

# A Corpus Study of Zero Pronouns in Hindi: An Account Based on Centering Transition Preferences

Rashmi Prasad

University of Pennsylvania  
Department of Linguistics  
619 Williams Hall  
36th and Spruce Street  
Philadelphia, PA 19104-6305  
rjprasad@babel.ling.upenn.edu

## Abstract

This paper presents a corpus-based investigation of the use of zero pronouns in Hindi. After establishing that the antecedents of these null arguments cannot be recovered syntactically (Rizzi, 1982; Jaeggli & Safir, 1989), I propose an account in terms of "Centering Theory" (Grosz et al., 1995). Given the Hindi-specific centering constraints proposed in Prasad & Strube (2000), I argue that the discourse constraint to license the felicitous use of Hindi zero pronouns should be formulated as a combination of preferences for sequences of transitions and a "zero pronoun rule", adapted from Rule 1 of Centering Theory. More generally, the proposed account will explain (a) why null elements are most frequently the subject, (b) why object drop in Hindi occurs only when the subject is also dropped.

## 1. Introduction

This paper presents an account of the constraints on the occurrence of zero pronouns in Hindi. Hindi allows major grammatical relations such as subject and object to be covert in finite clauses and these unexpressed relations function as pronouns. This raises the question of how the reference for these zero pronouns is determined. In Section 2, I show that Hindi zero pronouns are not recoverable via identification with rich AGR, as is argued for languages like Italian and Spanish (Rizzi, 1982; Rizzi, 1986; Jaeggli & Safir, 1989). This is despite the fact that Hindi has morphologically uniform and rich inflectional paradigms for verbal agreement. In such a case, the recoverability of these pronouns for reference is assumed to be constrained by rules of discourse. In previous work, Butt & King (1997) also make this assumption and further, specify a discourse constraint that relates the use of zeros to word order and information structure, in particular, to topicalization. I provide a brief discussion of this in Section 3. In Section 4, I show that the proposal made by Butt & King (1997), namely that arguments can be dropped when they are the continuing topic, with the topic equated with the topicalized constituent (if any) of a sentence, cannot account for all the observed facts. Finally, in Section 5, I provide an alternative (corpus-based) account of the discourse constraints on zero pronouns in Hindi tensed main clauses.<sup>1</sup> The account is couched within the framework of Centering Theory (Grosz et al., 1995) and the constraint is stated in terms of the *transition pair preferences* and the *zero pronoun rule*, adapted from Rule 1 of Centering Theory.

<sup>1</sup> Arguments are rarely dropped in subordinate clauses. The constraints on their occurrence are not addressed in this paper.

## 2. Hindi in a Typology of Null Argument Languages: On Identification via Agreement

Kameyama (1985) groups Hindi with Type II languages such as Italian and Spanish with respect to the extent to which major grammatical functions can be non-overt in tensed clauses. Following Rizzi (1982), Rizzi (1986) and Jaeggli & Safir (1989), such languages have a verbal morphology that is sensitive to one or more grammatical features (person, number, gender, aspect, etc..). Furthermore, since these languages have only subject-verb agreement, object drop is disallowed, and this fact is immediately explained by the requirement that the person/number/gender features of the zero pronoun should be recoverable by agreement, in order for referent identification to take place. This behavior contrasts with what is seen in languages like Japanese and Chinese, which display a syntactically unconstrained use of zero pronouns, in that person, number and gender features of null arguments are not associated with verbal agreement. Recoverability of these arguments then is argued to rely on discourse factors and not on syntax at all.

Hindi is like Spanish, Italian and Latin in having a rich verbal agreement system but is different from them in that the verb can agree with both the subject (1) and the object (2):<sup>2</sup>

- (1) *malay kitaab paRh rahaa hai*  
malay-3sg.M book-3sg.F read-INF PROG-M  
BE-3sg.PRES  
"Malay is reading the book"

<sup>2</sup> The following abbreviations are used in this paper: M = masculine, F = feminine, ERG = ergative, PER = perfect, PROG = progressive, QPL = question particle, FUT = future, INF = infinitive, PRES = present, ho = honorific, PAST = past, TOP = topic marker, SUBJ = subject, HAB = habitual

- (2) *malay ne kitaab paRhii*  
 malay-3sg.M ERG book-3sg.F read-PERF-3sg  
 “Malay read the book”

Given the classification of languages in terms of the rich AGR licensing condition, Kameyama (1985) claims that Hindi behaves like Italian and that it allows an argument to be dropped if the verb was inflected for its person/number/gender features. So, for example, if the verb agreed with the subject, as in (1), the subject should be able to appear as null and if the verb agreed with the object, as in (2), there would be nothing ruling out the object from being realized as null. At the same time, however, what she claims is not possible for Hindi is for the subject to be null when the verb agrees with the object, and vice versa. This is illustrated in the following two examples (adapted from Kameyama (1985)):

- (3) **Q:** *kyaa malay ne kitaab paRhii?*  
 QPL malay-3sg.M ERG book-3sg.F read-3sg.F.PERF?  
 “Did Malay read the book?”

**A:** *malay/\*0 ne kitaab/0 paRhii*  
 Malay/\*0-3sg.M ERG book/0-3sg.F read-3sg.F.PERF  
 “Malay read the book”

- (4) **Q:** *kyaa malay kitaab paRh rahaa hai?*  
 QPL malay-3sg.M book-3sg.F read-INF  
 PROG-M BE-3sg.PRES  
 “Will Malay read the book?”

**A:** *malay/0 kitaab/\*0 paRh rahaa hai*  
 Malay/0-3sg.M book/\*0-3sg.F read-INF  
 PROG.M BE-3sg.PRES  
 “Malay is reading the book”

Counterexamples to these expectations, however, are abundant in naturally occurring data, as examples (5) and (6) show:

- (5) a. *fanTuush ne aadmiyoN kaa gussaa saamaan par utaaraa.*  
 FanTush-3sg.M ERG men-3pl.M of anger furniture-3sg.M on took-down  
 “Fantuush took out his anger with the men on the furniture.”

b. *0 vahaaN kii sab kursiyaaN toR daaliN.*  
 0 (=FanTuush-3g.M) there of all chairs-3pl.F break-INF put-3pl.F.PERF  
 “(Fantuush) broke all the chairs there.”

- (6) a. *unhone (=Gajadhar Babu) shahar me makaan banvaa liyaa thaa*  
 he-3HO (=Gajadhar Babu) city in house make-CAUS PER-3sg.M BE-PAST  
 “He had got a house made in the city”

- b. *0 baRe laRke amar aur laRkii kaantii kii shaadiyaaN kar dii thiiN*  
 0 (=Gajadhar Babu-3HO) older son amar and daughter kaantii of marriages do-INF PERF-3pl.F BE-PAST  
 “[Lit.] (Gajadhar Babu) got the marriages done of his older son Amar and daughter Kaantii”

In both (5b) and (6b), the person, number and gender features of the null subjects cannot be determined by the verb because the verb agreement is with the object. So despite the rich agreement inflection on the verb, Hindi cannot be said to impose an agreement based licensing constraint on the use of zeros.<sup>3</sup> We therefore need to look elsewhere for the constraints that govern the use of null pronouns. In this sense, Hindi groups together more closely with Japanese and Chinese, although the use of zeros in the latter is more unrestricted than in Hindi.<sup>4</sup>

### 3. Previous Research on the Discourse Constraints Licensing Hindi Zero Pronouns

In their study of null elements in Hindi and Urdu discourse, Butt & King (1997) also argue that the interpretation of null elements in Hindi lies outside the realm of syntax and that the felicitous use of zero pronouns is instead governed by the discourse context in which the utterance is used. In particular, they attempt to relate argument drop in Hindi with the free word order that is characteristic of the language. The different word orders in Hindi have been shown to relate to different discourse functions (Gambhir, 1981), and Butt & King (1997)’s account attempts to relate word order, information structure (Vallduví, 1990) and referentiality. They draw on Gambhir (1981)’s account of the discourse functions of word order variants in Hindi, but focus primarily on four discourse functions: (a) Topic (b) Focus (c) Background and (d) Completive Information. They also claim that discourse functions in Hindi/Urdu are encoded syntactically, i.e., in order to receive a particular discourse function interpretation, a constituent must appear in the appropriate licensing position (otherwise the structure is infelicitous). So, topics appear sentence initially, foci immediately before the verb, and backgrounded material postverbally.<sup>5</sup> In particular, the *topic* in their account is identified with the clause-initial constituent in matrix clauses. Structurally, this position is identified as [Spec, IP]. According to this

<sup>3</sup> Note that Georgian is one language that patterns like Hindi in its verb agreement paradigm (see Kameyama, (1985)), and also respects the licensing constraint of null argument in terms of identification by agreement.

<sup>4</sup> What the extent is to which this difference exists and why it exists is a matter that I will not go into here. But the results presented in this paper on the constraints in Hindi provides a starting point for fruitful research in this direction.

<sup>5</sup> The backgrounded material provides more detailed information as to how the new information fits in with the already known information.

proposal then, *topicalizations* are assumed to be hosted by [Spec, IP] and are therefore the *topic* of the clause. This is shown in (7):<sup>6</sup>

- (7) [IP [SPEC,IP *hassan ko-TOPIC*] *naadiyaa ne Tofii dii*]  
 “To Hassan Naadiyaa gave a toffee”

As for the licensing constraint on zero pronouns, Butt & King essentially claim that an argument can be dropped if it is a *continuing topic* (i.e., if it is the topic of the current as well as the previous utterance) or if it is the *backgrounded information*. Their example (8) below gives an example of a *continuing topic*: the topic is the same from utterance (a) to (b) and can thus be felicitously dropped.

- (8) a. *main-TOPIC baais baras se yahaan rah rahaa huun*  
 “I-TOPIC have been living here for 22 year”  
 b. *0/main-TOPIC rozaanaa is hii saRak se guzartaa huun*  
 “0/I-TOPIC go by this street daily”

(9) gives an example of a *shifting topic*: the topic in the (b) utterance has been shifted to a different entity than in the (a) utterance and thus cannot be dropped.

- (9) a. *to (hum)-TOPIC uspe ek naaTak likhte hain.*  
 “So let (us)-TOPIC write a play about that”  
 b. *main/\*0-TOPIC erfors kaa aadmii huun*  
 “I/\*0-TOPIC am an airforce man”

#### 4. Dissociating Word Order and Information Structure from the Form of Referring Expressions

While the account proposed by Butt & King (1997) is an attractive one in that it relates word order, information structure and the form of referring expressions used in discourse, especially in light of such established correlations for other languages such as German (Rambow, 1993; Strube & Hahn, 1999), this three way correlation cannot be maintained for Hindi. Further motivation for the absence of such correlations comes from Prasad & Strube (2000). They propose a general method to determine the relative salience of discourse entities, which has an effect on pronominalization in discourse and on discourse coherence (Grosz et al., 1995), and on applying this method to Hindi by coding for different grammatical and discourse features such as grammatical function, word order and information status, show that word order and information status has no significant effect on discourse salience and therefore

also on pronominalization. The study in Prasad & Strube (2000), however, did not investigate the interpretation and constraints on the use of zero pronouns.<sup>7</sup> In what follows, I argue that word order and information structure (associated with the word order) has no bearing on the felicitous use of zero pronouns in Hindi.

A look at the examples in Butt & King (1997) containing a *continuing topic* shows that in each case, the *topic* coincides with the subject. Since there is no reliable way of telling whether the subject is indeed in the TOPIC position in the clause, the real test of the effect of the *topic* on null anaphora lies in clauses where some non-subject constituent has been topicalized. The following examples taken from the corpus used for this study show that a null argument can be licensed when it does not refer to the topicalized element. In (10a), the prepositional object “Alladiya” is the topicalized entity (where topicality is indicated by the topic marker *to* (Kidwai, 1997; Prasad, 1997a) in addition to its clause initial position). The zero pronoun in (10b) however cannot refer to this entity and instead picks the grammatical subject denoted entity as its antecedent.

- (10) (A group of people are talking about the appalling behavior of a man called Alladiya and suddenly recall Alladiya’s father, Hamiidaa..)  
 a. [*is haraamzaade allaadiyaa se to*]-TOPIC [*hamiidaa*]-SUBJ *laakh darze acchaa thaa* [this bastard alladiya from TOP]-TOPIC [*hamiidaa*]-SUBJ million times better was [Roughly] “Hamiidaa was a million times better than this bastard Alladiyaa”  
 b. *0 fakat ek hii baar kafan churaataa thaa*  
 0 EMPH once only time coffin steal did  
 “(0 = Hamiidaa/#Allaadiyaa) stole the coffin only once”

An important point to note here is that both the “topicalized constituent” as well as the “subject” are equally likely candidates for antecedents of the zero pronoun. So, the zero pronoun not only can be licensed when it refers to some non-topicalized constituent, but in fact cannot refer to this clause-initial topicalized element, even when the person/number/gender features are perfectly compatible with it. This point will be discussed further in the later sections.

In the next example (11), the prepositional phrase is topicalized and though the topicalized constituent is not a likely candidate for the zero pronoun, the example still illustrates that the zero can be licensed when it refers to the non-topicalized element.

<sup>6</sup> All the examples in this section are taken from Butt & King (1997).

<sup>7</sup> It is legitimate to assume that the constraints on overt pronouns are not the same as the ones on zero pronouns, primarily because they are not in free variation.

- (11) a. *[apnii duukaan se]-TOP [Kishan]-SUBJ param santushT va sukhi tha*  
[self's store with]-TOP [Kishan]-SUBJ totally satisfied and happy was  
"Kishan was totally satisfied and happy with his store"
- b. **0** *raat khaaT par jaataa to na jaane kitne sapne aate*  
**0** night cot on went-HAB then (he) didn't know how many dreams came (to him)  
"When (**0** = Kishan) went to his bed at night, there was no telling how much he dreamt"

Examples such as those given above indicate that word order/topicalization cannot be a strong licenser of zero pronouns in Hindi. In the next section, I turn to my own analysis of a hand collected corpus and propose an alternative account of null arguments in Hindi.

## 5. The Discourse Constraint on Hindi Zero Pronouns: A Centering Account

So far, I have established that Hindi Zero pronouns are neither identifiable by syntactic constraints such as the requirement for identification by rich AGR, nor by discourse constraints that attempt to relate their felicitous use to word order and information structure. In this section, I present a Centering analysis of a Hindi corpus which I have hand created and annotated for purposes of this study. In what follows, I first present a brief overview of Centering Theory (Grosz, et al., 1995) and then present a corpus analysis of zero pronouns in terms of Centering Theory. I have analyzed the corpus for the effect of transition preferences between utterances on whether a zero pronoun was used or not, and the results show a distinct correlation between certain transition pairings and the occurrence of the zero pronouns.

### 5.1 Centering Theory: A Brief Overview

The Centering model relates focus of attention, choice of referring expression and local discourse coherence. The model consists of two data structures: a set of discourse entities associated with each utterance  $U_i$  (the list of *forward-looking centers*, or the  $Cf(U_i)$ -list), and a unique entity in each utterance  $U_i$  (the *backward-looking center*, or the  $Cb(U_i)$ ). The  $Cb(U_i)$  is a member of the  $Cf(U_i)$ -list and is meant to represent the entity that the utterance is most centrally about. The set of forward-looking centers is ranked according to discourse salience. The most highly ranked element of the  $Cf(U_i)$ -list is called the *preferred center*, the  $Cp(U_i)$ , and the most highly ranked element of  $Cf(U_i)$  that is *realized* in  $U_{i+1}$  is the  $Cb(U_{i+1})$ .

The theory defines transition relations across pairs of adjacent utterances. The transitions differ from each other according to (a) whether the  $Cb$ 's of successive utterances are equal or not, and (b) whether the  $Cb$  of an utterance corresponds to the  $Cp$  of that

utterance or not. Explicit definitions of the transitions are as follows:

- **CONTINUE:**  $Cb(U_{i+1})$  is the  $Cb(U_i)$   
OR the  $Cb(U_i)$  is *undefined*  
**AND**  
 $Cp(U_{i+1})$  is the  $Cb(U_{i+1})$
- **RETAIN:**  $Cb(U_{i+1})$  is the  $Cb(U_i)$   
**AND**  
 $Cp(U_{i+1})$  is *not* the  $Cb(U_{i+1})$
- **SMOOTH-SHIFT:**  $Cb(U_{i+1})$  is *not* the  $Cb(U_i)$   
**AND**  
 $Cp(U_{i+1})$  is the  $Cb(U_{i+1})$
- **ROUGH-SHIFT:**  $Cb(U_{i+1})$  is *not* the  $Cb(U_i)$   
**AND**  
 $Cp(U_{i+1})$  is *not* the  $Cb(U_{i+1})$

The theory also proposes two rules. *for each utterance,  $U_{i+1}$ , in a discourse segment  $U_p, \dots, U_n$ :*

**Rule 1.** If some element of  $Cf(U_i)$  is realized as a pronoun in  $U_{i+1}$ , then so is the  $Cb(U_{i+1})$ .

**Rule 2.** Transition sequences are ordered: *continue* > *retain* > *smooth-shift* > *rough-shift*.

Following Prasad & Strube (2000), we assume that the elements of the  $Cf$  list in Hindi are ranked by grammatical function (*subject* > *direct object* > *indirect object* > *other subcategorized functions* > *adjuncts*).

### 5.2 Corpus and Coding

The corpus used for this study was a collection of 6 short stories and 3 newspaper articles. The total number of sentences was 2192. Each sentence in the corpus was coded for:

- Clause type (declarative, imperative, interrogative).
- Main vs. subordinate clause.
- Position of the grammatical functions of subject, object and indirect object (when they are overt).
- Overt/null realization of the grammatical functions.
- The cospecifier of the null elements with respect to grammatical function and position.
- Three Centering variables: (a) the  $Cp$ , (b) the  $Cb$ , (c) and the transition marked by the clause.

Clause types such as imperatives and interrogatives were excluded from the analysis, as were direct speech segments (which most often included the previous two types). Following Kameyama (1997), I assume direct speech segments to be inaccessible to the utterances in the next higher level of segmented discourse, so their exclusion does not upset the computation of the

transitions, which is the main point of this study. After exclusion of these clause types the sentences in the corpus totaled 1332. In coding for the elements that are potential candidates for subsequent reference, I have also excluded the annotation of things such as events, states, propositions, or phrases. Null arguments in subordinate clauses were found to be very rare (I counted only 4) and I have excluded these from the annotation too.

### 5.3 Analysis and Results

Out of the 1332 clauses that were finally coded for the Centering variables and for which the transitions were computed, there were 466 clauses in which there was no continuing reference from the previous utterance.<sup>8</sup> In the remaining 866 clauses, only 209 had one or more zero pronouns whereas the remainder had either a full NP or a overt pronominal form. A quick look at the distribution of zero pronouns with respect to the major grammatical functions showed that most of these zero pronouns were subjects (199), very few were direct objects (10) and there were none corresponded to any other grammatical function. While I will not attempt to answer the question of why no grammatical function other than the subject and the object are realized as null, the constraint that will be formalized below does explain the low number associated with zero objects. In fact, in the corpus, the zero objects occurred only when there was also a zero subject in the clause. I return to this presently.

After computing the transitions between all the utterances in the corpus, the first preliminary finding was that all the utterances containing zero pronouns were marked with the CONTINUE transition. Though this finding was interesting, it did not say much about what kinds of entities in the previous utterance could be realized as zero in the current one. A CONTINUE transition is obtainable after all the other three transitions, which means that any entity in the Cf list of the previous utterance has the potential of being realized as a zero in the current utterance. Furthermore, this also does not answer the puzzle indirectly posed in Section 4, namely, that if is not the topicalized constituent of an utterance that can be realized as a zero in the next one, then what is?

The next step of the analysis therefore involved extraction of *transition pairs* rather than just single transitions. So if an utterance had a zero pronoun, I recorded the transition marked on it as well as the one marked on the previous utterance. The distribution of transition preferences this time was the following:

- There were only three kinds of transition pairs associated with the zero pronouns:

- (a) CONTINUE + CONTINUE
- (b) SMOOTH-SHIFT + CONTINUE and

- (c) RETAIN + CONTINUE

The approximate frequency distribution of each pair for the 199 zero subject pronouns was as follows:

- CONTINUE + CONTINUE: 106 (53%)
- SMOOTH-SHIFT + CONTINUE: 85 (42%)
- RETAIN + CONTINUE: 8 (4%)

Note that in determining the frequencies above, the zero objects have not been excluded intentionally from the total number of zero pronouns, but rather for statistical reasons: as mentioned above, zero objects always occurred with the zero subjects, and furthermore, only with the CONTINUE + CONTINUE or the SMOOTH-SHIFT + CONTINUE transition pairs. Since the subjects are ranked higher than the objects in the ranking hierarchy of the Cf list, these zero objects will never figure in the computation of the transitions. I have therefore excluded them to prevent any distortion of the frequencies (even though this may not be significant enough).

### 5.4 Discussion

One of the first things that is obvious from the distribution seen above is the significance of the *preferred center*, the Cp, in Hindi for the realization of zero pronouns in discourse. The percentages of the Continue transitions following a Continue and following a Smooth shift are roughly the same. The low percentages of a Retain followed by a Continue also indicates that in contrast to languages like Yiddish (Prince, 1994), the *continuing topic*, if defined in Centering terms as the *backward looking center*, has very little likelihood of being realized as a zero pronoun unless it is also the *preferred center*.

The figures also explain why the subject is the grammatical function most often realized as null. Given the Cf-list ranking criterion in Prasad & Strube (2000), where the subject is always most highly ranked in the Cf list, and the categorical preference for a Continue transition that we have seen above, the subject is the only grammatical function that can generate a Continue transition. For the same reason, the objects cannot be dropped, because it would lead to either a Retain or a Rough shift transition which is dispreferred. At the same time, we can also now explain why objects can be dropped when the subjects are also dropped. This follows from Rule 1 of Centering Theory which we can reformulate for zero pronouns for Hindi:

- **Zero pronoun rule for Hindi:** If anything is realized as a zero pronoun in the utterance, then the Cp must be.

According to this rule, if there is a single zero pronoun in the utterance, then it must be the Cp (thus generating a Continue transition). The object drop is observed in the

<sup>8</sup> This could be said to mark the beginning of a new discourse segment in Centering terms. However, I do not address this question here.

data because there is nothing ruling it out as long as the Cp is also dropped. However, it is less likely to occur by itself because it would then violate the zero pronoun rule. Finally, we can also address the absence of the correlation between word order, namely, topicalization, and the realization of zero pronouns. This is because topicalizations, unless the subject itself is topicalized, do not rank as the highest entity in the Cf list in Hindi and can therefore never generate the two preferred transition pairs if they realized as null in the next utterance.

We can now state the discourse constraint that licenses the occurrence of zero pronouns in Hindi.

- Optionally drop an argument in  $U_{i+1}$  if:
  - (a) *the transition marking the previous utterance  $U_i$  is a CONTINUE or a SMOOTH-SHIFT,*
  - (b) *the transition marked by the current utterance,  $U_{i+1}$ , is a CONTINUE, and*
  - (c) *the Zero Pronoun Rule is not violated.*

One question still remains, however. There were a large number (247) of the preferred transition pairings (Continue + Continue and Smooth Shift + Continue) listed above that did not realize the Cp as null. A detailed study of these exceptions has been beyond the scope of this paper, but is the focus of current research.

## 7. Conclusion

In this paper, I have shown that the interpretation of and the licensing conditions on Hindi zero pronouns cannot be done in terms of syntactic constraints and that despite the rich agreement inflectional paradigm of the language, arguments may be dropped even when the identification via agreement requirement is not met. The statement of the constraints for the interpretation of these null elements needs to be made in terms of the discourse context. I discussed previous efforts in this direction and showed that these accounts were insufficient in explaining all the observed facts about argument drop. In particular, I argued against a discourse constraint that motivated a correlation between word order, information structure and the form of referring expressions. Finally, I presented a corpus based Centering analysis of Hindi texts and showed that zero pronouns occurred most often with certain transition pairings over utterances. In addition, I explained the difference in the frequency of occurrence found between subject and object pronouns in terms of the “zero pronoun rule for Hindi”.

## References

Butt, M., King, T.H. (1997). Null Elements in Discourse. *NULLS seminar*, University of Delhi, India.

Gambhir, S. (1981). Syntactic Restrictions and Discourse Functions of Word Order in Standard Hindi. *PhD Thesis*. Department of Linguistics, UPenn. Philadelphia.

Grosz, B.J., A.K. Joshi, and S. Weinstein. (1995). Centering: A framework for modelling the local coherence of discourse. In *Computational Linguistics*, 21(2): pp 203-225

Jaeggli, O. and K. J. Safir. (1989). *The Null Subject Parameter*. Kluwer. Dordrecht.

Kameyama, M. (1985). Zero Anaphora: The Case of Japanese. *Ph.D Thesis*, Stanford, California: Stanford University, Linguistics Department.

Kameyama, M. (1997). Intrasentential Centering: A Case Study. In M. Walker., A. Joshi and E. Prince, eds., *Centering Theory in Discourse*, Oxford University Press, Oxford.

Kidwai, A. (1997). Scrambling and Binding in Hindi-Urdu. *PhD Thesis*. Jawaharlal Nehru University, New Delhi.

Prasad, R. (1997a). The Topic Marker *to* in Hindi and Information Packaging. ms. Linguistics, University of Pennsylvania.

Prasad, R and M. Strube. (2000) Discourse Salience and Pronoun Resolution in Hindi. In *Penn Working Papers in Linguistics*, 6.3. UPenn: pp 189-208

Prince, E. (1994). Subject-Prodrop in Yiddish. In In Bosch, P., ed. *Focus and Natural Language Processing*. Vol 1. Intonation and Syntax: Working Papers of the IBM Institute for Logic and Linguistics 6, December. pp. 159-74.

Rambow, O. (1993). Pragmatic aspects of scrambling and topicalization in German. Ms. *Univ. of Pennsylvania, May 1993*.

Rizzi, L. 1986. Null objects in Italian and the theory of pro". *Linguistic Inquiry* 17: pp 501-557

Strube, M. and U. Hahn. (1999). Functional Centering: Grounding Referential Coherence in Information Structure. In *Computational Linguistics*, 25(3): pp 309-344

Vallduví, E. (1990). The Informational Component. PhD Thesis. Department of Linguistics, University of Pennsylvania. Philadelphia.