

Chapter 5

Transfer Grammar

5.1 Introduction

Hindi is a relatively free word-order language. The relations between constituents are marked explicitly with the help of postpositions. The morphological richness of the language allows the constituents to change their positions freely within a clause [72, 33]. In spite of that, not all the words can occur freely at any position [72].

Like many other languages, English word order differs from Hindi. English belongs to a class of languages in which the head is mostly placed before its complement [105, 106, 39]. These languages are called head first languages. In such languages, verbs, prepositions, adjectives and nouns precede the complements associated with them [37, 38]. Hindi, on the other hand belongs to the class of head last languages. It means that the order of English verbs and their complements etc. would be just the reverse of Hindi. This can be seen clearly through the examples in 53. In 53, the head is shown in bold letters for English and their equivalent Hindi translations.

- (53) a. **write** a story
likha.IMPR kahānī
'kahānī **likhiye**'
- b. **in** the story
meṃ kahānī
'kahānī **meṃ**'
- c. **nice** to see you
acchā INF_MRKR milanā āpa
'āpa se milanā **acchā** [rahā]'
- d. **students** of English
vidyārthī.PL kā angrejī
'angrejī ke **vidyārthī**'

A. P. Anantpur (2009) argues that due to the lack of any explicit morpheme to mark accusative case, English freezes the subject and object positions to encode the grammatical relationships which brings

about several structural divergences between English and Hindi. Language divergence includes lexical, structural and conflational divergences [52]. D. Gupta and N. Chatterjee (2003) talk about structural and categorial divergences for English-Hindi MT system from EBMT perspective. In this chapter, we will describe the structural divergences like head first vs head last, subject-verb-object (SVO) vs subject-object-verb (SOV) etc. for English-Hindi Anusāraka which is a rule based MT system. In Anusāraka, categorial divergences are handled through various WSD mechanisms which are not covered in this thesis. Table 5.1 summarizes the major structural divergences between English and Hindi.

English	Hindi
SVO	SOV
head first	head last
prepositional	postpositional
subject is sacrosanct	subject may be dropped

Table 5.1: Major structural divergences between English and Hindi

In order to handle the major structural differences shown in Table 5.1 and many more described in Section 5.4.1, besides appropriate lexical substitutions, reordering of words plays a crucial role in machine translation.

Statistically trained systems give good results in less amount of time and manual effort. Statistical systems work well upto a point, but when it comes to reordering in long sentences they start to fail. On the other hand, in comparison to statistically extracted rules, hand-crafted rules give better results for reordering and morphological tasks [41]. Apart from that, in order to overcome some of the challenges, any statistically trained system requires large amount of parallel corpus. Indian languages lack a reasonable parallel corpus size for English and Indian language pairs. On the other hand, given the nature of English and Hindi as mentioned in terms of head first and head last languages, the reordering can be handled by just a few rules. Therefore, we explore a rule based approach for English-Hindi word ordering, where we try to address reordering by using concepts of *pada* and *samasta-pada* and other insights from PG. However, there are limitations to this approach as well.

5.2 Our Approach for Transfer Grammar

We present a rule based approach which is based on PG, particularly on the concept of *pada* and *samasta-pada*.

Through example 30 shown in chapter 4, we have seen that it is both the root and the affix that jointly denote the meaning¹. We have seen that the *vibhaktis*, the inflectional affixes are a mechanism to

¹*prakṛtipratyayau sahārtham brūtaḥ*

mark the desired information explicitly. Recall that while defining *pada* and *samasta-pada* (described in Chapter 3) for English, we have identified the overt and covert mechanisms such as prepositions, auxiliaries and position of the words with respect to other words like subject and object positions that English uses as *vibhaktis* to mark the grammatical information. Thus the notion of *pada* and *samasta-pada* helps us make the implicit information explicit. Once the implicit information is made explicit, we get tremendous freedom for word ordering.

Our main focus while reordering a source language sentence into target language sentence is faithful transference of the source language information into the target language. This means that we might occasionally compromise with the fluency of the target language word order but not with the loss of information.

Apart from structural transfer, we will also explore how this approach can help in target language word generation. We will describe the usefulness of the notion of *pada* with respect to handling of direct and oblique cases in Hindi.

We would also like to mention that though this chapter presents an English-Hindi reordering approach, we claim that the same system is generic enough to be used for any English-Indian language pair with some modifications.

5.3 Representation of *Padas* and Other Grammatical Information Associated with Them

After identification of the *padas* and *samasta-padas* in source language, English in our case, the *padas* are then manipulated according to the target language syntactic structure. Henceforth, we will refer to both *padas* and *samasta-padas* as *padas* only, until and unless some specific information pertaining to them is to be mentioned.

Inside the System, the *padas* are represented in CLIPS' (C Language Integrated Production System) facts [58]. A fact is a piece of information. Facts can be ordered or unordered in CLIPS. Ordered facts are good for single slot facts. In ordered facts, the values of the fields are referenced positionally. The first field is a relation name which applies to the remaining fields in an ordered fact, as we have already seen through the relation labels in chapter 4. In order to access the information in an ordered facts, the knowledge of the position of the field also becomes important [42, 57, 58].

An unordered fact is also called a deftemplate in CLIPS. In a non-ordered fact atomic values are referenced by field or slot names which can be abstracted using the field names. A deftemplate is similar to the record or structure definition in Pascal and C [42, 57, 58].

In our system, the *padas* are represented using deftemplates. Enclosed in opening and closing parenthesis, a deftemplate construct starts with the name of the template along with zero or more fields or slots. Similarly, a slot, enclosed in an opening and closing parenthesis, contains the name of the slot fol-

lowed by zero or more fields. Since the deftemplates are non-ordered facts, the slots may be referenced in any order by their names. Let us look at the *pada* templates for sentence 54.

(54) Queen Victoria opened Blackfriars Bridge in November 1869.

We know that sentence 54 contains four different *padas*:

1. 'Queen Victoria', a *subanta pada* which contains subject position as a *vibhakti*
2. 'opened', a *tinanta pada* which contains verbal inflection '-ed' as a *vibhakti*
3. 'Blackfriars Bridge', a *subanta pada* which contains object position as a *vibhakti*, and
4. 'November 1869', a *subanta pada* which contains the preposition 'in' as a *vibhakti*

So, there would be four different unordered facts representing these *padas*. Each fact consists of a fact name 'pada.info' and multiple slots with slot names associated with their values. The facts representing the *padas* for sentence 54 would look like follows:

- (pada.info (pada.head opened²) (pada.type tinanta) (pada.members opened) (vibhakti -ed) (gender 0) (number 0) (case 0) (person 0))
- (pada.info (pada.head Victoria) (pada.type subanta) (pada.members Queen Victoria) (vibhakti subject_position) (gender 0) (number 0) (case 0) (person 0))
- (pada.info (pada.head Bridge) (pada.type subanta) (pada.members Blackfriars Bridge) (vibhakti 0) (gender 0) (number 0) (case 0) (person 0))
- (pada.info (pada.head 1869) (pada.type subanta) (pada.members November 1869) (vibhakti in) (gender 0) (number 0) (case 0) (person 0))

Notice that there are many slots in these facts like (gender 0), (number 0), (case 0), (person 0) and many more not shown here for simplicity. These slots carry all the grammatical information that is needed for transferring the source language grammatical structures into the target language, including the order of the members of the *padas* that is necessary for target language word generation. Initially, most of the slot values are '0' (zero). These values get filled by the subsequent modules.

Also make a note that the *vibhaktis* which are marked through various mechanisms such as the explicit ones like prepositions, postpositions, verbal inflections and auxiliaries or through various implicit mechanisms such as pre-verbal and post-verbal positions like subject and object positions in the sentence, are all mentioned explicitly in the 'vibhakti slot'. This provides the system an ability for transferring the syntactic structure of one language into another faithfully. So that if the target language is morphologically richer than the source language like Hindi and English, the target language can substitute appropriate *vibhaktis* to the *padas* without any fear of ungrammaticality and loss of information.

²We are showing the words here for simplicity, but system internally, the words are represented with their ids.

5.4 Formation of Reordering Rules

Due to the structural divergences like ‘head first’ vs ‘head last’, as discussed in Section 5.1, Hindi follows ‘mirror structure’ of English in terms of structural word order. It means that the daughters of English verb phrase (VP) come in mirror image when translated into Hindi. In other words, the SL VP looks like the mirror image in the TL VP and vice versa.

The SL to TL reordering happens in three stages in our reordering module. A sentence need not pass through all three stages. In most of the cases, we achieve a fluent target language word order at ‘Stage 1’ itself. Such sentences also passed through all three stages, but these sentences do not undergo any further reordering because no rules trigger in the later stages. The flow chart in Figure 5.1 gives an overview of these three stages.

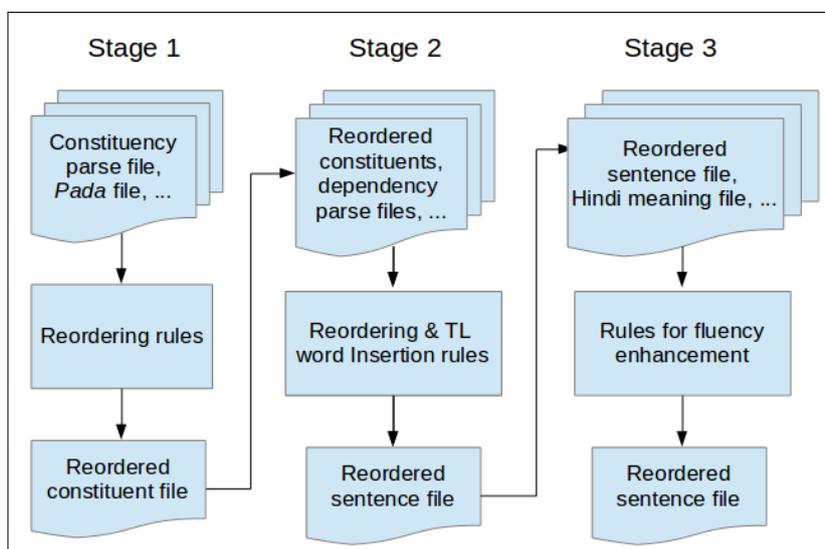


Figure 5.1: Flow chart illustrating the three stage reordering process

In this section, we will describe some of the major operations involved in these stages.

1. **Stage 1:** At ‘Stage 1’, the input to the reordering rules is the ‘*pada-information* and constituency parse of the input sentence. At this input, a set of reordering rules that are written in CLIPS are run. These rules manipulate the SL parse tree according to the TL word order and produce re-ordered constituents. These rules mainly do mirroring of the English structures according to Hindi sentence structure. For instance, let us take the sentence 54 to illustrate the ‘mirror structure’, and look at its constituency parse tree shown in Figure 5.2.

The mirror image of the VP of the constituency parse given in Figure 5.2 is shown in Figure 5.3. In 5.3, the information related to the *padas* is enclosed within opening and closing parentheses where ‘sup.GV.subj’ stands for *subanta pada* which has a ‘generalized vibhakti’ as subject position, ‘sup.GV.obj’ stands for *subanta pada* which has a ‘generalized vibhakti’ as object position and

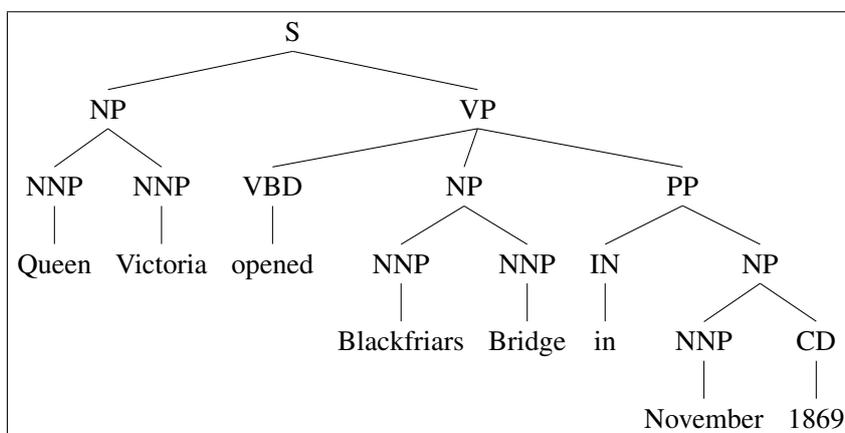


Figure 5.2: Showing constituency parse for 54

‘tiñ.-ed’ stands for *tiñanta pada* which has the verbal inflection ‘-ed’ as a *vibhakti*. The words that fall under the *pada* information enclosed in parentheses maintain the order similar to English, except the *vibhaktis* which are moved to the end of the *pada* in Hindi. In few cases, ‘intra-pada movement’ also needs to be done. Such cases are described in Section 5.4.1. Otherwise mirror image of the VP gives a perfectly ordered Hindi sentence.

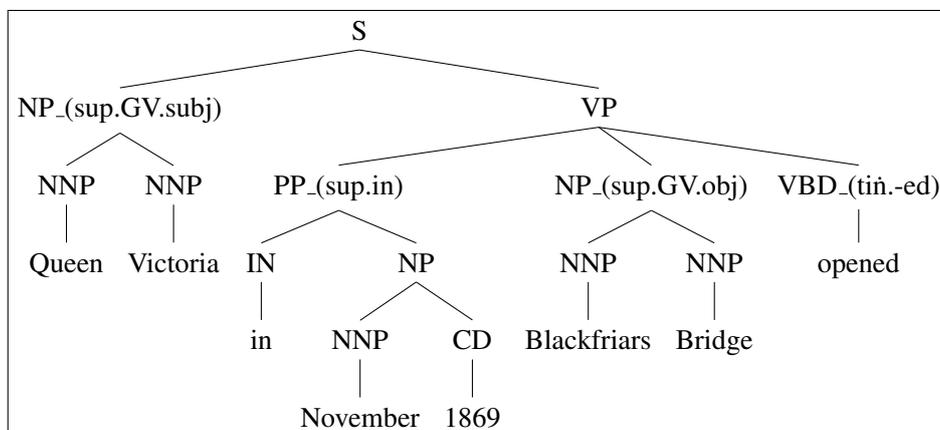


Figure 5.3: Showing mirror image of the constituency parse for 54

2. **Stage 2:** The output file of ‘Stage 1’ and dependency parse file become the input to the ‘Stage 2’. A second set of reordering rules is applied at this stage that rearranges the constituents for fluency enhancement. For example, when mere ‘mirror structure’ does not suffice to arrive at a fluent Hindi word order, we take advantage of dependency parse to handle such peculiarities in order to put the phrase at the desired place in TL. For example, indirect object tends to precede the direct object in Hindi. In 55a, the indirect and direct object reversal is restricted using dependency relations, since 55a sounds more natural than 55b to the native speakers of Hindi.

- (55) a. Hari ne use eka billī dī
 Hari.PROPN,NOM he.DAT a cat.ACC,SG give.PT,SG,FEM
 ‘Hari gave her a cat.’
- b. Hari ne eka billī use dī
 Hari.PROPN,NOM a cat.ACC,SG he.DAT give.PT,SG,FEM

The other important task that happens at this stage is insertion of the words that are not present in the source language such as insertion of relative pronouns and subordinating conjunctions. For example, in sentence 56, relative pronoun *jise* (who) has to be inserted in Hindi translation.

- (56) Rama saw the man you love.
 Rama.NOM dekha.PT vaha ādamī.SG āpa prema kara.PR
 *Rama ne vaha ādamī dekhā āpa prema karate haiṃ.
 ‘Rama ne vaha ādamī dekhā **jise** āpa prema karate haiṃ.’

3. **Stage 3:** In some cases reordering is done based on the target language translation. For instance, let us look at the translation of the English verb ‘want’ in sentences 57a and 57b. In sentence 57a, the verb ‘want’ is translated into a single word as ‘*cāha*’. On the other hand, in sentence 57b, the same verb is translated into a conjunct verb, a noun and a verbalizer as ‘*icchā kara*’. In such cases placement of negation morpheme ‘*nahī*’ varies according to the TL word translation. In 57a, it precedes the verb ‘*cāha*’. Whereas in sentence 57b, it precedes the verbalizer.

- (57) a. She does not **want** this.
 vaha.FEM.SG AUX.3,SG nahīṃ **cāha** yaha.SG
 ‘vaha ise nahīṃ **cāhatī**.’
- b. She does not **want** this.
 vaha.FEM.SG AUX nahīṃ **icchā kara** yaha.SG
 ‘*vaha isakī nahīṃ **icchā karatī** hai.’
 ‘vaha isakī **icchā** nahīṃ **karatī** hai.’

5.4.1 Some Special Cases

Apart from the cases described in Section 5.1, there are constructions which are not so flexible in terms of word order in Hindi. Therefore, special attention should be given to them. This section describes some such examples.

- **Starting Point and End Point:** While talking about a range or a span, the phrase describing an initial point precedes the phrase that describes the end point. This phenomenon seems to be common in all languages. Therefore, the order of such constructions should not be changed in target language. Only proper attachment of the inflections would be sufficient. For instance, the boldfaced expressions shown in 58 are some such examples from English. These would maintain the same order in Hindi as well.

- (58) a. In Kashmir, fishing is a good business and the ideal season is **from April to October**.
 b. I shall continue to work **from 6 a.m. till midnight**, even if it kills me.
 c. **October to March** is the best time to visit the Jaipur city.
 d. In the south, Jammu is a transition zone **from the Indian plains to the Himalayas**.

- **Relative Order of Arguments of a Verb:** The internal order of theme and recipient roles in English can be expressed in two ways: (i) through position as shown in 59 and (ii) through preposition as shown in 60.

(59) John gave **Harry** a book.

(60) John gave a book **to Harry**.

In Hindi, 59 and 60 both the sentences can be translated as in 61 and 62 respectively. Since Hindi is morphologically rich, it marks relations through morphemes. Therefore, as mentioned above, the order can be relatively free. Hence, both the sentences are possible.

(61) John ne **Harry ko** eka kitāba dī.

(62) John ne eka kitāba **Harry ko** dī.

However, 61 sounds more natural. This is the default order in Hindi. This is also consistent with the SOV parameter that the direct object tends to precede the verb in Hindi.

- **Source and Destination Phrases:** Source and destination phrases tend to occur immediately before their verbal head as shown in 63 and 64. For instance, “*vaha use skūla le gayā*” sounds more natural than “*vaha skūla use le gayā*” in Hindi.

(63) He **took** her **to the school**.

vaha **le gayā** use **skūla**

‘vaha use **skūla le gayā**.’

(64) He **picked** her **from the party**

vaha **le āyā** use **se pārtī**

‘vaha use **pārtī se le āyā**.’

However, if source and destination both occur as arguments of the same verb, then destination phrase tends to occur immediately before the verb as shown in 65.

(65) He **took** her **home** from the party.

vaha **le āyā** use **ghara** se **pārtī**

‘vaha use pārtī se **ghara le gayā**.’

- **Marking Negation:** The lexical items ‘no’, ‘not’ and ‘never’ mark negation in English. ‘No’ precedes nouns, ‘not’ comes between verbs and auxiliaries and ‘never’ can either negate actions or adjectives in English. However, when translated into Hindi, they all tend to precede the verbs. ‘No’ translates into **koi nahīm**. Where **koi** stands with the nominal head but **nahīm** leaves the nominal head and sits immediately before the verb as shown in 66.

- (66) a. No politician is completely honest.
koi nahīm rājanetā hai pūrī taraha se īmānadāra
 ‘***koi nahīm** rājanetā pūrī taraha se īmānadāra hai’
 ‘**koi** rājanetā pūrī taraha se īmānadāra **nahīm** hai’
- b. We seek no reward.
 hama cāhate haiṃ **koi nahīm** ināma
 ‘*hama **koi nahīm** ināma cāhate haiṃ’
 ‘hama **koi** ināma **nahīm** cāhate haiṃ’
- c. Bringing up a child is never easy.
 pālanā - baccā.ACC ho.PR **koi nahīm** āsāna
 ‘*baccā pālanā **koi nahīm** āsāna [kāma] hai.’
 ‘baccā pālanā **koi** āsāna [kāma] **nahīm** hai.’
- d. He never really recovered from the shock.
 vaha **kabhī nahīm** vāstava meṃ ubhara pā.PT se sadamā
 ‘*vaha **kabhī nahīm** vāstava meṃ sadame se ubhara pāyā’
 ‘vaha vāstava meṃ sadame se **kabhī nahīm** ubhara pāyā’

Some times an English verb translates into a conjunct verb in Hindi. A conjunct verb is a combination of a nominal and a verbalizer like *karanā*, *honā*, *lenā*, *denā* etc. [9]. In such cases, the negation marker comes before the verbalizer, the helping verb, not before the nominal (*kriyāmūla*) as shown in 67.

- (67) He did not wait for me.
 vaha PAST **nahīm pratīkṣā kara** liye mere
 ‘*usane mere liye **nahīm pratīkṣā kī**’
 ‘usane mere liye **pratīkṣā nahīm kī**’

- **Yes/No Interrogation:** The yes/no interrogative morpheme is missing in English [3]. Therefore, it inverts the positions of subject and verb/auxiliary to mark yes/no interrogation. Whereas in Hindi, yes/no interrogative question marker *kyā* is lexicalized and normally occurs at sentence initial position. Therefore, One needs to identify the yes/no question type sentence and then insert the morpheme *kyā* at appropriate position in Hindi, as shown in 68.

(68) Did you eat?
 PAST,yes/no āpane khā
 ‘**kyā** āpane khāyā?’

The word *kyā* should not be positioned immediately before the verb, because, Hindi also uses the word *kyā* as a wh-question morpheme which precedes the verb as shown in 69.

(69) Apane **kyā** khāyā?
 you.NOM **what.ACC** eat.PT
 ‘**What** did you eat?’

- **Intra-pada Movement:** While translating an English *pada* into Hindi, no movement takes place among the components of a *pada*. The components are transferred into Hindi in the same English order. But, sometimes movement within a *pada* is required according to the target language word order. We call this ‘intra *pada* movement’ or ‘within *pada* movement’. For example, in sentence 70, the positions of the words ‘all’ and ‘those’ have been interchanged according to Hindi word order.

(70) All those girls play chess.
 saba ve ladakiyāṃ khela śataranja
 ‘ve saba ladakiyāṃ śataranja khelatī haiṃ.’

- **Merging Padas:** In some cases, while translating English *padas* into Hindi, two *padas* form a single *samasta-pada*. For example, in 71, ‘females’ and ‘of birds’ are two different *subanta padas*. But, in Hindi, they form a single *samasta-pada* as ‘female birds’ (*mādā pakṣī*) and the internal preposition *of* gets deleted.

(71) Females of birds produce eggs.
 mādā kā cidiyā.PL de.PR.3.SG aṇḍā.PL
 ‘mādā cidiyā aṇḍe detī haiṃ.’

- **Information Sharing and Repetition:** We noticed an interesting phenomenon where the shared morphemes have to be repeated during translation in Hindi sometimes or morphemes that have been repeated in English do not repeat themselves in Hindi. For instance, in 72a, the verb ‘swim’ is shared by the *tin* inflections ‘can’ and ‘will’ in English. But in Hindi, the shared verb *swim* has to be repeated to make two *tinanta padas* with each *tin*, ‘*taira sakatī hai*’ (can swim) and ‘*tairegī*’ (will swim). In sentence 72b on the infinitive marker ‘to’ is shared with ‘observe’, ‘look’, ‘build’ and ‘use’ in English, whereas, it has to be repeated with every verb in Hindi.

(72) a. She can and will swim.
 vaha.FEM,3,SG **BE ABLE TO** aura **FT taira**
 ‘Vaha **taira sakatī hai** aura **tairegī**.’

- b. One kind of response from the earliest times has been **to observe** the physical environment carefully, **look** for any meaningful patterns and relations in natural phenomena, and **build** and **use** new tools to interact with nature.

Let us discuss another phenomenon of morpheme sharing where English repeats only some portion of the *tin* but Hindi does not do that. For instance, in sentence 73, “is -ing” together behaves as a verbal inflection (*tin vibhakti*) in English where only the auxiliary ‘is’ is being shared but the inflection ‘-ing’ is repeated with every verb. Whereas the equivalent *tin* is not repeated in Hindi translation.

- (73) The God **is** **seeing**, **listening** and
 paramātmā *AUX* dekha.**PR_CONT,3,SG** suna.**PR_CONT,3,SG** aura
 knowing everything.
 jāna.**PR_CONT,3,SG** sabakucha
 ‘Paramātmā sabakucha dekha, suna aura jāna **rahā hai**

What we notice from all these English sentences and their Hindi translations, is that if a morpheme forms one word with another morpheme, then, it has to be repeated every time, otherwise, it can be shared with other morphemes.

Section 5.4.2 and 5.4.3 describe some important reordering rules which are sufficient to give an overview of our approach.

5.4.2 Rules Based on ‘Mirror Structure’

The rules based on ‘mirror structure’ are very productive. These rules give around 70% constituent reordering accuracy. Inside the system, the rules are written in CLIPS where higher precedence rules are given higher salience. Thus, salience assigns priority to the rules. In general, specific rules have higher salience than a general rule. By default, the salience value is 0 (zero) which need not be specified in the rules. In this section, we will describe the pseudo-code of some of the rules.

The rules are written using the following convention: *[.mother child1 child2 child3] → [.mother child3 child2 child1]*, where the part before ‘→’ denotes SL parse tree and the later part is the manipulation of the SL parse tree according to the TL word order. The item marked with dot (.) represents the mother node followed by its child nodes. We have followed Penn tag-set to represent constituents. See Appendix B for the description of the tags.

Based on the two major types, the rules are named as follows: *Rules Based on Mirror Structure (RBMS) plus a Number*, e.g. RBMS1, RBMS2 etc. and *Rules for Restricting Mirror Structure (RRMS) plus a Number*, e.g. RRMS1, RRMS2 etc. The rules are illustrated in following order: we give the pseudo-code of a rule, its brief description and then an example or two describing the rule are given with glosses and translations.

- **RBMS1:** VP is reversed to allow ‘mirror structure’ as discussed in Section 5.2. See example 54, repeated here as 74 for convenience.

(74) Queen Victoria **opened** **Blackfriars Bridge** **in** **November** **1869**.
 mahārānī Victoria **udghātana kara**.PT **Blackfriars Bridge** **meṃ** **navambara** **1869**
 ‘mahārānī Victoria ne **navambara 1869** **meṃ** **Blackfriars Bridge** **kā** **udghātana** **kiyā**.’

- **RBMS2:** [.ADVP RB NP] → [.ADVP NP RB]

RB should be the last child of an ADVP.

(75) ... the Dow was **down about** **35 points**.
 ... - Dow thā **nīce** **lagabhaga** **35 aṃka**
 ‘... Dow **lagabhaga 35 aṃka** **nīce** thā.’

- **RBMS3:** [.SBAR IN S] → [.SBAR S IN] iff *IN* != sentential conjunction

An SBAR, introduced by a preposition is reversed as shown in sentence 76, except when it is not a sentential conjunction as in sentence 77.

(76) I expect a rough market **before prices** **stabilize**.
 maiṃ āśā kara.PR,3,SG **asthira** **bājāra** **pahale mūlya**.PL **sthira** **ho**.PR,3,PL
 ‘maiṃ **mūlya sthira hone se** **pahale** **asthira** **bājāra** **ki** **āśā** **karatā** **hūṃ**’

(77) It was so dark **that** I could not see anything.
 thā itanā andherā **ki** maiṃ saka **nahīṃ** dekha.PT kucha
 ‘itanā andherā thā **ki** maiṃ kucha **nahīṃ** dekha **sakā**’

- **RBMS4:** [.NP [.NP *] [.PP *] [.* *]] → [.NP [.* *] [.PP *] [.NP *]]

A ‘noun phrase plus prepositional phrase post-modifier’ sequence is reversed with respect to the head initial and head final nature of English and Hindi.

(78) The boy **in blue shirt** is here.
 ladakā.SG **vālā nīlī** **kamīja** ho.PR,3,SG yahāṃ
 ‘**nīlī kamīja vālā** **ladakā** **yahāṃ** **hai**’

5.4.3 Rules for Handling Exceptional Cases

Even after applying ‘mirror structure’ on VP, some cases remain non-fluent and/or incomprehensible because of the inflexibility of some of the constituents to precede or follow the other constituents as pointed out in Section 5.1. To put it in other words, in some case, ‘mirror structure’ leads to over reordering of some of the constructions which tend to have fixed word order. This section describes some of the rules that are exceptions to ‘mirror structure’.

- **RRMS1:** [.VP * V] → [.VP * V]

Since Hindi is a verb final language, we restrict VP inversion iff verb is its last child, as in 79.

- (79) The prices of winter wheat **now being planted** will not fall soon.
 mūlya.PL kā śarada geṃhūṃ **aba -jā rahe bo.PSSV FUT** nahim gira jaldī
 ‘**aba boye jā rahe** śaradīya geṃhūṃ ke mūlya jaldī nahim gireṃge’

- **RRMS2:** [.ADJP RB JJ PP] → [.ADJP PP RB JJ]

An ADJP having a prepositional phrase modifier reversed keeping the head last nature of Hindi.

Ex.

- (80) She is **very good at her** work.
 vaha.FEM,SG ho.PR,3.SG **bahuta acchā meṃ apanā.FEM,3,GEN kārya**
 vaha **apne kārya meṃ bahuta acchī** hai

- **RRMS3:** [.SQ WHNP VBD NP VP] → [.SQ VBD NP WHNP VP]

Wh-constituent in a sentence/clause should be placed before the verb of the sentence/clause.

- (81) **How many people** did you see?
kitane loga PT āpa mila
 ‘āpa **kitane logom** se mile?’

5.4.4 Illustration of Reordering Rules

An overview of the reordering rules is given in Table 5.2 through sentence 82.

- (82) In 2006, the Lebanese Hezbollah militia seemed able to rain rockets at
 meṃ 2006, Lebānī Hijabullāha ladāke dikha.PT sakṣama barasānā rauketa.PL para
 will during a monthlong conflict with the Jewish state.
 icchā ke daurana mahīnā bhara lambā saṃgharṣa ke sātha Yahūdī rājya
 ‘2006 meṃ, Lebānī Hijabullāha ladāke Yahūdī rājya ke sātha mahīnā bhara lambe saṃgharṣa
 ke daurana icchānusāra rauketa barasāne meṃ sakṣama dikhe.’

In Table 5.2, the first column lists the steps. The steps can increase or decrease based on the number of the rules fired and the length of the input sentence. The second column shows the reordering rules or procedure. A rule need not fire only once in a sentence. A rule can fire as many times as it meets the required conditions. In 82, the rule, RBMS1 has fired three times. The third column shows the effect of that particular rule on the given input.

S0	Input	(S (PP (IN In) (NP (CD 2006))) (, ,) (NP (DT the) (JJ Lebanese) (NNP Hezbollah) (NN militia)) (VP (VBD seemed) (ADJP (JJ able) (S (VP (TO to) (VP (VB rain) (NP (NNS rockets)) (PP (IN at) (NP (NN will)))))) (PP (IN during) (NP (NP (DT a) (JJ monthlong) (NN conflict)) (PP (IN with) (NP (DT the) (JJ Jewish) (NN state)))))))))))
S1	Pada formation	(S 2006 In , the Lebanese Hezbollah militia (VP (VBD seemed) (ADJP (JJ able) (S (VP (VP (VB rain) to (NP (NNS rockets)) (PP will at)) (PP (NP (NP a monthlong conflict during) (PP (NP the Jewish state with))))))))))
S2	RBMS1	(S 2006 In , the Lebanese Hezbollah militia (VP (VBD seemed) (ADJP (JJ able) (S (VP will at rockets rain to (PP (NP (NP a monthlong conflict during) (PP (NP the Jewish state with))))))))))
S3	RBMS1	(S 2006 In , the Lebanese Hezbollah militia (VP (VBD seemed) (ADJP (JJ able) (S (NP (NP a monthlong conflict during) (PP the Jewish state with)) will at rockets rain to)))
S4	RRMS2	(S 2006 In , the Lebanese Hezbollah militia (VP (VBD seemed) (ADJP (S (NP (NP a monthlong conflict during) (PP the Jewish state with)) will at rockets rain to able)))
S5	RBMS1	(S 2006 In , the Lebanese Hezbollah militia (NP (NP a monthlong conflict during) (PP the Jewish state with)) will at rockets rain to able seemed)
S6	RBMS4	2006 In , the Lebanese Hezbollah militia the Jewish state with a monthlong conflict during will at rockets rain to able seemed
S7	lexical substitution	2006 meṃ, Lebānī Hījabullāha ladāke Yahūdī rājya ke sātha mahīnā bhara lambe saṃgharṣa ke daurana icchānusāra rauketa barasāne meṃ sakṣama dikhe

Table 5.2: Illustration of the reordering rules

5.4.5 Target Language Word Generation and the Notion of *Pada*

Apart from word ordering, one of the important tasks in MT is correct target language word generation. If the forms of the TL words are not generated correctly, the sentence translation may sound very bad even after correct word ordering. In this section, we would see how the concept of *pada* and *samasta-pada* can help in TL word generation.

As mentioned before, we make the implicit *vibhaktis* that are marked through positions of the participants explicit while formation of the *padas*. This helps in insertion of appropriate target language *vibhaktis* and generation of the oblique or direct case accordingly.

Identification of the *samasta-padas* and *ardhasamāsas* in source language and their structural counterparts in the target language give tremendous power for target language word generation.

In chapter 3, we have said that all languages show spectrum of flexibility in compounding, Let us see this using Hindi expressions as well. This spectrum in Hindi is quite similar to English. Let us explore it through some examples.

Similar to Sanskrit, nouns and adjectives inflect for case (direct and oblique) and number (singular and plural) in Hindi. In the presence of a *vibhakti*, nouns and adjectives inflect for number and case both, but in the absence of a *vibhakti* they inflect only for number. For instance, the noun *ladakā* (boy) is seen with following forms: *ladakā* (direct singular), *ladake* (direct plural), *ladake ne* (oblique singular), and *ladakoṃ ne* (oblique plural).

Similarly, the inflected forms for the adjective *acchā* (good) are: *acchā* (direct singular), *acche* (direct plural) *acche ne* (oblique singular), and *acchoṃ ne* (oblique plural). The paradigms shown in Table 5.3 and 5.4 give an overview of the inflected forms of these two words respectively.

Table 5.3: Paradigm table for the noun *ladakā*

Case	Number	
	Singular	Plural
Direct	<i>ladakā</i>	<i>ladake</i>
Oblique	<i>ladake</i>	<i>ladakoṃ</i>

Table 5.4: Paradigm table for the adjective *acchā*

Case	Number	
	Singular	Plural
Direct	<i>acchā</i>	<i>acche</i>
Oblique	<i>acche</i>	<i>acchoṃ</i>

Having looked at the paradigm tables, let us examine some complex expressions by combining these two words as shown through examples in 83.

- (83) a. *acchā* *ladakā*
good.SG,DIR boy.SG,DIR
‘good boy’
- b. *acche* *ladake*
good.PL,DIR boy.PL,DIR
‘good boys’
- c. *acche* *ladake* *ne*
good.SG,OBL boy.SG,OBL NOM_MRKR
‘good boy’

- d. *acche ladakoṃ ne*
 good.PL,OBL boy.PL,OBL NOM_MRKR
 ‘good boys’

The expressions 83a and 83b have no *vibhaktis*, thus these are simple substitutions of the words from the paradigm tables. But as we move to 83c and 83d some interesting grammatical phenomenon is found. There too, our focus of attention is expression 83d.

The forms of the adjective *acchā* are same in both 83c and 83d. But, as we know, it is a modifier of the head noun *ladakā*, its number and case should be in alignment with the head. As per the paradigm table 5.4, shouldn't it be *acche* in 83c and *acchoṃ* in 83d?

In order to understand this, let us first explore what type of expressions are these. Are these compound expressions or non-compound expressions or something else?

Since, these expressions are composed of more than one word, and each expression has only one postposition attached to them, their behavior appears to be quite similar to compounds. Therefore, let us pass these expressions through the compound tests.

Since no overt *vibhakti* (postposition) is present after the adjective *acche*, we can say that *sublopa* (elision of internal *sup*) has taken place in these expressions.

Placing an external element such as verbs, adverbs etc., makes the expressions ungrammatical or sometimes, brings some different meaning to the expressions, for instance, **acche bolā ladake ne* or **acche bolā ladakoṃ ne*, hence, there is *avyavadhāna* (unintervention by other words, *pada*).

Changing the order of the words **ladake acche ne* or **ladakoṃ acche ne* makes the expressions ungrammatical, hence, there is *niyatapaurvāparya* (fixed word order).

“The question of ‘stress’ in Hindi is rather controversial” [98], thus, stress cannot be justified as a characteristic property for Hindi.

From these observations, we argue that expressions like the ones shown in 83 are a kind of compound expressions. But then also, the mystery of the forms of the adjective *acchā* in 83c and 83d is not yet resolved. If we go by the Table 5.4, the adjective *acche* can be either a direct plural form or oblique singular form of *acchā*. But, as we have already said, elision of the internal *sup* (postposition) has taken place in these expressions, by the principle of *pratyayalakṣaṇa* the operations pertaining to *sup* should take place. Recall that *pratyayalakṣaṇa* is the process for operations pertaining to a suffix that has been deleted. Therefore, the words comprising such expressions should be in oblique case. The information regarding the number would be inferred from the head *ladakā*.

These expressions exhibit characteristics of compounds, but also maintain multiple word status. Such expressions lie between the border of a simple word group and a compound. Hence, we say these are partly like compounds. Therefore, we call these *ardhasamāsa* (quasi-compounds).

This also brings into our notice the flexibility spectrum of expressions in Hindi. The compounds like *yudhiṣṭhira* etc. are examples of *aluksamāsa* identical to Sanskrit compounds; *bailagādī*, *siradarda*, etc. are compounds where internal *vibhaktis* are vanished. Such compounds form one word and can be paraphrased. There are compounds like *ghudasavāra* (horseman), *ghudadauda* (horse race) etc., where

the a component of the compounds is seen only in compound forms. In these compounds, the component *ghuda* refers to *ghodā* (horse). But the word *ghuda* is never found in non-compound expressions.

The compounds like *nakatā* (a dishonored person) will fall in *nīyasamāsa* class. Because such expressions give a specialized meaning that cannot be derived from their paraphrase. Table 5.5 gives an overview of the spectrum in Hindi compounds.

Table 5.5: Showing spectrum in Hindi expressions

	<i>ardhasamāsa</i>	<i>aluksamāsa</i>	<i>samāsa</i>	<i>nīyasamāsa</i>
	<i>acche ladakoṃ ne</i>	<i>yudhiṣṭhira</i>	<i>bailagādī,</i> <i>ghudasavāra,</i> <i>siradarda</i>	<i>nakatā</i>
<i>sup</i> -deletion (<i>sublopa</i>)	✓	x	✓	✓
uninterruption (<i>avyavadhāna</i>)	✓	✓	✓	✓
fixed order (<i>niyatapaurvāparya</i>)	✓	✓	✓	✓
one accent (<i>aikasvarya</i>)	✓	✓	✓	✓
one word	x	✓	✓	✓
multiple word	✓	x	x	x
paraphrasing	✓	✓	✓	x

Thus, identification of *padas* and *samasta-padas*, compounds and quasi-compounds helps in transferring the correct case information. Therefore, it helps in correct grammatical form generation of the target language words.

5.5 Conclusion

In this chapter, we talked about the transfer grammar approach for English-Hindi word ordering and generation. We talked about the structural divergences between English and Hindi and discussed how the concepts of *sup*, *tin* and *pada* from Pāṇinian Grammar can help in handling these two tasks. We showed how the notion of *pada* and ‘mirror structure’ of the verb phrase take care of minimal reordering for English-Hindi. We also deliberated on the rules which handle the exceptional cases that cannot be captured by the notion of *pada* and mirror image of English verb phrases.

We have also shown the usefulness of the notion of *pada* and *samasta-pada* in target language word generation, especially, how it helps in generating the appropriate grammatical cases: the direct and oblique cases for Hindi.