North East Linguistics Society

Volume 30 *Proceedings of the North East Linguistic Society 30 -- Volume Two*

Article 6

2000

The Emergence of the Unmarked Order in Hindi

Hanjung Lee
Stanford University

Follow this and additional works at: https://scholarworks.umass.edu/nels

Part of the Linguistics Commons

The Emergence of the Unmarked Order in Hindi[*]

Hanjung Lee

Stanford University

1. Introduction

One of the most challenging aspects of the analysis of 'free word order languages' like German, Hindi and Korean lies in motivating various possible constituent orders. In many of these free word order languages, it is also not uncommon to find fixed word order phenomena. But to date no general theory has been proposed to explain both the freedom of word order and the loss of the word order freedom of constituents, referred to as *freezing* (Mohanan 1992; Mohanan and Mohanan 1994).

This paper presents an Optimality Theory (OT: Prince and Smolensky 1993) account of word order in Hindi that can account for both the free ordering and fixed ordering of constituents. I propose that the OT theory of markedness provides a fundamental explanation for why the canonical word order becomes fixed in certain circumstances of marked associations of prominence scales. In particular, harmonic alignment of prominence hierarchies and local conjunction of constraints (Prince and Smolensky 1993; Smolensky 1995; Aissen 1999) offer exactly the formal devices needed to capture the marked associations of grammatical function, thematic role, and discourse functions of arguments which underlie 'the worst of the worst' type of freezing.

2. Word Order Freezing in a Free Word Order Language

2.1 Basic Clause Structure

Hindi is a right-headed language with SOV canonical order. However, unlike Japanese and Korean, the surface order of elements is not strictly head-final. The possible permutations of a simple Hindi sentence are shown in (1). The three elements in the sentence can appear in any order. The various possible orders have a primarily pragmatic effect in that they are

470   Hanjung Lee

(1)   Free Word Order in Hindi
a.   Anu-ne   caand   dek^{h}aa.
        Anu-ERG   moon-NOM   see/look at-PERF[1,2]
        'Anu saw the moon.'
b.   Caand Anu-ne dek^{h}aa.
c.   Anu-ne dek^{h}aa caand.
d.   Caand dek^{h}aa Anu-ne.
e.   Dek^{h}aa Anu-ne caand.
f.   Dek^{h}aa caand Anu-ne.

Hindi is discourse configurational in that discourse functions are encoded syntactically and thus affect word order (Kiss 1994). Following Butt and King (1996) and Sharma (1999), I assume that in Hindi topics occur clause initially in a position which is situated above the canonical argument positions. In particular, a topicalized constituent is assumed to be licensed in SpecIP (cf. King (1995) for Russian; Dwivedi (1994), Mohanan and Mohanan (1994) and Sharma (1999) for Hindi). In sentences like those in (1b) and (1d) in which an object is topicalized, its appearance in SpecIP results in non-canonical word order in which the object precedes the subject. On the other hand, sentences such as (1a), in which the subject is in initial position, can have two structures: one in which the subject is a topic and hence in SpecIP, and one in which the subject is not in SpecIP and hence is not interpreted as a topic.

Aside from topic, the two most commonly employed discourse functions in Hindi are focus and postposed background. The major function of focus is to provide new information relevant for the discourse structure. If there is only one focused constituent in the sentence, then it must appear immediately preverbally, in particular in SpecVP (Butt and King 1996; Sharma 1999a). This is illustrated in (2).

(2)   Focus in Hindi
a.   Niina-ne   Mohan-ko   [tofii]$_{FOC}$ d-ii.
        Niina-ERG   Mohan-DAT   toffee-NOM   give-PERF
        'Nina gave TOFFEE to Mohan.'
b.   #Niina-ne   [Mohan-ko]$_{FOC}$tofii   d-ii.
        Nina-ERG   Mohan-DAT   toffee-NOM   give-PERF
        'Nina gave toffee to MOHAN.'

Based on this correlation between word order and discourse function interpretation, I assume that the basic clause structure is flat, with discourse neutral arguments under S. Like most discourse configurational languages, Hindi employs endocentric configurations but only to express discourse information (see Sharma (1999) for extensive discussion of evidence for this view).

---

[1] The abbreviations used in this paper are as follows: ACC 'accusative', ADJ 'adjunct', BCK 'background', CAUS 'causality', DAT 'dative', ERG 'ergative', FOC 'focus', FUT 'future', GEN 'genitive', INST 'instrumental', LOC 'locative', NF 'non-finite', NOM 'nominative', OBJ 'object', OBL 'oblique', PERF 'perfective', PRON 'pronoun', SUBJ 'subject', TOP 'topic', VOL 'volitionality'.

[2] The label NOM here refers to the nominal form without a case clitic. A widely known generaliztion with regard to objects in Hindi is that the canonical case for animate objects is ACC, and the canonical case for inanimate objects is NOM. Verbs that are neutral to the animacy of their objects like dek^h 'see' can take either ACC or NOM depending on the animacy of their objects.

The Emergence of the Unmarked Order in Hindi   471

2.2   Word Order Freezing in Sentences with Non-volitional Subjects

With 'unaccusative transitives' or 'non-volitional transitives', the experiencer argument is marked with dative case, as in (3) and (4).[3] While the ergative subject in (1) carries the meaning of volitional action, the dative-marked arguments in (3) and (4) are nonagentive and non-volitional. Unlike the objects of volitional transitives, the theme arguments in (4) and (4) must be nominative even if animate: they cannot be accusative because they do not have the semantic property of being an entity toward which an action or event is directed by a volitional inceptor of the action or event (Mohanan 1994).

(3)   Anu-ko   caand   dik^{h}ii.
      Anu-DAT   moon-NOM   appear-PERF
      'Anu saw the moon.' (Lit. 'To Anu the moon appeared/became visible')

(4)   Vijay-ko   Ravii   milaa.
      Vijay-DAT   Ravi-NOM   find/encounter-PERF
      'Vijay met Ravi unexpectedly.'

Non-volitional transitives in Hindi exhibit two sets of puzzling behavior that are not found in volitional transitives. The first puzzle centers around the optionality in linking arguments to grammatical functions: non-volitional transitives in Hindi allow grammatical function alternation of their arguments. That is, either argument of these verbs can be mapped to the subject while the other is mapped to the object.[4] The second puzzle concerns word order freezing found when the theme is the grammatical subject. In a clause with a non-volitional experiencer subject and a theme object, all word orders except theme-experiencer-verb (OSV) are possible (Mohanan and Mohanan 1994: 175). The examples in (5), taken from Mohanan and Mohanan (1994) and Mohanan (1992), show that the non-volitional experiencer argument *Anu* is the grammatical subject: the reflexive takes it as its antecedent (5a,c,e);[5] the pronoun cannot be coreferent with it (5b,d,f).

(5)   SUBJ: non-volitional experiencer
a.   Annu-ko   Niinaa   apnii   bastii-me   dik^{h}ii. (SOV)
      Anu-DAT   Nina-NOM   self-GEN   neighbourhood-LOC   appear-PERF
      'Anu$_i$ saw Nina$_j$ in self's$_{i/*j}$ neighbourhood.'
b.   Anuu-ko   Niinaa   uskii   bastii-me   dik^{h}ii. (SOV)
      Anu-DAT   Nina-NOM   PRON-GEN   neighbourhood-LOC   appear-PERF
      'Anu$_i$ saw Nina$_j$ in her$_{j/*i}$ neighbourhood.'
c.   Niinaa   apnii   bastii-me   dik^{h}ii   Anuu-ko. (OVS)
      Nina-NOM   self-GEN   neighbourhood-LOC   appear-PERF   Anu-DAT
      'Anu$_i$ saw Nina$_j$ in self's$_{i/*j}$ neighbourhood.'
d.   Niinaa   uskii   bastii-me   dik^{h}ii   Anuu-ko. (OVS)
      Nina-NOM   PRON-GEN   neighbourhood-LOC   appear-PERF   Anu-DAT

---

[3] The dative subject construction in Hindi has been studied in detail in Bahl (1967), Mohanan (1994) and Verma and Mohanan (1990), among others. DAT case on the subject may be induced by any of three types of predicates. The first is a small set of 'non-volitional transitives' as in (3) and (4). Belonging to the second type are noun+verb complex predicates. A third source of DAT case on the subject are modality meanings such as urge and obligation, derived from complex verbals involving auxiliaries (Mohanan 1994: 142).

[4] The Marathi counterparts of Hindi non-volitional transitives also show this property; see Joshi and Asudeh (1999).

[5] For many Hindi speakers, the reflexive can take as its antecedent a subject, grammatical or logical, but no other argument (Mohanan 1994a: 122).

472   Hanjung Lee

'Anu$_i$ saw Nina$_j$ in her$_{j/*i}$ neighbourhood.'

e.	Dik^{h}ii	Anuu-ko	Niinaa	apnii	bastii-me.	(VSO)
	appear-PERF	Anu-DAT	Nina-NOM	self-GEN	neighbourhood-LOC	

'Anu$_i$ saw Nina$_j$ in self's$_{i/*j}$ neighbourhood.'

f.	Dik^{h}ii	Anuu-ko	Niinaa	uskii	bastii-me.	(VSO)
	appear-PERF	Anu-DAT	Nina-NOM	PRON-GEN	neighbourhood-LOC	

'Aun$_i$ saw Nina$_j$ in her$_{j/*i}$ neighbourhood.'

The examples in (6) show how the order of the subject and the object becomes frozen in SOV when the subject is a theme and the object is a non-volitional experiencer.

(6)   SUBJ: theme

a.	Niinaa	Anuu-ko	apnii	bastii-me	dik^{h}ii. (SOV)
	Nina-NOM	Anu-DAT	self-GEN	neighbourhood-LOC	appear-PERF

'Anu$_i$ saw Nina$_j$ in self's$_{ij}$ neighbourhood.'

b.	Niinaa	Anuu-ko	uskii	bastii-me	dik^{h}ii. (SOV)
	Nina-NOM	Anu-DAT	PRON-GEN	neighbourhood-LOC	appear-PERF

'Anu$_i$ saw Nina$_j$ in her$_{i/*j}$ neighbourhood.'

c.	*Anuu-ko	Niinaa	uskii	bastii-me	dik^{h}ii. (*OSV)
	Anu-DAT	Nina-NOM	PRON-GEN	neighbourhood-LOC	appear-PERF

'Anu$_i$ saw Nina$_j$ in her$_{i/*j}$ neighbourhood.'

The association between the thematic role and the grammatical function in Hindi non-volitional transitives is shown in (7), with the pattern of grammatical and ungrammatical orderings of the two arguments. We see that all orders except OSV order in (a2) are possible with the unmarked linking pattern shown in (7a) where the experiencer is the subject and the theme is the object. In other words, the unmarked linking is avoided in Hindi non-volitional transitives just in case the non-volitional experiencer is focus, which appears immediately preverbally, and the theme is topic, in SpecIP. In this situation the marked linking is employed instead but only with fixed SOV word order (as in (b1) in (7)).

(7)   Table 1. Word Order and Linking in Hindi non-volitional transitives

			a1.	S$_{exp}$ O$_{th}$ V
			a2.	*O$_{th}$ S$_{exp}$ V
a.	SUBJ	OBJ	a3.	S$_{exp}$ V O$_{th}$
	$\updownarrow$	$\updownarrow$	a4.	O$_{th}$ V S$_{exp}$
	< exp	th >	a5.	V S$_{exp}$ O$_{th}$
			a6.	V O$_{th}$ S$_{exp}$
			b1.	S$_{th}$ O$_{exp}$ V
b.	SUBJ	OBJ	b2.	*O$_{exp}$ S$_{th}$ V
			b3.	*S$_{th}$ V O$_{exp}$
	< exp	th >	b4.	*O$_{exp}$ V S$_{th}$
			b5.	*V S$_{th}$ O$_{exp}$
	TOP	FOC	b6.	*V O$_{exp}$ S$_{th}$

The descriptive generalization that emerges from this pattern is clear:

(8)   Generalization: The marked linking of experiencer to object and theme to subject
       is allowed only when the non-volitional experiencer is focus and the theme is topic.
       In this situation word order is frozen in unmarked order (i.e., SOV).[6]

There is evidence that supports this generalization. Consider the following examples in (9).
Suppose that a speaker A asked another speaker B the question in (9) and that the examples
in (10) are possible answers to it in that they provide the hearer with information as to who
saw Nina, namely Anu. The *what about Nina?* phrase, following Valldluví (1992) and
others, is used to identify the topic, namely the prominent old information which is the
center of interest in the current discourse. Since *Anu-ko* 'to Anu' provides the information
which answers the question, it is focused.

(9)   Aur Niinaa? Niinaa   kisko   dik^{h}ii?
       and Nina?   Nina-NOM   who-DAT   appear-PERF
       'What about Nina? Who saw Nina?' (Lit. 'To whom did Nina appear?')

(10)
a.   [Niinaa]$_{TOP}$ [Anu-ko]$_{FOC}$ dik^{h}ii.
       Nina-NOM   Anu-DAT   appear-PERF
       'ANU saw Nina.' (Lit. 'Nina appeared to ANU.')
b.   *Anu-ko Niinaa dik^{h}ii.
c.   *Niinaa dik^{h}ii Anu-ko.
d.   *Anu-ko dik^{h}ii Niinaa.
e.   *Dik^{h}ii Niinaa Anu-ko.
f.   *Dik^{h}ii Anu-ko Niinaa.

Among the six sentences in (10), only the sentence in (10a) is an appropriate answer since
the topic appears canonically sentence initially, and the focus immediately before the verb.
In contrast, in the sentences marked as ungrammatical the topic and focus are not in their
canonical position, and as a result they are ungrammatical in the context of (9) as expected
given the requirement that topic is clause initial and focus is immediately before the verb.[7]

Now, suppose that the sentences in (11b) are uttered by the speaker B as a response
to the speaker A's question in (11a), followed by the utterences in (10).

(11)
a.   Kahan?
       'Where?'
b. (i) Apnii bastii-me.
       self-GEN   neighbourhood-LOC
       'in self's neighbourhood.' (Nina (theme) = apnii, Anu (experiencer) ≠ apnii)
   (ii) Uskii bastii-me.
       PRON-GEN   neighbourhood-LOC
       'in her neighbourhood.' (Nina (theme) ≠ uskii, Anu (experiencer) = uskii)

As (11b) shows, for the speakers of Hindi that I have consulted, Nina is the only eligible
antecedent of the reflexive *apnii* within the context of (9) and (11b); it is also the only
element in the clause that cannot be coreferent with the pronoun *uskii*. We must therefore
conclude that the theme argument *Niinaa*, not the experiencer *Anu*, is the grammatical

---

[6] Thanks to Peter Sells for aid in developing this idea.

[7] The sentences in (10c) and (10d) are felicitous only in the context in which the preverbal NP is
topicalized with the verb as an information unit.

474   Hanjung Lee

subject when the former is topic and the latter is focus. In short, the facts on word order and coreference in (9)-(11) confirm the generalization stated in (8) that once linking arguments to grammatical functions is marked, their order cannot be marked as well.

In short, the word order freezing phenomena in Hindi examined in this section strongly suggest that there is another important dimension that constrains word order independently of discourse prominence, namely the relative *markedness* of subjects and objects.[8] A central problem then is how to formally relate the relative markedness of subjects and objects along multiple dimensions to word order. In section 4 I will show that a nonderivational analysis of syntactic markedness like the one offered by OT captures the markedness generalization about the freezing effects in Hindi in a way that acknowledges the universal basis of these effects and at the same time accounts for the language-particular ways in which these effects are realized.

3.   Deriving Free Word Order in OT-LFG

OT as a general theory of constraint interaction has been applied to a number of areas of linguistic research since its extraordinary success in the domain of phonology. For the domain of syntax, a growing body of work shows that many of the motivations for the OT approach to phonology are paralleled in syntax. Throughout this paper, I assume the formal framework of Lexical-Functional Grammar (LFG) recast within the OT framework (OT-LFG) (Bresnan 1998; Choi 1999; Kuhn 1999; Sells 1999, forthcoming).

In OT a grammar is a function mapping each linguistic input to its correct structural description or output. Within OT-LFG framework inputs are taken to be a (possibly underspecified) feature structure representing (i) a predicator, (ii) the proto-role properties of its argument(s), Proto-Agent (P-A) and Proto-Patient (P-P) (Dowty 1991), represented with features [VOL(itionality)], [CAUS(ality)], etc (Asudeh 1999), and (iii) other morphosyntactic and semantic information (e.g., TENSE and ASPECT) in a language independent form (Bresnan 1998). The universal input is modelled by sets of f(unctional)-structures. Following Choi (1999), I further assume that the input also contains a description of the informational status of the verb arguments represented with the two features [PROM] and [NEW]. An example (that has *Anu-ne caand dek^{h}aa* 'Anu saw the moon' (=(1a)) in Hindi as its optimal realization) is (12).

$$(12) \quad \begin{bmatrix} \text{GF1} & \begin{bmatrix} \text{PRED} & \text{'Anu'} \\ \text{PROM} & + \\ \text{NEW} & - \\ \text{VOL} & + \end{bmatrix}_x \\ \\ \text{GF2} & [\text{PRED} \quad \text{'moon'}]_y \\ \text{ASP} & \text{PERF} \\ \text{PRED} & \text{'see (P-A}_x\text{,P-P}_y\text{)'} \end{bmatrix}$$ (P-A=topic, P-P=discourse neutral)

Given an underspecified input f-structure, a set of output candidates are generated by the generator GEN. Here I assume that candidate sets consist of pairs of a c(onstituent)-structure and its corresponding f-structure, which is subsumed by the input f-structure (Bresnan 1998; Kuhn 1999). Candidates are evaluated against the input with respect to a

---

[8] A different type of word order freezing is found in sentences with ambiguous case marking (e.g. double nominative constructions and sentences with multiple nominals marked with the same case ending). See Lee (1999a,b) for an analysis of freezing effects involving morphological ambiguity, based on bidirectional optimization in OT (Smolensky 1996, 1998).

set of ranked constraints, and all constraints are universal and violable. Consequently, languages cannot differ in their constraint inventories; they differ only in the relative ranking of the constraints.

Clause structure and word order are constrained by potentially conflicting constraints in several parallel structures of grammar. To derive the canonical word order and deviations from this order in Hindi, I adopt the constraints proposed in previous works based on OT-LFG, in particular by Choi (1999) and Sells (1999, forthcoming). Those that are particularly relevant for the present purposes are given below. The interaction of two alignment constraints in (13) gives basic subtypes of clausal phrase structure without problematic recourse to complementary Left and Right alignment constraints (see Sells (1999, forthcoming) for details). For example, ranking Spine-R over HD-L will give right-branching languages. Head-final languages like Hindi, Japanese and Korean, which lack the structural functional heads of C and I, instantiate fully right-branching, with a single co-head at the bottom (Sells 1999).

(13) Constraints on Clausal Skeleton (Sells 1999, forthcoming):
a. Spine-R: co-head aligns right in its local subtree.[9]
b. H(EA)D-L: X^0 co-head aligns left in its local subtree.

The ordering of a verb's arguments in Hindi results from the interacting competing sets of constraints on word order: constraints on canonical ordering based on the hierarchies of grammatical functions and thematic roles (14); and information structuring constraints (15) distinguishing the contextual dimensions of discourse prominence and novelty, each marked by a binary feature. Here information structuring constraints proposed by Choi (1999) (i.e. PROM and NEW) are reinterpreted as f-structure alignment constraints à la Sells (1999, forthcoming).

(14) Canonical Phrase Structure Constraints CANON (Choi 1999):
a. CANON$_{GF}$ (f-s/c-s correspondence): Grammatical functions align with their canonical argument positions in c-structure according to the function hierarchy.
(SUBJ > D.OBJ > LOBJ > OBL > ADJUNCT (Bresnan 1994; Bresnan forthcoming))

b. CANON$_\theta$ (a-s/c-s correspondence): Non-verbal arguments at c-s align according to the thematic hierarchy.
(Agent > Beneficiary > Experiencer/Goal > Instrument > Patient/Theme > Locative (Bresnan and Kanerva 1989; Bresnan and Zaenen 1990; Bresnan forthcoming))

(15) Information Structuring Constraints:
a. TOP [PROM+, NEW−]-LFT: Topic aligns left in the clause.
b. FOC [PROM+, NEW+]-LFT: Focus aligns left in the clause.
c. BACKGROUND [PROM−, NEW−]-RT: Background information aligns right in the clause.

The discourse motivation for locating [+PROM] at one end of the clause and [−PROM] at the other is transparent. For Hindi, the dominance ranking is as (16). With this ranking, a topic will always be more to the left than a focus in the same clause; and a background will always be more to the right than spine elements in the same clause.

---

[9] The 'co-head' of the clause is any node which is part of the Extended projection (Grimshaw 1991, 1997), including V, V', VP, I, I', IP, C, and C'. See Bresnan (1998, forthcoming) and Sells (1999, forthcoming).

476   Hanjung Lee

(16)   {BCK-RT, TOP-LFT} » FOC-LFT » CANON$_\theta$ » CANON$_{GF}$ » Spine-R » HD-L

Crucially, the ranking for Hindi in (16) can predict that when the arguments do not differ in informational status, the canonical constraints will take effect, leading to SOV order; when there are differences, the canonical SOV order will however violate information structuring constraints, such that competitors with a noncanonical ordering can win out. In a discourse context in which the experiencer *Anu* is topic (i.e., prominent given information) and the theme *caand* 'moon' is focus (i.e., prominent new information), the input is as (12). This results in the optimal output (17a) (=(1a)), going through the constraint competition in (17) (violations of ordering constraints are computed by counting constituents from the left[10]). The candidates are schematically represented, and CANON$_{GF}$ is omitted here, since it has the same effect as CANON$_\theta$ in this case.

(17) Tableau 1. Volitional Transitives in Hindi[11]

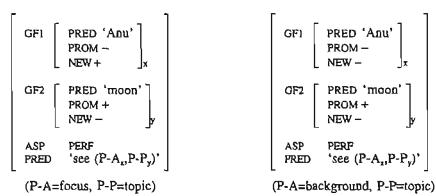| CANDIDATES:	BCK-R	TOP-L	FOC-L	CANON$_\theta$	Spine-R	HD-L
⇒ a. [$_{IP}$ S [$_{VP}$ OV]]]		0	1	0	0	1
b. [$_{IP}$ [S [$_{VP}$ V]]$_{IP}$ O]		0	2	0	1	0
c. [$_{IP}$ O [$_S$ S [$_{VP}$ V]]]		1	0	1	0	0
d. [$_{IP}$ [O [$_{VP}$ V]]$_{IP}$ S]		2	0	1	1	0
e. [$_{IP}$ [$_{VP}$ V] SO]		1	2	0	1	0
f. [$_{IP}$ [$_{VP}$ V] OS]		2	1	1	1	0

      As noted previously, the varied word orders in Hindi are optional from a purely syntactic point of view: non-canonical orderings are more restricted through context, and can be used only to mark a special information structure. In the present framework this can be captured by considering the role of the input (Choi 1999). For example, the candidate (17c) with OSV order corresponds more faithfully to (18a); the candidate (17d) with OVS order, to (18b). In other words, according to this analysis OSV becomes optimal for expressing the topical status of the object and the newness of the subject under the same ranking that yields (17a) as the optimal output; OVS is optimal for expressing the non-salient status of the subject as background information.

---

[10] SOV sentences such as (1a) can have three structures in Hindi. The first, and dominant, reading is one in which the subject is topic and the object is discourse neutral. In this case the subject must appear in SpecIP and the object in S, as in (17a). The second reading is one in which both the subject and the object are discourse neutral, and hence both are simply generated in S. The third reading is one in which the object is focus and hence is in SpecVP. The tableau in (17) considers only candidates in which the subject and the object are both an argument function and a discourse function (e.g. TOP and FOC).

[11] Violations of HD-L are counted within VP. As argued convincingly in Sharma (1999), no arguments appear within VP in Hindi, whether in specifier or complement position, and all arguments are generated directly under S. In other words, a VP does not contain the verb and its complements. Instead, the only VP-internal elements are those which are preverbally focused.

(18) a. Input yielding OSV as the optimal output b. Input yielding OVS as the optimal output

$$
\begin{bmatrix}
\text{GF1} & \begin{bmatrix} \text{PRED 'Anu'} \\ \text{PROM --} \\ \text{NEW +} \end{bmatrix}_x \\[2ex]
\text{GF2} & \begin{bmatrix} \text{PRED 'moon'} \\ \text{PROM +} \\ \text{NEW --} \end{bmatrix}_y \\[2ex]
\text{ASP} & \text{PERF} \\
\text{PRED} & \text{'see (P-A}_x\text{,P-P}_y\text{)'}
\end{bmatrix}
\qquad
\begin{bmatrix}
\text{GF1} & \begin{bmatrix} \text{PRED 'Anu'} \\ \text{PROM --} \\ \text{NEW --} \end{bmatrix}_x \\[2ex]
\text{GF2} & \begin{bmatrix} \text{PRED 'moon'} \\ \text{PROM +} \\ \text{NEW --} \end{bmatrix}_y \\[2ex]
\text{ASP} & \text{PERF} \\
\text{PRED} & \text{'see (P-A}_x\text{,P-P}_y\text{)'}
\end{bmatrix}
$$

    (P-A=focus, P-P=topic)   (P-A=background, P-P=topic)

To summarize, the varied word orders in Hindi and their related interpretations find a natural analysis in the OT-LFG framework combining the ideas of imperfect correspondence and violable constraints. The next step is to explain why orderings in Hindi sentences with more marked types of subjects (i.e., non-volitional subject) are more restricted, and word order is even frozen in SOV order in the most marked situation (see (8) in 2.2).

4. Markedness and The Emergence of the Unmarked Word Order

In this section I demonstrate that the 'worst of the worst' type of the freezing effects in Hindi outlined in 2.2. follows naturally from the general model of harmonic alignment proposed in (Prince and Smolensky 1993) and adopted in Aissen (1997, 1999), an important source of constraints in OT. The formal definition of harmonic alignment is given in (19) (Prince and Smolensky 1993: 136).

(19)   Suppose a binary dimension D_1 with a scale $X > Y$ on its element {X, Y}, and another dimension D_2 with a scale $a > b > ... > z$ on its elements. The harmonic alignment of D_1 and D_2 is the pair of Harmony scales:

H_x: $X/a > X/b > ... > X/z$
H_y: $Y/z > ... > Y/b > Y/a$

The *constraint alignment* is the pair of constraint hierarchies:

C_x: $*X/z \gg ... \gg *X/b \gg *X/a$
C_y: $*Y/a \gg *Y/b \gg ... \gg *Y/z$

Harmonic alignment was introduced originally to express the relation between syllable position and sonority: the more prominent position (the nucleus) attracts segments which are more sonorous, while less prominent positions (the margins) attract less sonorous segments. This paper extends an idea first proposed in Aissen (1997) and developed further in Artstein (1998) and Aissen (1999), that harmonic alignment plays a central role in the domain of morphosyntax to express the relative markedness of different associations of morphosyntactic prominence hierarchies. Our concern here is the relative markedness which characterizes the associations of grammatical function (GF) with the dimensions such as semantic role and topicality. The basic idea is that subject function plays a role in the clause analogous to that played by the peak in syllable structure: it is the most prominent

478   Hanjung Lee

grammatical function and thereby attracts elements which are relatively prominent on other dimensions such as semantic role and topicality.

Now, we can apply the definition above to the three prominence scales in (20) – the GF scale, the sematic role scale and the discourse information scale. I will adopt the scale Subject > Nonsubject (Aissen 1999) in (20a) and the scale Proto-Agent > Proto-Patient (P-A > P-P) in (20b), where ">" means "more prominent than". For present purposes, I use the last two elements of the decomposed prominence scale $P\text{-}A_{vol} > P\text{-}A_{-vol} > P\text{-}P$ (Asudeh 1999). In the predicates I discuss here, the experiencer argument corresponds to a non-volitional Proto-Agent ($P\text{-}A_{-vol}$), and the theme argument to a Proto-Patient (Dowty 1991). Furthermore, I use the additional scale of discourse information in (20c) to capture the effects of the relative newness (topicality) of the arguments on word order in Hindi. It has long been noted that both discourse 'prominence/importance' and 'newness' are salient properties of topic. These two dimensions related to topicality need to be distinguished, but for present purposes, it suffices to distinguish topic from non-topic by their relative newness in discourse.

(20)   Universal Scales
a.   GF: SUBJ > Non-SUBJ
b.   Semantic Role: $P\text{-}A_{-vol} > P\text{-}P$
c.   Topicality: [–NEW] > [+NEW]

If the scales in (20) are harmonically aligned, we obtain the pairs of combined harmony scales in (21):

(21)   Harmony Scales derived through Harmonic Alignment
a. H_1: $SUBJ/P\text{-}A_{-vol} > SUBJ/P\text{-}P$
b. H_2: $Non\text{-}SUBJ/P\text{-}P > Non\text{-}SUBJ/P\text{-}A_{-vol}$
c. H_3:   SUBJ/[–NEW] > SUBJ/[+NEW]
d. H_4:   Non-SUBJ/[+NEW] > Non-SUBJ/[–NEW]

The first two harmony scales concern the association between grammatical function and semantic role, and assert that the unmarked situation is for subject to be Proto-Agent, and for object to be Proto-Patient (The connective ">" is read as "more harmonic than".). The last two harmony scales involve the alignment of the grammatical function hierarchy and the topicality hierarchy. The basic insight is that the unmarked situation is for subject to be old information, and for non-subject to be new information. The corresponding constraint alignment are the pairs of structural markedness constraint hierarchies in (22):

(22)   Constraint Subhierarchies
a. C_1: $*SUBJ/P\text{-}P \gg *SUBJ/P\text{-}A_{-vol}$
b. C_2: $*\sim SUBJ/P\text{-}A_{-vol} \gg *\sim SUBJ/P\text{-}P$
c. C_3: $*SUBJ/[+NEW] \gg *SUBJ/[-NEW]$
d. C_4: $*\sim SUBJ/[-NEW] \gg *\sim SUBJ/[+NEW]$[17]

Each subhierarchy in (22) expresses the universal markedness relation (e.g., a clause with a Proto-Patient subject will lose out to a clause with a Proto-Agent subject). The important property of the constraint hierarchies in (22) is that while the ranking of constraints within a subhierarchy is fixed (e.g., *SUBJ/P-P always outranks $*SUBJ/P\text{-}A_{-vol}$), they must be ranked with respect to other constraints.

However, showing that a Proto-Patient subject and a non-topical subject are more marked than a Proto-Agent subject and a topical subject is not enough, because they are still

allowed in Hindi. In order to capture the idea that if the subject is both a Proto-Patient and non-topical (hence occurring in a non-initial position at phrase structure), it is the worst of the worst, we can use the mechanism of *local conjunction* (Smolensky 1995: 4).

(23)   The Local Conjunction of C_1 and C_2 in domain D, C_1 & C_2, is violated when there is some domain of type D in which both C_1 and C_2 are violated. Universally, C_1 & C_2 » C_1, C_2.

To derive the pattern of universal markedness reflected in freezing effects in Hindi, let us consider conjunction of the two subhierarchies C_1 and C_3 in (22). This results in two new constraint subhierarchies in (24). The high-ranked constraint in (24a) expresses the basic idea that if subject is a highly marked Proto-Patient argument, it should not be associated with marked types of non-topical discourse functions. This most marked configuration excluded by this constraint obtains in a sentence like (6c), repeated here as (25), where the subject 'Nina' is a Proto-Patient (e.g., theme) *and* a focus, and is indeed the case of the worst of the worst that is not tolerated in Hindi.

(24)   Constraint Conjunction
a.   Conjoining *SUBJ/P-P with C_3:
       C_5: *SUBJ/P-P&*SUBJ/[+NEW] » *SUBJ/P-P&*SUBJ/[−NEW]
b.   Conjoining *SUBJ/P-A$_{-vol}$ with C_3:
       C_6: *SUBJ/P-A$_{-vol}$&*SUBJ/[+NEW] » *SUBJ/P-A$_{-vol}$&*SUBJ/[−NEW]

(25)   *Anuu-ko   Niinaa   uskii   bastii-me   dik^{h}ii.(*O$_{exp/top}$S$_{th/foc}$V)
       Anu-DAT   Nina-NOM   PRON-GEN neighbourhood-LOC   appear-PERF
       'Anu$_i$ saw Nina$_j$ in her$_{i/*j}$ neighbourhood.'

Recall from 2.2 that the theme-experiencer-V order is not possible even if the subject is canonically associated with a non-volitional Proto-Agent and the object with a Proto-Patient. Here the situation is one in which the unmarked linking of arguments to grammatical functions is employed, but the association between grammatical function and discourse function is marked. This situation, exemplified in (26), is excluded by the higher-ranked constraint in (24b), which expresses the intuition that the subject cannot be both a non-volitional Proto-Agent and new information (i.e., focus).

(26)   *Niina   Annu-ko   uskii   bastii-me   dik^{h}ii.(*O$_{th/top}$S$_{exp/foc}$V)
       Nina-NOM   Anu-DAT PRON-GEN   neighbourhood-LOC   appear-PERF
       'Anu$_i$ saw Nina$_j$ in her$_{j/*i}$ neighbourhood.'

Applying the same scheme to the conjunction of the subhierachy C_4 from (22) with the higher-ranked constraint *~SUBJ/P-A$_{-vol}$ from C_2, we get the following new subhierarchy in (27), which concerns the markedness of non-subject:

(27)   Conjoining *~SUBJ/P-A$_{-vol}$ with C_4:
       C_7: *~SUBJ/P-A$_{-vol}$&*~SUBJ/[−NEW] » *~SUBJ/P-A$_{-vol}$&*~SUBJ/[+NEW]

The higher-ranked constraint expresses the basic idea that the most marked situation obtains when the object is associated both with a marked semantic role (i.e., Proto-Agent) and with a marked discourse function (i.e., topic or background). The effect of this is that marked types of objects must occur in unmarked position, namely in immediately preverbal position, a position which licenses focus in Hindi.

480   Hanjung Lee

The ranking that emerges for Hindi is given in (28).

(28)   Ranking for Hindi:
        Conjoined markedness constraints:
        {*SUBJ/P-P&*SUBJ/[+NEW], *~SUBJ/P-A$_{-vol}$&*~SUBJ/[-NEW]} »
        *SUBJ/P-A$_{-vol}$ &*SUBJ/[+NEW] »
        Linking constraints : {*SUBJ/P-P, *~SUBJ/P-A$_{-vol}$} » Information structuring
        constraints » Canonical phrase structure constraints »

Crucially, the higher ranking of the three conjoined constraints on markedness of argument
types over the information structuring constraints (e.g., TOP-L and FOC-L) has the effect
of restricting the word order freedom motivated by the discourse prominence and newness
of arguments (see section 3): marked argument types (e.g., Proto-Patient subject and non-
volitional Proto-Agent object) must occur in unmarked position in the clause.

        Now, let's assume a discourse context in which the Proto-Agent is topic and the
Proto-Patient is focus. This particular context renders an input like (29). This input then
results in the optimal output (a1), going through the constraint competition in (30). In the
tableau, candidates are again schematically represented, and faithfulness constraints and
component constraints of the high-ranked conjoined constraints are omitted. Candidates
(a1) to (a6) are associated with the same f-structure, where the Proto-Agent argument
'Anu' is canonically mapped to the subject, and the Proto-Patient 'Nina' to the object.
Similarly, candidates (b1) to (b6) are paired with the same f-structure with the opposite
linking. Also, candidates labeled the same number have the c-structure string. For example,
both candidate (a1) and (b1) share the same string *Anuu-ko Niinaa dik^{h}ii*.

(29)   Input:
        $$\begin{bmatrix} GP1 & \begin{bmatrix} PRED & 'Anu' \\ PROM & + \\ NEW & - \end{bmatrix}_x \\ GP2 & \begin{bmatrix} PRED & 'Nina' \\ PROM & + \\ NEW & + \end{bmatrix}_y \\ ASP & PERF \\ PRED & 'appear (P-A_x, P-P_y)' \end{bmatrix}$$
        (P-A=topic, P-P=focus)

(30)   Tableau 2:   Linking and word order in Hindi non-volitional transitives

CANDIDATES:	*SUBJ-P-P/[+NEW]	*~SUBJ-P-A$_{-vol}$/[-NEW]	*SUBJ-P-A$_{-vol}$/[+NEW]	TOP-L	FOC-L	CANON$_{GF}$	Spine-R	HD-L
⇒a1. S$_{P-A}$O$_{P-P}$V				0	1		0	1
a2.*O$_{P-P}$S$_{P-A}$V			*!	1	0	*	0	1
a3. S$_{P-A}$VO$_{P-P}$				0	2		1	0
a4. O$_{P-P}$VS$_{P-A}$				2	0	*	1	0
a5. VS$_{P-A}$O$_{P-P}$				1	2		2	0
a6. VO$_{P-P}$S$_{P-A}$				2	1	*	2	0
b1.*O$_{P-A}$S$_{P-P}$V	*!	*!		0	1	*	0	1
b2. S$_{P-P}$O$_{P-A}$V				1	0		0	1
b3.*O$_{P-A}$VS$_{P-P}$		*!		0	2	*	1	0
b4.*S$_{P-P}$VO$_{P-A}$		*!		2	0		1	0
b5.*VO$_{P-A}$S$_{P-P}$		*!		1	2	*	2	0
b6.*VS$_{P-P}$O$_{P-A}$		*!		2	1		2	0

Due to the two high ranking constraints that penalize highly marked types of arguments, candidates (b1), (b3), (b4), (b5) and (b6) are ruled out immediately: (b1) has a focal Proto-Patient subject, violating *SUBJ/P-P&*SUBJ/[+NEW], and (b3), (b4), (b5) and (b6) have Proto-Agent objects which represent [–NEW] information (e.g., topic or background). Candidate (a2) does have a Proto-Agent subject and Proto-Patient object, thus not violating the two high-ranked conjoined constraints, but it is eliminated by *SUBJ/P-A_vol &*SUBJ/[+NEW], since it has a non-volitional Proto-Agent subject which is focal. Among (a1), (a3), (a4), (a5), (a6) and (b2), (a1) is the best; it satisfies more higher-ranking constraints than any other candidates. Thus, the constraint ranking in (30) correctly predicts that when the non-volitional Proto-Agent is topical and Proto-Patient is focal, the optimal output is one that realizes the non-volitional Proto-Agent as clause-initial subject and Proto-Patient as object, in the focus position.

In a discourse context in which the non-volitional Proto-Agent is focal and Proto-Patient is topical, the input is as (31). In this context, however, non-canonical linking becomes optimal under the same ranking, as illustrated in (32).

(31)   Input:

$$
\begin{bmatrix}
\text{GF1} \begin{bmatrix} \text{PRED} & \text{'Anu'} \\ \text{PROM} & + \\ \text{NEW} & + \end{bmatrix}_x \\[2em]
\text{GF2} \begin{bmatrix} \text{PRED} & \text{'Nina'} \\ \text{PROM} & + \\ \text{NEW} & - \end{bmatrix}_y \\
\text{ASP} \quad \text{PERF} \\
\text{PRED} \quad \text{'appear (P-A}_x\text{,P-P}_y\text{)'}
\end{bmatrix}
$$

(P-A=focus, P-P=topic)

(32)   Tableau 3:   Non-canonical linking becomes optimal under the same ranking

CANDIDATES:	*SUBJ-P-P/[+NEW]	*-SUBJ-P-A_vol/[-NEW]	*SUBJ-P-A_vol/[+NEW]	TOP-L	FOC-L	CANON_GF	Spine-R	HD-L
a1. $S_{P-A}O_{P-P}V$				1	0		0	1
a2. *$O_{P-P}S_{P-A}V$			*!	0	1	*	0	1
a3. $S_{P-A}VO_{P-P}$				2	0		1	1
a4. $O_{P-P}VS_{P-A}$				0	2	*	1	0
a5. $VS_{P-A}O_{P-P}$				2	1		2	0
a6. $VO_{P-P}S_{P-A}$				1	2	*	2	0
b1. *$O_{P-A}S_{P-P}V$	*!	*!		1	0	*	0	1
⇒b2. $S_{P-P}O_{P-A}V$				0	1		0	1
b3. *$O_{P-A}VS_{P-P}$		*!		2	0	*	1	1
b4. *$S_{P-P}VO_{P-A}$		*!		0	2		1	0
b5. *$VO_{P-A}S_{P-P}$		*!		2	1	*	2	0
b6. *$VS_{P-P}O_{P-A}$		*!		1	2		2	0

Thus, the constraint ranking proposed here accounts for the fact that in a clause with a P-P subject and a non-volitional P-A object word order is fixed as SOV for expressing the content in (31), capturing the basic generalization that highly marked argument types occur only in unmarked position.

In sum, I have argued that harmonic alignment in OT can fruitfully be applied in word order freezing found in Hindi: the 'worst of the worst' generalization follows naturally from the general model of harmonic alignment. I have also shown that local

482   Hanjung Lee

conjunction of markedness constraints is highly appropriate to dealing with the relative markedness of argument types and its interaction with word order.

5.   Conclusion

This paper has presented an OT-LFG account of 'the emergence of the unmarked order' in Hindi.[12] Marked associations of morphosyntactic hierarchies, which provide an important source of the 'worst of the worst' type of freezing, have been formally modelled as harmonic alignment and constraint conjunction: the most marked associations of grammatical function with other prominence hierarchies are expressed in the unmarked word order. Yet this generalization is not captured within most current formal syntactic frameworks, because they give no theoretical role to markedness, as opposed to purely structural syntactic aspects of grammar (e.g. transformational derivations). Furthermore, word order freezing effects in Hindi examined here show that concepts that have been successfully modelled in phonology—markedness hierarchies, harmonic alignment, etc.—also play a key role in the syntactic domain of constituent order.

References

Aissen, Judith. 1999. Markedness and Subject Choice in Optimality Theory. To appear in *NLLT* and Géraldine Legendre, Sten Vikner and Jane Grimshaw (eds.), *Optimality-theoretic Syntax*. Cambridge: MIT Press.

Aissen, Judith. 1997. Person and Subject Choice in Optimality Theory. Paper presented at the Hopkins Optimality Workshop, Inner Harbor, Baltimore, May 9-12, 1997.

Artstein, Ron. 1998. Hierarchies. MS., Rutgers University.

Asudeh, Ash. 1999. Linking, Optionality, and Ambiguity in Marathi: An Optimality Theory Analysis. MS., On-line, Stanford University: http://www.stanford.edu/~asudeh.

Bahl, K. C. 1967. *A Reference Grammar of Hindi*. Chicago: University of Chicago Press.

Bresnan, Joan. Forthcoming. *Lexical-Functional Syntax*. Oxford: Blackwell.

Bresnan, Joan. 1998. Optimal Syntax. To appear in Joost Dekkers, Frank van der Leeuw and Jeroen van de Weijer (eds.), *Optimality Theory: Phonology, Syntax and Acquisition*. Oxford: Oxford University Press.

Bresnan, Joan. 1994. Linear Order vs. Syntactic Rank: Evidence from Weak Crossover. In *CLS* 30-I.

Bresnan, Joan and Annie Zaenen. 1990. Deep Unaccusativity in LFG. In K. Dziwirek et al. (eds.), *Grammatical Relations. A Cross-Theoretical Perspective*: 45-57. Stanford: CSLI Publications.

Bresnan, Joan and Jonni Kanerva. 1989. Locative Inversion in Chichewa: A Case Study of Factorization in Grammar. *Linguistic Inquiry* 20: 1-50.

Butt, Miriam and Tracy Holloway King. 1996. Structural Topic and Focus without Movement. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG96 Conference*, On-line, CSLI Publications: http://csli-www.stanford.edu/publications.

Choi, Hye-Won. 1999. *Optimizing Structure in Context: Scrambling and Information Structure*. Stanford: CSLI Publications.

Dowty, David. 1991. Thematic Proto-roles and Argument Selection. *Language* 67: 547-619.

---

[12] Interestingly, 'the worst of the worst' type of word order freezing discussed in this paper are also found in Korean; see Lee (1999b) for an analysis of freezing effects in Korean in the present framework.

The Emergence of the Unmarked Order in Hindi   483

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

Grimshaw, Jane. 1997. Projection, Heads, and Optimality. *Linguistic Inquiry* 28: 373-422.

Grimshaw, Jane. 1991. Extended Projection. MS., Rutgers University.

Joshi, Smita. 1993. *Selection of Grammatical and Logical Functions in Marathi*. Doctoral Dissertation, Stanford University.

King, Tracy Holloway. 1995. *Configuring Topic and Focus in Russian*. Stanford: CSLI Publications.

Kiss, Katalin É (ed.). 1994. *Discourse Configurational Languages*. Oxford: Oxford University Press.

Kuhn, Jonas. 1999. Two Ways of Formalizing OT Syntax in the LFG Framework. MS., On-line, Universität Stuttgart: http://www.ims.uni-stuttgart.de/~jonas.

Lee, Hanjung. 1999a. The Emergence of the Unmarked Order. On-line, Rutgers University: Rutgers Optimality Archive. ROA-323-0699, http://ruccs.rutgers.edu/roa.html.

Lee, Hanjung. 1999b. The Emergence of the Unmarked Order in Hindi. Longer version of the paper presented at NELS 30. On-line, Stanford University: http://www-lfg.stanford.edu/lfg/archive.

McCarthy, John J. and Alan Prince. 1994. The Emergence of the Unmarked: Optimality in Prosodic Morphology. In *NELS* 24, 333-379. GLSA, University of Massachusetts, Amherst.

Mohanan, K.P. and Tara Mohanan. 1994. Issues in Word Order. In M. Butt, T.H. King and G. Ramchand (eds.), *Theoretical Perspectives on Word Order Issues in South Asian Languages*, 153-184. Stanford: CSLI Publications.

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

Mohanan, Tara. 1992. Word Order in Hindi. Paper presented at the Syntax Workshop, Stanford University.

Prince, Alan and Paul Smolensky. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. RuCCS Technical Report #2. Piscateway: Rutgers University Center for Cognitive Science.

Sells, Peter. Forthcoming. *Structure, Alignment and Optimality in Swedish*. Stanford: CSLI Publications.

Sells, Peter. 1999. Constituent Ordering as Alignment. To appear in Susumu Kuno et al. (eds.), *Harvard Studies in Korean Linguistics*, Vol. 8. Seoul: Hanshin Publishing Co.

Sharma, Devyani. 1999. Sentential Negation and Focus in Hindi. MS., Stanford University.

Smolensky, Paul. 1998. Why Syntax is Different (But Not Really): Ineffability, Violability and Recoverability in Syntax and Phonology. Handout of the talk given at the Stanford/CSLI OT Workshop. Stanford University, December 12-13, 1998.

Smolensky, Paul. 1996. On the Comprehension/Production Dilemma in Child Language. *Linguistic Inquiry* 27: 720-731.

Verma, Manindra K. and K.P. Mohanan (eds.). 1990. *Experiencer Subjects in South Asian Languages*. Stanford: CSLI Publications.

Department of Linguistics
Stanford University
Stanford, CA 94305-2150
USA

hanjung@csli.stanford.edu