

# Null Elements in Discourse Structure<sup>0</sup>

Miriam Butt and Tracy Holloway King

## Abstract

In recent work on discourse structure, several proposals have been made as to the *discourse configurationality* (É. Kiss 1995) of languages characterized by a freedom in word order. Parallel to this, Vallduví's (1992) idea of *information structure*, in which differing possible word orders correspond to differing possible ways of packaging information, allows a finer grained, syntactically based, identification of theme-rheme/topic-focus. A question that has not been addressed at any length within these newer approaches to discourse has been the role of null elements. Within our framework the generalization as to which noun phrases may be realized as null in discourse emerges to be that only old information is dropped: in particular, only continuing topics (Yokoyama 1986) and the background information that is related to a continuing topic may be dropped.

## 1 Introduction

Two well known facts about Hindi/Urdu<sup>1</sup> are the relative “free word order” and the ability to optionally “drop” every argument of a clause, as illustrated in (1).

- (1) a. tum=ne yasiin=ko vo aam de di-yaa?  
 you=ERG Yassin.M=DAT that mango.M.NOM give give-PF.MSG  
 ‘Did you give Yassin that mango?’
- b. jii, aam de di-yaa  
 yes.RESP mango.M.NOM give give-PF.MSG  
 ‘Yes, (I) gave the mango (to Yassin).’
- c. jii, de di-yaa  
 yes.RESP give give-PF.MSG  
 ‘Yes, (I) gave (the mango to Yassin).’

---

<sup>0</sup>We would like to thank the members of the audience of the Seminar on Nulls for interesting questions and criticisms and Christine Kaschny for helping with the formatting and the proofreading of the paper.

<sup>1</sup>Modern colloquial Hindi and Urdu differ significantly in terms of vocabulary in that Hindi draws heavily on Sanskrit while Urdu draws on Persian. However, barring dialectal variation, the languages cannot be said to exhibit significant structural differences (i.e., in terms of the morphology, syntax, semantics). As such, the two languages are referred to in this paper as Hindi/Urdu. Most of the data presented here are drawn from Hindi linguistic literature and movies; a portion of the data are drawn from the modern colloquial Urdu spoken in Lahore.

Each of these properties has been studied in its own right, although scrambling has received a great deal more attention than the question of argument-drop: in the linguistic literature specific to Hindi/Urdu a substantial body of literature on scrambling exists; however, next to nothing has been done with regard to *pro-drop*.

One well known direction of research on scrambling in Hindi/Urdu has concentrated on finding purely syntactically motivated explanations for the differing word order possibilities (see Mahajan 1990). However, several researchers have suggested that word order in Hindi/Urdu can only be fully understood through an acknowledgment and explication of the connection between word order and discourse functions (e.g., Verma 1970, Gambhir 1981, Kidwai 1997) and additional semantic factors such as referentiality (Dwivedi 1994).

This paper builds on the body of work already done for Hindi/Urdu with respect to word order and discourse, but seeks to go a little further by offering an analysis of argument drop which takes discourse structure into account.<sup>2</sup> In order to consider a possible treatment of these NULL elements, we first review some of the core work on word order and discourse functions in Hindi/Urdu, present our own view as realized within the framework of Lexical-Functional Grammar (LFG), then look at null elements in the context of discourses taken from several Hindi movies, and finally propose an analysis of null elements as instantiated by argument drop in terms of our proposals for information-structure.

## 2 Word Order and Discourse Functions in Urdu

In some early work on Hindi word order, Verma (1970) proposes that the position of a given constituent in a clause directly reflects its function in discourse. In his view, the Hindi clause has three components and three possible placement positions:

---

<sup>2</sup>The basic data for Hindi/Urdu presented here also confirm data from Turkish as to the relationship between word order and discourse functions (Hoffman 1995), indicating that the emergent pattern has crosslinguistic application.

- (2)
- INITIAL PLACE: Theme, Topic
  - MEDIAL PLACE: Emphasis
  - FINAL PLACE: Neutral

Gambhir (1981), on the other hand, argues for a more complex account of Hindi word order. In a very detailed and systematic investigation, Gambhir takes into account the discourse context in which a given sentence is uttered, and posits that the primary purpose of the variation in word order is to express the (pragmatic) discourse effects of CONTRAST, EMPHASIS, DEEMPHASIS, AFTERTHOUGHT, and TOPICALIZATION.

The unmarked word order is taken to be S IO O ADV V (as it is indeed taken to be by most researchers on Hindi/Urdu, with Mahajan (1990) as a notable exception). A simplistic one-to-one correlation between position and function is rejected. Rather, Gambhir posits the correlations in (3).

- (3)
- TOPIC: Informs hearer what speaker is talking about. Given or old information is shared by speaker and hearer. Topicalization is taken to be distinct from scrambling. The topicalized constituent is always moved out of and to the left of the S (evidence based on data from embedded clauses).
  - CONTRAST: Contrast between two or more members of a given set. Either sentence initial, postverbal or preverbal (the latter is what we refer to as focus in this paper). However, the notion is dependent on other notions (such as topic), and is therefore derivative.
  - EMPHASIS: Sentence initial or postverbal (this discourse function is dependent on further contextual factors and the notion topic).
  - DEEMPHASIS: Postverbal or deleted, depending on the degree of deemphasis.
  - AFTERTHOUGHT: Postverbal, characterized by intonational break.

Gambhir also takes a position on the effects of word order on truth conditions. In contrast to other strands of research such as that represented by Mahajan 1990, Gambhir believes that a change in word order does not effect a change in truth conditions. Rather, the only semantic effects observable must be analyzed as “narrow” semantic effects of *focus* or *contrast* (cf. Rooth 1985). This is also the basic position taken by Vallduví (1992), and the one adopted in this paper.

Dwivedi (1994) focuses on relative clauses in Hindi and proposes a set of generalizations relevant to word order. She distinguishes between three kinds of leftward movement to initial position, detailed in (4), and argues for a topic position in SpecTopP, which dominates CP.

- (4)
- LEFT DISLOCATION: The NP is base-generated in SpecTopP and is coindexed with a pronoun in the sentence (“resumptive topics”).
  - TOPIC DISLOCATION: Referential NPs in SpecTopP are coindexed with a *pro* in the sentence.
  - TOPICALIZATION: Any phrase in SpecTopP, coindexed with a phrase in the sentence.

We follow King’s (1995) analysis of Russian in analyzing the material in Dwivedi’s SpecTopP as an instance of “external topics” (Aissen 1992, Banfield 1973) and leave the issue at that. In this paper, we focus on the problem of discourse functions “internal” to the sentence, not ones that are extracted or generated out of it.

Kidwai (1997) independently proposes many of the same generalizations with respect to word order and discourse presented here, and her observations and semantic work with regard to focus and topic are much more detailed. However, her approach differs from ours in that differing word orders are assumed to be a combination of base-generated adjunction and XP-adjunction driven by the checking of features such as [+TOPIC] and [+FOCUS]. Within her minimalistic framework, subject and object agreement occur via Spec-head relations in AgrP both VP-internally and VP-externally. These assumptions as to UG syntax lead to further assumptions about Hindi/Urdu syntax that we do not share. It should nevertheless be emphasized that we do share a discourse based view of Hindi/Urdu word order with Kidwai, that many of our assumptions are entirely compatible with Kidwai’s approach, and that our respective proposals were developed independently of one another.

### 3 Basic Data

We posit four distinct discourse functions: TOPIC, FOCUS, BACKGROUND and COMPLETIVE INFORMATION. These are formalized in section 4.

### 3.1 Topic

Topics occur in clause initial position in matrix clauses, as in (5a), and in second position in clauses with complementizers, as in (5b).

- (5) a. [hassan=ko]<sub>T</sub> naadyaa=ne Toffi d-ii  
 Hassan.M=DAT Nadya.F=ERG toffee.F.NOM give-PF.FSG  
 ‘To Hassan Nadya gave toffee.’
- b. [anjum=ne]<sub>T</sub> dek<sup>h</sup>-aa [ki [hassan=ko]<sub>T</sub> naadyaa=ne  
 Anjum.F=ERG see-PF.MSG that Hassan.M=DAT Nadya.F=ERG  
 Toffi d-ii]  
 toffee.F.NOM give-PF.FSG  
 ‘Anjum saw that to Hassan Nadya gave toffee.’

We propose that topics be situated in SpecIP. This entails that not all clause initial elements are necessarily in topic position, as they could have been generated at the same level as the other internal arguments. In fact, sentences as in (6) confirm this view. Here *Nadya* is ambiguous between a topic and a non-topic reading.

- (6) [naadyaa=ne]<sub>(T)</sub> hassan=ko Toffi d-ii  
 Nadya.F=ERG Hassan.M=DAT toffee.F.NOM give-PF.FSG  
 ‘Nadya gave toffee to Hassan.’

In the non-topic reading (6) either has no topic at all, or has a non-overt (continuing) topic from a previous utterance.

### 3.2 Focus

If there is only one focused constituent in the sentence,<sup>3</sup> then it must appear in the immediately preverbal position (see Kidwai 1996, 1997 for more detail on the phonology associated with preverbal focus).<sup>4</sup>

<sup>3</sup>In a complete analysis of discourse functions and information structure, the discourse function of the verb would also have to be taken into account. This area of research is unfortunately often neglected. See King 1997 for discussion of the technical difficulties posed by the inclusion of verbs in i(nformation)-structure.

<sup>4</sup>In a number of languages, wh-phrases are also found in focus position (see É. Kiss 1995 and articles therein for examples and discussion). In Hindi/Urdu the immediately preverbal position is also the preferred position for wh-words, patterning with Turkish in this respect (Kornfilt 1995).

- (7) a. naadyaa=ne hassan=ko [Tofii]<sub>F</sub> d-ii  
 Nadya.F=ERG Hassan.M=DAT toffee.F.NOM give-PF.FSG  
 ‘Nadya gave TOFFEE to Hassan.’
- b. #naadyaa=ne [hassan=ko]<sub>F</sub> Tofii d-ii  
 Nadya.F=ERG Hassan.M=DAT toffee-F.NOM give-PF.FSG  
 ‘Nadya gave toffee to HASSAN.’

In addition to the preverbal focus position, *in situ* focusing of a phrase is possible with multiple foci. This *in situ* focus is a case of *contrastive* focus, as illustrated in (8), and is often also expressed by means of the focus particles *hii* or *b<sup>h</sup>ii* (see Kidwai 1996, 1997 for a discussion of the interaction of *hii* with focus). In (8), the focus on *Hassan* is only permissible in a context in which Hassan is contrasted with another possible recipient. We do not consider *in situ* contrastive focus in this paper.

- (8) (aadnaan=ke-liiye nahī), naadyaa=ne [hassan=ke-liiye]<sub>CF</sub> [Tofii]<sub>F</sub>  
 Adnan.M=FOR not Nadya.F=ERG Hassan.M=FOR toffee.F.NOM  
 xariid-ii  
 buy-PF.FSG  
 ‘Nadya bought TOFFEE for HASSAN (not for Adnan).’

### 3.3 Background Information

The interpretation of backgrounded information is akin to topicalized information in that both have the status of “old” or “known” information. The difference between the two from a discourse structure point of view can be described as follows: while topics are the pointer to the relevant information (file card in the Heimian 1982 metaphor) to be accessed by the hearer, backgrounded material only provides more detailed information as to how the new information fits in with the already known information. That is, the backgrounded material provides the information that may be necessary for a good understanding of the new (focused) information supplied (also see Hoffman 1995:141–143 for some data on the variable salience of backgrounded information).

Backgrounded phrases in Hindi/Urdu occur postverbally,<sup>5</sup> as shown in (9).<sup>6</sup>

- (9) to aapne aap=se ek vadaa kar lii-yaa t<sup>h</sup>-aa  
 so self self=from one vow make take-PF.MSG be-PST.MSG  
 [mãĩ=ne]<sub>Back</sub>  
 I=ERG  
 ‘So I made a vow to myself.’ (Dilwale Dulhania Le Jayenge)
- (10) mer-ii acc<sup>h</sup>-ii=val-ii bacc<sup>h</sup>-ii hai [tu]<sub>Back</sub>  
 my-FSG nice-FSG=one-FSG child-FSG.NOM be.PRST.SG you.NOM  
 ‘You are my nice child.’ (Dilwale Dulhania Le Jayenge)

### 3.4 Completive Information

At first glance, it appears that some background information is found preverbally. That is, not all preverbal information is topicalized or focused. Hoffman (1995) suggests for Turkish that there are subtle informational structural differences which give rise to slight distinctions in interpretation between preverbal and postverbal backgrounded information. In particular, postverbal backgrounded material is required to be referential, as (12) shows within the context of (11).

- (11) naadyaa kahãã=se aa rah-ii hai  
 Nadya.F.NOM where=from come Stat-FSG be.PRST.SG  
 ‘Where is Nadya coming from?’
- (12) a. naadyaa to abhii [Tofii]<sub>CI</sub> [bazaar=se]<sub>F</sub> xariid  
 Nadya.F.NOM indeed just now toffee.F.NOM market.M=from buy

---

<sup>5</sup>However, as pointed out by Gambhir, the postverbal position cannot be exclusively identified with backgrounding. Consider the example in (i), taken from Gambhir 1981:318, which must be identified as a case of postverbal presentational highlighting.

i. sun-o ek t<sup>h</sup>-aa raajaa us=kaa naam  
 hear-IMP one be-PST.MSG king.M.NOM pron=M.GEN name.M.NOM  
 t<sup>h</sup>-aa jaysĩg<sup>h</sup>  
 be-PST.MSG Jaysingh  
 ‘Listen! There was a king. His name was Jaysingh.’

Again, we do not discuss this discourse strategy, but confine ourselves to first attempting to provide a theory of discourse structure for the rather basic and robust word order data displayed by Hindi/Urdu in conjunction with null elements.

<sup>6</sup>Note that in examples (9) and (10) the postverbal material is not preceded by an intonational break. While such examples do exist, they are true *afterthought* instances, or cases in which the speaker is *repairing* some part of the previous utterance.

rah-ii t<sup>h</sup>-ii  
 Stat-FSG be-PST.FSG  
 ‘Nadya was just buying toffee AT THE MARKET.’

- b. # naadyaa to abhii [bazaar=se]<sub>F</sub> xariid rah-ii t<sup>h</sup>-ii  
 Nadya.F.NOM indeed just now market.M=from buy Stat-FSG be-PST.FSG  
 [Tofii]<sub>Back</sub>  
 toffee.F.NOM  
 ‘Nadya was just buying toffee AT THE MARKET.’

The sentences in (12) are possible answers to (11) in that they provide the hearer with information as to where Nadya was, namely at the market. Since *bazaar=se* ‘from the market’ provides the information which answers the question, it is focused. However, consider what happens with the object, *Tofii*. In (12a) the object appears preverbally, but is neither topicalized or focused, and the sentence is an appropriate answer. In contrast, the utterance in (12b) presumes the toffee to be a familiar entity, as indicated by its postverbal position, hence one that can be referred to. This is infelicitous. While a more detailed investigation of these differing shades of meanings are material for future research, we tentatively assume that completive information must be interpreted as new, but not prominent. Unlike topics, foci and backgrounded material, completive information is not licensed in a particular phrase structure position. Instead, it is the result of arguments remaining *in situ*, and this interpretation can be considered to be a default.

## 4 Syntactic Mapping

We conclude from the above data that discourse functions in Urdu/Hindi are encoded syntactically; in order to receive a particular discourse function interpretation, a constituent must appear in the appropriate licensing position; otherwise the structure is illicit. Topics appear sentence initially, foci immediately before the verb, and backgrounded material is postverbal. In this section we present an analysis of the relationship between syntactic position and discourse structure in terms of LFG’s



projection architecture (Bresnan and Kaplan 1982) and then go on to consider the place of null elements within this analysis in the next section.

## 4.1 Basic Syntactic Structure

In the projection-based architecture of LFG, a grammar is encoded as several (mathematically defined) projections which represent mutually constraining but essentially independent levels of linguistic representation. The core levels of representation, or projections, in “classic” LFG have been c(onstituent)-structure, which encodes linear word order and constituency, and f(unctional)-structure, which primarily encodes predicate-argument relations in terms of grammatical functions (SUBJ, OBJ, OBL, etc.) and head-modifier relationships. In addition to these two, a s(emantic)-structure and an a(rgument)-structure form the core projections within most current LFG analyses. Here we also put forward the idea of an i(nformation)-structure which encodes the discourse functions of the sentence in context.

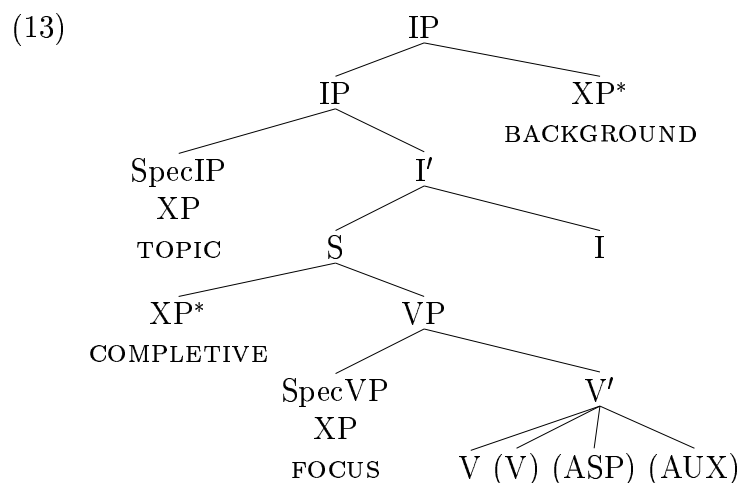
For the purposes of this paper, we assume the treatment of phrase structure presented by Bresnan (1997, 1998) and King (1995). There is a basic  $X'$  syntax with its specifier, head, and complement structure, as well as adjunction to maximal projections. Specifiers are filled either by traditional grammatical functions, e.g., subject or object, or by the prominent discourse functions, i.e., topic or focus. In addition, there is a lexocentric category  $S$  which behaves like a small clause in that it does not project according to the  $X'$  schema. This  $S$  category captures the non-configurational portion of the phrase structure, as will be seen below.<sup>7</sup> Positions which are not lexically realized in a given sentence are not projected in the c-structure.

For Urdu/Hindi we posit the structure in (13) based on the data in section 3 and the principles of phrase structure projection. Arguments are taken to be generated under  $S$  where they receive default discourse function interpretation, and the specifiers are associated with topic (SpecIP) and focus (SpecVP). Right-adjoined to the IP is

---

<sup>7</sup>We adopt a version of the VP-internal subject hypothesis (Fukui and Speas 1986) in that all arguments can be generated internally, in this case internal to  $S$ .

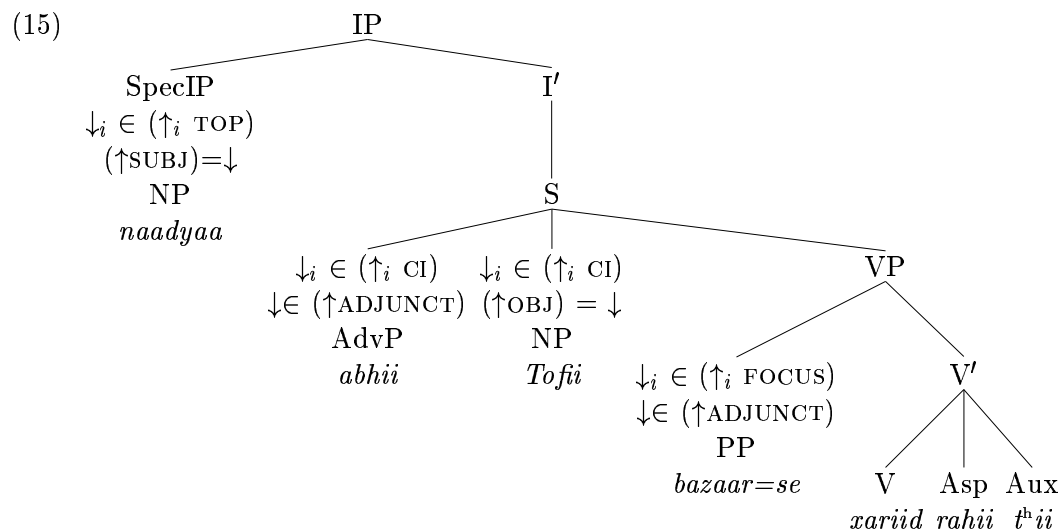
backgrounded information.<sup>8</sup>



While the syntactic positions are associated with discourse functions, they do not themselves serve as the representation of the discourse structure of an utterance. That is done by a separate level of representation we posit in terms of the projection architecture: the *i*(nformation)-structure (Kaplan 1987, Choi 1996, King 1997). Just as syntactic positions are associated with grammatical functions via FUNCTIONAL EQUATIONS, so are these syntactic positions associated with the appropriate discourse functions. This is illustrated in (15) for the example in (14). The *f*-structure and *i*-structure associated with (i.e., projected from) the *c*-structure tree in (15) are shown in (16). (Plain arrows indicate projection to the *f*-structure; arrows annotated with *i* indicate projection to the *i*-structure.)

- (14) [naadyaa]<sub>T</sub> (to) [abhii]<sub>CI</sub> [Tofii]<sub>CI</sub> [bazaar=se]<sub>F</sub> xariid  
 Nadya.F.NOM indeed just now toffee.F.NOM market.M=from buy  
 rah-ii t<sup>h</sup>-ii  
 Stat-FSG be-PST.FSG  
 ‘Nadya was just buying toffee at the market.’

<sup>8</sup>The I<sup>0</sup> position is filled by the copula, while auxiliaries appear under V'. See King 1995 on the (non-)overt realization of head positions.



(16) **Functional structure:**

PRED	'buy<SUBJ,OBJ> '
SUBJ	[PRED 'NADYA ']
ADJUNCT	{ [PRED 'MARKET '] [PRED 'NOW ']
OBJ	[PRED 'TOFFEE ']

**Information-structure:**

TOPIC	{ [PRED 'NADYA ']
FOCUS	{ [PRED 'MARKET ']
COMP.INF	{ [PRED 'NOW '] [PRED 'TOFFEE ']

The use of these functional equations in conjunction with functional uncertainty paths (Kaplan and Zaenen 1989) allows for a great deal of flexibility. The SpecIP position, for example, can be, but need not be, associated with a subject: given the appropriate formulation of a functional uncertainty path, an object or adjunct could just as well appear in this position.<sup>9</sup> In (15) we have not spelled out the functional

<sup>9</sup>Further note that the functional equations for discourse structure could in principle also range over functional uncertainties, and in fact, such a treatment will be needed for long distance dependencies.

uncertainty equations; rather, for the sake of simplicity, we have treated them as “resolved” and have simply provided the appropriate grammatical function.

Note that the f- and i-structures of the sentence are by no means isomorphic. For example, in the i-structure it is the completive information (‘now’ and ‘toffee’) which has been placed into a set, while in the f-structure the adjuncts ‘market’ and ‘now’ formed a set. As multiple topics (e.g., in Russian, see Yokoyama 1986, King 1995) and foci are in principle possible, we consider each of the discourse functions to have sets as a value. In (16) it so happens that the completive information (COMP.INF) is the only discourse function with more than one element in the set.

In many LFG treatments of discourse, the discourse functions are represented as part of f-structure (e.g., Bresnan 1998, King 1995) as there are some phenomena which necessitate the interaction of these two. However, because the phenomena sensitive to discourse do not exclusively (or indeed most heavily) interact with grammatical functions, but also interact with semantic interpretation and the realization of c-structure (as we show in our consideration of null elements in the next section), we believe that information structure should be accorded its own independent status within the theory (see Choi 1996 and King 1997 for a similar conclusion).

## 5 Null Elements

Having briefly presented the overall treatment of discourse structure which we assume, we now turn to an examination of the behavior of null elements in discourse and seek to explain the observed phenomena in terms of discourse structure.

### 5.1 Previous Work

While a substantial body of formal linguistic research exists on the question of *pro-drop* in formal linguistics, there is very little for Hindi/Urdu in particular. One clue to the behavior of null elements can be found in Dwivedi (1994). As part of her examination of relative clauses in Hindi, Dwivedi observes that only referential

NPs may be dropped. However, as not every referential NP is necessarily dropped, Dwivedi's observation does not go all the way towards explaining when null elements are used.

Our proposal is to take Dwivedi's observation and connect it to the notion of discourse structure. In terms of the four distinct discourse functions we have identified (topic, focus, background, completive information), only continuing topics and background represent discourse functions which could encompass referential NPs, i.e., NPs which do not introduce a new discourse entity. Our proposal is that only NPs which are functioning either as continuing topics and background information may be dropped.<sup>10</sup>

As shown below via an examination of discourses taken from Hindi movies, this rather simple proposal goes a long way in explaining the occurrence of null elements in discourses.

## 5.2 Data

### 5.2.1 Switch Topic (overt)

Arguments which function as a topic within their clause, but which simultaneously indicate a change (switch) in topic from the preceding utterance cannot be realized as null. This is particularly evident in the discourses in (17) and (18), where the topic is a first person pronoun, an entity that is considered to be particularly susceptible to being dropped.

- (17) a. [vo]<sub>T</sub>        zindegi aapn-ii    marzii=se        jiye-gii        ...  
           Pron.NOM life.FSG self-FSG choice.F=from live-FUT.FSG  
           'She<sub>Topic</sub> will live life according to her own wishes.'
- b. magar [māĩ]<sub>T</sub>    valat    t<sup>h</sup>-ii,        simrat  
           but    I.NOM wrong be-PST.FSG simrat  
           'But, I<sub>switch.topic</sub> was wrong, Simrat.' (Dilwale Dulhania Le Jayenge)

---

<sup>10</sup>Unlike canonical pro-drop, which is obligatory in that having an overt pronoun results in strong emphasis and focusing of that pronoun (Italian subjects are a typical example), Hindi/Urdu demonstrates argument-drop as a more powerful and flexible discourse governed phenomenon.

- (18) a. to [us=pe]<sub>T</sub> ek naaTak          lik<sup>h</sup>-te          hãĩ  
 so Pron=on on drama.M.NOM write-IMP.F.OBL be.PRST.PL  
 ‘So let us write a drama about that.’ (Silsilla)
- b. [mãĩ]<sub>T</sub> ek erfors=kaa          aadmii          hũ          ...  
 I.NOM one airforce-GEN.MSG man.M.NOM be.PRST.1.SG  
 ‘I<sub>switch.topic</sub> am an airforce man.’ (Silsilla)

The larger context for the utterance in (17) is a mother recounting her shattered hopes for bringing up her daughter as a human being free to choose how to live her life. The context for (18) is a conversation about the fact that trains are always late in India and how one should write about it.

### 5.2.2 Null Continuing Topic

As opposed to switch topics, continuing topics, i.e., entities which are the topic of the current utterance *and* of the previous utterance, can be dropped and in general do not occur overtly. This is illustrated in (19) with a first person pronoun.

- (19) a. [mãĩ] bais          barf=se          yahãã rah rah-aa          hũ  
 I.NOM twenty-two winter=from here live Stat-MSG be.PRST.1.SG  
 ‘I<sub>topic</sub> have been living here for 22 years.’
- b. rozaanaa is hii sarak=se          guzar-taa          hũ  
 daily this EMPH street.F=from pass-IMP.F.MSG be.PRST.1.SG  
 ‘Daily (I<sub>cont.topic</sub>) go through this street.’ (Dilwale Dulhania Le Jayenge)

In fact, most of the instances of continuing topic drop involve exactly the first person pronoun, though this may be an artifact of the discourses that were examined. As a result, one might initially be led to believe that an explanation in terms of pro-drop being tied in with verb agreement (Rizzi 1986) might be attempted for Hindi/Urdu as well. However, (20) shows very clearly that verb agreement is orthogonal to the issue of null topics.

In (20) the current topic is some pigeons (=ye in (20a)). In (20b) the overt realization of this topic would be in the ergative, as opposed to the nominative in (20a) and (20c). While there would be verb agreement in (20c), if the topic were

realized, there can be no verb agreement in (20b): verbs in Hindi/Urdu do not agree with non-nominative subjects but instead show object agreement if the object is unmarked (nominative), as is the case in (20b).

- (20) a. [ye]<sub>T</sub> b<sup>h</sup>ii mer-ii=kii tarãã hãĩ  
 they also my-FSG=GEN.FSG like be.PRST.PL  
 ‘They<sub>topic</sub> are also like me.’
- b. jahãã daanaa dek<sup>h</sup>-aa  
 where seed.MSG see-PF.MSG  
 ‘where ever (they<sub>cont.topic</sub>) see a seed’
- c. udar ga-ye aor peT bar kar uR ga-ye  
 there go-PF.MPL and stomach fill having rise go-PF.MPL  
 ‘there (they<sub>cont.topic</sub>) go and having filled (their) stomach (they<sub>cont.topic</sub>) fly away.’ (Dilwale Dulhania Le Jayenge)

Thus, an explanation of the occurrence of null elements cannot be formulated in terms of a relationship between pronouns and verb agreement. Instead, generalizations in terms of discourse structure properties provide a more promising avenue of analysis for languages like Hindi/Urdu.

### 5.2.3 Null Background

Finally, let us consider background information. This is known information (unfocused, untopicalized) from the previous utterance that is backgrounded from a discourse structure point of view. An example of null background arguments is given in (21). Here *mard* ‘man’ in (21a) presumably furnishes the antecedent for the continuing null topic in (21b) which is dropped. The remainder of the arguments that are dropped (*kurbaani* ‘sacrifice’ and *aorat=ke liye* ‘for a woman’) have clear antecedents in the previous utterance in (21a) and are thus analyzed as backgrounded in (21b).

- (21) a. kyũke [mard]<sub>T</sub> to aorat=ke liye kab hii [kurbaanii]<sub>F</sub>  
 because man.M.NOM indeed woman.F=for ever EMPH sacrifice.F.NOM  
 de-taa hii nahĩ  
 give-IMPF.MSG EMPH not  
 ‘Because man<sub>topic</sub> does not ever make a sacrifice<sub>focus</sub> for a woman.’

- b. aor nahĩ kab hii de-gaa  
 and not ever EMPH give-FUT.MSG  
 ‘nor will (he<sub>cont.topic</sub>) ever make (a sacrifice for a woman)<sub>Back</sub>.’  
 (Dilwale Dulhania Le Jayenge)

All instances of argument drop found in the discourses examined so far could be analyzed in terms of the distinctions made here. It therefore appears that this analysis is on the right track and should contribute to a more complete analysis of discourse structure and null elements in general.

## 6 Analysis

In this section, we present the generalizations arrived at above in terms of an LFG analysis by taking a sample discourse fragment and working through it.

As described in section 4, we assume an information projection at which the discourse functions are represented. While this information projection is independent, it constrains and is in turn constrained by the other projections. In particular, it is placed into a relationship with the f-structure via general well-formedness constraints. That is, each grammatical function at f-structure must be linked to a discourse function at i-structure.<sup>11</sup> This ensures that each and every argument is indeed represented at both the i-structure and f-structure. While a given argument may be null in the c-structure, it is still realized at the level of argument structure and can be assumed to be linked to a grammatical function via the principles of Lexical Mapping Theory (see Bresnan and Zaenen 1990). In addition, as shown in section 5, elements which are null in the c-structure play a significant role in terms of i-structure and must be represented there.

Given that null elements must be represented at both f-structure and i-structure, it is imperative that the two levels of representation be placed into a correspondence with one another so that each grammatical function is linked to a discourse function.<sup>12</sup>

---

<sup>11</sup>In fact, in order to include both arguments and verbal predicates in the i-structure, the formulation should be that each predicate PRED containing element at f-structure should be linked to a discourse function (King 1997).

<sup>12</sup>This correspondence may be direct or may be realized via the s(emantic)-structure.



This prevents the possibility of an unconstrained and incorrect proliferation of functions at either structure: the information at both levels of representation must be mutually compatible.

The sample discourse in (22) is rather simple (as it should be for illustrative purposes) in that there is only one element that is dropped: the continuing topic. The corresponding f- and i-structures are shown in (23) and (24) respectively.

- (22) a. [māĩ] bais            barf=se            yahāã rah rah-aa    hũ  
 I.NOM twenty-two winter=from here    live Stat-MSG be.PRST.1.SG  
 ‘I<sub>topic</sub> have been living here for 22 years.’
- b. rozaanaa is    hii    sarak=se            guzar-taa            hũ  
 daily    this EMPH street.F=from pass-IMPF.MSG be.PRST.1.SG  
 ‘Daily (I<sub>cont.topic</sub>) go through this street.’ (Dilwale Dulhania Le Jayenge)

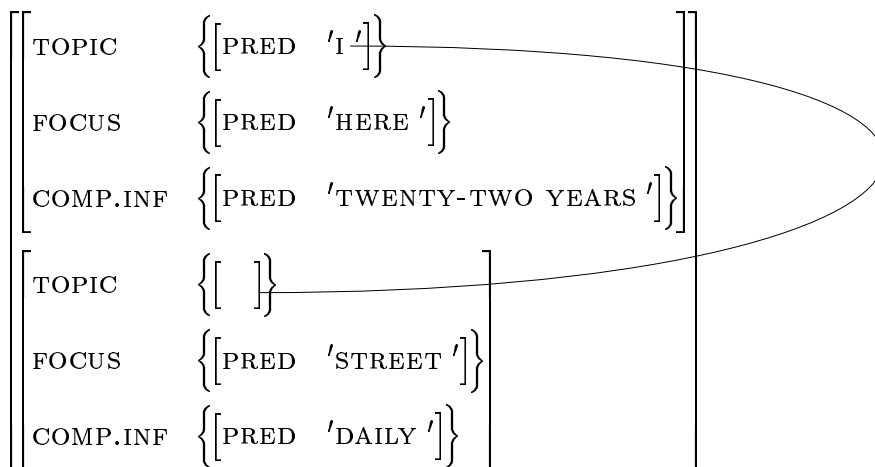
(23) **Functional structure** for (22a):

$$\left[ \begin{array}{ll} \text{PRED} & \text{'live<SUBJ,OBL> '}' \\ \text{SUBJ} & \left[ \text{PRED 'I '}' \right] \\ \text{ADJUNCT} & \left\{ \left[ \text{PRED 'TWENTY-TWO YEARS '}' \right] \right\} \\ \text{OBL} & \left[ \text{PRED 'HERE '}' \right] \end{array} \right]$$

**Functional-structure** for (22b):

$$\left[ \begin{array}{ll} \text{PRED} & \text{'pass<SUBJ> '}' \\ \text{SUBJ} & \left[ \text{PRED PRO} \right] \\ \text{ADJUNCT} & \left\{ \left[ \text{PRED 'STREET '}' \right] \right\} \\ & \left\{ \left[ \text{PRED 'DAILY '}' \right] \right\} \end{array} \right]$$

(24) **Information-structure** for (22a) and (22b):



Given the interconnection between f-structure and i-structure, it is possible to determine which discourse functions have been realized as null. This is possible even when more than one argument has been dropped. In (21), for example, the backgrounded information is null as well as the continuing topic. Here, as in our sample example in (22), linking from argument-structure to f-structure yields the information that two extra unexpressed arguments are present. A look back at the previous discourse utterance shows that these unexpressed arguments are compatible with the ones that have been introduced previously and should thus be analyzed as background information. Formally, the empty value for the background discourse function is linked to the appropriate values in the preceding i-structure, as illustrated in (24) for the topic.<sup>13</sup>

Note that the precise nature of the algorithm needed to determine null continuing topics and background information remains to be worked out. It is conceivable that some of the strategies developed by Grosz and Sidner (1986) in their centering approach to anaphora resolution should be adapted.

<sup>13</sup>Note that we have not dealt with embedded clauses. A representation and treatment of embedded clauses should follow along the lines presented here, with the extra provision that the i-structure corresponding to the entire utterance would display levels of embedding corresponding to the syntactic embedding.

## 7 Conclusion

We hope to have provided further evidence that word order permutations in Urdu/Hindi indeed reflect discourse functions, as has been suggested by Gambhir (1981) and, more recently, Kidwai (1997). Following the work of King (1995) for Russian, we further proposed that these discourse functions be encoded in terms of syntactic positions, and, loosely based on the work of Vallduví (1992) and Choi (1997), posited four types of discourse functions: *topic*, *focus*, *background* and *completive information*.

Given this view of discourse structure, we examined the role of null elements and concluded that argument drop is licensed at the level of discourse structure and that only continuing topics or backgrounded information may be omitted.

## 8 Appendix: Full Sample Discourse

(from Dilwale Dulhania Le Jayenge)

- (25) a. [ye]<sub>T</sub> [landan]<sub>F</sub> hai  
 this London be.PRST.SG  
 ‘This is London.’
- b. dunyaa=kaa sab=se bar-aa šeher  
 world.F=GEN.MSG all=from big-MSG city.M.NOM  
 ‘The world’s biggest city.’
- c. [mãĩ]<sub>T</sub> bais barf=se [yahãã]<sub>F</sub> rah rah-aa hũ  
 I.NOM twenty-two winter=from here live Stat-MSG be.PRST.1.SG  
 ‘I’ve been living here for 22 years.’
- d. rozaanaa [is hii sarak=se]<sub>F</sub> guzar-taa hũ  
 daily this EMPH street.F=from pass-IMPF.MSG be.PRST.1.SG  
 ‘Daily (*I<sub>cont.topic</sub>*) go through this street’
- e. aor rozaanaa [ye sarak]<sub>T</sub> mujh=se [mer-aa naam]<sub>F</sub> puc<sup>h</sup>-tii hai  
 and daily this street.F.NOM I=from my-MSG name.M.NOM ask-IMPF.FSG be.PRST.SG  
 ‘and daily this street<sub>switchtopic</sub> asks my name.’
- f. [ye]<sub>F</sub> puc<sup>h</sup>-tii hai ke [Chaudhry Banjir Singh]<sub>T</sub> [kon]<sub>F</sub> hai  
 this ask-IMPF.FSG be.PRST.SG that Chaudhry Banjir Singh who be.PRST.SG  
 ‘(*It<sub>cont.topic</sub>*) asks this: who is Chaudhry Banjir Singh<sub>switchtopic</sub>.’
- g. [kahãã=se]<sub>F</sub> a-yaa hai [kyũ]<sub>F</sub> a-yaa hai  
 where=from come-PF.MSG be.PRST.SG why come-PF.MSG be.PRST.SG  
 ‘Where has (*he<sub>cont.topic</sub>*) come from? Why has (*he<sub>cont.topic</sub>*) come?’
- h. [yahãã]<sub>T</sub> mujhe [koi nahĩ]<sub>F</sub> jaan-taa  
 here I.DAT some not know-IMPF.MSG  
 ‘Here<sub>switchtopic</sub> nobody knows me’
- i. sawaii in kabutrõ=ke kyũke [ye b<sup>h</sup>ii]<sub>T</sub> [mer-ii=kii tarãã]<sub>F</sub> hãĩ  
 except these pigeon.PL=of because they also my-FSG=of like be.PRST.PL  
 ‘except for these pigeons, because they<sub>switchtopic</sub> are also like me.’
- j. jahãã daanaa dek<sup>h</sup>-aa udar ga-ye  
 where seed.M.NOM see-PF.MSG there go-PF.MPL  
 ‘where ever (*they<sub>cont.topic</sub>*) see a seed, there (*they<sub>cont.topic</sub>*) go’
- k. aor peT bar kar uR ga-ye  
 and stomach.M.NOM fill having rise go-PF.MPL  
 ‘and having filled (*their*) stomach (*they<sub>cont.topic</sub>*) fly away.’

## References

- Aissen, J. 1992. Topic and focus in Mayan. *Language* 68:43–80.
- Banfield, A. 1973. Narrative style and the grammar of direct and indirect speech. *Foundations of Language* 10:1–39.
- Bresnan, J. 1998. *Morphology Competes with Syntax: Explaining Typological Variation in Weak Crossover Effects*. In P. Barbosa, et al. (eds.) *Is the Best Good Enough? Proceedings from the Workshop on Optimality in Syntax*, 59–92. Cambridge, MA: The MIT Press.
- Bresnan, J. 1997. *Lexical-Functional Syntax*. Unpublished book manuscript, Stanford University.
- Bresnan, J., and R. Kaplan. 1982. Introduction: Grammars as Mental Representations of Language. In J. Bresnan (ed.) *The Mental Representation of Grammatical Relations*, xvii–lii. Cambridge, MA: The MIT Press.
- 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 Publications.
- Choi, H. 1996. *Optimizing Structure in Context: Scrambling and Information Structure*. PhD thesis, Stanford University.
- Dwivedi, V. 1994. *Syntactic Dependencies and Relative Phrases in Hindi*. PhD Thesis, University of Massachusetts at Amherst.
- 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, Philadelphia.
- Grosz, Barbara J., and Candance L. Sidner 1986. Attention, intention, and the structure of discourse. *Computational Linguistics* 12(3):175–204.
- 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.
- Kaplan, R. 1987. Three seductions of computational psycholinguistics. In P. White-lock et al. (eds.) *Linguistic Theory and Computer Applications*, 149–181. London: Academic Press.

- 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: University of Chicago Press.
- Kidwai, A. 1996. Word Order and Focus Positions in Universal Grammar. Paper presented at the *Table ronde internationale sur la Grammaire du focus*. U. Paris II and U. Paris X, February.
- Kidwai, A. 1997. *Scrambling and Binding in Hindi-Urdu*. PhD thesis, Jawaharlal Nehru University, New Delhi.
- King, T.H. 1995. *Configuring Topic and Focus in Russian*. Stanford, CA: CSLI Publications.
- King, T. H. 1997. Focus domains and information-structure. In M. Butt and T. H. King (eds.) *Proceedings of the LFG97 Conference*. CSLI Publications. <http://www-csli.stanford.edu/publications/>.
- É. Kiss, K. 1995. *Discourse Configurational Languages*. Oxford University Press.
- Kornfilt, J. 1995. *On Focusing and De-Focusing in Turkish*. Paper presented at the Generative Grammatik im Süden (GGS), Universität Jena.
- Mahajan, A. 1990. *The A/A-bar Distinction and Movement Theory*. PhD thesis, MIT.
- Rizzi, L. 1986. Null objects in Italian and the theory of *pro*. *Linguistic Inquiry* 17:501–558.
- Rooth, M. 1985. *Association with Focus*. PhD thesis, University of Massachusetts, Amherst.
- Vallduví, E. 1992. *The Informational Component*. New York: Garland Press.
- Verma, S.K. 1970. *Word Order in Hindi*. *Achiv Orientalnini*, Number 38:29–32.
- Yokoyama, O. 1986 *Discourse and Word Order*. Amsterdam, The Netherlands: John Benjamins.

(Butt)  
Universität Konstanz  
FG Sprachwissenschaft <D186>  
78457 Konstanz  
Germany  
`miriam.butt@uni-konstanz.de`

(King)  
NLTT/ISTL  
Xerox Corporation  
Palo Alto Research Center  
3333 Coyote Hill Road  
Palo Alto, CA 94304  
U.S.A.  
`thking@parc.xerox.com`