Structural Topic and Focus without Movement[1]

Miriam Butt and Tracy Holloway King

1 Introduction

One of the challenges in presenting a syntactic analysis of free word order languages like Urdu and Turkish lies in motivating the various possible permutations of a given sentence. For example, differing word orders have been motivated in terms of case theory, adjunction, or head movement (e.g., Mahajan (1990)). Within a Minimalist approach, a possible analysis could even be posited in terms of an optional [scr(amble)] feature which must be checked (Müller (1995)). Alternatively, in declarative, constraint-based theories like Lexical-Functional Grammar, word order permutations have been taken to be merely a statement about a language particular c(onstituent)structure flexibility. Grammatical functions are dissociated from word order and are coded in terms of f(unctional)-structure (see e.g., T. Mohanan (1994) and Butt (1995) on Hindi/Urdu). However, a deeper explanation and understanding of word order effects in these languages elude these approaches.

We propose to analyze the free word order in Urdu[2] and Turkish[3] as base-generated possibilities which reflect differing information structures of a sentence. We propose that the varied word orders are optional from a purely syntactic point of view: they are not motivated by case or agreement. Instead, they are motivated both by semantic factors, such as specific vs. nonspecific interpretations and discourse considerations, such as topic and focus. In particular, we investigate an intriguing interaction between preverbal structural focus and the nonspecific interpretation of preverbal objects in Urdu and Turkish.

2 Word Order and Discourse Functions

Urdu and Turkish are right-headed (SOV) languages. However, unlike in Japanese or Korean, the surface order of elements is not strictly head-final; so, a sentence with a subject, verb, and object can have all six possible orders. We take the position that languages like Urdu and Turkish are discourse configurational in that discourse functions are encoded syntactically and thus affect word order (É. Kiss (1995)). The various word orders in Turkish and Urdu primarily have a pragmatic effect in that they express differing ways of conveying some particular information (Gambhir 1981; Erguvanlı 1984; Hoffman 1995). As such, in order to determine the syntax and phrase structure of these languages, it is necessary to analyze how discourse functions are licensed, i.e., which positions correspond to which discourse function interpretations of the arguments found there.

2.1 Theoretical Considerations

Vallduví (1992, 1993) for Catalan and English and King (1995) for Russian argue that the traditional bipartite divisions of a sentence drawn in terms of *topic-focus*, *theme-rheme*, *old information-new information* are best understood in terms of a tripartite distinction. Vallduví

[1]We would like to acknowledge the invaluable help we received from Beryl Hoffman, Ümit Turan and Tara Mohanan in their willingness to answer questions about Turkish and Urdu data, and for their enthusiastic participation in discussions about focus, topic, and incorporation in general. We would also like to thank Joe Eskenazi, Veerle van Geenhoven, Michael Inman, Louise McNally, Chris Piñón, Peter Sells, Robert Underhill and the audiences at the 1996 LSA and a Stanford University Colloquium for comments and valuable input.

[2]The South Asian language Urdu is closely related to Hindi, which is mostly spoken in India. In this paper, we primarily draw our data from the dialect of Urdu spoken in Lahore, Pakistan and the dialect of Hindi-speaking informants from New Delhi, India.

[3]The data and claims for Turkish are taken mainly from Hoffman (1995), who bases her findings on a corpus gleaned from the CHILDES database (Mac Whinney and Snow (1995)), transcribed colloquial speech, and contemporary novels. Further data is taken from Kornfilt (1995) and additional fieldwork conducted independently.

defines the relevant notions in new terms, in order to avoid potential confusion with existing terminologies. He views the information structure of a sentence as instructions to the hearer on how to update his/her current knowledge store. He couches the idea of a knowledge store in terms of a Heimian collection of file-cards (Heim (1982)). The FOCUS part of a sentence can be seen as an instruction to update a given file-card or to add an entirely new one. The GROUND represents the information that is already known. However, a distinction is made between the kind of information that represents a LINK, and the kind that is contained in the TAIL. The link points the hearer to the file card that is to be updated, while the tail further specifies how the new information fits onto the given file card. Here, we will refer to the link as TOPIC and the tail as BACKGROUND.

Choi (1996) proposes an extension of Vallduví's system whereby there is a four-way distinction based on two features: [±New] and [±Prominent].[4] Following Choi, we assume that the information structure must be further divided and make the following proposal based on her two features. Topic and focus share the discourse function feature [+Prominent] which differentiate them from their relatively less prominent pairs, background and completive information respectively. Completive information and focus share the feature [+New] since they both introduce new information into the discourse. While focus's major function is to fill the informational gap between the speaker and the hearer,[5] completive information provides information which is new to the hearer, and hence [+New], but which is not of primary importance to the information structure of the discourse at hand, and hence [−Prom].

(1) [+New] = focus [+Prom]
completive information [−Prom]
[−New] = topic [+Prom]
background information [−Prom]

In this paper, we incorporate the extension of Vallduví's proposal into the structural approach to topic and focus (É. Kiss (1995), King (1995)). The remainder of this section discusses how Turkish and Urdu encode these discourse functions.

2.2 Topic

In Urdu and Turkish the first element in an utterance is interpreted as TOPIC. Topics occur in clause initial position in matrix clauses, as in (2), and in clauses with complementizers, as in (3).[6]

(2) a. [hassan=ko]$_T$ naadyaa=ne tofii d-ii
Hassan.M=Dat Nadya.F=Erg toffee.F.Nom give-Perf.F.Sg
'To Hassan Nadya gave toffee.' (Urdu)

b. [bu kitab-ı]$_T$ Hasan ban-a ver-di
this book-Acc Hasan.Nom I-Dat give-Past.3Sg
'This book Hasan gave to me.' (Turkish)

[4]Choi's use of these features is different than ours in that her division of focus differentiates between CONTRASTIVE FOCUS and COMPLETIVE FOCUS, as in (i). We do not discuss contrastive focus in this paper.

(i) S = focus, ground
focus [+New] = contrastive [+Prom]
completive [−Prom]
ground [−New] = topic [+Prom]
background [−Prom]

[5]This is what is often refered to as new-information and presentational focus (Dik et al. (1981); Rochemont (1986); Rochemont and Culicover (1990)).

[6]The '=' in the Urdu examples is intended to encode that the case markers act more like clitics than bound morphemes (T. Mohanan (1994)), which are indicated by '-'.

(3) a. [anjum=ne]$_T$ dek^h-aa [ki [hassan=ko]$_T$ naadyaa=ne tofii
Anjum.F=Erg see-Perf.M.Sg that Hassan.M=Dat Nadya.F=Erg toffee.F.Nom
d-ii]
give-Perf.F.Sg
'Anjum saw that to Hassan Nadya gave toffee.' (Urdu)

b. Zeynep [[çocuk-lar-ı]$_T$ ben-im okul-a götür-düğ-üm-ü] bil-iyor.
Zeynep.Nom [child.Pl-Acc I-Gen school-Dat take-Ger.1Sg-Acc] know-Prog.3Sg
'Zeynep knows that the children, I took to school.' (Turkish)

We propose that these topics be represented in a position which is situated above the default position of the arguments. In particular, the topicalized constituent is in SpecIP (compare Bresnan (1996) and King (1995) for Russian, Dwivedi (1994) for Hindi). In sentences like those in (2) and (3) in which a non-subject argument is topicalized, its appearance in SpecIP results in non-canonical word order in which the non-subject argument precedes the subject. On the other hand, sentences such as (4), in which the subject is in initial position, can have two structures.

(4) [naadyaa=ne]$_{(T)}$ hassan=ko tofii d-ii
Nadya.F=Erg Hassan.M=Dat toffee.F.Nom give-Perf.F.Sg
'Nadya gave toffee to Hassan.' (Urdu)

Topic: [$_{IP}$ naadyaa=ne [$_{I'}$ ...]] No topic: [$_{IP}$ [$_{I'}$ [naadyaa=ne ...]]]

The first, and dominant, reading is one in which the subject is topicalized and hence is in SpecIP. Since the subject already precedes the other non-subject arguments, this movement does not result in an overt difference in word order, although it must apppear in SpecIP in order for the subject to be licensed as a topic. The second reading is one in which the subject is not in SpecIP and hence is not interpreted as a topic. In this case, (4) has no topic or has a non-overt continuing topic from a previous utterance in the discourse.

2.3 Focus

Focus's major function is to fill the informational gap between the speaker and the hearer and hence provides new information relevant for the discourse structure. If there is only one focused constituent in the sentence, then it must appear immediately preverbally. This is illustrated in (5) for Urdu and (6) for Turkish.[7] We propose that focus is licensed in SpecVP. As discussed in section **??** this licensing interacts with that of nonspecific objects.

(5) a. naadyaa=ne hassan=ko [tofii]$_F$ d-ii
Nadya.F=Erg Hassan.M=Dat toffee.F.Nom give-Perf.F.Sg
'Nadya gave TOFFEE to Hassan.' (Urdu)

b. #naadyaa=ne [hassan=ko]$_F$ tofii d-ii
Nadya.F=Erg Hassan.M=Dat toffee.F.Nom give-Perf.F.Sg
'Nadya gave toffee to HASSAN.' (Urdu)

(6) [bu kitab-ı]$_T$ Hasan [ban-a]$_F$ ver-di
this book-Acc Hasan.Nom I-Dat give-Past.3Sg
'This book Hasan gave to ME.' (Turkish)

[7] In addition to the preverbal focus position, *in situ* focusing of a phrase is possible when there are multiple foci. This *in situ* focus is always a case of contrastive focus relative to the preverbal focus, as illustrated in (i). In this example, the focus on *Hassan* is only permissible in a context in which Hassan is contrasted with another possible recipient. We will not consider multiple foci, and hence *in situ* contrastive focus, in this paper.
(aadnaan=ke-liye nahĩ) naadyaa=ne [hassan=ke-liye]$_{CF}$ [tofii]$_F$ xarid-ii
Adnan.M=for not Nadya.F=Erg Hassan.M=for toffee.F.Nom buy-Perf.F.Sg
'Nadya bought TOFFEE for HASSAN (not for Adnan).' (Urdu)

2.4 Background Information

The interpretation of BACKGROUND information is akin to topicalized information in that both have the status of "old" or "known", i.e., [−New], information. The difference between the two from an informational structural point of view can be described as follows: while topics are the pointer to the relevant information (file card in the Heimian metaphor) to be accessed by the hearer, backgrounded material only provides more detailed information as to how the new information provided fits in with the already known information and is hence [−Prominent]. That is, the backgrounded material provides the background that may be necessary for a good understanding of the new, focused information supplied.

Syntactically, backgrounded phrases in Urdu and Turkish occur postverbally, as in (7). For example, the Turkish (7b) is uttered in the context of the speaker just having been told to give a ball to a cat (from Hoffman (1995:140)).[8] (8) demonstrates that the postverbal backgrounded material is part of the clause and not simply an afterthought.

(7) a. [tiin hii ciṭʰiyãã]$_F$ ḍaal-ii hãĩ [us=ne]$_{Back}$
three only letter.F.Pl.Nom put-Perf.F.Sg be.Pres.Pl pron=Erg
'Only three letters (she) has sent.' (Urdu)

b. yok, [Funda'nın top-u-nu]$_F$ ver-me-m [kedi-ye]$_{Back}$
no, Funda-Gen ball-Poss3-Acc give-Neg-1Sg cat-Dat
'No, (I) won't give Funda's ball to the cat.' (Turkish)

(8) Zeynep [çocuk-lar-ı okul-a götür-düğ-üm-ü ben-im]
Zeynep.Nom [child-Pl-Acc school-Dat take-Ger-1S-Acc [I-Gen]$_{Back}$]
bil-iyor.
know-Prog.3Sg
'Zeynep knows that I took the children to school.' (Turkish)

These constituents occur postverbally in a position that is right adjoined to IP. This analysis is in keeping with traditional analyses of extraposition in Hindi (Srivastav (1991)), and is further supported by data and arguments from German (Büring and Hartmann (1994)) which find parallels in Urdu. Constituents which are neither focus nor topic, but represent known backgrounded information, must appear in this postverbal position.

2.5 Completive Information

At first glance, it appears that some background information is found preverbally. That is, not all preverbal information is topicalized or focused in preverbal position, although in naturally occurring speech this is usually the case. Hoffman (1995) suggests for Turkish that there are subtle informational structural differences which give rise to distinctions in interpretation between preverbal and postverbal "backgrounded" information. In particular, postverbal backgrounded material is required to be referential, i.e., [−New], as (10) shows within the context of (9).

(9) Naadyaa kahãã=se aa rah-ii hai
Nadya.F.Nom where=from come Stat-F.Sg be.Pres.3.Sg
'Where is Nadya coming from?' (Urdu)

[8] Note that in examples (7) the postverbal material is NOT preceded by an intonational break. While such examples do exist, of course, such sentences must be considered to be true *afterthought* instances, or cases in which the speaker is proceeding to *repair* some part of the previous utterance. We are not considering examples of this sort.

(10) a. naadyaa to abhii tofii [bazaar=me]$_F$ xarid
Nadya.F.Nom indeed just now toffee.F.Nom market.M=in buy

rah-ii t^h-ii
Stat-F.Sg was-F.Sg
'Nadya was just buying toffee AT THE MARKET.' (Urdu)

b. # naadyaa to abhii [bazaar=me]$_F$ xarid rah-ii t^{h}ii [tofii]$_{Back}$

The sentences in (10) are possible answers to (9) in that they provide the hearer with information as to where Nadya was, namely at the market. Since *bazaar=me* 'at the market' provides the information which answers the question, it is focused. However, consider what happens with the object, *tofii* 'toffee'. In (10a) the object appears preverbally, but is neither topicalized or focused, and the sentence is an appropriate answer. In contrast, the utterance in (10b) presumes the toffee to be a familiar entity, as indicated by its postverbal position, and as a result is infelicitous as expected given the requirement that postverbal material be [−New, −Prominent].

While a more detailed investigation of these differing shades of meanings are material for future research, we will tentatively assume that these preverbal constituents have the interpretation of being [+New] but [−Prominent]. Since we are primarily concerned with the information status of these preverbal phrases as nontopics and nonpreverbal foci, we will refer to them as COMPLETIVE INFORMATION. Unlike topics, contrastive foci, and backgrounded material, these constituents are not licensed in a particular phrase structure position. Rather, they are simply generated in S, and their nterpretation can be considered to be a default.

To summarize, in languages like Urdu and Turkish discourse functions are encoded syntactically; in order to receive a particular discourse function interpretation, a constituent must be generated in the appropriate position; otherwise the structure is illicit. We posit four types of discourse functions which are relevant for the discussion at hand: Topics appear sentence initially, foci immediately before the verb, and backgrounded material is postverbal. Material which is generated within S, but outside of the VP (see (**??**)) is interpreted as completive information.

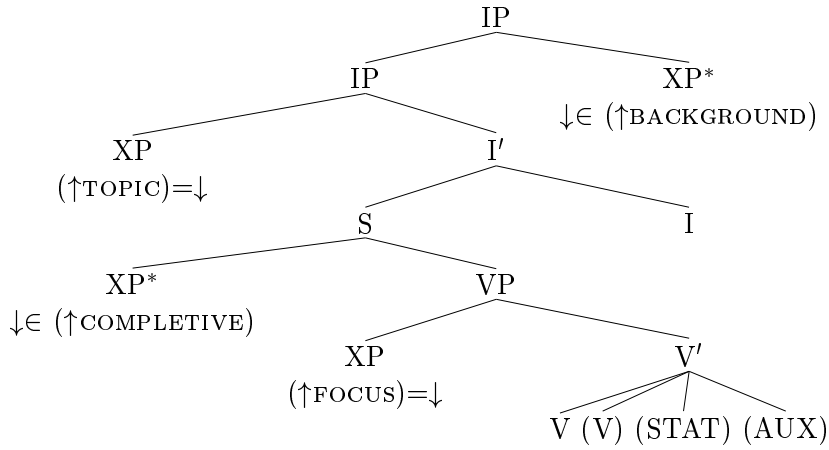
3 Encoding Discourse Functions

3.1 C-structure

The correlation between word order and discourse function interpretation supports a structure like (11) for Urdu and Turkish.[9] Certain phrase structure positions are associated with particular discourse functions via functional uncertainty. In general, the phrase structure rules follow an X′ schema. Although all nodes are optional, the rules are constrained so that each node must be overtly realized in at least one construction in the language.

[9]The structure of V′ is as proposed in Butt (1995) for Urdu. We make no claims about the internal structure of the verbal complex in Turkish.

(11)



The association of particular phrase structure positions with discourse functions captures the intuition that word order reflects the discourse functions of constituents. These syntactic positions act as licensers: in order for a constituent to be interpreted as having a particular discourse function it must appear in the appropriate position. In some cases, only a single constituent can be licensed in a given position (topic and focus). These are represented as Specifier positions. Specifier positions are syntactic function positions: SpecIP for topic and SpecVP for focus.[10] Note that discourse functions can be associated both with functional projections (IP) and lexical projections (VP). In others cases, the positions are associated with less prominent discourse functions. These positions allow multiple constituents to bear the licensed discourse function (completive and background information). These are represented as flat structure adjunctions.[11] These multiple constituents form a discourse function field, and the internal ordering of the field is determined by pragmatic factors. Since the word order of a clause is derivative from the phrase structure, the motivated appearance of constituents into these positions results in the desired orderings, without resorting to 'stylistic' PF reorderings.

3.2 F-structure

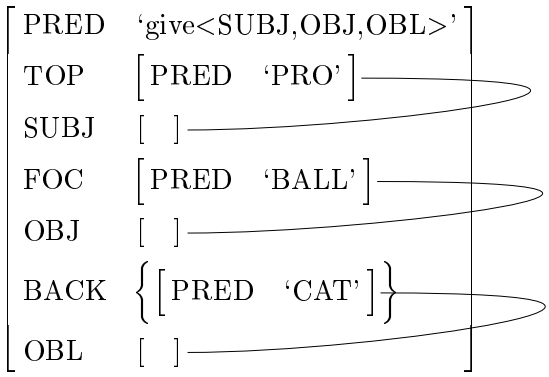
The distribution of grammatical functions in phrase structure is regulated by the interaction of functional uncertainty with well-formedness conditions on the functional-structure of the sentence. Much of this information interacts with the case marking of the arguments (see T. Mohanan (1994) and Butt (1995) for Hindi/Urdu and Neidle (1988) and King (1995) for Russian). We will not discuss the details of this interaction for Urdu and Turkish, except briefly for objects.

First consider the Turkish sentence in (12). It contains a preverbal focus, *Funda'nın topunu* 'Funda's ball'; the indirect object, *kediye* 'cat', is backgrounded. There is no overt topic; for purposes of explication we assume that the pro-dropped subject is the understood topic (see Hoffman (1995) on understood topics in Turkish).

(12) yok, [Funda'nın top-u-nu]$_F$ ver-me-m [kedi-ye]$_{Back}$
no, Funda-Gen ball-Poss3-Acc give-Neg-1Sg cat-Dat
'No, (I) won't give Funda's ball to the cat.' (Turkish)

[10] Here we leave aside the issue of whether all sentences have a topic and focus. It is generally assumed that all informative sentences have a focus, although perhaps not all utterances do. Whether all sentences have a topic is more controversial since there are sentences which are described as containing all new information, e.g., *Funda arrived, it is raining.* Hoffman (1995) argues for Turkish that in the discourse function structure all sentences have a topic, although this topic may not be overt in the sentence (see also Turan (1995)).

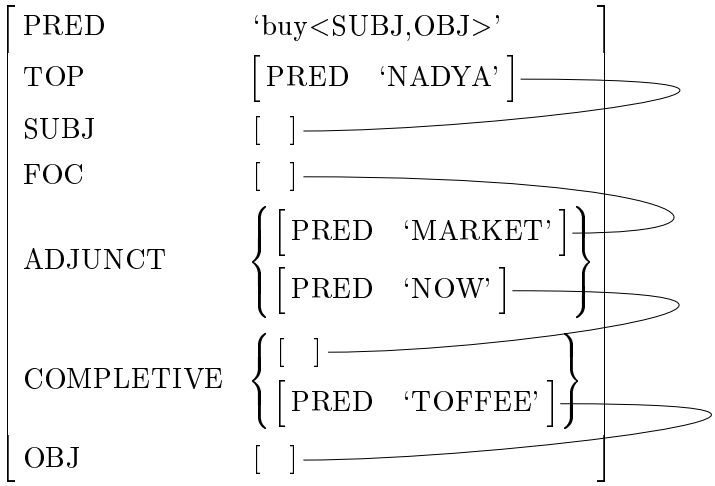
[11] We suggest that flat structures be restricted in distribution: to certain nodes, e.g., S, and to adjunction structures.

(13)


Next consider the Urdu sentence in (14) in which there is a topicalized subject, *naadyaa* 'Nadya', a focused adjunct, *bazaar=me* 'in the marker', and a completive information object, *tofii* 'toffee'.

(14) [naadyaa]$_T$ to [abhii]$_{CI}$ [tofii]$_{CI}$ [bazaar=me]$_F$ xarid
Nadya.F.Nom indeed just now toffee.F.Nom market.M=in buy

rah-ii t^h-ii
Stat-F.Sg was-F.Sg
'Nadya was just buying toffee AT THE MARKET.' (Urdu)

(15)


3.3 Where is Discourse Function Information?

One issue touched on here is where discourse function information belongs in LFG. The fact that certain phrase structure positions are associated with discourse functions requires phrase structure to make reference to discourse functions. For example, the annotation on SpecIP states that that constituent is the topic of the clause. Where is this information realized?

We assume that any theory, including LFG, will need a discourse function structure/representation separate from the syntax. In addition, in some languages topic and focus interact with the (morpho)syntax directly, not just in terms of c-structure, but also in terms of f-structure. As such topic and focus information has been placed in the f-structure in LFG. However, an alternative scenario is also possible.

Although the presence of topic and focus information in the f-structure causes no harm, in many languages, including Urdu and Turkish (see also King (1995) on Russian) is never used there, and it seems at best redundant to place it there. Given the argument that discourse information is instead directly relevant to s(emantic)-structure, but not necessarily to f-structure, the discourse information could be projected either directly from the c-structure to

an s-structure, or could be projected onto a another level of representation (e.g., d(iscourse)-structure)) altogether.[12] We remain agnostic on the various possibilities within the scope of this paper and represent discourse information at the level of f-structure in the interests of simplicity.

4 Object Interpretation and Focus

The picture with regard to word order and discourse information is, of course, more complex than can be done complete justice to within the analysis presented above. However, we hope to have laid a strong enough foundation to allow the analysis of further discourse and word order related phenomena. In the remainder of the paper, we thus examine a particularly intriguing interaction between object interpretation and focus. Recall that the focus position was seen to be immediately preverbal. We show below that nonspecific objects can also only appear immediately preverbally. We argue that both focus and nonspecific objects are in fact licensed in the same position, but that the interaction between focus and object realization necessitates a clearer understanding of the interaction between object position and object interpretation.

4.1 The Interpretation of Objects

We follow Enç (1991) in distinguishing between indefinite NPs that are specific, and those that are non-specific. Definite NPs are taken to be always specific since they refer to known discourse entities. Our assumption as to the distinction between specific and nonspecific objects can most concisely been stated as in Eskenazi (1996:3).[13]

> ... *nonspecific* nominals include only those indefinites that are neither referential nor have wide scope; *specific* nominals are simply the complement of this set, which includes nominals that either refer to or presuppose a familiar entity/entities.

In Urdu, as in Turkish (Enç (1991)), specificity is encoded by case marking through the accusative case (Butt (1995)). Furthermore, in Urdu/Hindi a "non-canonical" positioning of the object also forces a specific interpretation (T. Mohanan (1992)). Within structural approaches this correlation between object position, case-marking and interpretation has been analyzed in terms of a distinction between WEAK and STRONG Case positions within the VP (de Hoop (1992)). Within an LFG approach, such a structural distinction is not necessarily motivated. However, work on a variety of languages from Scottish Gaelic (Ramchand (forthcoming)) to West Greenlandic (van Geenhoven (1996)) has shown that two types of object types must be distinguished. We show that this must be true of Urdu and Turkish as well and distinguish between objects that are required to appear immediately preverbally (nonspecific objects), and those that may appear anywhere else in the clause (specific objects). Paralleling the strong vs. weak Case of de Hoop (1992), we further propose that these two interpretations reflect different grammatical functions, OBJ and OBJ_θ, respectively.

4.2 Case Marking and Position

4.2.1 Urdu

In Urdu, accusative objects are always specific, while nominative objects can be either specific or nonspecific. The context in (16) requires a nonspecific reading of *murɣii* 'chicken' in (17). The

[12] Another, broader, issue concerns how discourse as a whole is to be represented. One possibility would be to have it as a part of the semantics (or perhaps what is usually thought of as semantics and pragmatics is a part of the larger representation of the discourse). Since the continuity of the discourse must be represented, this structure should provide information about whether a constituent is a viable topic or focus. These possibilities are not mutually exclusive. There could be separate semantic and discourse modules or representations, but certain information would be present in both.

[13] The definition of specific/nonspecific as assumed here differs from that used by Enç (1991), who equates partitivity with specificity (see Farkas (1995) for a discussion of differing notions of specificity).

nominative in (17a) is felicitous in this context, while the accusative in (17b), which is specific, is not.[14]

(16) adnaan aaj raat=kii salen ke-liye muryii cah rah-aa
Adnan.M.Nom today night.F=Gen.F curry for chicken.F.Nom want Stat-M.Sg
t^{h}-aa
was-M.Sg
'Adnan wanted chicken$_{NS}$ for tonight's curry.' (Urdu)

(17) a. us=ke xansaame=ne bazaar=se [muryii] xarid-ii
pro=Gen.Obl cook.M.Obl=Erg market.M=from chicken.F.Nom buy-Perf.F.Sg
'His cook bought a chicken$_{NS}$ from the market.' (Urdu)

b. #us=ke xansaame=ne bazaar=se [muryii=ko] xarid-aa
pro=Gen.Obl cook.M.Obl=Erg market.M=from chicken.F=Acc buy-Perf.M.Sg
'His cook bought a particular/the chicken$_{S}$ from the market.' (Urdu)

Nominative objects in immediately preverbal position can be specific or nonspecific, as in (18a), while nominative objects elsewhere in the clause can only be specific, as in (18b/c).

(18) a. naadyaa=ne hassan=ko [xat] di-yaa
Nadya.F=Erg Hassan.M=Dat letter.M.Nom give-Perf.M.Sg
'Nadya gave Hassan a (particular) letter$_{S/NS}$.' (Urdu)

b. naadyaa=ne [xat] hassan=ko di-yaa
Nadya.F=Erg letter.M.Nom Hassan.M=Dat give-Perf.M.Sg
'Nadya gave Hassan a particular letter$_{S}$.' (Urdu)

c. naadyaa=ne hassan=ko [xat] jaldii=se di-yaa
Nadya.F=Erg Hassan.M=Dat letter.M.Nom quickness=Inst give-Perf.M.Sg
'Nadya gave Hassan a particular letter$_{S}$ quickly.' (Urdu)

4.2.2 Turkish

The Turkish data is similar, but not identical to that in Urdu. In Turkish, nominative NP objects may not generally appear in a position other than the immediately preverbal one, despite the otherwise relatively free word order, as shown in (19).

(19) a. Nadya Hasan'a çabucak [mektup] yaz-dı.
Nadya.Nom Hasan.Dat quickly letter.Nom write-Past.3Sg
'Nadya quickly wrote (a) letter(s) to Hasan.' (Turkish)

b. *Nadya Hasan'a [mektup] çabucak yaz-dı.

c. Nadya Hasan'a [mektub-u] çabucak yaz-dı.
Nadya.Nom Hasan.Dat letter-Acc quickly write-Past.3Sg
'Nadya quickly wrote the letter to Hasan.' (Turkish)

However, when there is a preverbal focus, certain nominative NPs can appear postverbally felicitously, as in (20) (from Göksel (1995)), while others cannot, as in (21).

[14] The NS and S in the glosses indicate nonspecific and specific readings, respectively.

(20) a. dün [ben]$_F$ gör-düm [bir filim]$_{Back}$
yesterday I.Nom see-Past.1Sg one film.Nom
'Yesterday I saw one (particular) film$_S$.' (Turkish)

b. bu adamlardan [kaçı]$_F$ gör-müş dün [bir filim]$_{Back}$
these men-Abl how many see-Past yesterday a film.Nom
'Of these men, how many have seen a film?' (Turkish)

(21) ??dün [ben]$_F$ gör-düm [filim]$_{Back}$
yesterday I.Nom see-Past.1Sg film.Nom
'Yesterday I saw some film(s)$_{NS}$.' (Turkish)

The generalization for Turkish appears to be that truly bare NPs like *filim* do not lend themselves to a specific interpretation and hence only felicitously appear in immediately preverbal position. In contrast, modified nominative NPs like *bir filim* 'one film' can be interpreted as referential, and are therefore not tied exclusively to the immediately preverbal position. As in Urdu, there thus seems to be a three-way distinction among object NPs: accusative NPs that are unquestionably specific, immediately preverbal nominative NPs that are nonspecific, and nominative NPs that cannot be interpreted as nonspecific due to their nonpreverbal position.

The correlations between specificity, case-marking and position in the two languages are summarized in (22).

(22)

Object Case	**Position**	**Specific**	**nonspecific**	**Language**
Accusative	Anywhere	√	*	Turkish, Urdu
Nominative	Anywhere	√	*	Urdu
	Postverbal (+F)	√	*	Turkish
Nominative	Immed. Preverbal	√	√	Urdu
	Immed. Preverbal	*	√	Turkish

The facts with regard to the scrambling possibilities of nominative NPs in the two languages are similar enough to warrant a unifying analysis, especially considering the general parallelism as to the position of focus, topic, and backgrounded information in Turkish and Urdu (section **??**). However, these facts also lead to a puzzle: how can one and the same position — the immediately preverbal one — license *both* nonspecific objects and focused, specific NPs? In building up to our analysis of focus licensing in these languages, we first present an analysis of bare NPs and the interaction between specificity and case for Urdu and Turkish. The typology in (22) suggests that finer-grained semantic distinctions than have been made so far need to be formulated. An approach addressing exactly the kind of issues raised by Urdu and Turkish nominative NPs is presented by van Geenhoven (1996), who proposes a theory of SEMANTIC INCORPORATION.

4.3 Strong/Weak Case and Semantic Incorporation

De Hoop (1992) sees Turkish as an instance of a more general crosslinguistic pattern and proposes that there are two types of object case assigned in the VP: STRONG case, which gives rise to specific interpretations, and WEAK case, which is associated with nonspecific interpretations. De Hoop's proposal comes with the syntactic assumption that strong case is assigned in a derived position; it also provides a semantic interpretation in terms of type-shifting. We do not follow the particulars of her approach here, but do base our analysis on her fundamental insight that there are two distinct object positions which give rise to differing object interpretations.

Van Geenhoven (1996) lists a number of strikingly similar semantic properties that characterize both West Greenlandic noun incorporation and West Germanic bare plurals: both always

take narrow scope with respect to negation, neither can be used to pick up a salient referent that was established in previous discourse, and neither can be used partitively. These nouns are still discourse transparent, as is wellknown from the literature on noun incorporation. However, when the incorporated nouns or bare plurals are embedded under negation, they become opaque for discourse purposes, indicating that these constructions do not introduce a referent of their own. Thus, van Geenhoven (1996:137) assumes that semantically incorporating predicates absorb one (or more) predicates as restrictions on the variable representing its internal argument, as in (23a). Nonincorporating predicates are represented as ordinary n-place predicates, as in (23b). Note that in the incorporating predicate in (23a), the predicate introduces an existential quantifier for its internal object, so that its argument is already satisfied. In (23b), on the other hand, the object is treated as a variable which must be bound at some point in the derivation.[15]

(23) a. Incorporating: $\lambda P \lambda x \exists y$ [eat(x,y) $\wedge$ P(y)]

b. Nonincorporating: $\lambda y \lambda x$ [eat(x,y)]

In addition to this semantic distinction, van Geenhoven argues for a three-way syntactic distinction between types of NPs. NPs corresponding to arguments of nonincorporating predicates are essentially strong case NPs which are not restricted in position. In contrast, nonspecific arguments of incorporating predicates can either be in a position adjacent to the verb or in a syntactically incorporated structure; in either case they are assigned weak case and cannot appear in any other position. The difference between the two is taken to be motivated by language-dependent factors: West Germanic, for example, requires syntactic adjacency for semantic incorporation. In sum, a three-way syntactic distinction which corresponds to a two-way semantic distinction, as in (24).

(24)

Semantics	**Case**	**Syntactic Position**
incorporated	weak	adjacent to verb morphologically incorporated
nonincorporated	strong	free

The fact that objects in the weak case position are assigned a predicative, and hence nonspecific, interpretation is exactly in line with the Urdu and Turkish data. In fact, this way of interpreting bare or incorporated objects is reminiscent of the type of constructions that have been discussed by T. Mohanan (1995) for Hindi from the point of view of noun incorporation; an example is the Urdu 'horse-selling' illustrated in (25).

(25) anil [g^{h}oṛe bec-taa hai]
Anil.M=Nom horse.M.Pl sell-Impf.M.Sg be.Pres
'Anil is selling (some) horses.'
'Anil does horse-selling.' (Urdu/Hindi)

T. Mohanan (1995) provides independent evidence that in Hindi the semantic difference in interpretation between 'did horse-selling' vs. 'sold horses' corresponds to differences in syntactic and phonological behaviour. Rather than adopting her proposal that these structures are the syntactic equivalent of West Greenlandic type of noun incorporation, we adopt van Geenhoven's

[15]While this approach has the consequence that most verbs in the lexicon will be listed as both incorporating and nonincorporating, i.e., *eat apples* vs. *eat the apple*, it allows for a clear and intuitive account of the various semantic properties observed in conjunction with noun incorporation and bare plurals (narrow scope, nonpartitivity, discourse transparency, modification of the nouns, scope effects (split topicalization in German), interaction with quantification and the resulting intermediate readings). Furthermore, it mirrors the intuition behind Carlson's (1977) original idea that it is the verbal predicates themselves which give rise to the existential interpretation of the English bare plural.

view: the distinction between semantically incorporating and nonincorporating predicates in languages like Hindi/Urdu and Turkish, in which the requirement for a predicative incorporating interpretation is syntactic adjacency, can be accounted for by the syntactic and semantic distinction between two differing types of objects.

Within an LFG analysis, we propose that the differing object types are encoded in terms of grammatical functions. We take the semantically restricted OBJ$_\theta$ to be the counterpart of the semantically richer strong Case, and the OBJ to correspond to the weak Case. This involves a shift in perspective within LFG's linking theory, but we would like to argue that this shift is not in fact a particularly radical one.

Linking from argument structure to grammatical functions is mediated via LFG's Lexical Mapping Theory (LMT). Grammatical functions are taken to be decomposable into the features [± r(estricted)] and [± o(bject)]. LMT provides a systematic assignment of these features to thematic roles, thus establishing the linking between argument structure and grammatical functions. Within this theory, the OBJ is generally taken to be semantically unrestricted, while the OBJ$_\theta$ is taken to be semantically restricted (e.g. Bresnan and Moshi (1990)). That is, OBJ was intended to represent the more general case, and OBJ$_\theta$ a more restricted scenario. However, in light of the semantic view on object interpretation presented above, we propose that in the more general case is exactly one in which the object carries more semantic information, rather than less. Thus, rather than seeing the OBJ$_\theta$ as being semantically *restricted*, we propose to view it as semantically *enriched*, while the OBJ is only associated with a nonspecific semantic interpretation. This shift in perspective on the encoding of objects also involves a slight reworking of passivization and other object related syntactic phenomena, however, as shown in Butt and King (1996), the actual revisions needed are negligible and entirely in keeping with the overall framework.

Thus, strong objects are associated with a particular semantics and thematic role, while the weak objects only satisfy an argument position of the predicate. This difference in grammatical function between OBJ and OBJ$_\theta$ can be exploited to capture the syntactic distribution of objects in Urdu and Turkish.[16]

(26)

	Position	**Role**	**Specific**	**Case**
Strong	Anywhere	OBJ$_\theta$	+	Acc/Nom
Weak	Adjacent to Verb	OBJ	−	Nom

The question that now remains to be investigated is how the restriction of nonspecific objects to preverbal position may be captured, and whether there is an interaction with the interpretation of preverbal focus.

4.4 Objects and Focus

Specific objects may be focused, as shown in (27b). When a constituent other than the object is in focus, as in (27a) and (27c), the object can only be interpreted as specific. In our terms, it must be analysed as an OBJ$_\theta$. Additionally, (27c) shows that topicalization is possible with specific objects. Crucially, in all of the examples in (27) the object cannot be interpreted as nonspecific, even when it is immediately preverbal, as in (27b).

(27) a. naadyaa=ne [xat] [hassan=ko]$_F$ di-yaa
Nadya.F=Erg letter.M.Nom Hassan.M=Dat give-Perf.M.Sg
'Nadya gave Hassan a particular letter$_S$.' (Urdu)

[16]There is little semantic difference between accusative and nominative specific objects in Urdu. Accusative *ko* marked objects are more generally compatible with explicitly telic constructions (e.g., aspectual complex predicates) and give rise to a more "affected" reading than nominative specific objects. Tentatively, we propose that the case marking is a reflex of the specific aspectual role assigned to the object. For example, *ko* marked objects are generally dispreferred with verbs like *write*, i.e., verbs of creation and consumption (T. Mohanan (1994:81)).

b. naadyaa=ne hassan=ko [xat]$_F$ di-yaa
Nadya.F=Erg Hassan.M=Dat letter.M.Nom give-Perf.M.Sg
'Nadya gave Hassan a particular letter$_S$.'

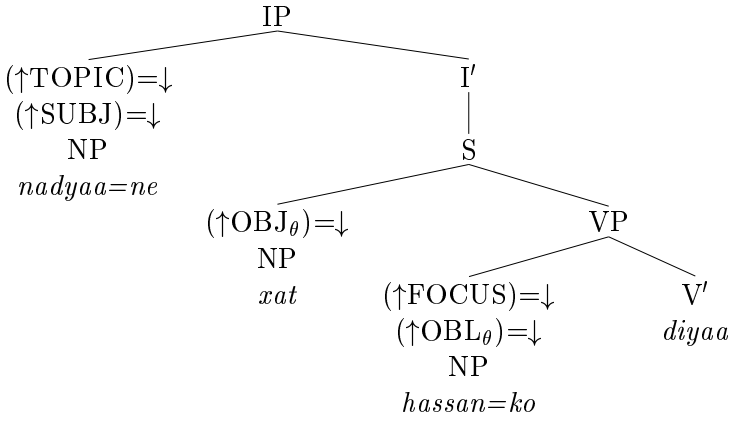
c. [xat]$_T$ naadyaa=ne [hassan=ko]$_F$ di-yaa
letter.M.Nom Nadya.F=Erg Hassan.M=Dat give-Perf.M.Sg
'The letter$_S$ Nadya gave Hassan.'

The fact that nonspecific objects and focus are mutually exclusive suggests that focus and OBJ are licensed in one and the same position, namely SpecVP (see Horvath 1995 on the similarity of focus and case licensing mechanisms). Material appearing in SpecVP is necessarily focused. In addition, the OBJ function can be assigned here, which due to its semantic impoverishment, is set apart from the other grammatical functions.

(28) VP $\longrightarrow$ XP V′
(↑FOCUS)=↓ ↑=↓
((↑OBJ)=↓)

As an illustration of the analysis, take (29), where the object is in the nominative. However, since it is not in SpecVP (the indirect object occupies this position), it cannot be assigned the OBJ function and hence can only be interpreted as specific, as seen in (30).

(29) a. naadyaa=ne [xat] [hassan=ko]$_F$ di-yaa
Nadya.F=Erg letter.M.Nom Hassan.M=Dat give-Perf.M.Sg
'Nadya gave Hassan a particular letter$_S$.'

b.


(30) a. # naadyaa=ne [xat] [hassan=ko]$_F$ di-yaa
Nadya.F=Erg letter.M.Nom Hassan.M=Dat give-Perf.M.Sg
'Nadya gave Hassan a letter$_{NS}$.'

b. NO WELL-FORMED C- AND F-STRUCTURE

Note that is possible to find focus on nonspecific objects, as in (31). However, here it is not the object itself which is focused, rather the entire predicate is focused. Since nonspecific objects are semantically incorporated, focusing one is similar to focusing the predicate itself.

(31) naadyaa har roz [g$^{\text{h}}$oṛe]$_F$ bec-tii hai
Nadya.F.Nom every day horse.M.Pl.Nom sell-Impf.M.Sg be.Pres
'Nadya [sells horses$_{NS}$] every day.'

Thus, the interaction between immediately preverbal focus and object interpretation can be accounted for by positing a distinction between two object types, and allowing focus to be licensed in the same position as nonspecific objects.

5 Conclusion

In sum, we have presented a a "discourse configurational" view of Urdu and Turkish (É. Kiss 1995), whereby differing word orders may be *optional* from a strictly syntactic point of view, but in fact are motivated by discourse factors such as topic, focus, and background information. Furthermore, we examined the interaction between focus and the distribution of objects in Urdu and Turkish. It was seen that discourse functions and grammatical functions interact, but that only one kind of object may be licensed in the same position as focus, thus leading to a pattern of mutual exclusion between focus and nonspecific objects.

References

Bresnan, J. 1996. Lexical-functional syntax. Ms., Stanford University.

Bresnan, J., and L. Moshi. 1990. Object asymmetries in comparative Bantu syntax. *Linguistic Inquiry* 21:147–186.

Büring, D., and K. Hartmann. 1994. Doing the right thing – extraposition as a movement rule. Arbeitspapier 13, Universität Frankfurt.

Butt, M. 1995. *The Structure of Complex Predicates in Urdu.* Stanford, California: CSLI Publications.

Butt, M., and T. H. King. 1996. Distribution of objects in Urdu. Presented at the Stanford University Linguistics Colloquium.

Carlson, G. 1977. *Reference to Kinds in English.* PhD thesis, Univeristy of Massachusetts, Amherst.

Choi, H.-W. 1996. *Scrambling: Optimality-theoretic Interaction between Syntax and Discourse.* PhD thesis, Stanford University.

de Hoop, H. 1992. *Case Configuration and Noun Phrase Interpretation.* PhD thesis, Rijksuniversiteit Groningen.

Dik, S., e. 1981. On the typology of focus phenomenon. In T. Hoekstra, H. van der Hulst, and M. Moortgat (Eds.), *Perspectives on functional grammar*, 41–74. Dordrecht: Foris.

Dwivedi, V. 1994. *Syntactic Dependencies and Relative Phrases in Hindi.* PhD thesis, University of Massachusetts at Amherst.

É. Kiss, K. (Ed.). 1994. *Discourse Configurational Languages.* Oxford: Oxford University Press.

Enç, M. 1991. The semantics of specificity. *Linguistic Inquiry* 22(1):1–25.

Erguvanlı, E. 1984. *The Function of Word Order in Turkish Grammar.* University of California Press.

Eskenazi, J. 1996. Bare objects and the NP/DP distinction. Unpublished Ms., University of California, Santa Cruz.

Farkas, D. 1995. Specificity and scope. In *LRC Working Papers.* To appear in Proceedings of Langues et Grammaire 1.

Gambhir, V. 1981. *Syntactic Restrictions and Discourse Functions of Word Order in Standard Hindi.* PhD thesis, University of Pennsylvania, Philadelphia.

Göksel, A. 1995. Linearity, focus, and the preverbal position in Turkish. Unpublished manuscript, SOAS.

Heim, I. 1982. *The semantics of definite and indefinite noun phrases.* PhD thesis, University of Massachusetts, Amherst.

Hoffman, B. 1995. *The Computational Analysis of the Syntax and Interpretation of "Free" Word Order in Turkish.* PhD thesis, University of Pennsylvania, Philadelphia.

Horvath, J. 1995. Structural focus, structural case, and the notion of feature-assignment. In K. É. Kiss (Ed.), *Discourse Configurational Languages*, 28–64. Oxford: Oxford University Press.

King, T. H. 1995. *Configuring Topic and Focus in Russian.* Stanford, California: CSLI Publications.

Kornfilt, J. 1995. On focusing and de-focusing in Turkish. Paper presented at the Generative Grammatik im Süden (GGS), Universität Jena.

MacWhinney, B., and C. Snow. 1985. The child language data exchange system. *Journal of Child Language* 12:271–296.

Mahajan, A. 1990. *The A/A-Bar Distinction and Movement Theory.* PhD thesis, MIT.

Mohanan, T. 1992. Word order in Hindi. Presented at Syntax Workshop, Stanford University.

Mohanan, T. 1994. *Argument Structure in Hindi.* Stanford, CA: CSLI Publications.

Mohanan, T. 1995. Wordhood and lexicality: Noun incorporation in Hindi. *Natural Language and Linguistic Theory (NLLT)* 13(1).

Müller, G. 1995. Cyclicity and economy: Movement theory in the Minimalist Program. Course Reader, LOT Winterschool, Tilburg University.

Neidle, C. 1988. *The Role of Case in Russian Syntax.* Kluwer.

Ramchand, G. forthcoming. *Aspect and Predication: The Semantics of Argument Structure.* Oxford: Oxford University Press.

Rochemont, M. 1986. *Focus in Generative Grammar.* Amsterdam: John Benjamins.

Rochemont, M., and P. Culicover. 1990. *English Focus Constructions and the Theory of Grammar.* Cambridge: Cambrige University Press.

Srivastav, V. 1991. *WH Dependencies in Hindi and the Theory of Grammar.* PhD thesis, Cornell University.

Turan, U. 1995. *Overt vs. Null Subjects in Turkish Discourse.* PhD thesis, University of Pennsylvania, Philadelphia.

Vallduví, E. 1992. *The Informational Component.* New York: Garland Press.

Vallduví, E. 1993. Information packaging: A survey. Research Paper HCRC/RP-44.

van Geenhoven, V. 1996. *Semantic Incorporation and Indefinite Descriptions.* PhD thesis, University of Tübingen.

Please send comments to:
mutt@ims.uni-stuttgart.de
thking@csli.stanford.edu