* 'do', cited as supporting evidence for Argument Transfer, is actually problematic for the Argument Transfer analysis. Notice that in the causative construction in (28) all the arguments bear verbal case.

(28) Mary-ga John-ni Bill-to **aiseki-o** saseta
Mary-Nom John-Dat Bill-with table-sharing-Acc do.Cause.Past
'Mary made John share a table with Bill.'

According to Grimshaw and Mester, Argument Transfer does not require a completely empty argument structure, only an incomplete one. Causativization is analyzed as adding an argument to light *suru* 'do'. The resulting incomplete argument structure of causativized light *suru* 'do' is shown in (29). To this argument structure, Argument Transfer is taken to apply in the usual way.

(29) *saseru* (Agent,)<acc>

Consider, however, what the effects of Argument Transfer actually are. The derivation below shows that if Argument Transfer indeed applies in the usual way, there are suddenly two Agents to be accounted for in the combined structures. Is one demoted? And if so, how? And what accounts for the dative case on the demoted agent? While these are problems which arise with the causativization of any transitive verb, note that the special problem here is that causativization of the verb *suru* 'do' has already taken place. The questions of which agent is to be demoted thus cannot be accounted for by a general theory of causativization. Rather, something extra must be said about what happens in the case of Argument Transfer.

aiseiki 'table-sharing' (Agent, Comitative)
saseru 'cause-do' (Agent,)<acc>
aiseki ( ) + *suru* (Agent, Agent, Comitative)<acc>

I argue that an analysis which provides light *suru* 'do' with a non-empty, albeit bleached, argument structure, and which views complex predicate formation in terms of Argument Fusion provides a successful account not only of the patterns described in Grimshaw and

Mester and Sells, but can also provide an account of the Japanese causative in (28) which is in tune with the theory of causatives formulated by Alsina (1993).

Grimshaw (1990) explicitly argues for a representation of argument structure which contains only slots marked by "x" or "y", where only hierarchical, but not semantic information is encoded. Within this framework, the argument that light *suru* 'do' places no selectional restrictions on its subject would not seem to be enough to motivate an entirely empty argument structure. Furthermore, Isoda (1991) and Matsumoto (1992) both argue that light *suru* 'do' must have at least an agentive argument. Some evidence for this comes from the fact that light *suru* 'do' only combines with unergative, but not unaccusative, θ-nouns.

Indeed, Isoda (1991) presents an analysis of light *suru* 'do', which is quite close to the Argument Fusion presented for Urdu. He formulates an argument structure account within LFG in which the agent argument of *suru* 'do' and the agent argument of the θ-noun are fused. The two distinct argument structures are thus merged and then placed in correspondence with a monoclausal f-structure. As this proposal is exactly the approach I take, it should come as no surprise that I simply adopt Isoda's fundamental idea here. However, Isoda cannot adequately explain the appearance of "mixed" case marking in the examples Grimshaw and Mester consider crucial for the motivation of Partial Transfer. He also does not provide an explanation for the appearance of accusative case on the θ-noun. In what follows, I attempt to motivate these facts within an elaborated argument structure account.

7.3.3.2 Accounting for the Facts

The elaborated argument structure for light *suru* 'do' I propose is shown in (30). I take light *suru* 'do' to differ only minimally from heavy *suru* 'do'. One difference, of course, is that light *suru* 'do' has a transparent Event argument, while heavy *suru* 'do' takes a 'real' Event. It has furthermore been noted that heavy *suru* 'do' somehow imposes stricter selectional restrictions on agents than light *suru* 'do' (Isoda 1991). I propose to express this in terms of the [+cc] (conscious choice) feature on the Action Tier: heavy *suru* 'do' selects for [+cc] agents, while light *suru* 'do' just requires an agentive argument.

(30)

$$\left[\begin{array}{l} \text{suru 'do'} \\ \left[\begin{array}{l} \text{DO}([\alpha], \{\quad\}^{\beta}_{E_T}) \\ \text{AFF}([\ ]^{\alpha}, [\ ]^{\beta}) \\ \text{ASP}(\_\ \_\ \_) \end{array}\right]_E \end{array}\right]$$

As also proposed by Isoda (1991), light *suru* 'do' thus takes two arguments: a doer and a done Event. Note that I posit a second argument at the Action Tier which is coindexed with the transparent Event argument. The fact that there is a second argument of AFF, the AFF (, $[\ ]^{\beta}$), ensures that the θ-noun will be marked with accusative case once linking is performed. This is quite unlike the Urdu permissive, for example, where the transparent Event was not represented on the Action Tier, but only on the Thematic Tier. Correspondingly, the Urdu permissive infinitive can never bear case, while the θ-noun in a *suru* 'do' construction can.

As in Isoda (1991), the combination of light *suru* 'do' with a θ-noun is taken to be just another instance of complex predicate formation through Argument Fusion. Only slight additions to my theory are needed in order to account for all the facts. Consider the entries for the two θ-nouns *aiseki* 'table-sharing' and *keikoku* 'warning' in (31) and (32). Recall that these nouns differ from other nouns in that they have an argument structure (in Grimshaw's (1990) terms, only process nominals have an argument structure). I represent this by positing an Action Tier for process nominals. Nouns which do not have an argument structure will not contain an Action Tier, only a Thematic Tier, in their lexical entry. This corresponds to the idea expressed by Grimshaw and Mester that process nominals are like transitive verbs in some sense, and to the idea in Sells that these nominals have the ability to subcategorize for arguments. In addition, the fact that these nouns are Events rather than Things also identifies them as process nominals.

(31)

$$\left[\begin{array}{l} \text{aiseki 'table-sharing'} \\ \left[\begin{array}{l} \text{BE}([\alpha], AT[TABLE], WITH[\quad]^{\beta}) \\ \text{AFF}(\quad, [\ ]^{\alpha}) \end{array}\right]_E \end{array}\right]$$

(32)

$$\left[\begin{array}{l} \text{keikoku 'warning'} \\ \left[\begin{array}{l} \text{WARN}([\alpha], TO[\beta], [\quad]^{\gamma}_{E}) \\ \text{AFF}^{+}([\ \ ]^{\alpha}, [\quad]^{\beta}) \end{array}\right]_{E} \end{array}\right]$$

Notice that although these θ-nouns are Events, I have not provided them with an Aspectual Tier. I propose that in the combination of a θ-noun with light *suru* 'do', the θ-noun gains access to the Aspectual Tier of *suru* 'do'. This is roughly equivalent to Sells's idea that *suru* 'do' assigns the feature [+asp] to a θ-noun. Although the interaction between aspect and θ-nouns is thus formalized, I do not at the present time make use of it. In order to account for the larger picture of Japanese and complex predicates, it nevertheless is desirable to be able to account for the interaction.

Like the other light verbs presented in this dissertation, the transparent Event argument of light *suru* 'do' requires that it must combine with an argument taking entity, such as a θ-noun. Exactly what governs whether some light verbs combine with other verbs, with nouns, or with adjectives, remains a topic for investigation, but I suspect that the light verb simply retains the original subcategorization information of the full, or heavy, version of the verb.

The structure in (33) represents the combination of the θ-noun *aiseki* 'table-sharing' with light *suru* 'do'. It is a complex predicate: the transparent Event argument of *suru* 'do' has triggered Argument Fusion and created a single argument structure, which must be linked to the syntax. While the structure in (33) looks very much like the complex structures for the Urdu complex predicates, there is a slight difference in how Argument Fusion is applied in (33). Recall that Argument Fusion was defined as operating on the lowest matrix argument and the highest embedded argument.

(33)

$$\left[\begin{array}{l} \text{aiseki suru 'table-sharing do'} \\ \left[\begin{array}{l} \text{DO}([\alpha], \left\{\begin{array}{l} \text{BE}([\gamma], AT[TABLE], WITH[\quad]^{\delta}) \\ \text{AFF}(\quad, [\alpha]^{\gamma}) \end{array}\right\}^{\beta}_{E_T}) \\ \text{AFF}([\ \ ]^{\alpha}, [\ \ ]^{\beta}) \\ \text{ASP}(\_\ \_\ \_) \end{array}\right]_{E} \end{array}\right]$$

In (33) ostensibly the lowest matrix argument is the β marked argument on the Action Tier. However, this is not the argument that has fused with the highest embedded argument. Rather, the matrix argument marked α has fused with the highest embedded, γ marked, argument. The selection of the matrix α rather than the β argument for Argument Fusion is a direct result of the fact that the β argument is already coindexed with the transparent Event argument and is therefore not available for Argument Fusion. The principle of Argument Fusion must therefore be revised and stated as in (34).

(34) **Argument Fusion:** The highest embedded argument is fused with the lowest available matrix argument.

Thus, the arguments to be linked in (33) are: the doer AFF ([ ]$^{\alpha}$,), the done thing AFF (, [ ]$^{\beta}$), and the WITH [ ]$^{\delta}$. According to Grimshaw and Mester, (35a) is the only well-formed result of linking and case marking (the data is presented as reported by them). Under an elaborated argument structure approach, (35a) is arrived at quite straightforwardly from (33). *Bill-to* is linked to the argument of WITH, the θ-noun to the second argument of AFF because ths argument is coindexed with the transparent Event argument, and the remaining NP *John-wa* is linked to the remaining argument slot: the first argument of AFF. The example in (35c) is ruled out because the effect of complex predicate formation does not allow any of the NPs to be marked with genitive case. This contrasts with the example in (35b), which I would consider to be an instance of heavy *suru* 'do'. In this case, there is no transparent Event at argument structure, so complex predicate formation is not triggered and genitive case on the argument of the noun *aiseki* 'table-sharing' is acceptable. Grimshaw and Mester's analysis, on the other hand, predicts that (35b) should be ruled out completely. I return to this point further on.

(35) a. John-wa Bill-to **aiseki-o** shita
John-Top Bill-with table-sharing-Acc do.Past
'John shared a table with Bill.'

b. ?John-wa [Bill-to-no **aiseki**]-o shita
John-Top Bill-with-Gen table-sharing-Acc do.Past
'John shared a table with Bill.'

c. *John-no Bill-to-no **aiseki**-o shita
John-Gen Bill-with-Gen table-sharing-Acc do.Past
'John shared a table with Bill.'

The application of LMT, which links a skeletal f-structure to the a-structure in (33) is shown in (36). In addition, I have listed the case marking possibilities for each of the, as yet unfilled, slots in the skeletal f-structure. Subjects may be marked either with nominative or topic. Similarly, direct objects may be either accusative or topic. The OBJ_θ, on the other hand, can only be marked with the semantic case marker *to*. This OBJ_θ, which is already linked to the argument of WITH, must thus be linked with the c-structure NP *Bill-to* 'with Bill' in (35a). Since the NP *aiseki-o* 'table-sharing' is the only one with accusative case, it must be identified as the direct object. That leaves the topicalized *John-wa*, which is identified with the subject. Notice that in (36) the argument corresponding to the θ-noun has least priority in terms of linking: it comes last in the argument list. This is because its status as a transparent Event argument overrides the principles of linearity and depth of embedding which ordinarily determine the linking hierarchy.

(36)

	AFF ([ ]$^\alpha$,)	WITH [ ]$^\delta$	AFF (, [ ]$^\beta$)
	(ag)	(com)	(Ev_T/th)
	|	|	|
intrinsic	[-o]	[+o]	[-r]
default	[-r]	[+r]	
GF	SUBJ	OBJ_θ	OBJ
Case	nom/top	with	acc/top

The end result of linking is shown in (37) at argument structure.

(37)

$$\left[\begin{array}{l}\text{aiseiki suru 'table-sharing do'} \\ \left[\begin{array}{l}\text{DO}([\alpha], \left\{\begin{array}{l}\text{BE}([\gamma], AT[TABLE], WITH[Bill]^\delta) \\ \text{AFF}(\ \ , [\alpha]^\gamma)\end{array}\right\}^\beta_{E_T}) \\ \text{AFF}([John]^\alpha, [table-sharing]^\beta) \\ \text{ASP}(\_\ \_\ \_)\end{array}\right]_E\end{array}\right]$$

Under the Argument Fusion approach to light *suru*, the sentence in (35a) is thus the only possible outcome of complex predicate formation. Complex predicate formation always entails Complete Transfer in Grimshaw and Mester's terms and therefore the appearance of genitive case on arguments is precluded. Thus, neither (35b) and (35c) are predicted to be possible constructions with light *suru* 'do'. On the other hand, (35b) should be a perfectly acceptable construction with heavy *suru* 'do'. Recall that I analyze heavy *suru* 'do' as a

transitive main verb, which requires an agent and a thing/event as arguments. As this is precisely what (35b) consists of, this sentence should be good. The somewhat questionable nature of (35b) might be explained by the strict conscious choice requirement of heavy *suru*: sharing a table might not constitute enough of a conscious act to be compatible with heavy *suru* 'do'.

Having thus dealt successfully with the *aiseki-o suru* 'table-sharing do' construction, it is now time to move on to the occurence of mixed case-marking displayed by a combination of ditransitive θ-nouns with *suru* 'do'. The combination of *keikoku* 'warning', a nominal with a ditransitive argument structure, and light *suru* 'do' is shown in (38).

(38)

$$\left[\begin{array}{l}\text{keikoku suru 'warning do'}\\ \left[\begin{array}{l}\text{DO}([\alpha], \left\{\begin{array}{l}\text{WARN}([\gamma], TO[\delta], [\quad]^{\epsilon}_{E})\\ \text{AFF}^{+}([\alpha]^{\gamma}, [\quad]^{\delta})\end{array}\right\}^{\beta}_{E_T})\\ \text{AFF}([\ ]^{\alpha}, [\ ]^{\beta})\\ \text{ASP}(\_\ \_\ \_)\end{array}\right]_{E}\end{array}\right]$$

At first glance, it would appear that the results of LMT in combination with the case-marking possibilities shown in (39) also allow a straightforward mapping of argument structure positions to the c-structure constituents in (40). The *villagers* must be linked to the OBJ_{go} ($\text{AFF}^{+}(\ \ ,[\quad]^{\delta})$), the complement phrase to direct object (the Event argument of WARN), and the θ-noun with the transparent Event argument. Which again leaves the topicalized *John-wa* to be linked to the subject at f-structure. The ultimate result is shown in (41) for the level of argument structure. Notice, however, that the grammatical function of the θ-noun *keikoku* 'warning' cannot be determined clearly in (39). As the complement *ookami-ga kuru-to* 'that wolf come' is linked to the direct object slot at f-structure, this grammatical function is then not available to the θ-noun.

(39)

	AFF ([ ]$^{\alpha}$,)	AFF^{+} (, [ ]$^{\delta}$)	WARN (, , [ ]$^{\epsilon}{}_{E}$)	AFF (, [ ]$^{\beta}$))
	(ag)	(ben/go)	(Ev/th)	(Ev$_T$/th)
	|	|	|	|
intrinsic	[-o]	[+o]	[-r]	[-r]
default	[-r]	[+r]		
GF	SUBJ	OBJ_{go}	OBJ	??
Case	nom/top	dat	acc	acc/top

(40) [John-wa] [murabito-ni] [ookami-ga kuru-to] [**keikoku-o**] shita
John-Top villagers-Dat wolf-Nom come-Comp warning-Acc do.Past
'John warned the villagers that the wolf was coming.'

(41)

$$\left[\begin{array}{l} \text{keikoku suru 'warning do'} \\ \left[\begin{array}{l} \text{DO}([\alpha], \left\{\begin{array}{l} \text{WARN}([\gamma], TO[\delta], [wolf\ coming]^{\epsilon}_{E}) \\ \text{AFF}^{+}([\alpha]^{\gamma}, [villagers]^{\delta}) \end{array}\right\}^{\beta}_{E_T}) \\ \text{AFF}([John]^{\alpha}, [warning]^{\beta}) \\ \text{ASP}(\_\ \_\ \_) \end{array}\right]_{E} \end{array}\right]$$

One could argue that perhaps the way I have determined the order in which LMT is applied to the arguments in (39) is wrong. However, Sells (1989) clearly shows that the complement *ookami-ga kuru-to* 'that wolf come' must indeed have the status of a direct object because it can be passivized.

Note that the undetermined status of the θ-noun *keikoku* 'warning' under the Argument Fusion approach is exactly in accordance with the data. All of the strange properties of light *suru* 'do', which prompted Grimshaw and Mester towards an Argument Transfer analysis, are found exactly in the situation where a ditransitive θ-noun combines with *suru* 'do': exactly in those situations where an argument other than the θ-noun is eligible to function as the direct object of the sentence. In fact, when one considers the generalizations about mixed case established by Grimshaw and Mester, Sells, and Isoda carefully, it turns out that it is always only the Theme argument of the θ-noun which behaves differently from the other arguments. Consider again the sentences in (42).

(42) a. John-wa murabito-ni [[ookami-ga kuru-to]-no **keikoku**]-o shita
John-Top villagers-Dat wolf-Nom come-Comp-Gen warning-Acc do.Past
'John warned the villagers that the wolf was coming.'

b. ?*John-wa [ookami-ga kuro-to] [murabito-e-no **keikoku**]-o shita
John-Top wolf-Nom come-Comp villagers-Dat-Gen warning-Acc do.Past
'John warned the villagers that the wolf was coming.'

c. *John-wa [murabito-e-no [ookami-ga kuru-to]-no **keikoku]-o**
John-Top villagers-Dat-Gen wolf-Nom come-Comp-Gen warning-Acc
shita
do.Past
'John warned the villagers that the wolf was coming.'

It is only ever the Theme argument, and never the Goal or Agent argument, which is allowed to appear with genitive case. Therefore there seems to be something exceptional about Themes. In light of the Double-o constraint discussed by Poser (1989) and by Sells (1989) with respect to light *suru* 'do', it is actually not surprising that the appearance of a Theme argument should cause trouble. Only when another Theme argument besides the θ-noun is present at argument structure does the possibility of violating the Double-o constraint arise.

I propose that Japanese allows two differing strategies of avoiding a violation of the Double-o constraint. One strategy is to have the Theme argument of the θ-noun function as the only direct object of the sentence. The θ-noun is then not assigned any grammatical function, but must form some sort of a constituent with the verb. This accounts for the properties noted by Matsumoto (1988) that the θ-noun cannot be modified, scrambled, passivized, or anaphorically replaced whenever its Theme argument is functioning as a direct object. The θ-noun must still be marked with the accusative *o* because it is linked to the second argument slot at the Action Tier. Since there are is a first argument slot, an agent, at the Action Tier as well, the θ-noun here is technically not a Theme, but a Patient, and must be marked with accusative case.

The alternative strategy is to identify the θ-noun as the direct object after all through the application of LMT.[12] This then leaves the Theme argument of the θ-noun in limbo – it is not assigned a grammatical function and can only receive 'nominal' case: the genitive.

All of the data presented above now follows. In particular, both (40) and (42a) are good because they represent the two differing strategies of avoiding a violation of the Double-o constraint. In (40) the complement *ookami-ga kuru-to* 'that wolf come' is the direct object, while *keikoku* 'warning' cannot be scrambled or passivized. In (42a), *keikoku* 'warning' is the direct object of the sentence and the complement must appear in the genitive. In this case, *keikoku* can be the subject of passivization, and it can be scrambled. Examples (42b) and (42c) are bad because there is no justification for the genitive case on the goal argument *murabito* 'villagers'.

Although I do not show it here, the causative data presented earlier can now also be explained straightforwardly through a combination of Argument Fusion and the theory of causatives formulated by Alsina (1993). The argument structure of the Japanese causative

[12]In this case, the θ-noun would be linked before the Theme argument. For a justification of such an alternation in Linking, see Bresnan and Moshi (1989) on Bantu languages.

contains a causer, a causee, and a caused event. When it combines with a *suru* 'do' complex predicate, the causeee fuses with the agent of *suru* 'do' and must be marked in the dative.

Finally, sentences like the ones in (43) remain to be accounted for (the example in (43a) was discussed earlier). Here there are two accusative marked NPs in one sentence. Recall that Sells (1989) used sentences like (43a) to argue that the theme argument *Hanako-no ronbun-o* 'Hanako's paper' could not have been transferred, but was being assigned accusative Case directly by the θ-noun.

(43) a. Taroo-ga [Hanako-no ronbun]-o kibishiku **hihan**-o shimashita
Taroo-Nom Hanako-Gen paper-Acc severely criticism-Acc do.Past
'Taroo severely criticized Hanako's paper.'

b. *Taroo-wa **hihan**-o [Hanako-no ronbun]-o kibishiku shimashita
Taroo-Top criticism-Acc Hanako-Gen paper-Acc severely do.Past
'Taroo severely criticized Hanako's paper.'

As (43b) shows, the θ-noun in this construction cannot be scrambled or passivized (Sells 1989). The θ-noun *hihan* 'criticism' in (43) is thus not functioning as a direct object. This sentence is thus exactly like the *keikoku* 'warning' construction in (40), except for the fact that the direct object *Hanako-no ronbun-o* 'Hanako's paper' here carries an overt accusative case marker. The well-formedness of (43a) is expected under my analysis, but the sentence should actually be ruled out under consideration of the Double-o constraint. However, I would argue that neither the deep, nor the surface, Double-o constraint apply to (43). The Double-o constraint in its deep manifestation is formulated as a prohibition against two direct objects within the domain of a single predicate (PRED at f-structure). Under my analysis only *Hanako-no ronbun-o* 'Hanako's paper' may function as the direct object, so the Double-o constraint is ultimately not violated. The surface Double-o constraint does not apply in this particular example either because the two *-o* marked NPs are not immediately adjacent to one another. Thus, a complex predicate analysis of light *suru* 'do' in terms of an elaborated a-structure is able to correctly account for (43), as well as the other data presented in this section.

7.3.4 Summary

Under Grimshaw and Mester's Argument Transfer account of light *suru* 'do', the assignment of accusative case by *suru* was not motivated, neither could a treatment of the interaction of causativization and light *suru* 'do' be ultimately successful within Argument Transfer. In

Sells's account, it was not clear what determined the subcategorization information that was crucial to Transfer, and it was not clear why certain combinations were not possible. For example, why could both the goal and the theme argument of the θ-noun not appear with genitive case? The above treatment of light *suru* 'do' within a theory of complex predicate formation that relies on an elaborated argument structure and a process of Argument Fusion, on the other hand, allows an analysis which accounts for the data in a quite natural way. The ability of *suru* to assign accusative case is motivated by the fact that the θ-noun is an argument of *suru*. The fact that only themes have the option of being marked with the genitive is accounted for by the interaction of the linking process and the Double-o constraint in Japanese. Furthermore, the theory of complex predicate formation presented here has the advantage of not confining the composition of argument structures to the lexicon, as was necessary in Grimshaw and Mester's approach. Given that Poser (1991) shows that the "incorporated" *suru* 'do' construction cannot be analyzed as an instance of lexical incorporation, it seems likely that light *suru* 'do' can similarly not be treated as combining with its θ-noun in the lexicon. Rather, as with the Urdu complex predicates, complex predicate formation with Japanese *suru* 'do' must take place in the syntax.

7.4 Serial Verbs

7.4.1 Looking beyond the Label

Zwicky (1990) notes that there are many linguistic terms for which a clearly established *meaning* cannot be identified. He cites the term *clitic* as one such example. It is generally realized that there are several different types of clitics, but nobody seems to be quite sure what essentially defines a clitic. Rather, the term *clitic* is loosely used as a cover-term to describe various types of phenomena that appear to be related. Zwicky (1990) considers the term *serial verb* as underdetermined as the term *clitic* and proceeds to discuss the various possible constructions that might be deemed serial verbs.

While nothing speaks against the usage of general cover terms in linguistics, the lack of a clear notion of what characterizes a serial verb, as opposed to a complex predicate, or a compound verb is frustrating for any endeavour which seeks to formulate a theory of complex predicates, or serial verbs, or compound verbs. A quick comparison of two typical serial verbs in (44) with one of the Urdu Aspectual complex predicates in (45) would make

a concrete distinction seem desirable.[13] A theory of complex predicates, as formulated here, would then not be immediately responsible for providing an account of the constructions in (44). Conversely, a theory of serial verbs would then not need to puzzle over constructions like (45). As (44) illustrates, serial verbs can stack several events in a single clause. This is not possible for complex predicates. The light verb *le* 'take' in (45) merely contributes aspectual information to the extracting event, while in (44a) there are two events: a pulling event and a removing event. I shall return to this point. Furthermore, as (44b) illustrates, each member of the serial verb may display agreement features. This has not been possible in the Urdu complex predicates examined.

(44) a. kofi hari a ston puru na ini a olo (Sranan)
Kofi pull the stone remove Loc in the hole
'Kofi pulled the stone out of the hole.'

b. iire rehe-sooni vakilii rehe-haa (Paamese)
1Pl.Incl 1Pl.Incl-Distant.Throw canoe 1Pl.Incl-Distant.Go
'We will go, putting (throwing) our canoe to sea.'

(45) anjum=ne pat^{h}ar=ko bahar nikaal li-yaa
Anjum.F=Erg stone.M=Acc out extract take-Perf.M.Sg
'Anjum got the stone out.'

On the other hand, there are also many similarities between complex predicates and serial verbs. For example, many of the lexical items used in serial verbs and complex predicates are the same: *come*, *go*, *take*, *give*, and *put*. The similarities between the two constructions have the effect that in the literature the same construction can be found variously labeled a complex predicate, a serial verb, or even a compound verb, sometimes even by the same author. In Pandharipande (1989), for example, a Marathi construction almost identical to the Urdu Aspectual complex predicates, both in form and behaviour, is referred to as a compound verb. In Pandharipande (1990), on the other hand, this construction is described as a serial verb. I take the defining characteristic of a compound verb to be that it is formed in the lexicon, rather than being put together in the syntax. This particular construction, an example of which is shown in (46), can be shown not to form a compound verb. In fact, I would argue that it can be neither a compound verb, nor a serial verb, but must be a complex predicate.

[13]The Sranan example in (44a) is from Baker (1989), the Paamese example in (44b) from Durie (1993), who quotes Crowley (1987:47) as the source.

(46) sud^{h}aa kaam karuun baslii (Marathi)
Sudha work do sat
'Sudha did the work (inadvertently).'

The liberal labeling of constructions which form one kind of verbal complex or another does not constitute a problem in and of itself. However, when a construction which looks very much like a complex predicate is labeled as a serial verb, and then used to argue against a particular theory of serial verbs (for example Nagarajan (1990)), the situation becomes quite murky. I strongly believe that a precise characterization of both serial verbs and complex predicates is necessary for an ultimate formulation of a comprehensive theory of differing types of verbal complexes. The work presented in this dissertation is intended as a step towards a precise characterization of the properties of complex predicates, which will allow a comprehensive theory of complex predicates, which in turn will interact with a theory of serial verbs.

7.4.2 Characterization of Serial Verbs

The literature on serial verbs is vast. There are also many works which attempt a formal characterization of the properties of serial verbs. In particular, the generalizations stated in Sebba (1987) are often taken as a base for subsequent work. I again have neither the time nor the space to discuss all the relevant literature, but will instead focus on the ideas presented in Baker (1989) and, particularly, in Durie (1993).

Baker (1989) attempts to formulate a theory of serial verbs within a purely syntactic approach. He takes the defining property of serial verbs to be that they share not only subjects, but also objects. Baker's analysis is couched within GB, but essentially proposes a form of argument structure fusion through which two verbs can θ-mark the same direct object NP. Although the theory naturally allows for more complexity than the basic idea of object-sharing, Durie (1993) shows in considerable detail that Baker's analysis cannot successfully account for the fuller range of serial verb constructions found in the literature. In particular, Durie shows that the property of object-sharing cannot be taken to be a defining characteristic of serial verbs. Durie summarizes a range of generalizations about serial verbs that emerge in the literature through a comparison of serial verbs from various differing languages. The generalizations, taken from Durie (1993:2) are listed below.

(47) i. A single serial verb complex describes a single conceptual event: this is repeatedly reported to be a clear intuition of native speakers, and can be demonstrated

through semantic analysis. It follows from this that a serial verb complex can be best translated into a non-serializing language using a single, mono-verbal clause.

ii. The serial complex has shared tense, aspect, modality, and polarity: this is often reflected in a single morphological realization, or in obligatory concord across the verbs.

iii. Serial verbs 'share' at least one and possibly more arguments.

iv. One verb is not embedded within a complement of the other.

v. Intonational properties of a clause with serialization are those of a mono-verbal clause (Givón 1990/91).

vi. The complex takes only one subject/external argument.

vii. When serialization results in a complex of more than two arguments, the configuration of arguments corresponds closely to the kinds of configurations of argument and adjuncts found for single clauses in non-serializing languages.

viii. There is a very strong diachronic tendency to lexicalization and grammaticalization of the meaning of serial complexes: this can involve treating the whole serial complex as a single lexical(ized) item, or 'demotion' of the meaning and grammatical status of one of the verbs to that of a modifier or case-marker.

The properties of serial verbs, as listed above, in comparison with the properties of complex predicates reveal considerable overlap, but also highlight significant differences. The properties (47iii.), (47v.), (47vi.), and (47viii.) are common to both complex predicates and serial verbs: at least one argument is shared or fused, there is only a single subject, in fact, the complex behaves like simple sentence both with regard to intonation and grammatical function information. Furthermore, the complex predicates as well as serial verbs have a strong tendency towards lexicalization.

Complex predicates do share a single marking of tense and aspect, as described in (47ii.), however, they do not allow this information to appear on more than one of the items in the complex predicate. Similarly, while complex predicates always have the structure of a simple clause, as detailed in (47vii.), complex predicates are not capable of the instrumentality, resultativeness, or benefaction expressed by serial verbs in quite the same way. That is, arguments contributed by various verbs in the complex do not take on an adjunct status, as they do in serial verbs.

These subtle differences between serial verbs and complex predicates in my view are a direct consequence of the properties listed in (47i.) and (47iv.). As described in (47i.) for serial verbs, complex predicates are used to express a single conceptual event. However,

there is a difference in that complex predicates generally express a single event, such as a *permitting* event, while the several events described in a serial verb complex are stacked together to form a single, complex, conceptual event. This difference is more clearly expressed by the property in (47iv.). In serial verbs one verb can never be considered to function as the complement to another verb in the complex. My treatment of complex predicates, in contrast, takes a defining characteristic of light verbs to be that they contain a transparent Event at argument structure, and therefore can be said to subcategorize for an argument-taking predicate in some way. This does not seem to be the case for serial verbs. While it is true for serial verbs, as it is for complex predicates, that each verb in the complex may also be used as a full verb in a simple clause, the serializing versions of verbs cannot be analyzed as containing a transparent Event.

Durie (1993) in fact proposes that serial verbs are subject to semantic and pragmatic conditions which govern the formation of complex events. He shows that serial verbs can only be used to describe events which are "normal" events. For example, *buy take fish* is an expected sequence of events, while *sell take fish* is not. Correspondingly, it can be shown in Sranan, for example, that the former exists as a serial verb, while the latter does not.

The crucial difference between serial verbs and complex predicates, in my view, thus lies at the level of argument structure. Complex predicates are formed because of the presence of a transparent Event in the argument structure of a light verb: the light verb demands an argument structure composition. In serial verbs, on the other hand, one predicate is not subcategorized for by another predicate. Rather, the verbs can be gathered together into a complex under certain circumstances. The precise nature of these circumstances must be determined by a sophisticated theory of Complex Event Types, which is as yet lacking. The arguments of the predicates thus subsumed under a Complex Event must be mapped to a simple grammatical functional structure. The exact details of the mechanisms and constraints involved in this linking process must be determined through a careful analysis of the various possible serial verb constructions. However, generally two agents from seperate argument structures might be mapped to the same subject slot, while two theme arguments might be mapped to the same direct object slot, thus giving rise to the effect of object-sharing detailed by Baker (1989).

7.5 Conclusion

This chapter has shown that the treatment of complex predicates within an elaborated argument structure approach allows a unifying analysis of both Romance restructuring verbs and Japanese *suru* 'do' constructions. While other approaches, such as Rosen (1989) and Grimshaw and Mester (1988) have presented theories of complex predicate formation based on argument structure processes, these approaches have not been able to account for the range of properties discussed in this chapter. Furthermore, if complex predicate formation is taken to be triggered by the presence of a transparent Event in the argument structure of a light verb, as proposed here, a formal characterization of the crucial differences between serial verbs and complex predicates may be arrived at. No member of a serial verb complex is a light verb containing a transparent Event argument. Therefore, Argument Fusion triggered by complex predicate formation can never take place. Rather, the sharing of arguments observed in serial verb constructions must be the result of mapping separate, non-interacting, argument structures within a Complex Event onto a single grammatical functional structure.

Chapter 8

Conclusion

The main thread of argumentation presented in this dissertation is essentially quite a simple one. I examined two differing types of complex predicates in Urdu, and have proposed a theory of complex predicate formation which allows a unifying account of their diverse properties. In particular, evidence from the permissive showed that the status of a complex predicate must not necessarily be encoded structurally. A comparison of the permissive with the superficially very similar instructive in Chapter 3 showed that the permissive must be analyzed as a complex predicate. The two predicates of the permissive combine to function as a single unit, with a single matrix subject. The instructive, on the other hand, consists of a finite matrix verb and an embedded complement. There are correspondingly two subject positions: the subject of the matrix clause, and the subject of the embedded infinitival clause.

Despite these differences between the permissive and the instructive, I also showed that the two constructions in fact behave identically in terms of constituent structure. Evidence from coordination, negation, and scrambling showed that each of the two construction allowed two differing phrase structure realizations: one in which the two predicates form a unit, and one in which the infinitival predicate and its object argument form a unit. I therefore proposed that the essential defining characteristic of a complex predicate cannot easily be expressed at the level of phrase structure, but lies in the fact that a complex predicate has two or more semantic heads which correspond to a single PRED at the level of f-structure. I articulated the following definition of complex predicates.

(1) **Definition of a Complex Predicate:**

• The argument structure is complex (two or more semantic heads contribute arguments).
• The grammatical functional structure (f-structure) is that of a simple predicate. It is *flat*: there is only a single predicate (a nuclear PRED) and a single subject.
• The phrase structure (c-structure) may be either simple or complex. It does not necessarily determine the status of a complex predicate.

In Chapter 4, I examined Urdu Aspectual complex predicates, whose properties confirm the definition of complex predicates proposed in (1), and furthermore provide evidence for an elaborated level of argument structure. An Aspectual complex predicate was seen to be well-formed only if constraints on semantic properties such as volitionality and inception/completion were met: a main verb negatively specified for one of these domains cannot combine with a light verb positively specified for the same domain.

I therefore proposed an elaborated level of argument structure based on Jackendoff's (1990) theory of Conceptual Semantics, in which the systematic representation of fine-grained semantic information is possible. It should be noted, that I only represent such semantic information as is directly relevant to the overt syntactic realization of a clause, i.e., information relevant for case marking, complementation relations, complex predicate formation, etc. A theory of argument structure based on traditional θ-role labels, or indeed, based on the notion that argument structure representation must be as abstract as possible (Grimshaw 1990, S. Rosen 1989, Ritter and S. Rosen 1993, Alsina 1993) does not allow a similar systematic formulation of constraints on complex predicate formation.

While Aspectual complex predicates have been well studied by South Asian linguists (again, see Hook (1974) for a thorough survey), no comprehensive theory for the constraints on complex predicate formation has so far been formulated. The elaborated level of argument structure I propose accounts for a large number of possible and impossible combinations. I believe it will furthermore be the key factor towards a successful formulation of a comprehensive in-depth, systematic analysis of Aspectual complex predicates.

In addition, the level of elaborated argument structure can, and should, be put to use for an insightful account of many other phenomena discussed in the literature. A theory of unaccusativity, as already argued for by Van Valin (1990) and others, must make reference to semantic properties, which are best represented at a level of elaborated argument structure. The notions of external causation vs. internal causation discussed by B. Levin and

Rappaport Hovav (1993), which interact with unaccusativity, can also receive a straightforward account under an elaborated a-structure account. In fact, the notions of external vs. internal causation would appear to correspond in some sense to the volitionality parameter I articulated with respect to Aspectual complex predicates. This issue remains to be explored further.

An account of psych predicates can also be presented in a fairly straightforward fashion through the level of elaborated argument structure. Jackendoff (1990) presents an account of psych predicates by positing a function REACT in addition to AFF at the Action Tier. The function REACT in essence simply serves to invert the order of arguments at the Action Tier so as to come up with the right order of arguments for the linking to syntax. The effect of the function REACT, however, is not convincing and leads to some inconsistencies within the formalism. I believe that an analysis of psych predicates can instead follow simply from an interaction of the Thematic Tier with the Action Tier, along the lines of Grimshaw's (1990) proposal, where two argument structure hierarchies interact with one another: a hierarchy of θ-roles, and a hierarchy of causation, where a cause is ranked higher than a causee. This is in effect similar to the interaction between the Thematic Tier and the Action Tier. The former expresses θ-role like information, while the latter encodes relations between actor and patient/beneficiary, i.e., between causer and causee.

An elaborated a-structure also allows a ready account of causativization. This is illustrated in particular by Alalou and Farrell's (1994) account of Berber causatives, which is based largely on Jackendoff's (1990) theory of Conceptual Semantics as well. Since Alalou and Farrell's account is also within the spirit of Alsina's (1993) account of Romance and Bantu causatives, and my approach is largely compatible with Alsina's account, it is clear that the level of elaborated a-structure provides an account of causatives as well.

The issue of causatives also illustrates an important feature of the theory of complex predicate formation presented here and in Alsina (1993). Alsina examines both Romance causatives, which must be combined in the syntax (as the Urdu permissive is), and Bantu causatives, which are formed morphologically, i.e., in the lexicon. The theory of complex predicate formation at argument structure that Alsina presents, and which I have argued for here, does not distinguish between complex predicates which are formed in the lexicon and ones which are put together in the syntax. Alsina's theory of Predicate Composition, and my proposals for Argument Fusion cover both in a unifying manner. This follows from the fact that the status of a complex predicate cannot be determined from its structural properties.

Rather, a complex predicate is a construction in which two more more semantic heads combine to function as a single unit in terms of grammatical relations that are predicated (no embedded grammatical relations). This is regardless of whether that head is an independent predicate like the Italian *fare* 'make', or a morpheme like the Urdu causative *-aa*.

As was demonstrated, the theory of LFG allows a successful characterization of the properties of complex predicates because differing syntactic information is factored out into independent levels of representation. Also recall that despite the clear advantages of an approach within LFG, the permissive, as well as the Romance causatives, require that the theory allow the unification of two independent heads into a single discontinuous head at phrase structure. This issue is a problem for any theory of syntax. Within LFG, Alsina proposed the theory of Predicate Composition to account for Romance causatives. As already mentioned, I adopt this theory for Urdu complex predicates as well, but additionally make use of case marking information to account for the greater word order variability found in Urdu. In Chapter 6, I relied on the notion of semantic as well as grammatical case and proposed a mechanism for linking. This linking mechanism relates the various levels of representation to one another by relying on semantic and functional information expressed in the lexical entries of case markers.

This view of case is close to Kiparsky's (1987) view of case in his formulation of Linking Theory and also reflects a current trend in GB Theory, in which two types of Case are recognized: inherent and structural (Mahajan 1990). Similarly, the recognition that differing types of information must be represented in a distinct fashion is also reflected in current work within GB, especially by researchers working on complex predicate constructions: it is primarily in these cases where mismatches in information are evident. The definition of complex predicates I present above could be thought of as follows within a Government-Binding approach. There are two semantic heads with distinct argument structures. These two distinct argument structures, however, do not form seperate clauses, but must be contained not only within the same IP, but also within the same VP. Only one subject position may be realized. Under a VP-internal subject approach, as in Mahajan's (1990) approach to Hindi, this would mean that there could only be one Spec of VP. The realization of S-structure representation does not reflect the status of a complex predicate, as the permissive showed.

Many approaches to complex predicates within GB already incorporate some of these ingredients, but an explicit formulation of what exactly characterizes a complex predicate

remains to be expressed. The work on Romance causatives (S. Rosen 1989), Japanese *suru* 'do' (Grimshaw and Mester 1988), Chinese resultatives (Li 1990), and resultatives in polysynthetic languages (Baker 1993) assume a level of argument structure. Grimshaw (1990) and S. Rosen (1989) discuss it as an independent level of representation and assume that it is related to phrase structure by the Projection Principle. Other researchers like Li (1990) and Baker (1993) choose to make use of theta-grid representations, which percolate upwards in a phrase structure tree, and are discharged as they encounter relevant NPs. A combination of these appraoches, along with the proposals I make for complex predicate formation in terms of Argument/Event Fusion would allow the formulation of a unifying, cross-linguistic theory of complex predicate formation within GB.

A theory of complex predicate formation at argument structure alone, however, does not in and of itself solve the particular problem the Urdu permissive in comparison with the instructive poses. In both cases there are two independent predicates. In the case of the permissive, these two predicates combine into a single, discontinuous head. In the case of the instructive they do not. Within LFG this difference is directly expressed in terms of different f-structure representations for the instructive and the permissive. Within a GB approach, this would correspond to two differing D-structure representations which reflect that the instructive contains an embedded VP, while the permissive does not. While this is entirely possible, of course, the challenging part of the analysis would be to derive the identical coordination, negation, and scrambling facts from the two differing D-structure representations. I believe this can only be achieved by recognizing that D-structure indeed is akin to LFG's f-structure in that it serves as a representation of the grammatical relations in a clause. This level of representation does not interact with scrambling and movement in a straight-forward way, i.e., does not indeed have a structural reality, as has been assumed so far for D-structure. Furthermore, it must be recognized that the relationship between argument structure and grammatical relations (D-structure) is not one-to-one, as is required by the Uniformity of Theta Assignment Hypothesis (Baker 1988). Rather, a more complex theory of linking between argument structure and D-structure must be formulated.

Another issue which has been debated in this light is the issue of raising verbs. Within GB approaches, raising phenomena are analyzed structurally. Bresnan (1982a), on the other hand, argues that raising is essentially an instance of functional control (the Urdu instructive is an example of functional control), which is represented at f-structure. In this dissertation, I have extended the notion of control and complementation to the level of

argument structure. The control relations encoded at argument structure are reflected at f-structure by the process of linking a-structure up with f-structure representations. This approach to control is in concord with the theory of semantic control proposed by Sag and Pollard (1991) and should be explored further.

In conclusion, I have presented an in-depth examination of the structure and properties of two differing Urdu complex predicates, the permissive and the Aspectual complex predicates. In the process of formulating a theory of complex predicate formation which accounted for the properties of both of these types of complex predicates, I addressed issues concerned with argument structure, linking, and case marking. Finally, I showed that the theory of complex predicates I present not only allows a successful account of Romance restructuring verbs and Japanese *suru* 'do', but can also be used as a firm base of comparison for an analysis of serial verb constructions.

Bibliography

Adger, D., and C. S. Rhys. 1993. Eliminating disjunction in lexical specification. In D. Adger and C. S. Rhys (Eds.), *Argument Structure, Functional Categories and Parametric Variation: Papers from the workshop on parametric variation and grammar specification*, Vol. 8. Edinburgh, Scotland: EUCCS/WP.

Aissen, J., and D. Perlmutter. 1983. Clause reduction in Spanish. In D. Perlmutter (Ed.), *Studies in Relational Grammar 1*, 360–403. Chicago: The University of Chicago Press.

Alalou, A., and P. Farrell. 1994. Argument structure and causativization in Berber. To appear in *Journal of African Langauges and Linguistics*.

Allen, W. S. 1951. A study in the analysis of Hindi sentence-structure. *Acta Linguistica Hafniensia* 6:68–86.

Alsina, A. 1990. Where's the Mirror Principle? Evidence from Chichewa. Paper presented at the 13th GLOW Colloquium at St. John's College, Cambridge University.

Alsina, A. 1993. *Predicate Composition: A Theory of Syntactic Function Alternations*. PhD thesis, Stanford University.

Alsina, A., and S. Joshi. 1991. Parameters in causative constructions. In L. Dobrin, L. Nichols, and R. M. Rodriguez (Eds.), *Papers from the 27th Regional Meeting of the Chicago Linguistic Society*, 1–16.

Alsina, A., and S. Mchombo. 1988. Lexical Mapping in the Chichewa applicative. Paper presented at the 19th Annual Conference on African Linguistics.

Alsina, A., and S. Mchombo. 1993. Object asymmetries in the Chichewa applicative construction. In S. Mchombo (Ed.), *Theoretical Aspects of Bantu Grammar*, 17–45. Stanford, California: CSLI Publications.

Andrews, A. D., and C. D. Manning. 1993. Information spreading and levels of representation in LFG. Technical Report 176, Center for the Study of Language and Information.

Bahl, K. 1974. *Studies in the Semantic Structure of Hindi: synonymous nouns and adjectives with* karaanaa*, Volume 1.* Delhi: Motilal Banarsidass.

Baker, M. 1985. Syntactic affixation and english gerunds. In J. Goldberg, S. MacKaye, and M. Wescoat (Eds.), *Proceedings of the Fourth Annual West Coast Conference on Linguistics*, 1–11.

Baker, M. 1988. *Incorporation: A Theory of Grammatical Function Changing*. Chicago: University of Chicago Press.

Baker, M. 1989. Object sharing and projection in serial verb constructions. *Linguistic Inquiry* 20:513–553.

Baker, M. 1993. Complex predicates and agreement in a polysynthetic language. In A. Alsina, J. Bresnan, and P. Sells (Eds.), *Complex Predicates*. Stanford, CA: CSLI. To Appear.

Bashir, E. 1988. *Topics in Kalasha Syntax: an Areal and Typological Perspective*. PhD thesis, University of Michigan.

Bashir, E. 1989. Causal chains and compound verbs. Paper presented at the Eleventh South Asian Languages Analysis Roundtable, University of Wisconsin, Madison.

Belletti, A. 1988. The case of unaccusatives. *Linguistic Inquiry* 19(1):1–34.

Bender, E. 1967a. *Hindi Grammar and Reader*. Philadelphia: University of Pennsylvania Press.

Bender, E. 1967b. *Urdu Grammar and Reader*. Philadelphia: University of Pennsylvania Press.

Bhatia, T. K. 1973. On the scope of negation in Hindi. *Studies in the Linguistic Sciences* 3(2):1–27.

Bittner, M., and K. Hale. 1993. Ergativity: Towards a heterogeneous class. Ms., Rutgers University and Massachusetts Institute of Technology.

Bobaljik, J. 1992. Nominally absolutive is not absolutely nominative. In J. Mead (Ed.), *Proceedings of the Eleventh Annual West Coast Conference on Formal Linguistics*. To Appear.

Bresnan, J. 1978. A realistic transformational grammar. In M. Halle, J. Bresnan, and G. Miller (Eds.), *Linguistics Theory and Psychological Reality*. Cambridge, MA: The MIT Press.

Bresnan, J. 1982a. Control and complementation. In J. Bresnan (Ed.), *The Mental Representation of Grammatical Relations*, 282–390. Cambridge, MA: MIT Press.

Bresnan, J. (Ed.). 1982b. *The Mental Representation of Grammatical Relations*. Cambridge, MA: MIT Press.

Bresnan, J. 1991. Locative case vs. locative gender. In L. Sutton, C. Johnson, and R. Shields (Eds.), *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society*, 53–68.

Bresnan, J., and J. Kanerva. 1989. On locative inversion in Chichewa. *Linguistic Inquiry* 20:1–50.

Bresnan, J., and L. Moshi. 1990. Object asymmetries in comparative Bantu syntax. *Linguistic Inquiry* 21:147–186.

Bresnan, J., and A. Zaenen. 1990. Deep unaccusativity in LFG. In K. Dziwirek, P. Farrell, and E. Mejías-Bikandi (Eds.), *Grammatical Relations: A Cross-Theoretical Perspective*, 45–57. Stanford, CA: CSLI.

Burzio, L. 1986. *Italian Syntax: A Government-Binding Approach*. Dordrecht: Reidel.

Butt, M. 1991. Acquisition of the ergative in Urdu. Ms., Stanford University.

Butt, M. 1993a. Complex predicate scrambling in Urdu. In M. Butt, T. King, and G. Ramchand (Eds.), *Theoretical Perspectives on Word Order Issues in South Asian Languages*. Stanford, CA: CSLI. To appear.

Butt, M. 1993b. Object specificity and agreement in Hindi/Urdu. In *Papers from the 29th Regional Meeting of the Chicago Linguistic Society*. To Appear.

Butt, M. 1994. Complex Predicates and Machine Translation. In *Proceedings of the Konferenz zur Verarbeitung natürlicher Sprache (KONVENS)*. Österreichische Gesellschaft für Artificial Intelligence (ÖGAI). To Appear.

Butt, M., M. Isoda, and P. Sells. 1990. Complex Predicates in LFG. Ms., Stanford University.

Butt, M., and T. H. King. 1991. Semantic case in Urdu. In L. Dobrin, L. Nichols, and R. M. Rodriguez (Eds.), *Papers from the 27th Regional Meeting of the Chicago Linguistic Society*, 31–45.

Carrier-Duncan, J. 1985. Linking of thematic roles in derivational word formation. *Linguistic Inquiry* 16:1–35.

Cattell, R. 1984. *Composite Predicates in English.* Sydney: Academic Press Australia.

Comrie, B. 1981. *Language Universals and Linguistic Typology.* Chicago: The University of Chicago Press.

Crowley, T. 1987. Serial verbs in Paamese. *Studies in Language* 11(1):35–84.

Dalrymple, M. 1990. *Syntactic Constraints on Anaphoric Binding.* PhD thesis, Stanford University.

Dalrymple, M. 1991. Against reconstruction in ellipsis. Ms., Xerox-PARC, Palo Alto, CA.

Dalrymple, M., J. Lamping, and V. Saraswat. 1993. LFG Semantics via Constraints. In *Proceedings of the Sixth Meeting of the European Chapter of the Association for Computational Linguistics.*

Davies, W., and C. Rosen. 1988. Unions as multi-predicate clauses. *Language* 64:52–88.

Davison, A. 1985. Case and control in Hindi-Urdu. *Studies in the Linguistic Sciences* 15(2):9–23.

Davison, A. 1988. Constituent structure and the realization of agreement features. In L. MacLeod, G. Larson, and D. Brentari (Eds.), *Papers from the 24th Regional Meeting of Chicago Linguistic Society*, 41–53.

Davison, A. 1990. Long distance syntactic anaphors in Hindi-Urdu. Paper presented at the Delhi University Conference on Pronouns and Anaphors.

Davison, A. 1991a. Feature percolation and agreement in Hindi-Urdu. Paper presented at the South Asian Conference, University of Wisconsin.

Davison, A. 1991b. Finiteness and case in Hindi-Urdu complements. Paper presented at the Eleventh South Asian Languages Analysis Roundtable, University of Illinois, Urbana-Champaign.

Dayal, V. 1993. Binding facts in Hindi and the scrambling phenomenon. In M. Butt, T. H. King, and G. Ramchand (Eds.), *Theoretical Perspectives on Word Order Issues in South Asian Languages*. Stanford, CA: CSLI. To Appear.

Dayal, V. S. 1992. The singular-plural distinction in Hindi generics. In *Semantics and Linguistic Theory II*.

de Hoop, H. 1992. *Case Configuration and Noun Phrase Interpretation*. PhD thesis, Rijksuniversiteit Groningen.

DeLancey, S. 1986. Evidentiality and volitionality in Tibetan. In W. Chafe and J. Nichols (Eds.), *Evidentiality: the Linguistic Coding of Epistemology*. Ablex.

Dixon, R. M. W. 1979. Ergativity. *Language* 55(1):59–138.

Dowty, D. 1979. *Word Meaning and Montague Grammar*. Dordrecht: Reidel.

Dowty, D. 1991. Thematic proto-roles and argument selection. *Language* 67(3):547–619.

Durie, M. 1993. Grammatical structures in verb serialization: some preliminary proposals. In A. Alsina, J. Bresnan, and P. Sells (Eds.), *Complex Predicates*. Stanford, CA: CSLI. To Appear.

Dwivedi, V. 1991. Negation as a functional projection in Hindi. In K. Hunt, T. Perry, and V. Samiian (Eds.), *Proceedings of the Western Conference on Linguistics*, 88–101. California State University of Fresno: Department of Linguistics.

Dwivedi, V. 1993. *Syntactic Dependencies and Relative Phrases in Hindi*. PhD thesis, University of Massachusetts at Amherst.

Enç, M. 1991. The semantics of specificity. *Linguistic Inquiry* 22(1):1–25.

Fodor, J. 1970. Three reasons for not deriving 'kill' from 'cause to die'. *Linguistic Inquiry* 1:429–438.

Foley, W. A., and R. D. Van Valin. 1984. *Functional Syntax and Universal Grammar*. Cambridge: Cambridge University Press.

Fortescue, M. 1985. Learning to speak Greenlandic: A case study of a two-year-old's morphology in a polysynthetic language. *First Language* 5:101–114.

Fukui, N., and M. Speas. 1986. Specifiers and projections. *MIT Working Papers in Linguistics* 8:128–172.

Gambhir, V. 1981. *Syntactic Restrictions and Discourse Functions of Word Order in Standard Hindi*. PhD thesis, University of Pennsylvania.

Garrett, A. 1990. The origin of NP split ergativity. *Language* 66(2):261–296.

Gazdar, G., and G. K. Pullum. 1981. Subcategorization, constituent order and the notion "head". In M. Moortgart, H. van der Hulst, and T. Hoekstra (Eds.), *The Scope of Lexical Rules*. Dordrecht: Foris Publications.

Gazdar, G., and G. K. Pullum. 1982. Generalized phrase structure grammar: A theoretical synopsis. Technical report, Bloomington: Indiana University Linguistics Club.

Givón, T. 1979. *On Understanding Grammar*. New York: Academic Press.

Givón, T. 1990. Verb serialization in Tok Pisin and Kalam. In J. W. Verhaar (Ed.), *Melanesian Pidgin and Tok Pisin*, 19–55. Amsterdam: John Benjamins.

Givón, T. 1991. Some substantive issues concerning verb serialization: grammatical vs. cognitive packaging. In C. Lefebvre (Ed.), *Serial verbs: grammatical, comparative and cognitive approaches*, 137–184. Amsterdam: John Benjamins.

Glassman, E. H. 1977. *Spoken Urdu*. Lahore: Nirali Kitaben.

Grimshaw, J. 1987. Unaccusatives — an overview. In J. McDonough and B. Plunkett (Eds.), *North Eastern Linguistic Society 17*, 244–258.

Grimshaw, J. 1979. Complement selection and the lexicon. *Linguistic Inquiry* 10:279–326.

Grimshaw, J. 1990. *Argument Structure*. Cambridge, MA: The MIT Press.

Grimshaw, J. 1991. Extended projection. Ms., Brandeis University.

Grimshaw, J., and A. Mester. 1988. Light verbs and θ-marking. *Linguistic Inquiry* 19(2):205–232.

Guilfoyle, E., H. Hung, and L. Travis. 1992. SPEC of IP and SPEC of VP: Two subjects in Austronesian languages. *Natural Language and Linguistic Theory* 10:375–414.

Gumperz, J. 1971. *Language in Social Groups*. Stanford, CA: Stanford University Press.

Gurtu, M. 1985. *Anaphoric Relations in Hindi and English*. PhD thesis, Central Institute of English and Foreign Languages.

Halvorsen, P.-K. 1983. Semantics for Lexical-Functional Grammar. *Linguistic Inquiry* 14(4):567–615.

Halvorsen, P.-K. 1987. Situation semantics and semantic interpretation in constraint-based grammars. Technical Report 101, Center for the Study of Language and Information.

Halvorsen, P.-K., and R. Kaplan. 1988. Projections and semantic description in Lexical-Functional Grammar. In *Proceedings of the International Conference on Fifth Generation Computer Systems*, 1116–1122.

Harbert, W., and V. Srivastav. 1988. A complement/adjunct asymmetry in Hindi and other languages. *Cornell Working Papers in Linguistics* 8:79–106.

Holisky, D. A. 1987. The case of the intransitive subject in Tsova-Tush (Batsbi). *Lingua* 71:103–132.

Hook, P. 1979. *Hindi Structures: Intermediate Level*. University of Michigan: Center for South and Southeast Asian Studies.

Hook, P. E. 1974. *The Compound Verb in Hindi*. Center for South and Southeast Asian Studies: The University of Michigan.

Hook, P. E. 1991. The emergence of perfective aspect in Indo-Aryan languages. In E. Traugott and B. Heine (Eds.), *Approaches to Grammaticalization*, 59–89. Amsterdam: Benjamins.

Huang, C.-J. 1992. Complex predicates in control. In R. Larson, S. Iatridou, and U. Lahiri (Eds.), *Control and Grammar*, 109–147. Kluwer Academic Publishers.

Iida, M. 1987. Case assignment by nominals in Japanese. In M. Iida, S. Wechsler, and D. Zec (Eds.), *Working Papers in Grammatical Theory and Discourse Structure: Interactions of Morphology, Syntax, and Discourse*, 93–138. Stanford, CA: CSLI.

Inman, M. 1993. *Semantics and Pragmatics of Colloquial Sinhala Involitive Verbs.* PhD thesis, Stanford University.

Inoue, K. 1976. *Henkei Bunpoo to Nihongo [Transformational Grammar and the Japanese Language].* Vol. 1. Tokyo: Taishukan.

Ishikawa, A. 1985. *Complex Predicates and Lexical Operators in Japanese.* PhD thesis, Stanford University.

Isoda, M. 1991. The light verb construction in Japanese. In L. Dobrin, L. Nichols, and R. M. Rodriguez (Eds.), *Papers from the 27th Regional Meeting of the Chicago Linguistic Society*, 261–275.

Jackendoff, R. 1983. *Semantics and Cognition.* Cambridge, MA: MIT Press.

Jackendoff, R. 1990. *Semantic Structures.* Cambridge, MA: The MIT Press.

Jackendoff, R. 1991. Parts and boundaries. *Cognition* 41:9–45.

Jespersen, O. 1954. *A Modern English Grammar on Historical Principles.* London: Allen and Unwin.

Johns, A. 1992. Deriving ergativity. *Linguistic Inquiry* 23:39–76.

Kachru, Y. 1966. *An Introduction to Hindi Syntax.* Urbana-Champaign: University of Illinois.

Kachru, Y. 1980. *Aspects of Hindi Grammar.* New Delhi: Manohar Publications.

Kachru, Y. 1981. Transitivity and volitionality in Hindi-Urdu. *Studies in the Linguistic Sciences* 11(2):181–193.

Kachru, Y. 1987. Ergativity, subjecthood and topicality in Hindi-Urdu. *Lingua* 71:223–238.

Kachru, Y., and R. Pandharipande. 1978. On ergativity in selected South Asian languages. *Studies in the Linguistic Sciences* 8(1):111–127.

Kaplan, R., and J. Bresnan. 1982. Lexical-Functional Grammar: A formal system for grammatical representation. In J. Bresnan (Ed.), *The Mental Representation of Grammatical Relations*, 173–281. Cambridge, MA: MIT Press.

Kaplan, R. 1987. Three seductions of computational psycholinguistics. In P. Whitelock, H. Somers, P. Bennett, R. Johnson, and M. M. Wood (Eds.), *Linguistic Theory and Computer Applications*, 149–188. London: Academic Press.

Kaplan, R., and J. Wedekind. 1993. Restriction and correspondence-based translation. In *Proceedings of the Sixth Meeting of the European Chapter of the Association for Computational Linguistics*. University of Utrecht, April.

Kaplan, R., and A. Zaenen. 1989. Long-distance dependencies, constituent structure, and functional uncertainty. In M. R. Baltin and A. S. Kroch (Eds.), *Alternative Conceptions of Phrase Structure*, 17–42. Chicago, IL: Univ. of Chicago Press.

Khan, B. 1987. The ergative case in Hindi-Urdu. *Studies in the Linguistic Sciences* 17(1):91–101.

King, T. 1993. *Configuring Topic and Focus in Russian*. PhD thesis, Stanford University.

Kiparsky, P. 1987. Morphology and grammatical relations. Ms., Stanford University.

Krishnamurti, B. 1989. Complex predicates in Telugu. Paper presented at the Eleventh South Asian Language Analysis Roundtable, University of Wisconsin, Madison.

Kroeger, P. 1993. *Phrase Structure and Grammatical Relations in Tagalog*. Stanford, CA: CSLI Publications.

Kural, M. 1993. V-to(-I-to)-C in Turkish. *UCLA Occasional Papers in Linguistics* 11.

Lascarides, A., and A. Copestake. 1999. Default representation in constraint-based frameworks. *Computational Linguistics* 25.

Levin, B. 1983. *On the Nature of Ergativity*. PhD thesis, Massachusetts Institute of Technology.

Levin, B., and M. Rappaport Hovav. 1993. Are all causative verbs equal? Paper presented at the Meeting of the Sonderforschungsbereich 340, Stuttgart.

Levin, L. 1986. *Operations on Lexical Forms: Unaccusative Rules in Germanic Languages*. PhD thesis, MIT.

Li, Y. 1990. On V-V compounds in Chinese. *Natural Language and Linguistic Theory* 8:177–208.

Lord, C. 1976. Evidence for syntactic reanalysis: from verb to complementizer in Kwa. In S. Mufwene, C. Walker, and S. Steever (Eds.), *Papers from the 12th Regional Meeting of the Chicago Linguistic Society, Parassesion on Diachronic Syntax*, 179–91.

Mahajan, A. 1989. Agreement and agreement phrases. *MIT Working Papers in Linguistics* 10:217–252.

Mahajan, A. 1990. *The A/A-Bar Distinction and Movement Theory*. PhD thesis, MIT.

Mahajan, A. 1991. Clitic doubling, object agreement and specificity. In T. Sherer (Ed.), *North Eastern Linguistic Society 21*, 263–277.

Mahajan, A. 1992. The specificity condition and the CED. *Linguistic Inquiry* 23(3):510–516.

Mahajan, G. 1990. On complex verbs in Hindi. Paper presented at the Twelfth South Asian Languages Analysis Roundtable, University of California, Berkeley.

Manning, C. D. 1993. Analyzing the verbal noun: Internal and external constraints. In S. Choi (Ed.), *Japanese/Korean Linguistics 3*, 236–253. Stanford Linguistics Association, CSLI.

Manning, C. D. 1992. Romance is so complex. Technical Report 168, Center for the Study of Language and Information.

Marantz, A. 1984. *On the Nature of Grammatical Relations*. Cambridge, MA: MIT Press.

Masica, C. 1976. *Defining a Linguistic Area: South Asia*. Chicago: The University of Chicago Press.

Matsumoto, Y. 1988. Still more on light verbs. Ms., Stanford University.

Matsumoto, Y. 1992. *On the Wordhood of Complex Predicates in Japanese*. PhD thesis, Stanford University.

McKay, T. 1985. *Infinitival Complements in German.* Cambridge: Cambridge University Press.

Milsark, G. 1988. Singl-ing. *Linguistic Inquiry* 19.

Miyagawa, S. 1989. Light verbs and the ergative hypothesis. *Linguistic Inquiry* 20:659–668.

Moens, M., and M. Steedman. 1988. Temporal ontology and temporal reference. *Computational Linguistics* 15–28.

Mohanan, K. P. 1982. Grammatical relations in Malayalam. In J. Bresnan (Ed.), *The Mental Representation of Grammatical Relations*, 504–589. Cambridge, MA: MIT Press.

Mohanan, K. P. 1989a. Compositionality of complex predicates in Malayalam. Paper presented at the Eleventh South Asian Languages Analysis Roundtable, University of Wisconsin, Madison.

Mohanan, K. P. 1989b. On the representation of thematic role information. Ms., Stanford University.

Mohanan, T. 1990. *Arguments in Hindi.* PhD thesis, Stanford University.

Mohanan, T. 1992. Wordhood and lexicality: Noun incorporation in Hindi. To appear in *Natural Language and Linguistic Theory.*

Mohanan, T. 1993a. Case alternation on objects in Hindi. *South Asian Language Review* 3(1).

Mohanan, T. 1993b. Case OCP in Hindi. In M. Butt, T. King, and G. Ramchand (Eds.), *Theoretical Perspectives on Word Order Issues in South Asian Languages.* Stanford, CA: CSLI. To appear.

Mohanan, T. 1993c. Multidimensionality of representation: Noun-verb complex predicates in Hindi. In A. Alsina, J. Bresnan, and P. Sells (Eds.), *Complex Predicates.* Stanford, CA: CSLI. To Appear.

Nagarajan, S. 1990. Tamil serial verbs. In A. Zwicky and B. Joseph (Eds.), *When verbs collide: Papers from the 1990 Ohio State Mini-conference on Serial Verbs*, 200–217.

Ochs, E. 1985. Variation and error: A sociolinguistic approach to language acquisition in Samoan. In D. I. Slobin (Ed.), *The Crosslinguistic Study of Language Acquisition, Vol. 1: The Data*. Hillsday, NJ: Lawrence Erlbaum.

Pandharipande, R., and Y. Kachru. 1977. Relational grammar, ergativity, and Hindi-Urdu. *Lingua* 41:217–238.

Pandharipande, R. 1989. Split properties of compound verbs in Marathi. Paper presented at the Eleventh South Asian Languages Analysis Roundtable, University of Wisconsin, Madison.

Pandharipande, R. 1990. Serial verb construction in Marathi. In A. Zwicky and B. Joseph (Eds.), *When verbs collide: Papers from the 1990 Ohio State Mini-conference on Serial Verbs*, 178–199.

Perlmutter, D. 1978. Impersonal passive and the Unaccusative Hypothesis. In J. Jaeger, A. Woodbury, F. Ackerman, C. Chiarello, O. Gensler, J. Kingston, E. Sweetser, H. Thompson, and K. Whistler (Eds.), *Proceedings of the Fourth Annual Meeting of the Berkeley Linguistics Society*, 157–189.

Porterfield, L., and V. Srivastav. 1988. (In)definiteness in the absence of articles: Evidence from Hindi and Indonesian. In H. Borer (Ed.), *Proceedings of the Seventh West Coast Conference on Formal Linguistics*.

Poser, W. 1989. What is the 'Double-o Constraint' a constraint on? Ms., Stanford University.

Poser, W. 1991. Japanese periphrastic verbs and noun incorporation. Ms., Stanford University.

Pye, C. 1990. The acquisition of ergative languages. *Linguistics* 28:1291–1330.

Ramchand, G. 1990. Complex predicate formation in Bangla. In A. Halpern (Ed.), *Proceedings of the Ninth West Coast Conference on Formal Linguistics*, 443–458.

Ramchand, G. 1993. *Aspect and Argument Structure in Scottish Gaelic*. PhD thesis, Stanford University.

Ritter, E., and S. Rosen. 1993. The independence of external arguments. In *Proceedings of the Twelfth West Coast Conference on Formal Linguistics*. To Appear.

Rizzi, L. 1982. *Issues in Italian Syntax*. Dordrecht: Foris Publications.

Rosen, C. 1990. Italian evidence for multi-predicate clauses. In K. Dziwirek, P. Farrell, and E. Mejías-Bikandi (Eds.), *Grammatical Relations: A Cross-Theoretical Perspective*. Stanford, CA: CSLI.

Rosen, C. 1993. Auxiliation and serialization: On discerning the difference. In A. Alsina, J. Bresnan, and P. Sells (Eds.), *Complex Predicates*. Stanford, CA: CSLI. To Appear.

Rosen, S. 1989. *Argument Structure and Complex Predicates*. PhD thesis, Brandeis University.

Sag, I., and C. Pollard. 1991. An integrated theory of complement control. *Language* 67(1):63–113.

Schieffelin, B. 1985. The acquisition of Kaluli. In D. I. Slobin (Ed.), *The Crosslinguistic Study of Language Acquisition, Vol. 1: The Data*. Hillsday, NJ: Lawrence Erlbaum.

Sebba, M. 1987. *The Syntax of Serial Verbs*. Amsterdam: John Benjamins.

Sells, P. 1985. *Lectures on Contemporary Syntactic Theories*. Stanford, CA: CSLI.

Sells, P. 1989. More on light verbs and θ-marking. Ms., Stanford University.

Sells, P. 1990. VP in Japanese: Evidence from *-te* complements. In H. Hoji (Ed.), *Proceedings of the First Conference on Japanese and Korean Linguistics*, 319–333.

Sells, P. 1991. Properties of Verbal Nouns and Category Status in Japanese. Ms., Stanford Unversity.

Seuren, P. 1991. The definition of serial verbs. In Byrne, Francis, and T. Huebner (Eds.), *Development and structure of creole languages: Essays in honour of Derek Bickerton*, 193–205. Amsterdam: John Benjamins.

Silverstein, M. 1976. Hierarchy of features and ergativity. In R. Dixon (Ed.), *Grammatical Categories in Australian Languages (AIAS Linguistic Series 22)*, 112–171. Canberra: Australian Institute of Aboriginal Studies.

Simpson, J. 1983. *Aspects of Warlpiri Morphology and Syntax*. PhD thesis, MIT.

Simpson, J. 1991. *Warlpiri Morpho-Syntax*. Dordrecht: Kluwer Academic Publishers.

Simpson, J., and J. Bresnan. 1982. Control and obviation in Warlpiri. In D. Flickinger, M. Macken, and N. Wiegand (Eds.), *Proceedings of the First Annual West Coast Conference on Formal Linguistics*, 280–291.

Singh, M. 1990. The aspectual content of compound verbs. In Y. No and M. Libucha (Eds.), *Proceedings of the Sevent Eastern States Conference on Linguistics*, 260–271.

Singh, M. 1993. Thematic roles, word order, and definiteness. In M. Butt, T. King, and G. Ramchand (Eds.), *Theoretical Perspectives on Word Order Issues in South Asian Languages*. Stanford, CA: CSLI. To Appear.

Slobin, D. (Ed.). 1993. *The Crosslinguistic Study of Language Acquisition, Volume 3*. Hillsday, NJ: Lawrence Erlbaum.

Smith, C. 1991. *The Parameter of Aspect*. Dordrecht: Kluwer Academic Publishers.

Smith, H. 1992. *Restrictiveness in Case Theory*. PhD thesis, Stanford University.

Speas, M. 1990. *Phrase Structure in Natural Language*. Dordrecht: Kluwer Academic Publishers.

Srivastav, V. 1991a. Subjacency effects at LF: The case of HindiΘ Wh. *Linguistic Inquiry* 22(4):762–769.

Srivastav, V. 1991b. The syntax and semantics of correlatives. *Natural Language and Linguistic Theory* 9(4):637–686.

Srivastav, V. 1991c. *WH Dependencies in Hindi and the Theory of Grammar*. PhD thesis, Cornell University.

Steever, S. 1979. Noun incorporation in Tamil, or what's a noun like you doing in a verb like this. In P. Clyne, W. Hanks, and C. Hofbauer (Eds.), *Papers from the 15th Regional Meeting of the Chicago Linguistic Society*, 279–291.

Stowell, T. 1981. *Origins of Phrase Structure*. PhD thesis, MIT.

Subbarao, K. V. 1984. *Complementation in Hindi Syntax.* Delhi, India: Academic Publications.

Tenny, C. 1987. *Grammaticalizing Aspect and Affectedness.* PhD thesis, MIT.

Tsujimura, N. 1990. Ergativity of nouns and case assignment. *Linguistic Inquiry* 21:277–287.

Vale, R. 1948. *Verbal Composition in Indo-Aryan.* Poona, India: Deccan College.

Van Valin, R. 1987. The unaccusative hypothesis vs. lexical semantics: Syntactic vs. semantic approaches to verb classification. In J. McDonough and B. Plunkett (Eds.), *North Eastern Linguistic Society 17*, 641–661.

Van Valin, R. 1990. Semantic parameters of split intransitivity. *Language* 66(2):221–260.

Van Valin, R. 1981. Grammatical relations in ergative languages. *Studies in Language* 5(3):361–394.

Van Valin, R. 1993. An overview of ergative phenomena and their implications for language acquisition. In D. I. Slobin (Ed.), *The Crosslinguistic Study of Language Acquisition, Vol. 3.* Hillsday, NJ: Lawrence Erlbaum.

Vendler, Z. 1967. *Linguistics in Philosophy.* Cornell University Press.

Verma, M. K. 1971. *The Structure of the Noun Phrase in English and Hindi.* Delhi: Motilal Banarsidass.

Watters, J. 1993. The morphosyntax of V-V constructions in Tepehua. Paper presented at the Annual Meeting of the Linguistics Society of America, Los Angeles.

Yatabe, S. 1993. *Scrambling and Japanese Phrase Structure.* PhD thesis, Stanford University.

Zaenen, A., J. Maling, and H. Thráinsson. 1985. Case and grammatical functions: the Icelandic passive. *Natural Language and Linguistic Theory* 3:441–483.

Zwicky, A. 1990. What are we talking about when we talk about serial verbs? In A. Zwicky and B. Joseph (Eds.), *When verbs collide: Papers from the 1990 Ohio State Mini-conference on Serial Verbs*, 1–13.