ANUSAARAKA
अनुसारक

Guidelines

for

Language Resource Development

for

Machine Translation

(For English to Hindi Translation)

Guidelines Version 1.8 (July 2009)

by

Akshara Bharati

CONTENTS

PREFACE

Anusaaraka is a Language Accessor - a software that will not only automatically translate an input text but also provide the basis for conversion of one language into another, in various steps. In other words, it shall allow the user to see a layered output of how the source language text is analyzed and converted to the target language.

Work of this project may broadly be classified as follows:

a) Software development – actual development of the software which would be utilized by the end-user.
b) Mapping grammatical structure of English and Hindi, taking insights from Paninis (Sanskrit) grammar. Similar concepts and patterns are observed in English and Hindi, in order to establish co-relations between the vastly different grammatical structures of both languages.
c) Language Resource development – a structured method wherein, the end objective is to provide 'logic' as to how a certain word should translate in varying context. The output from this task serves as input for software development.

These guidelines are an attempt to document the step-by-step procedure followed for Language Resource Development for Anusaaraka. This process document has been made possible due to inputs from the Anusaaraka contributors over the years.

The current document may further be modified and refined on the basis of feedback from users.

Please do send your comments / suggestions for improvement to: anusaaraka@chinfo.org

Participants are requested to go through the Frequently Asked Questions (FAQs) on Anusaaraka appended at the end, prior to reading these guidelines. The FAQs are also available at the links given below:

http://chinfo.org/anusaraka/toolsbox.asp
http://ltrc.iiit.net/~anusaaraka

The FAQs are critical for a basic understanding of Anusaaraka.

CHAPTER 1

INTRODUCTION

When it comes to translation from one language to another the quality of translation varies from individual to individual based on her/his proficiency in the languages concerned. The proficiency covers the vocabulary, grammar of both the source and target language, sensitivity to the nuances of language and the culture of the two languages involved.

As is evident, while building a Machine Translation System (i. e. getting a machine / computer to do the translation task), one needs to make the above sources of knowledge explicitly available to the machine. In other words, the human mind stores grammatical structures of languages and within split seconds uses this stored knowledge while translating. However, these structures need to be documented and stored in the machine memory so that it translates as clearly (to the extent possible) as a human being.

To put it differently, an individual attempting to learn a language formally needs to be taught the alphabet and some basic grammar of the language. Similarly, in order for a machine to be able to translate, it becomes necessary that appropriate rules and logic are given to the machine so as to generate the most appropriate output.

The challenges in making the machine capable of translation, are as follows:

‒ Absence of the faculty of 'common sense' in a machine.
   This would mean that the machine only follows the rules / logic that is available to it. The machine will not be able to resolve ambiguity in language unless appropriate rules and logic are provided to it.

‒ The machine does not have the ability to evaluate and improve the quality of translation.
   Given that the end-user does not have adequate or any knowledge of the source language, the user would not be able to tell if there has been an error in the machine translation. Nor does the machine have the ability to warn of possible errors in the translation. This can lead to misinterpretation.

   As we are all aware - the dangers of misinterpretation are far more serious than lack of interpretation.

   Therefore, it is important to provide adequate language related information to the machine in order to ensure greater degree of accuracy in translation.

Language related information for the machine:

(A) Development of 'Contextual Rules (Lakshancharts)' for selecting the appropriate sense of a polysemous word (i.e. words with more than one meaning) in a given context.

This task involves translation of sample sentences, understanding various senses (viz. meanings) in which a word is used and establishing a logical explanation for what exactly in a sentence determines the choice of a particular sense in a given context.

(B) Development of 'Transfer Verb Frames'. While translating we note that some words appear in the target language (translated text), although they may not have a corresponding word in the source language (original text). Often these words are dependent on the verb in the source language and its translation in the target language. Thus, this task involves creation of norms (technically called 'Transfer verb frames') which shall allow one to understand how verbs are translated.

(C) Development of 'Padasutras' for polysemous words. This task would involve arriving at the core meaning of a polysemous word and also establishing a link between various senses of the word. This is based on insights from Paninian Grammar.

The above 3 tasks are collectively referred to as Language Resource Development. The languages for this project of Anusaaraka are English as source and Hindi as target.

All the resources so developed will be available as a "free" resource under GPL (General Public License).

Step-by-step:

Participants will be assigned a set of English words (typically 4 pages each) from the Cambridge Advanced Learner's Dictionary, which is being used as the standard for this project.

The task would involve the following:

1. Collect English example sentences for every word assigned to them.

2. Translate the English example sentences into Hindi.

3. Develop Lakshancharts (contextual rules) for all the senses that a word has, when translated to Hindi. (Chapter 2 of these Guidelines).

4. For Verb entries, 'frames' will have to be developed for both the English and the translated sentences. (Chapter 4 of these Guidelines).

5. Develop padasutras for each polysemous word. This involves deriving the 'core meaning' of a polysemous word and framing a formula to represent its polysemy.

Participants are requested NOT to start work before reading these Guidelines.

The Guidelines are designed in such a way that the steps for the task are graded in terms of difficulty levels. It is strongly recommended that you spend sufficient time at each level and get comfortable with the task before moving to the next step.

For assistance at any stage, please write to: anusaaraka@chinfo.org

CHAPTER 2

Guidelines for Word Sense Disambiguation (WSD)
and
Framing Contextual Rules (Lakshancharts)

2.1 Translation of Example Sentences

Objective:

- ✓ Translation of English example sentences into Hindi.

  Sentences could illustrate:
  - Words
  - Multi word Expressions (MWEs)

  Note: Multi Word Expressions refer to the expressions which consist of more than one word but together they denote a single meaning. For example, 'take off' has two words but its meaning has nothing to do with 'take' or 'off' separately. Phrasal verbs, idioms etc. are MWEs.

Requirements:

Mandatory:

- ✓ Average Knowledge of both the source (English) and the target (Hindi) language.

Optional:

- ✓ Access to and working knowledge of a computer and the Internet. The participants are expected to get comfortable with using a computer and Internet over a period of time.

- ✓ A prerequisite for the subsequent steps is the installation of the Hindi keyboard on the machine.

  For installing the Hindi keyboard on one's own computer:

- ➢ If you are working on Windows platform:
    - ➢ please visit: http://www.bhashaindia.com
    - ➢ click on the 'Downloads' tab appearing on the left side
    - ➢ select the 'Indic IME' option and follow instructions for installation
    - ➢ please visit the Anusaaraka Language Resource Development site (http://sanskrit.uohyd.ernet.in/anusaaraka) and download the Hindi keyboard software usage key.

> ➢ If you are working on the Linux platform:
>   > ➢ please contact: **anusaaraka@chinfo.org** for installation instructions.

Procedure:

Each participant will be assigned 4 pages from the Cambridge Advanced Learners Dictionary (CALD), for translation, by the Anusaaraka Team; and will become part of the Anusaaraka Language Resource Development Community.

After you have completed the work on your words you can either submit the work online at: http://sanskrit.uohyd.ernet.in/anusaaraka or mail it to: anusaaraka@chinfo.org.

You must follow the steps given below while translating:

STEP 1: Collect the sentences for a given word from the Cambridge Advanced Learners Dictionary (CALD) and WordNet (Url is mentioned under 'Resources' for this section). In case you have other dictionaries, sentences from these can also be used as supplementary examples. For example,

English word: 'way' (Sentences collected from various sources)

1. I told you, we should have done it my **way**!
2. It is not his **way** to admit that he had made a mistake.
3. They grinned at her in a friendly **way**.
4. I hate the **way** she always criticizes me.
5. Get out of my **way** ! I'm in a hurry.
6. Kids were running this **way** and that.
7. I will fight all the **way.**
8. They decided to split the money four **ways**.

Note: These examples are used in all the steps hereafter. However, more exhaustive case-studies for some words (including 'way'), are provided at the end in the Appendices (Appendix 1.1).

STEP 2: Translate the English sentences into Hindi.

Please note that the attempt, while translating, is not to use fanciful terms or vocabulary, but more importantly, to do natural translations that capture the implied meaning (भावार्थ) of the sentence with simplicity.

English word : 'way'

1. I told you, we should have done it my **way**!
   मैंने तुमसे कहा था, हमें यह मेरे बताये तरीके से करना चाहिये था |

2. It is not his **way** to admit that he had made a mistake.
   यह कबूल करना कि उसने गलती की है, उसका तरीका नहीं है |

3. They grinned at her in a friendly **way**.
वे दोस्ताने ढंग से उसकी ओर मुस्कुरा रहे थे |

4. I hate the **way** she always criticizes me.
जिस तरह से वह हमेशा मेरी निन्दा करती है, मुझे पसन्द नहीं |

5. Get out of my **way** ! I'm in a hurry.
मेरे रास्ते से हट जाओ, मै जल्दी में हूं |

6. Kids were running this **way** and that.
बच्चे सब तरफ भाग रहे थे |

7. I will fight all the **way**.
मैं अन्त तक लड़ूँगा |

8. They decided to split the money four **ways**.
उन्होंने पैसों को चार भाग में बाँटने का निर्णय लिया |

See **Appendix 1.2** for more case-studies.

Step 3: At this stage it will be useful to do an exercise called **Information Book-Keeping.** It is a step-by-step process of translation from English to Hindi. This process will help you to understand how information is encoded differently in English and Hindi. This will also help you to get a Hindi translation which will be, as far as possible, close to the English structure without compromising with readability in Hindi. Information book-keeping will also give you insights that will come in handy when you formulate contextual rules (Section 2.3). The procedure for Information book-keeping is given below:

(i) Give the literal meaning of each word

That 's not the right way to   hold the pair of scissors.

वह है नहीं - सही तरीका के_लिये पकड़ना - जोड़ा का कैंची

Note: Since Hindi does not have an equivalent for the word 'the' of English, a hyphen has been put in the Hindi version.

(ii) Replace the words which have a literal sense in the Step (i) with the words that capture the contextual meaning.

वह है नहीं - सही तरीका का पकड़ने - जोड़ा का कैंची

(iii) Add/delete any word(s) which are necessary for the exact Hindi translation. In the following sentence 'जोड़ा का' has been removed:

वह है नहीं - सही तरीका का पकड़ने - - - कैंची

(iv) Reorder the word(s) within the relevant verb phrase, noun phrase etc. Here नहीं
_है and पकड़ने_का have been reordered:

वह - नहीं _है - सही तरीका पकड़ने_का - - - कैंची

(v) Reorder the word(s) within the sentence to get a proper Hindi sentence.
वह   कैंची पकडने_का सही तरीका नहीं_है

Note: The symbol "_" (underscore) between two words indicates that the two words together form a single grammatical unit.

The complete process of Information book-keeping will appear as follows:

That 's not the right way to   hold the pair of scissors.

(i) वह है नहीं -   सही तरीका के_लिये पकड़ना -   जोड़ा का   कैंची
(ii) वह है नहीं - सही तरीका का पकड़ने - जोड़ा का कैंची
(iii) वह है नहीं - सही तरीका का पकड़ने - - - कैंची
(iv) वह - नहीं _है - सही तरीका पकड़ने_का - - - कैंची
(v) वह   कैंची पकडने_का सही तरीका नहीं_है ।

For more examples see **Appendix 1.3**

<u>**Appendices for Section 2.1:**</u>

Appendix 1.1 - Example sentences
Appendix 1.2 - Translation of example sentences
Appendix 1.3 - Examples of Information book-keeping

<u>**Resources for Section 2.1:**</u>

Given below is a list of online resources that may be used for getting example sentences for a given word:

 (a) Online English-Hindi dictionaries:

http://www.cfilt.iitb.ac.in/~hdict/webinterface_user/dict_search_user.php
http://www.shabdkosh.com/
http://ltrc.iiit.ac.in/onlineServices/Dictionaries/Dict_Frame.html

 (b) Recommended Online English dictionaries:

http://www.merriam-webster.com/
http://dictionary.cambridge.org/

(c) English Wordnet:

http://wordnet.princeton.edu/perl/webwn

(d) Hindi Wordnet:

http://www.cfilt.iitb.ac.in/wordnet/webhwn/

(e) Corpus of English:

http://www.americancorpus.org
http://corpus.byu.edu/

Exercises for Section 2.1:

Given below are the example sentences of the English words 'apply' and 'state'.

a) Translate the sentences into Hindi.
b) Do information book-keeping for these sentences.
c) Please contact: anusaaraka@chinfo.org for feedback on completion of these exercises.

Exercise 1: Sample sentences for the English word 'apply'

1. By the time I saw the job advertised it was already too late to **apply**.
2. Please **apply** in writing to the address below.
3. We've **applied** to a charitable organization for a grant for the project.
4. Those were old regulations - they don't **apply** any more.
5. He wants a job in which he can **apply** his foreign languages.
6. The court heard how the driver had failed to **apply** his brakes in time.
7. **Apply** the cream liberally to exposed areas every three hours and after swimming.
8. The paint should be **applied** thinly and evenly.
9. You can solve any problem if you **apply** yourself.
10. I **apply** this rule to get good results.
11. This thinking was **applied** to many projects.
12. These rules **apply** to everyone.
13. She has **applied** for a job at college.

Exercise 2: Sample sentences for the English word 'state'

1. The building was in a **state** of disrepair.
2. She was found wandering in a confused **state** (of mind).
3. The kitchen was in its original **state**, with a 1920s sink and stove.
4. She got into a real **state** before her driving test.
5. The drought is worst in the central African **states**.
6. The creation of independent **states** has led to a resurgence of nationalism.
7. The Queen rode in **state** to the opening of Parliament.
8. They received **state** funding for the project.
9. The police couldn't arrest him because he'd fled across the **state** line.
10. Children in the **stated** (=named) areas were at risk from a lack of food, the

report said.

11. Our warranty clearly **states** the limits of our liability.
12. The procession moved through the mountain village at a **stately** pace.
13. He always walked with a stately bearing.
14. After the floods the government declared a **state** of emergency.
15. The machine uses all the newest technology and is considered **state**-of-the-art.
16. The Prime Minister has gone on a **state** visit.
17. The solid **state** of water is called ice.
18. Can you please **state** this distance in kilometers.
19. He was in such a **state** that one could not reason with him.

2.2 Sense Identification and Sense Minimization

Objective:

- ✓ Finding out the various senses (meanings) in which a word has been used.

- ✓ Grouping the sentences on the basis of the senses.

- ✓ Making separate note of specific usage of a word in a context; and even ambiguous usages.

Requirements:

Mandatory:

- ✓ English sentences translated into Hindi. (i.e. Input from section 2.1)

Optional:

- ✓ Orientation session on 'Word Sense Disambiguation (WSD)' conducted by Anusaaraka team.

Procedure:

There are many words in a language which are ambiguous, i.e. they have more than one 'sense' in which they are commonly used. Therefore, it is important to understand the different senses and identify the appropriate sense in a given context. Given below is a step-by-step procedure for the same:

2.2.1 SENSE IDENTIFICATION

Step 4: Examine the sentences which have been translated and identify the number of senses that a word has.

English word : 'way'

1. I told you, we should have done it my **way**!
मैंने तुमसे कहा था, हमें यह मेरे बताये तरीके से करना चाहिये था |

2. It is not his **way** to admit that he had made a mistake.
यह कबूल करना कि उसने गलती की है, उसका तरीका नहीं है |

3. They grinned at her in a friendly **way**.
वे दोस्ताने ढंग से उसकी ओर मुस्कुरा रहे थे |

4. I hate the **way** she always criticizes me.
जिस तरह से वह हमेशा मेरी निन्दा करती है, मुझे पसन्द नहीं |

5. Get out of my **way** ! I'm in a hurry.
मेरे रास्ते से हट जाओ, मै जल्दी में हूं |

6. Kids were running this **way** and that.
बच्चे सब तरफ भाग रहे थे |

7. I will fight all the **way**.
मैं अन्त तक लड़ूँगा |

8. They decided to split the money four **ways**.
उन्होंने पैसों को चार भाग में बाँटने का निर्णय लिया |

It appears that the English word 'way' has 7 senses, as per the sentences given in Section 2.1.

Step 5: Group the sentences which have the same sense. At the end of this step, you will have all the sentences classified into various groups, each representing one sense.

Grouping of sentences according to senses for the English word 'way' is given below:

Sense 1 – तरीका

1. I told you, we should have done it my **way**!
मैंने तुमसे कहा था, हमें यह मेरे बताये तरीके से करना चाहिये था |

2. It is not his **way** to admit that he had made a mistake.
यह कबूल करना कि उसने गलती की है, उसका तरीका नहीं है |

Sense 2 - ढंग

3. They grinned at her in a friendly **way**.
वे दोस्ताने ढंग से उसकी ओर मुस्कुरा रहे थे |

Sense 3 - तरह से

4. I hate the **way** she always criticizes me.
जिस तरह से वह हमेशा मेरी निन्दा करती है, मुझे पसन्द नहीं है |

Sense 4 – रास्ता

5. Get out of my **way** ! I'm in a hurry.
   मेरे रास्ते से हट जाओ, मै जल्दी में हूँ |

Sense 5 – तरफ

6. Kids were running this **way** and that.
   बच्चे सब तरफ भाग रहे थे |

Sense 6 – अन्त तक

7. I will fight all the **way**.
   मैं अन्त तक लड़ूँगा |

Sense 7 – भाग

8. They decided to split the money four **ways.**
   उन्होंने पैसों को चार भाग में बाँटने का निर्णय लिया |

2.2.2 SENSE MINIMIZATION

Step 6: One must keep in mind at this stage that although a word may have many 'senses', one should try and see if a sense can be 'clubbed' along with another sense, without compromising on the meaning of the sentence.

For example, the Senses 1, 2, and 3 for the word 'way' given above, can be clubbed into one group as they all imply 'तरीका'. Replacing 'ढंग' and 'तरह से ' in the given examples by 'तरीका' would still be comprehensible to the Hindi readers and would not change the intended meaning.

Sense 1- तरीका (2. ढंग ->तरीका, 3. तरह से ->तरीके से)

1. I told you, we should have done it my way!
   मैंने तुमसे कहा था, हमें यह मेरे बताये तरीके से करना चाहिये था |

2. It is not his way to admit that he had made a mistake.
   यह कबूल करना कि उसने गलती की है, उसका तरीका नहीं है |

3. They grinned at her in a friendly way.
   वे दोस्ताने ढंग से उसकी ओर मुस्कुरा रहे थे |

4. I hate the way she always criticizes me.
   जिस तरीके से वह हमेशा मेरी निन्दा करती है, मुझे पसन्द नहीं |

Sense 2 - रास्ता

5. Get out of my way ! I'm in a hurry.
  मेरे रास्ते से हट जाओ, मै जल्दी में हूँ |

Sense 3 - तरफ

6. Kids were running this way and that.
  बच्चे सब तरफ भाग रहे थे |

Sense 4 - अन्त तक

7. I will fight all the way.
  मैं अन्त तक लड़ूँगा |

Sense 5 - भाग

8. They decided to split the money four ways.
  उन्होंने पैसों को चार भाग में बाँटने का निर्णय लिया |

After clubbing the senses for the word 'way', we are left only with 5 senses, instead of 7.

IMPORTANT: Please 'club' senses ONLY if they can naturally be clubbed without changing the intended meaning in English.

For example, Senses 2 and 3 (given below) are not clubbed with sense 1; because replacing 'रास्ते' in the example sentence 5 and 'तरफ' in the example sentence 6 by 'तरीका' would completely change the intended meaning and would also give an inappropriate translation in Hindi.

Thus,

Sense 2 – रास्ता

5. Get out of my way ! I'm in a hurry.
  मेरे रास्ते से हट जाओ, मै जल्दी में हूँ |
* मेरे तरीके से हट जाओ, मै जल्दी में हूँ |

Sense 3 – तरफ

6. Kids were running this way and that.
  बच्चे सब तरफ भाग रहे थे |
* बच्चे सब तरीके भाग रहे थे |

The sign (*) indicates that the sentences are poorly translated or incorrect.
Thus, combining senses 2 and 3, in the above examples would yield undesirable results. Refer to **Appendix 2.1b for more examples of Sense Minimization.**

Step 7: Besides the commonly occurring senses, there are also cases where a word occurs in phrases and the entire phrase gives one meaning. These are referred to as Multi Word Expressions (MWEs).

For example, *'one way or the other'* , *'back and forth'*; *'change for better'*. Also, the expression 'Mind your language' in English, is a fixed expression and would be translated in Hindi as ' अपनी ज़बान पर लगाम दो' .

Therefore, all such expressions should be written separately so that the system always recognizes them as MWEs, and translates them accordingly.

Let us look at the examples given for the Senses 3 and 4 for the word 'way' above :

Sense 3 - सब तरफ

6. Kids were running **this way and that.**
   बच्चे सब तरफ भाग रहे थे |

Sense 4 - अन्त तक

7. I will fight **all the way.**
   मैं अन्त तक लड़ूँगा |

We find the sense 'तरफ ' for the word 'way' in example 6. However, a closer look at the example will indicate that it is not the word '*way*' alone which is giving the intended sense but the entire phrase 'this way and that' which is giving us the sense 'सब तरफ' . Similarly, for the example sentence 7, the phrase 'all the way' is giving us the sense 'अन्त तक'.
At this stage, we will separate these MWEs from the other senses of the word 'way'.

For more examples of multi word expressions see **Appendix 2. Multi Word Expressions.**

Refer to **Appendix 2.1a - Identification and minimization of word senses.**

Note : In case you do not find any Multi Word Expressions in your data, please proceed to the next step.

Appendices for section 2.2:

Appendix 2.1 a – Identification of senses
Appendix 2.1 b – Minimization of senses
Appendix 2.2 – Multi Word Expressions

Resources for section 2.2:

Same as section 2.1

Exercise for section 2.2:

      a) Take the sentences translated by you in section 2.1. Group them on the basis of their senses.
      b) Try to club the senses, if possible.

Please contact: anusaaraka@chinfo.org for feedback on the Exercises done by you.

Objective:

  ✓ To develop the logic and contextual rules that shall form the basis for the software to select the appropriate sense of a word in a given context.

Mandatory requirements:

  ✓ Understanding of basic grammar of both the source and the target languages. An introductory Grammar for WSD task has been provided (Chapter 3 of these Guidelines) for help.

  ✓ Having worked on the previous 2 sections.

  ✓ Working knowledge of computers.

Procedure:

Step 8: If a sense of a particular word appears in a large number of instances that sense should be taken as the 'default sense' of the word. For example, for the word *way* in English, 'तरीका' appears to be the default sense.

Step 9: You may find some definite elements in the context which determine a particular sense for the word in several instances and therefore you detect a pattern. These patterns should be written as formal statements ('if – then' statements).

To begin with you can simply note down which element(s) in the sentence are triggering a particular sense. You can then convert your observations into 'if-then' statement. For example, for the sense 'रास्ता' of the word 'way' we observe that all the sentences that indicate the sense 'रास्ता' have something in common: i.e., the word immediately before the word 'way' is (a) an adjective; (b) it belongs to the set of words which include 'back', 'long', 'full', 'short' etc and (c) the adjective preceding 'way' is modifying the noun 'way'.

So the 'if-then' statements based on the above observations would be written as follows:

Word: way
Example: They ran away by the back way.
        He kept on speaking the full way.
        We went the long way around.
Rule id: way 2
Saliency: medium
If the root is way
If way is a विशेष्य and it has ?विशेषण as its विशेषण
and

?विशेषण belongs to the list (long,full,back ...)
Then its Hindi meaning is: 'रास्ता'
Which is a root
Sentence type: any

Some possibilities of the contextual triggers are provided in the box below:

A. Context – The context for a word would be the other words in the sentence. For example, take the English sentence

He is on his way.

The presence of the word 'on' before the word 'way' triggers the sense of 'रास्ता '.

These contexts can be:

(i) Mere presence of certain words as in the above examples.

(ii) Grammatical relation (refer to Chapter 3 on Introductory Grammar for the WSD task) of the concerned word with other words in the sentence.

For example, look at the word 'catch' in the English sentence

Catch some sleep.

The above English sentence would be translated as

थोड़ा सो लो ।

Although the most frequent sense of 'catch' is 'पकड़ना' in Hindi, in the above example 'catch' is translated as 'लो'. The context that is causing this change in the meaning of the word 'catch' is the presence of the word 'sleep'. But it is not just the presence of the word 'sleep' in the same sentence which is causing this change.

We can see another example where sleep occurs in the same sentence as 'catch' – 'He catches butterflies in his sleep'. Here the sense of catch would remain 'पकड़' in Hindi. Thus, we need more information than the mere presence of a word in the context of the word under consideration. Going back to our example 'Catch some sleep'
 we realize that we need more information to arrive at the correct translation.

A closer study of the sentence shows us that when the word 'catch' occurs as a verb and has 'sleep' as its object, it translates as 'लो'. Thus, we may not only consider the other words in the context, but also the relations between these

---

words to arrive at the correct solution.

B. Our world knowledge – sometimes the selected sense is not evident from the words in the sentence. For instance, observe the following sentences:

1. Mary sold the cow because she did not give milk any more.
   मेरी ने गाय बेच दी क्योंकि उसने दूध देना बंद कर दिया था ।

2. Mary sold the cow because she needed the money.
   मेरी ने गाय बेच दी क्योंकि उसको पैसे चाहिये थे ।

When translated into Hindi the word 'she' in both (i) and (ii) will appear as 'उस'. and therefore pose no problem.

However, we understand the meaning of 'she' in (i) as referring to the 'cow' and the 'she' in (ii) as referring to 'Mary' because of our world knowledge which tells us that generally a cow 'gives milk' and Mary as a human being may 'need money'.

Such instances may occur but we shall ignore them for the rule making process.

See **Appendix 3.1a – 'if-then' statements**, for more examples.

Step 10: Having written the formal statements enter your inputs into the Rule Forms.

Note: It may appear repetitive to write the 'if-then' statements first and then make entries into the Rule Form. But it is important that we go through the process of 'if-then' statements in order to work out the logic for framing rules.

The Rule Form can be filled either online (http://sanskrit.uohyd.ernet.in/anusaaraka) which is a preferred option, or, on paper which can then be passed on to your contact person for entry online.

Writing rules using Rule-Form :

- Enter your observations into the Rule form – selecting the appropriate option from the drop down menus given.

- Save the rule(s) created by you. The form has a provision for testing your rule (See section 2.4 for details).

  – Rules that run well (i.e. give appropriate translations in Hindi on a given set of sample sentences) can be submitted to the Anusaaraka database. Rules which do not run well must be re-looked at / modified / improved and thereafter, tested again before submitting them to the Anusaaraka database.
  –

Given below is a sample of the rule form that you will be required to fill.

WSD Rule Form

Root/Word	
Rule File	Rule File :- .clp
Example Sentences	
Rule id	The ruleid is system generated. Read the message after posting the form to know the rule id.

Saliency Select from below ⬍

if the select appropriate option ⬍ is [   ]

Add/Remove   Add   Remove

if select appropriate option ⬍

And/Or Remove

Then its Hindi meaning is [   ] **which is** Select Appropriate Option ⬍

Sentence type Select Appropriate Option ⬍

submit

Abort

How to fill the Rule Form:

Entry no.1 - Root/Word

In the frame that follows the words 'root/word' enter the 'root'. Note that in majority of the cases you will need to enter the 'root'. However, in some cases you may need to enter the 'word'. This will apply in cases where two words with different roots have the same shape, e.g., **leaves1** which is the plural form of the noun 'leaf'; and **leaves2** which is the third person singular form of the verb 'leave'. The ambiguity is only on **leaves1** and **leaves2** and not in 'leaf' and 'leave'.

Note: 'root' is the minimal form of a word after it is stripped of all affixes. For example, in the word 'worked', '-ed' is an affix that indicates past tense. When this affix is removed we are left with the root 'work'. For more information on this see the section on verbs in Chapter 3 - Introductory Grammar for WSD task.

<u>Entry no.2</u> - Rule File

The machine will automatically generate this. You need not do anything.

<u>Entry no.3</u> - Example sentences

Enter as many example sentences as possible, for a particular sense of the word.

<u>Entry no.4 </u>- Rule id

Enter the word and identify it with a number. For instance, if a word (such as 'shell') has two senses, the rule id for the two senses will be 'shell1' and 'shell2'. The rules for the two senses will be entered in two separate forms.

<u>Entry no. 5</u> – Saliency

Here you have a drop-down menu which gives you the option to indicate the saliency of a 'sense' on a scale ranging from 'very high' to 'very low'.

The following rule of thumb may help you to fix the saliency of a 'sense'. The 'default' sense (which is the most general sense and is usually listed as the first meaning in the dictionary) has the least number of constraints stated in the rule. The saliency of the default sense would be 'very low'. As the constraints become more and more complex and specific for a sense, the saliency goes higher. In the event of there being an extremely specific condition for a sense, the saliency would be 'very high'.

Note that saliency is relative among the different senses of a word.

<u>Entry no.6 </u>– if the ....is.......

Select word/root/phrase from the drop-down menu for the first frame. Enter the concerned word in the second box. Example. 'if the <u>root </u>is <u>leaf</u>'...

<u>Entry no. 7</u> – if...

Here the drop-down menu gives a few options which indicate possible conditions that individually or in combination with other conditions trigger a particular sense of a word.

    a) <u>Related</u>:

You select this option when you feel that a particular sense is indicated because the word to be translated has a particular grammatical relationship with another word in the sentence. For instance, when the word 'catch', which is a verb, has a concrete object, it has the sense of पकड (as in 'I want to catch the ball.'); whereas when it has an object like 'sleep' it has the sense of 'ले' (as in 'Catch some sleep if you can.'). You will find that grammatical relatedness will account for larger number of senses than the other options listed in the menu for this item.

When you click on 'Related', a frame will open (see below) which will indicate the possible relationships between two words. The frame will have two text boxes and each text box will have a drop-down menu.

If .... is ..... and it has .... as its ....
....
....

For instance:

if <u>catch</u> is <u>क्रिया</u> and it has <u>sleep</u> as its <u>object</u>

   b) <u>Semantic</u>:

We shall begin by first explaining what we mean by the term 'semantic'. Words are composed of several components of meaning which together constitute the semantic properties of the word. For instance, the word 'man' has the following components of meaning:

animate, human, *male*, adult.

If we change any of the components, we will get another word. So, the following configuration will give us the word 'woman':

animate, human, *female*,adult.

So, the words 'man' and 'woman' are distinguished from each other on the basis of the semantic feature 'male/female'

To take another example, consider the following sets of nouns:

1. book, girl, banana, milk, computer etc.
2. honesty, beauty, friendship, marriage, anger etc.

The nouns in the first set share the semantic property of 'concreteness' and they differ from the nouns in the second set which share the property of 'abstractness'. The semantic features of the two sets are 'concrete' and 'abstract' respectively.

The drop-down menu gives a list of the possible semantic properties of a word, e.g., concrete, abstract, animate,inanimate, human, male, female etc.

For instance, in the following sentences:

The hut does not <u>afford</u> protection from the elements.
The seat <u>afforded</u> her a good view of the stage.

If 'afford' is the word for which we have to frame a rule. The default meaning of afford is 'समर्थ/हैसियत_हो' which does not seem to be the sense in which the word is used in the above sentences. Instead, we get a sense of 'give':'दे'.If we think this sense is triggered by the fact that the subject of the verb 'afford' has the semantic property 'inanimate', then our 'if...' statement will be

if <u>afford</u> is <u>क्रिया</u> and it has <u>?noun</u> as its <u>subject</u>
and
if <u>?noun</u> has the semantic property <u>inanimate.</u>

Note:The symbol '?' indicates that the entity following it is variable, i.e., it can be 'any' item from the category mentioned. In the above sentences, for example, first we state that the verb has 'any or ?' noun as its subject. Then in the next step we go on to give specific characteristics of the ?noun because we want to say that it is not enough to have any noun as the subject but it is the nature of the noun (here, the semantic property of being 'inanimate') that triggers the particular sense of 'दे' for afford word.

c) <u>category</u>:

The drop-down menu consists of grammatical categories like noun, verb, adjective, adverb etc. The Introductory Grammar for the WSD task (Chapter 3) gives definitions of terms and examples.

You may need to use this option when dealing with sentences like the following:

1. There is a <u>marked</u> improvement in her performance.
2. The manager <u>marked</u> the files he needed for the meeting.

In (1) <u>marked</u> is an adjective and its Hindi meaning would be उल्लेखनीय whereas in (2) it is a verb and would mean चिह्नित_कर_ . If we take (2) as the default sense, then our 'if...' statement for the sense उल्लेखनीय will be the following:
if <u>marked</u> is an <u>adjective</u>...
d) <u>list</u>:

In this option one can make a list of words that may be associated with our word and trigger a particular sense of a word.

For instance,

The default sense of 'switch on', which is a verb, is चलाना but when it has words like
'lamp, tube light, bulb...'
as its object, its Hindi meaning will be जलाना.

Our 'if-statement' will be

if <u>switch-on</u> is <u>क्रिया</u> and it has <u>?noun</u> as its <u>object</u>
and
if <u>?noun</u> belongs to the <u>list</u> 'lamp, tube light, bulb...'.

e) <u>position</u>:

Sometimes a word may acquire a different meaning because of another word used before or after it.

For example, the default meaning of the word 'badly' is बुरी तरह से as in

The boy was <u>badly</u> beaten
The house was <u>badly</u> damaged
The servant was <u>badly</u> treated

But when it is followed by the word 'needed', it has the meaning अत्यंत or बहुत as in
'Don't skip the party, your help is badly needed there'.
Our 'if-statement' will be:

if <u>needed</u> is after <u>badly</u>...

<u>Entry no.8</u> – Hindi meaning

Then its Hindi meaning is....... which is ....
....
....

Enter the Hindi meaning in the first slot in Devanagari/Hindi script (please refer to
'Installation of Hindi Keyboard in Chapter 1 of these Guidelines). The text box has a
drop-down menu which includes items like 'root, word form...'. You have to indicate
whether the Hindi meaning is a root, word...

For instance:
Then the Hindi meaning is <u>बहुत</u> which is a <u>root</u>

<u>Entry no.9</u> – Sentence type:

Here the drop-down menu will give a list of sentence types such as -- declarative
(सामान्य वाक्य), interrogative (प्रश्नात्मक वाक्य), 'any' etc. and you have to indicate the
sentence type of your example sentences. In case you feel that the sentence type
does not impact the meaning of a word, you may select the 'any' option.

(For more examples of filled out forms see **Appendix 3.2a.**)

<u>Step by step procedure for Word Sense Disambiguation (WSD) for a word.</u>

We will now take up a word and follow the WSD procedure from the beginning to the
point of filling up the rule form.

Our example word is 'dissolve'.

1: Collect example sentences.

(1) Dissolve two spoons of sugar in warm water.
(2) Why are you dissolving salt in your tea?
(3) I think the sugar has not dissolved completely.
(4) Oil does not dissolve in water.
(5) The tension in the office just dissolves when she walks out.
(6) Their marriage was dissolved in 1968.
(7) The Parliament has been dissolved.

2: The example sentences for the word are translated.

(1) Dissolve two spoons of sugar in warm water.
गरम पानी में दो चम्मच शक्कर <u>घोलिये</u> |

(3) Why are you dissolving salt in your tea?
आप अपनी चाय में नमक क्यों <u>घोल</u> रहे हैं?

(3) I think the sugar has not dissolved completely.
मुझे लगता है कि शक्कर पूरी तरह से <u>घुली</u> नहीं है |

(4) Oil does not dissolve in water.
तेल पानी में <u>घुलता</u> नहीं है|

(5) The tension in the office just dissolves when she walks out.
उसके बाहर चले जाते ही कार्यालय में तनाव समास हो जाता है |

(6) Their marriage was dissolved in 1968.
१९६८ में उनका विवाह <u>भंग</u> कर दिया गया था |

(7) The Parliament has been dissolved.
संसद <u>भंग</u> कर दिया गया है |

3: Sense identification and minimization.

On the basis of the examples given above, we are able to identify four senses of the word 'dissolve': घोल, घुल, भंग, and समास. We also find that it is not possible to minimize these senses to a smaller number because substitution of senses will yield unacceptable and absurd sentences in Hindi. For instance, we cannot say the following:

* तेल पानी में नहीं घोलता है|
* संसद घुल कर दिया गया है|
* उसके बाहर जाने से तनाव भंग हो गया |

4: Observation.

Having identified the different senses we proceed to write the contextual rules for each one of them. We begin by scrutinizing the sentences carefully to 'observe' the context which may have triggered one particular sense and not the other. In other words, what is it that gives us a clue that a word has sense X and not sense Y in a particular sentence.

Some of the observations are listed below:

(a) sense घोल.
घोल (in sentences 1 and 2) seems to be the default sense. The reasons for taking it as default are twofold. This is the meaning that comes to one's mind first and it also happens to be the meaning listed first in the dictionary. It may therefore, be the most general sense of the word 'dissolve'.

(b) sense घुल.
In sentences 1 and 2 we find that the word 'dissolve', which is a verb, has an Object – 'sugar' in 1 and 'salt' in 2. Since the verb requires an Object we may say that it is a Transitive Verb and translates in Hindi as घोल. We observe something different in

sentences 3 and 4. Here we find that the verb 'dissolve' does not have (require) an Object, i.e., the verb is Intransitive in this case and its Hindi translation is घुल.

c) sense समास_हो.
In sentence 5 the sense is समास_हो. As in sentences 3 and 4, the verb 'dissolve' has an Intransitive usage, i.e., it has no Object. How is it then different from घुल? Can we then look at the Subject for a clue? The Subject is 'tension' which is an abstract noun indicating an emotional state. We try to see if other such nouns can replace 'tension'. We find that 'fear', 'anger' and 'anxiety' can substitute 'tension'. For instance:

1. Her fear dissolved when the police arrived.
2. The leader's anxiety dissolved after the poll results were declared.

The difference between the senses of 'dissolve' in sentences 3,4 on the one hand and 5 on the other is that though the verb is Intransitive in all these sentences, the Subject of 3 and 4 is 'concrete' whereas the Subject of 5 is 'abstract'.

(d) sense भंग.
In sentences 6 and 7 we observe that both the sentences are in the passive voice. We may note here that while writing the WSD rules, we should consider the sentences as they would be in the active voice, e.g.,

> X dissolved their marriage in 1968.
> X dissolved the parliament.

'Marriage' and 'Parliament' which are occupying the Subject position in sentences 6 and 7 are, in fact, the Objects of the verb 'dissolve' in the active voice.
We observe that 'dissolve' is a Transitive Verb in these sentences, i.e., it requires an Object. How is it then different from sentences 1 and 2 with the sense घोल?
Is it possible that the nature of the Object has something to do with the difference?
We discover that the Objects in sentences 6 and7 (marriage, parliament) happen to be 'abstract' as opposed to sentences 1 and 2 where the Objects are concrete (sugar, salt).

5: Framing rules with 'if-then' statements.

At this stage we have to write our 'observations' as logical statements. In 4 above, we had noted observations that may have been responsible for obtaining different senses of a word. We shall now encode these observations by formulating rules for the machine in the format designed for this. Following are the 'contextual rules' for the four senses of the word 'dissolve'.

Word: dissolve
Rule File: dissolve.clp
Example: Dissolve two spoons of sugar in warm water.
         Why are you dissolving salt in your tea?
Rule id: dissolve1
Saliency:low
If the root is dissolve

if dissolve is a/an transitive verb
Then its Hindi meaning is घोल
Which is a root
Sentence type: declarative (सामान्य वाक्य)

Word: dissolve
Rule File: dissolve.clp
Example: I think the sugar has not dissolved completely.
          Oil does not dissolve in water.
Rule id: dissolve2
Saliency: low
If the root is dissolve
if dissolve is a/an Intransitive verb
Then its Hindi meaning is घुल
Which is a root
Sentence type: declarative (सामान्य वाक्य)

Word: dissolve
Rule File: dissolve.clp
Example: The tension in the office just dissolves when she walks out.
Rule id: dissolve3
Saliency:high
If the root is dissolve
if dissolve is क्रिया and it has ?noun as its subject
and
if ?noun has the semantic property abstract
Then its Hindi meaning is समास_हो
Which is a root
Sentence type: declarative (सामान्य वाक्य)

Word: dissolve
Rule File: dissolve.clp
Example: The Parliament has been dissolved.
          Their marriage was dissolved in 1968.
Rule id: dissolve4
Saliency: medium
If the root is dissolve
if dissolve is क्रिया and it has ?noun as its object
and
If ?noun has the semantic property abstract
Then its Hindi meaning is भंग
Which is a root
Sentence type: declarative (सामान्य वाक्य)

6: Filling up the Rule Forms.

Once the 'if-then' statements are ready filling up the Rule Forms is an easy task. All you need to do now is to enter your inputs (already worked out in 5) into the form systematically (as described in the section : How to fill the Rule Form.)

Rules once framed, must be tested and improved from time to time. The Anusaaraka team shall also give feedback on the rules framed and submitted by individuals. However, the onus for improving the rules rests with the Participant, for the words assigned to them.

This has been further elaborated in section 4 & 5 of these Guidelines.

Note: Those who are comfortable using CLIPS (an Expert System Shell). Refer to **Appendix 3.2b – Lakshancharts in CLIPS.**

Appendices for section 3:

Appendix 3.1 – Drafting 'if-then' statements.
Appendix 3.2a – Lakshanchart Rule forms.
Appendix 3.2b – Lakshancharts in CLIPS.

IMPORTANT:
In this section, one has to keep in mind that the focus should be on observing the patterns of sentence formations in the source language (i.e. English) and NOT the target language.

Exercise for section 3:

a) Write 'if-then' statements for the final senses (i.e. after clubbing), from the previous module.

b) Fill the 'Rule form' for the same.

Please contact: anusaaraka@chinfo.org for feedback on the Exercises done by you.

2.4 ANUSAARAKA OUTPUT GENERATION & TESTING

Objective:

- ✓ To verify the rules created by running them in Anusaaraka
- ✓ Analyze inappropriate translations and improve / revise the rules, if required.

Requirements:

Mandatory:

- ✓ Familiarity with reading Anusaaraka output**.**
- ✓ Substantial experience of having worked on the 3 previous sections.

Procedure:

The following steps are recommended to achieve this objective:

Step 11: Create a file having example sentences for each word that is being tested. Save this file with a name as per the following format: <Word>-test-input-<dd/mm/yy>. Here 'Word' in angular (<>) brackets would be the word for which the rules are being tested and 'dd/mm/yy' in the angular brackets stands for the date, month and year, when the file is created. Therefore, if your word is 'way' and you created a test file on May 1st 2009, the name of the test file would be: 'way-test-input-010509'.

This file will be continuously updated as and when new sentences for the word under consideration are added. The updation could be due to adding new rules in the rule file for a new sense, or, a new type of sentence, which you think should be included in the test-input file.

Whenever you make any changes in the file, please change the date in the filename. However, the earlier file will still be kept for reference. Thus, the file with a new date will be a new file which will be used for testing.

Step 12: Evaluate the suitability of the rules framed for a particular word, by 'loading' and 'running' the test-input file in CLIPS software. Please refer to the Appendices for this Module**.** One may also like to refer to the CLIPS 'help' function – available as an option with the software itself **(Appendix 4.1 CLIPS Usage Guide for Anusaaraka).**

Step 13: After running Anusaaraka on the test-input file, take the output and separate sentences which have correct output for the word under consideration; and sentences in which the output is wrong.

1. Store the **correct sentences** in a file with a name as per the following format: <Word>-ref-examples-<dd/mm/yy>. 'Word' angular (<>) brackets is the word for which the rules are being tested and 'dd/mm/yy' in the angular brackets stands for the date, month and year when the file is created. For example, if

one is testing sentences relating to the word 'God' on May 01, 2009; the sentences in which the translation is in the correct sense should be stored as a file titled: 'God-ref-examples-010509'.

2. Store the **sentences with incorrect** output in the file with a name as per the following format: <Word>-errors-<dd/mm/yy>. 'Word' angular (<>) brackets is the word for which the rules are being tested and 'dd/mm/yy' in the angular brackets stands for the date, month and year when the file is created. For example, if one is testing sentences relating to the word 'God' on May 1, 2009, the sentences which give an inappropriate sense on translation, should be stored as a file titled: 'God-errors-010509'.

Note: Step 13 will be repeated every time new sentences are added to the 'test-input' file.

Also, each time a text which has the word under study is tested on Anusaaraka, the sentences for which the output is appropriate (i.e the word under study is translated in the right sense) – are to be copied and pasted from the 'test-input' file to the file 'ref-examples' (i.e. file containing sentences with correct output).

The 'ref-examples' file will act as a reference to check if the addition or revision of a rule effects the output of the existing rules. If it does, a re-look would be required, both at the existing rules and the added/revised rule(s). Therefore, it is important to keep a file which has the sentences for which correct output was achieved.

Similarly, copy the sentence which do not give appropriate output in the file named <Word>errors<dd/mm/yy> (implying inappropriate output).

STEP 14: Go back to the rules file and analyze why the output is wrong / inappropriate. Make changes to the rules after this and check the output again (i.e go back to Step 4.2). Each time you make a change in the rule file, run Anusaaraka again, on the 'ref-examples' file to check its effect on the other rules.

STEP 15: Check the output and compare with the earlier correct outputs. If any of the earlier correct outputs are NOW wrong, re-look at the change made by you.

Appendices for section 4:

Appendix 4.1 – CLIPS usage guide for Anusaaraka.

Resources for section 4:

Version of CLIPS software. This is available for download under Resources at: (url)

Exercise for section 4:

a) Create '<Word>-test-input-<dd/mm/yy>' file containing the example sentences for the words that you have worked on in the previous sections.
b) Create two other files titled <Word>-ref-examples-<dd/mm/yy>; and <Word>-

errors-<dd/mm/yy>, to separately record sentences which are appropriately and inappropriately translated, respectively.

c) Test your rules on the '<Word>-test-input-<dd/mm/yy>' file.
d) Segregate the results of the test in the respective files.
e) Re-evaluate and edit rules, as required.

2.5 MAINTENANCE OF WORDS

Objective:

- ✓ To ensure that there is a regular and constant improvement in the output generated by Anusaaraka; by using varied sample sentences and testing output; and making qualitative improvements to the rules.

- ✓ A re-look at the rules formulated earlier, for revision/modification.

Requirements:

Mandatory:

- ✓ Substantial experience of having worked on the above modules.
- ✓ Framing of rules for some words.

Procedure:

You will now and then receive sentences having inappropriate output for the words for which you have framed the rules. For these sentences repeat the procedure given in sections 3 & 4 (i.e. Creation of Lakshancharts and Anusaaraka Output Generation & Testing).

Alternatively, you can test the rules created for the word(s) allocated to you, on sentences taken from the Corpus.

The Corpora that you may use for this purpose are available at the following links: (also referred to in section 1, under resources)

> http://www.americancorpus.org

or,

> http://corpus.byu.edu/

Every time you find the output not up to the mark, analyze, discuss these with others in the **Anusaaraka Discussion Forum**, and make changes to enhance the output.

You shall by default become a member of the Anusaaraka Discussion Forum on registering as a participant with the Anusaaraka Language Resource Development site. In case of any queries – please write to the administrator the the Forum at:(email id)

IM PORTANT: Individuals are, by this stage, expected to have 'adopted' words, in the sense that they take complete responsibility for the words assigned to them. These are the words that they have been working on this far.

'Adopting' a word would mean that individuals regularly test 'new' usages of the word; revise and improve rules, so that the Anusaaraka output for that word is with minimal error.

However, although the individual is 'responsible' for the 'maintenance' of the word, s/he does not have to do all improvements by herself / himself. S/he can discuss, seek ideas, suggestions, raise queries, and share thoughts with a group of Anusaaraka Participants working on the machine translation area. In other words, the **Anusaaraka Discussion Forum** will discuss the language related issues and work towards resolving them. This will lead to better output of the system.

Resources for section 5:

Parallel Corpora (links mentioned under Procedure in this section).
Membership of the Anusaaraka Discussion Forum.

CHAPTER 3

INTRODUCTORY GRAMMAR FOR ANUSAARAKA WSD TASK.

Note: Please note that this is an Introductory grammar for a specific purpose. It is possible that some of the terms used here are not explained in detail. In case you need further clarification, please refer to any standard grammar book.

Introduction

Sentences are formed using words and establishing their relations to each other. The type of relations the words have with each other have a bearing, not only on the meaning of the whole sentence but, some times, also on the meaning of the word in the context. Therefore, while developing rules for Word Sense Disambiguation(WSD), it is important to consider the relations of the words as it might have a role in the realization of one particular sense of a word out of its multiple senses. Described below are some grammatical notions and terms which can be of relevance while framing the WSD rules. In other words, here is a simple grammar for the WSD task for the Anusaaraka system which has only those concepts and terms which might be relevant for this task.

3.1 PARTS OF SPEECH (POS)

All words belong to some word class or the other. These classes are defined depending on the primary function the words perform in a sentence. For instance, in the following sentence:

A lonely child walked along the green hedge.

the words
'a' and 'the' function as **articles**;
'lonely' and 'green' are **adjectives**;
'walked' is a **verb**;
'child' and 'hedge' are **nouns**; and
'along' is a **preposition**.

'Article', 'adjective', 'verb', 'noun' and 'preposition' have different functions in the sentence and are classified as different word classes. These classes are known as Parts of Speech (POS).

Given below is a brief description of some of the major word classes/parts of speech with examples.

Verb (क्रिया)

The word in a sentence indicating performance of an action is called a Verb (क्रिया). A

Verb (क्रिया) is an action word. The words printed in capital letters in the following sentences are examples of Verb (क्रिया).

The dog CHASES the cat.
There ARE flowers in the garden.
I must not GO to the store.
ARE you ready for next week.
He RUNS fast.
I WONDER when he WILL COME!
He is apparently LEAVING.
He DIDN'T RUN fast enough.
It IS likely that Jane WILL GO.

That part of the Verb (क्रिया) which is present in all its forms is called the root (धातु). All words are composed of two parts – a base form and one or more suffixes. The base form of a verb is the root (धातु) and the base form of a noun is also the root, referred to as प्रातिपदिक.

For example, when all the suffixes are removed from the following forms of the verb:
works, worked, working
only 'work' is common in all of them.

Similarly, चल is present in all the following forms of the verb after all the suffixes are removed:
चल, चलना, चलो, चला, चली, चले, चलें, चलिए

'work' and 'चल' are **roots**.

We are often tempted to treat the verb in its infinitive form as the root. For instance,
चलना, बोलना, देखना etc.

But this is a mistake because the -ना at the end of these words is a suffix which marks the verb in its infinitive form. It is parallel to 'to' in the following English examples:
to walk, *to* speak, *to* see.

Thus, while writing contextual rules, we give the Hindi meaning of the word 'spoke' in the following sentence:
The teacher spoke for two hours.

as the root बोल and **not** बोला, बोली, बोले etc.

A verb may tell us:

(a) What a person or a thing does.
The cat SLEEPS.
(b) What is done to a person or thing.
He BROKE the window.
(c) What a person or a thing is.
This glass IS brittle. I AM a woman.

Verbs can be of different kinds:

(i) **Transitive:** A **verb which has an object** and the effect of the action is felt by the object is called a transitive verb. In the following sentences 'kick' and 'ate' are transitive verbs with the objects 'ball' and 'mouse' respectively.

> The boy **kicked** the ball**.**
> The cat **ate** the mouse.

(ii**) Di-transitive:** Some **verbs must have two objects** and are called 'di-transitive'. In the following examples 'give' is a di-transitive verb with the two objects 'me' and 'money' . Similarly, 'sold' is a di-transitive verb which has 'him' and 'farm' as its objects.

> She gave **me** some **money**.
> They sold **him** the **farm** at a very low price.

(iii) **Intransitive:** Verbs that **do not take an object** are intransitive verbs. The effect of the action is thus limited to the subject. 'Sleep' and 'smile' are intransitive verbs in the following sentences. The effect of sleeping and smiling is limited to the cat and mother respectively:

> The cat is **sleeping**.
> Mother **smiled** quietly.

(iv) **Stative:** Verbs indicating the **state** rather than action of a person, animal or thing are called stative verbs.

> The lion **was** hungry.
> The building **is** tall.
> I **know** the answer.

Verbs appear in a sentence as (a) Finite or (b) Non-finite:

(a)**Finite:** Verbs that **indicate tense** are called finite verbs. 'Write' and 'rise' in the following sentences are finite verbs indicating past and present tense respectively.

> I **wrote** a novel.
> The sun **rises** from the east.

(b) **Non-finite:** कृदन्त**.** These verbs **do not indicate tense**. 'Swimming' in both the following sentences is 'tense-less' and behaves somewhat like a noun. Such forms of the verb are called **gerunds**. Notice that it is the finite verbs 'help' and 'love' that indicate the present and past tense respectively, in the sentences given below.

> **Swimming** helps us to stay fit.
> The children loved s**wimming** .

Similarly, 'to see' (the **infinitive** form) and 'broken' (the **participle** form) are non-finite verbs. The verbs indicating tense in these sentences are 'want' and 'need'. The participle form 'broken' functions as an adjective in the last sentence.

> She wanted **to see** the doctor urgently.
> **Broken** windows need to be repaired.

> **Gerund:** Nouns derived from verbs.
> **Infinite:** Verbs when they occur with 'to'.
> **Participle:** Verbs when they occur with 'en' suffix.

Noun (संज्ञा)

The name of a person, place, object or emotions etc. is called a Noun **(संज्ञा).**

> **Radha** lived in **Brindavan**.
> The **book** is on the t**able**.
> We all seek **peace** and **happiness**.
> **Boredom** can lead to **depression**.

Nouns can be of several types:

(i) **Concrete:** These nouns refer to entities that exist in both space and time.
> **Mother** is cutting the **potatoes** with a **knife**.
> The **trees** are shedding **leaves**.

(ii) **Temporal:** These nouns code entities that exist only in time.
> It was a rainy **Sunday morning** in **July**.
> I shall meet you on **Monday** in the **evening**.

(iii) **Abstract:** These are nouns that code entities whose existence cannot be defined in terms of either time or space.
> I do not agree with the new government **policy**.
> We love **freedom**.
> Nehru was known for his **love** for children.

(iv) **Count:** These nouns code individuated entities and therefore can be counted.
> Three **men** walked towards the red **houses**.
> Please buy two **pens** for me.

(v) **Mass:** These nouns code groups or un-individuated entities and therefore cannot be counted or pluralized.
> **Water** is precious for all of us.
> These days we are breathing a lot of polluted **air**.
> It is not necessary that **blood** should be thicker than **water**.

Preposition / Postpotision (संबन्ध बोधक)

These are entities that indicate the relation between words in a sentence. The relation may be between a verb and a noun or between two nouns.

In English, these entities are positioned before the noun and are therefore called 'prepositions'. Notice that in (i), (ii) and (iii) below, the prepositions 'in','after' and 'of' occur before the nouns.

In Indian languages, these entities are placed after the nouns and are called 'postpositions'. In (iv), (v) and (vi) the postpositions occur after the nouns.

    (i) The child plays **in** the park.
    (ii) I will see you **after** the class.
    (iii) The father **of** the boy was very upset.

    (iv) बच्चा बगीचे में खेलता है ।
    (v) मैं तुम्हे घर पर मिलुंगी ।
    (vi) मोहन के पिता बच्चों को पढ़ाते हैं ।

Adjective (विशेषण)

The word that says something about/modifies a noun is called an adjective.
'Red', 'sweet' and 'little' indicate the qualities of the 'shirt', 'water' and 'boy' in the following sentences:

    Wear your **red** shirt today.
    The water of Godavari is **sweet**.
    The dog chased the **little** boy.

Adverb (क्रिया विशेषण)

The word that indicates the nature of a verb (action) is called an adverb.

    He read the letter **slowly**.
    She **quickly** finished her work.
    It is **likely** to rain today.

Adverbs can be of several types:

(i) **Manner adverb:** It adds to the meaning of a verb.
    He did it **thoughtlessly**.
    She spoke **softly**.

Manner adverbs can also be expressed as prepositional phrases.
    He fought **like a tiger**.
    I am here **by accident.**

(ii) **Time adverb:** These give information about the time, frequency, or other temporal aspects of the event.
    Why are you **always** late?
    Soon the leaves will start falling.

(iii) **Evaluative adverbs:** These adverbs convey the speaker's attitude.
    Unfortunately, you are no longer my friend.
    Hopefully the war has ended.

(iv) **Emphatic adverbs:**
    I **just** know he will not cheat me.
    She did **exactly** what she was told not to do.

Adverbs can also modify, quantify or intensify the meaning of adjectives.

> We are **very** lucky.
> This is an **unbelievably** weak strategy.

Determiner

In English, determiners precede the noun and in a sense modify them.

Following are some classes of words that function as determiners:

(i) **Articles:** English articles are of two types:

(a) **Definite: the -** this is used to indicate something specific/particular.

> Keep **the** book here.
> I loved **the** music.

(b) **Indefinite: a/an, some -** these are used to indicate something that is not specific.

> I saw **a** bird in the morning.
> He eats **an** egg every day.
> I want **some** money.

(ii) **Demonstratives:** English demonstratives - **this, that, these, those** - encode the features of 'distance from the speaker' and 'number'.

> **This** is my book.
> **That** was a bird.
> **These** are roses.
> **Those** books are yours.

(iii) **Possessor pronouns:**

> **My** sister sings.
> **Your** cat is filthy.
> I met **his** uncle.
> **Her** dress is neat.
> These are **our** students.

(iv) **Quantifiers** (like all, many, few etc.) and **numerals** are also considered to be demonstratives as they are noun modifiers.

> **All** humans are equal.
> Lord Ganesha has **four** arms.

Apart from the above mentioned word classes, conjuncts, particles etc. are some other parts of speech categories.

3.2 INFLECTIONS: NOMINAL AND VERBAL

Inflection expresses the grammatical information about a word. It shows the tense,

aspect, mood, number, gender, person, case etc. of a word. In English nouns and verbs express inflected forms.

3.2.1 Nominal Inflection

Nouns in English inflect for number. The plural is generally marked with a suffix, e.g.,
  school-**s**
  boy-**s**
  bus-**es**

However, a few nouns have irregular plurals, e.g.,
  woman > women
  ox > oxen
  wife > wives
  foot > feet
Very few nouns have a 'zero' plural marker, i.e., both the singular and the plural have the same form, e.g.,
  sheep > sheep
  deer > deer

3.2.2 Verbal Inflections

Verbal inflection in any language is the most complex part of the grammar. Apart from encoding information about Tense, Aspect and Modality (TAM) of an action, verbal inflections can also encode information about 'person/number/gender' in some languages. However, we shall restrict ourselves to TAM for our present purposes. Although TAM is relevant for all languages, we shall limit the examples to English as it is of immediate relevance to us.

3.2.2.A - Tense

The function of Tense is to encode the relation between two points along the linear dimension of time -
  a) the time of speech, and
  b) event time.

We can visualize the continuum of time in the following manner:

```
        past              present              future
         |                  |                    |
         |                  |                    |
---------X------------------X-X-X----------------X--------
                            |
                            |
                        speech time
```

The time of speech is the reference point for event time. Commonly, a verb relates to three tenses:
  (i)   **Past**: Here the event time precedes time of speech

(ii) **Future**: Event time follows the time of speech
(iii) **Present**: Event time and time of speech are simultaneous.

(i) Past

In English, the past tense is most of the time marked by the suffix -**ed**. For instance,

> talk-**ed**
> happen-**ed**
> look-**ed**
> pass-**ed**

However, English has many verbs that have irregular past forms which are unpredictable and the verb stem itself undergoes change. For instance,

> think > thought
> feel   > felt
> sit   > sat
> break > brought
> know > knew

(ii) Future

There are three forms to mark the future tense in English.

- the modal auxiliaries **will/shall**, e.g., I **will** meet the minister tomorrow.
  > I **shall** look into the matter.
- the complex auxiliary **be going to**, e.g., I **am going to** sleep now.
  > They **are going to** Delhi next week.
- the progressive auxiliary **be- -ing**, e.g., I **am** leav**ing** in the evening.
  > She **is** arriv**ing** tomorrow.

The three options seem to mark subtle shades of difference in the intention of the speaker. The use of 'will' perhaps signals more formality and more distant future than the use of 'be going to' and 'be- -ing'.

(iii) Present

There is no specific marker for the present tense in English. It is only with the third person singular nouns that the present tense is marked on the verb with the suffix '-*s*'. For instance,

> She love**s** to read poetry.
> He pray**s** every morning.
> It taste**s** sweet.

Present tense is also encoded by the auxiliary verbs. For instance,

> He **is** a brave soldier.
> They **are** our neighbors.
> Where **have** you kept the chocolates?

3.2.2.B - <u>Aspect</u>

Aspect is distinct from tense though it often combines with various tenses. If tense encodes an event time in relation to speech time, aspect relates to the nuances of the event/action at any given point of time. These nuances involve the temporal properties of the action – continuity, sequentiality or completion – whether in past, present or future tense.

Some major 'aspects' in language are:
   (i) Progressive
   (ii) Perfect
   (iii) Habitual

(i) Progressive (also called 'Continuous')

The progressive aspect relates to an action that is in progress, i.e., it captures an ongoing process – by implication then, the action has not been completed. In English the progressive aspect is marked by the auxiliary verb '**be**' before the main verb and the main verb takes the suffix -***ing***.
   ✓ Present progressive: The boys **are** sing**ing.**
   ✓ Future progressive: The boys will **be** sing**ing**.
   ✓ Past progressive: The boys **were** sing**ing**.

The progressive form can also signal a habitual action. For instance,
         Whenever I visit him he **is reading** a book.

This sentence may be paraphrased to mean – 'Every time I visit him I **always** find him reading a book'.

(ii) Perfect

Perfect is the most complex of the aspects. It involves three temporal points -
   ✓ Time of speech
   ✓ Event time
   ✓ Reference time (i.e., the event is seen with reference to another point of time).

The prefect aspect is used to indicate an event that has been completed before the reference time. It is marked by the auxiliary **have** and -**ed** suffix with the verb (or past participle form of the verb). The following diagrams may be helpful for clarity.

(a) Present perfect

_____**X**_____**O**_____
      event time   both
                                                              reference time and
                                                              time of speech

         Rama **has** finish**ed** his dinner.
         I **have** wash**ed** the clothes.

The present perfect indicates that the event is still relevant at the time of speech.

(b) Past perfect (also called 'Pluperfect')

```
___X_____#_____O_____
    event time            reference time          time of speech
```

By the time Mohana arrived the guests **had left**.

In the above example there is reference to two events at the time of speaking (a) the arrival of Mohana and (b) the guests leaving. Here, the event of the 'guests leaving' (event time) takes place with reference to the time of 'Mohana's arrival' (reference time). Event (b) has taken place before event (a), and both these events have taken place before the time of speech. Following are some more examples:

When I reached home, she **had gone** to sleep.
By the time he woke up the show **had ended**.

(·c) Future perfect

```
_____O_____X_____#_____
    time of speech            event time          reference time
```

I **will have finished** the book by morning.

The event time of 'finishing the book' and the reference time of 'morning' occur after the time of speech. The event time comes before the anticipated reference time.

Some more examples:
I **would have left** by then
They **will have constructed** the building by next week.
The police **will have caught** the thieves before long.

(iii) Habitual

There is a habitual aspect in which an event is *timeless* in the sense that the event time is either left *unspecified* or the event *always* occurs. The simple form of the present tense is used for this purpose. For instance,
The sun **rises** from the east.
The Revenue Department **collects** tax.
Monkeys **eat** peanuts.
You **lac**k commitment.

3.2.3 Modality

Modal verbs are auxiliary verbs in English. They capture the intention or attitude of a speaker. These attitudes include the speaker's judgment concerning the main verb in the clause. The judgments may relate to truth, probability, certainty, belief, evidence,

desirability, preference, ability, obligation etc. Each one of them can be seen as a sub-modality.

There are four major modal verbs in English along with their past tense forms:

present	past
can	could
shall	should
will	would
may	might

Most of these modal verbs can code more than one sub-modality. For instance, it is possible to derive three different senses from the modal verb 'can':

- ✓ Ability: If he applies himself he **can** solve these problems.
- ✓ Probability:The one who peeped through the window **could** be a thief.
- ✓ Permission: Yes, you **can** open the window now.

Similarly, notice the different senses captured by the modal 'should':

- ✓ Obligation: They **should** stop bombing Iraq.
- ✓ Probability: Rani **should** be here any moment now.

OR

'Must'

- ✓ Obligation:You **must** listen to your inner voice.
- ✓ Probability: It **must** be just accidental.

3.3. SOME GRAMMATICAL CONCEPTS

3.3.1 Subject

The concept of 'subject' is very complex. We shall refrain from going into the theoretical intricacies and limit our description of the term for our immediate purpose and only in relation to English.

In English **declarative** sentences, Subject is an element which **occurs immediately before the verb** .
The **DOG** chases the cat.
HE runs fast.
HE didn't run fast enough.
IT is likely that **Rekha** will go**.**
He will go to Rajasthan next year.

In **interrogative** sentences it **occurs immediately after the auxiliary verb** (an auxiliary verb is the supporting/helping verb as opposed to a main verb).
Will he go to London next year ?
May I come in?

Have you seen my red book?
Are you ready for the exams?

The Subject (in the third person) also **agrees with the verb in number**.

He runs fast.
They run fast.

3.3.2 <u>Dummy subject</u>

In English the **subject** occurs before the verb and it is a **position** which **cannot** be left **empty**. Therefore, sometimes you do find an element which occurs in this position but **does not contribute to the sentence meaning**. Such a subject is called a dummy subject.

THERE are flowers in the garden.
It is raining.
Is **it** going to rain tomorrow?
She loves to take a walk when **it** snows.

3.3.3 <u>Raised subject</u>

Since the subject position cannot be left empty, sometimes either the subject or the object of the verb of a subordinate clause is raised to the subject position of the main verb.

John is easy to please.
John seems to run fast.

'John' is in the subject position of 'is' but is actually the object of the verb 'please'. Similarly, in the second example, 'John' is in the subject position of the verb 'seems' but is actually the subject of verb 'run'.

The above sentences can be paraphrased as follows :

It is easy to please **John.**
It seems that **John** runs fast.

In both the paraphrased sentences, 'it' is a dummy subject. So there is a possibility of filling this position by an argument of the subordinating verb.

3.3.4 <u>Object</u>

An object is that element in a sentence which denotes somebody or something involved in the subject's performance of the verb. In English it occurs normally immediately after the verb.

The dog chased **the cat**.
I gave a **book** to Harry.
The monkey imitated **the child**.

Some verbs can take two objects. Thus, there are two types of objects – (a) **Direct object** and (b) **Indirect object.** In the following sentence,

I gave **Harry a book.**

both 'Harry' and 'a book' are objects. The object immediately after the verb is called the Indirect Object and the second object (here 'a book') is called the Direct Object.

3.4 GRAMMATICAL RELATIONS

As mentioned in the Introduction to this chapter, the words by themselves do not convey the full meaning of a sentence. The relation they hold with each other helps us in understanding the meaning of the sentence. Also, very often, a particular sense of a word is understood because it holds a certain relation with another word in the sentence in which it occurs. Therefore, it is important for us to get to know some of the important relations that the words hold with each other. The following section has the definitions and descriptions of some of the important relations which would be of help for our task.

Before we go to any particular relations between words in a sentence, it will be useful to note that Verb (क्रिया) plays a central role in any sentence. Therefore, the relation that a क्रिया has with the other elements in the sentence is of vital importance. We shall thus begin with the relations where क्रिया is involved.

A relation exists between two elements of which one modifies the other (modified). In our convention the modified element is on the left hand side (LHS) and the modifier on the right hand side (RHS). LHS and RHS elements are separated by a hyphen.

For example: क्रिया-अधिकरण
            (LHS)-(RHS)
This can also be read as क्रिया का अधिकरण ।

We use the "_" (underscore) sign between two elements to indicate that together they form a single unit(e.g., क्रियार्थ_क्रिया).

3.4.1 क्रिया-subject

It is the relation between the क्रिया and its subject. In the following sentences,
            The CAT CHASED the mouse.
            HARI DRANK the milk.
'cat' and 'Hari' are the subjects of the क्रिया 'chased' and 'drank' respectively.

3.4.2 क्रिया-object and क्रिया-object 2

In a sentence a verb can have more than one object, e.g.

            She SENT THE BOSS A LETTER explaining the new situation.

The object which comes immediately after the verb will be called **object** and the object that comes after that will be called **object 2**.

Thus, in the sentence above, 'the boss' is the **object** and **'a letter'** is **object 2** of Verb

(क्रिया) 'sent'**.**

Similarly, in
        Kunal GAVE SURESH A BOOK.

The relation between 'gave' and 'Suresh' is that of क्रिया- object; and the relation between 'gave' and 'a book' is that of क्रिया-object 2.

Note that the above sentence can also be framed as
        Kunal gave a book to Suresh.

Here we shall say that the object of the verb 'gave' is 'a book'. 'Suresh', which is preceded by a preposition, is part of the prepositional phrase 'to Suresh' and it will not be considered as object 2 of the verb 'gave'.

3.4.3 क्रिया-वाक्यकर्म

If the object of a verb is 'a sentence', then the relation between the main verb and the verb of the embedded sentence is that of क्रिया-वाक्यकर्म. For instance, in the following sentence,
        He SAID Mohana WILL COME to the dance class.

In the above sentence 'Mohana will come to the dance class' is a subordinate clause. The subordinate clause is the वाक्यकर्म of the verb 'said'. The verb of the subordinate clause is 'will come'. Thus the relationship between the main verb 'said' and the verb of the वाक्यकर्म 'will come' is that of क्रिया-वाक्यकर्म.

Similarly, in

        They CLAIMED Tendulkar had PLAYED well.

the sentence 'Tendulkar had played well' is the subordinate clause and the relation between the main verb 'claimed' and the verb of the subordinate clause 'played' is that of क्रिया-वाक्यकर्म.

3.4.4 क्रिया-अधिकरण (Location)

This is a relation that expresses the location of the activity denoted by the verb. The location can be either in time or place. In the following sentence,
        I WROTE the exams in LATE DECEMBER.

'late December' is the अधिकरण (time) of the क्रिया 'wrote'.

Similarly. In
        Birds are CHIRPING in the GARDEN.
'garden' is the अधिकरण (place) of the क्रिया 'chirping'.

Both the locational relations may occur in a sentence for the same verb. Thus, in the

following sentence,

I MET him in LONDON (place) LAST YEAR (time)

'London' and 'last year' are the place and time अधिकरण of the क्रिया 'met'.

3.4.5 क्रिया-क्रियार्थ_क्रिया

When there are two verbs (indicating two actions) in a sentence and one action is necessary for the other action to happen, the first action is the क्रियार्थ_ क्रिया of the main क्रिया. In the following sentence:

I RAN to CATCH the bus

'ran' is the क्रियार्थ_क्रिया of the क्रिया 'to catch'.

Similarly, in the sentence,

You must WORK HARD to REACH the top

'work hard' is the क्रियार्थ_क्रिया of the क्रिया 'reach'.

3.4.6 क्रिया- उपसर्ग (Particle)

Sometimes, certain particles when combined with a verb can change its meaning. For example, when we add the particle '**on**' to the verb '**switch**', we get the meaning of 'starting something' like a light , fan or generator.

SWITCH ON the light please.

It is time to SWITCH ON the lamp.

But when we add the particle '**off**' to the same verb 'switch' we get the meaning of 'ending/terminating something'.

Please SWITCH OFF the lights and fans when you leave the room.

When you are not watching the TV please SWITCH it OFF.

'On' and 'off' are the उपसर्ग of the क्रिया 'switch'.

3.4.7 क्रिया- प्रेरक_क्रिया

This relation concerns the use of a **causative** verb. A causative verb is used when we do not carry out an action ourselves, but are responsible for the action being performed. For example, in

The teacher MADE the students RUN after the class.

the teacher is not himself running but is responsible for or has caused the students to run.

Similarly, in the following sentence,

Priya HAD her children COOK dinner for her,

Priya does not do the cooking herself but is responsible for the act of cooking by the children.

'Made' and 'had' in the above sentences are the प्रेरक_क्रिया of the क्रिया 'run' and 'cook'

respectively.

3.4.8 क्रिया-क्रिया_विशेषण

As observed earlier, क्रिया_विशेषण is the modifier of the क्रिया. In the following sentence,

        The dancer MOVED his feet SLOWLY.

'slowly' is the क्रिया_विशेषण of the क्रिया 'moved'.

Or, in

        Unless you WRITE LEGIBLY you would fail.

'legibly' is the क्रिया_विशेषण of the क्रिया 'write'.

3.4.9 क्रिया-विभक्ति_ word_ सम्बन्ध

विभक्ति_word refers to a preposition. In the following sentence,

      The dog RAN IN the yard.

the relation between the verb 'ran' and the preposition 'in' is that of क्रिया-विभक्ति_word_सम्बन्ध.

Similarly, in the following sentences,

      PUT the glass ON the table.
      The class BEGINS AT five thirty.

'on' and 'at ' are विभक्ति_words that relate to the verbs 'put' and 'begins' respectively.

3.4.10 क्रिया-समानकालिक_ क्रिया

This is the relation between two activities which are happening simultaneously. For instance, in the following sentence,

      They SING while COOKING.

the singing and cooking are taking place at the same time.

Similarly, in

      She ENTERED the room HUMMING a lively tune.

the two activities of 'humming' and 'entering the room' are taking place simultaneously.

3.4.11 समानाधिकरण

समानाधिकरण (समान+अधिकरण) indicates that two entities have the **same locus**.
For example, in the following sentence,

      The teacher considers RAM INTELLIGENT.

Ram is the locus of both 'रामत्व' (i.e., being Ram or the essence of being Ram) and 'intelligent'.

Or,

      KRISHNA was the CHARIOTEER of Arjuna.

In the above sentence, both 'कृष्णत्व' and 'charioteer' are embodied in the same locus-Krishna.

Similarly, in
>MOHANA is EFFICIENT.
the essence of being 'Mohana' and 'efficient' are both located in Mohana.

3.4.12 subject-subject_समानाधिकरण

When the subject of a sentence is the locus of the समानाधिकरण the grammatical relation is called subject-subject_समानाधिकरण. Following are some examples:

>I am PROUD of you.
>SACHIN TENDULKAR is a GREAT CRICKETER.
>Are YOU a DOCTOR?
>How SMART can YOU be?

3.4.13 Subject_वाक्यसमानाधिकरण

A clause that follows a verb 'be' and adds some information about the subject is वाक्यसमानाधिकरण of the subject. For example,
>The QUESTION is WHO KILLED NICOLE.

Since the verb is the head of the sentence/clause, the relation will be shown between the subject and the verb of the वाक्यसमानाधिकरण (clause)

3.4.14 Object-object_समानाधिकरण

When the object of a sentence is the locus of the समानाधिकरण the relation is that of object-object_समानाधिकरण. For instance,
>I made HIM HAPPY
>I found HIM INTERESTING
>The teacher considers PADMA to be a GENIUS.
>The children left the ROOM CLEAN.

3.4.15 विशेष्य-विशेषण

This is the relation between a noun (विशेष्य) and an adjective (विशेषण). विशेषण is the modifier and   विशेष्य is the modified entity. For example, in
>Emperor Jahangir created many BEAUTIFUL GARDENS.

'beautiful' is the विशेषण of the विशेष्य 'gardens'. The विशेषण 'beautiful' modifies the विशेष्य 'garden'.

3.4.16 विशेष्य-det_विशेषण

As mentioned earlier, (see 'Determiners' in the section on Parts of Speech), 'determiners' like articles, demonstratives and quantifiers etc. also serve to modify the noun. Thus, although they do not fall in the category of adjectives, they in a sense have the same function of modifying the noun. For instance, in
>Can you lend me SOME money?

ALL men are born equal.
I would like you to watch THAT film.

'some', 'all' and 'that' are the det_विशेषण of the विशेष्य 'money', 'men' and 'film' respectively.

3.4.17 उपमेय-more_than_उपमान

This relation connects two nouns in comparative construction. Her the comparison can be of any kind, like 'more than' or 'less than'. For example,

We do this more for PLEASURE than for MONEY.
He has more MONEY than TIME.

Note: We have given above some of the grammatical relations that you may have to refer to very often. However, there are some more possible grammatical relations. See **Appendix 3.4** for a comprehensive list of grammatical concepts which are used in these relations. This is given in a tabular form for quick reference.

3.5. SENTENCE TYPES

Different kinds of sentences can be used to serve different communicative purposes. For instance, a sentence can be uttered to make a request, give a command, make a statement, ask a question, or prevent someone from doing something etc. Given below are examples of different kinds of sentences:

5.1 Declarative - सामान्य वाक्य : Such sentences generally consist of statements. For example,
a) The weather is fine today.
b) Saleem was the son of Emperor Akbar.
c) Nirmal Verma is not with us anymore.

5.2 Imperative - आज्ञार्थक वाक्य : These convey orders, requests, commands etc. The person(s) to whom these sentences are addressed are in the second person i.e. 'you'. This pronoun is thus generally omitted in the sentence. For example,
a) Give me a glass of water.
b) Close the door.
c) Please switch off all the lights before you leave the room.

5.3 Indirect question - प्रश्नात्मक वाक्य : It expresses a question in the mind of the speaker although it is not explicitly asked. For example,
a) I wonder why we hardly had any winter this year.
b) Anwar wanted to know when Mohan would be free to meet him.
c) They asked me how it was possible to read so many books.

5.4 Yes/No question - हां/ना प्रश्नात्मक वाक्य : These are questions that require to be answered with 'yes' or 'no'. For example,
a) Has the teacher given the reports?
b) Are you following what is being said?
c) Will the class continue after lunch?

5.5 Wh-question - wh-प्रश्नात्मक वाक्य : These questions cannot be answered with 'yes' or 'no'. They seek new information. For example,

      a) When is the session likely to end?
      b) Where have all the children gone?
      c) What do you expect to learn from this experience?

5.6 Tag question - tag-प्रश्नात्मक वाक्य : Here, a question is added (tagged) to a statement to elicit confirmation/support etc. from the listener. For example,

      a) She is intelligent, isn't she?
      b) You won't make this mistake again, will you?
      c) Sabina wanted to be a doctor, didn't she?

We have tried to give some essential grammatical concepts and relations which would be required for the WSD task. You may come across cases which may not be covered here. Please contact anusaaraka@chinfo.org whenever you face any difficulty while working on your words.

CHAPTER 4

TRANSFER VERB FRAMES

Some effort in the direction of making Transfer Verb Frames, was initially started as part of LERIL (Lexical Resources for Indian Languages),an open-source, collaborative initiative of several groups (and individuals) to create shareable resources for Indian languages.

The initial effort called TransLexGram was based on available free resources. It involved development of 'verb frames'. Kendriya Hindi Sansthan (Agra) played a major role in the development of this resource initially. In collaboration with IIIT-H, Kendriya Hindi Sansthan conducted a number of workshops and transfer verb frames were developed for all the verb entries in Shabdanjali (a resource wherein various senses of a word are available – somewhat like a Thesaurus).

The objective of this chapter is to introduce participants to the process of developing transfer verb frames between English and Hindi

---

4.1 Procedure:

It is assumed that you have already translated English example sentences into Hindi for a given word. After translating the sentences:

a) For 'verb' entries, 'frames' have to be made both for English and the corresponding Hindi verbs.

> **Note:** Please note that the sense id in the Lakshanchart should correspond to the sense id in the verb frames.

b) Please note that a single verb may have many sense. Separate frames have to be made for each sense. Frames for verbs:
  i)   should contain only the mandatory arguments of the verb;
  ii) should be in 'declarative sentence' form
  iii) should contain the prepositions (if any) on the English side and corresponding
       vibhakti (case markers/postpositions) on the Hindi side.

The frames would essentially capture the rules for structural transfer from English to Hindi. For example, for the English verb 'see', the following transfer verb frame will be created:

  FRAME_Eng::A sees B

  FRAME_Hin::A B [को] देखता है देखता

In the above example, Frame_Eng shows that normally the English verb 'see' has two

arguments: a subject and an object. The subject is shown by the letter A and object is shown with the letter B. When translated into Hindi, B is followed by an optional को vibhakti (postposition).

The optionality of को is indicated by placing को within square brackets.

The Hindi frame is shown against Frame_Hin.

4.2 List of Field Names

  ID-Sense - Every sense will have a unique id
  Headword - The lexical item for which the entry is being made
  Hin_Meaning - Indian language equivalent for the Headword
  Eng_Exp - Example sentence in English
  Hin_Trans - Natural translation
  Frame_Eng - Frame for the English sentence
  Frame_Hin - Frame for the Hindi translation
  Constraints – Conditions, if any, under which the 'Hindi vibhakti (case marker / postposition)' is selected

4.3 Some example entries

After translating the English example sentences (discussed in section 1.2) the frames will be built as shown in the example entries given below. The frames should be in simple habitual present tense irrespective of the mood, aspect and tense (Refer chapter 3 of Guidelines) of the verb, in the example sentences.

Example :

    Frame_Eng:: A goes to B
    Frame_Hin:: A B [को] जाता है

Standard format* for the verb frames is essential for automatic processing.

The final entry after adding the translation and grammar component would be as follows**:**

Entries (Forms 1 and 2) for English verb **'see' are on the next page.**

ID-Sense	see1a
Headword	see
Hin_Meaning	देखना
Eng_Example	I *see* the birds
Hin_Trans	मैं पक्षी देखता हूँ
Frame_Eng	A sees B
Frame_Hin	A B देखता है
Constraints	

Form 1

ID-Sense	see1b
Headword	see
Hin_Meaning	देखना
Eng_Example	**I** often **see** **Mohan** in the park
Hin_Trans	मैं अक्सर मोहन को पार्क में देखता हूँ
Frame_Eng	A sees B
Frame_Hin	A B को देखता है
Constraints	B has to be human

Form 2

Both the **Form 1** and **Form 2** above, have the same Hindi sense of English verb 'see'. However, note that the Hindi entries in Frame_Hin is different in the two cases. In Form 1, the Frame_Hin has no 'case marker (postposition)' whereas in Form 2, the Frame_Hin has a को after B. This difference is due to a rule of Hindi grammar. According to a Hindi grammar rule, if the object of verbs such as 'see' is human, it takes the को vibhakti. This rule of Hindi is covered by adding the information in the 'Constraints' field.

Also, notice that the sentence has the phrase 'in the park' also. But this has not been covered in the frame as it is NOT a mandatory argument of the verb.

As indicated above, we try to define verb frames for English and the corresponding post-position for Hindi. Linguists are the preferred people for doing this task, but the transfer verb frames is kept so natural that a person without linguistic training, but, with a language sensitivity can also contribute to it. A short period of training or self-learning is needed. Though this grammar is expected to be used by the machine directly, the framework has been kept rather simple. A few more examples are given in the Appendix.

CHAPTER 5

PADASUTRAS

When we translate a word from the source language to the target language we often find that the target language has several senses with apparently no connection between these senses. When a word has several related senses it is called polysemous and the phenomenon is known as polysemy.

Polysemy is a major issue that one has to dealt with carefully while building bilingual language resources for the machine use. The question arises is whether polysemy is a mere coincidence or can one capture the relatedness of the senses in a polysemous word?

The concept of Padasutra is an attempt to capture the underlying thread which relates various meanings of a polysemous word.

Both Padasutras and Lakshancharts can be useful for any machine translation system. They are particularly useful for the Anusaaraka system.

There are two possible reasons for a word to have more than one sense:

i) different words with the same form

 Sometimes two words from different roots become similar in form due to the process of historical change. The Hindi word बस is a good example of this. One meaning of बस is enough and the other meaning is the English 'bus' i.e. a vehicle. Another example of this phenomenon is the English word 'lie':-

'lie' 1. झूठ~बोलना

   It is obvious that he was lying to her.

   2. लेटना

    He was lying on a hard bed.

However, interestingly English does not have the same word for the actions of झूठ बोलना and लेटना.

These are two words with a similar form. When used in the past tense, it becomes clear that they are two distinct words.

       1a. It is obvious that he lied to her.

       2a. He lay on a hard bed.

ii) many senses of the same word

It also happens that the same word may have different senses when used in different contexts. Take for example, the English word 'issue'.

'issue'

1. विषय

    The committee is discussing the issue.

समिति इस विषय पर चर्चा कर रही है ।

2. संस्करण

   The article is published in the latest issue.

   यह प्रपत्र नवीनतम संस्करण में छपा है ।

3. उद्गम स्थान

   The point of issue of Ganges is in the Himalayas.

   गंगा का उद्गम स्थान हिमालय में है ।

4. संतान

   Even after five years of marriage she had no issue.

   शादी के पाँच साल बाद भी उसके कोई संतान नहीं थी ।

5. ज़ारी~करना

   Orders are issued to the effect.

   इस आशय के आदेश जारी कर दिये गये हैं ।

6. प्रकाशित~करना

   A revised version of the book is going to be issued soon.

   किताब का एक संशोधित संस्करण जल्दी ही प्रकाशित किया जाएगा.

Here, for the word 'issue' we require different words in Hindi. These words also appear to be totally unrelated with each other.

Does it mean that 'issue' is in fact six different words?

In the historical process it is not unusual for two or three words to acquire the same form. But it seems quite improbable that six or more words should acquire identical forms.

Linking different senses of a word:

How did one word come to have different senses when used in different contexts?

Are these senses somehow related?

If they are related, what relates them?

When we ponder over these questions, we find that all these senses are at some deeper level 'strung' by a common 'underlying thread ' i.e. a 'sutra'.

This relationship can also be formally expressed as a formula (sutra).

The formal statement that indicates the underlying link between the various senses of a word may be called a '*padasutra*'.

Following are some examples of the attempt to derive a Padasutra:

Example 1: 'issue'

On analyzing the sample sentences relating to the word 'issue' given above it appears that the core meaning of the word is production. And this core meaning is present in all the senses of the word as an 'underlying thread' (अंतर्निहित सूत्र).

Let us re-examine the different senses of the word 'issue'. The sample sentences are given below once again.

The note after the sign '<--' indicates how the different senses of 'issue' are linked by a common thread.

"issue","V"

"1.ज़ारी~करना"   <-निकालना <-निष्पादन

Orders are issued to the effect.

"2.प्रकाशित~करना"   <-- पुस्तक प्रकाशित करना यानि <--पुस्तक निकालना <-यानि पुस्तक का निष्पादन

Please issue a revised edition of the book.

"issue","N","

"1.निकलने का स्थान[समय]"   <-- निष्पादन का स्थान

(Any thing that is produced has a time and place of production. This time and place of production has got attached to the meaning of the word 'issue'.)

   He buys stamps for his collection on the day of issue.

   The point of issue of Ganges is in the Himalayas.

   Issue of blood from the wound has been quite a bit.

"2.संस्करण"   <-- जो निष्पादित हुआ वह

   The article is published in the latest issue.

"3.संतान"   <-- जिसने जन्म लिया <-जो निष्पादित हुआ

   Even after five years of marriage she had no issue.

"4.विषय"   <- निष्पादन की प्रक्रिया से जो सम्बन्धित है वह

   The committee is discussing the issue.

The relationship between these different senses can be formally stated as a '*sutra*' . In the process of formulating a *sutra*, it becomes clear that though the senses appear to be different, they are linked by a common core meaning. But before formulating a *sutra*, let us see the underlying thread that links the different senses.

The *sutra* for the word 'issue' can be formulated in the following way.

निष्पादन (अस्तित्व~में~लाना/आना)--> निष्पत्ति~का~स्रोत -->

निष्पत्ति (संतान,विषय,संस्करण)

Examples for various stages:

issue orders... point of issue ..had no issue after marriage,

latest issue is out, issue under discussion,

The सूत्र for the word 'issue' can be made as follows:

"issue","विषय[~~निष्पादन]

In this *sutra* the core meaning of the word 'issue', i.e. 'coming into being' (निष्पादन), has been placed in brackets. The signs '~~' before 'निष्पादन' are indicative of the fact that there are several steps to arrive at the meaning given outside the brackets.

In the sutra, the word 'विषय' is given outside the brackets because it is used in this sense (issue under discussion) very commonly. This is an attempt at formulating a *padasutra*. In the same *sutra*, the core meaning 'निष्पादन' can be placed outside the brackets and other related senses inside the brackets.

On the other hand, we may place the most commonly used sense outside the brackets and the core meaning inside the brackets.

Example 2: 'case'

Let us see how different senses can be linked and strung by one *sutra* by taking another example from English -'case'

"case","N"

1. मामला

   It was obviously a case of blackmail.

2. डब्बा

    Keep the ornaments in that jewellery case.

3. अल्मारी"

   Objects in the museum have been displayed in big glass cases.

4. बड़े अक्षर (अंग्रेज़ी के कैपिटल अक्षर)

    Write your name in the upper case.

5. खोल

I need three pillow cases.

In the above examples 'case' has many senses.

Are all these senses unrelated?

To answer this question, examine the above sample sentences once again. Upon reconsideration, it begins to appear that the different senses are somehow connected. There is a thread running through these senses which connects the different senses of the word 'case'.

To understand this better, let us start with example 2. In this sentence the meaning of the word 'case' is 'box' (डिब्बा). Besides, we get the sense of a 'box' in other uses like 'hot-case' 'suit-case' 'brief-case' etc.. Thus, one suggestion could be that we take 'box' as a meaning of the word 'case' because it seems to be an appropriate meaning in many examples.

But English has another word 'box'. And डिब्बा looks like a good translation for 'box'. So, should we use 'डिब्बा' as a Hindi meaning for both 'case' and 'box'? And then, if both 'box' and 'case' mean the same thing in English, then why does English have two words for the same concept?

To answer this question we have to raise another question and that is, what is the difference between 'box' and 'case' in English?

From the contexts in which these two words are used, it seems 'box' refers to an 'ordinary box' and 'case' to a 'special box'. What is special about 'case' is that it takes the form/shape of the object for which it is meant. Thus, for different objects there are different 'cases', e.g., 'bangle-cases' are specially made for bangles;'comb-case' takes the shape of a comb. Similarly, 'spectacle-case', 'pencil-case', and 'suit-case' etc. are all boxes made specifically to match the shapes of the objects they are meant to contain.

Therefore, in this sense, 'case' comes closer to the word 'खोल' in Hindi.

It the fourth sentence, the 'showcase' is a special type of cupboard which is specially made for the proper display of the objects in a museum.

If we go back to 'bangle-case' once again, it would appear that a sense of 'classification' is also implied in the word 'case' because certain kinds of objects will go into certain cases.'Bottle-case' is another good example of this. The holes in a bottle case are made according to the shapes and sizes of the bottles – bottles of the same size are kept in the same case.

Here, a 'case' is meant for all those objects which are similar and are kept in a box specifically made for this purpose.

In printing presses, the letters of the English alphabet were divided and kept in two boxes (cases).These were probably kept one on top of the other. Thus, those letters kept in the upper box became 'upper cases' and those in the lower box became 'lower

cases'. This also gives a sense of classification.

This is how, 'case' has acquired the sense of 'classification'.

It is this sense of 'classification' that helps us to tie up with an apparently unrelated meaning of मामला (court-case).मामला also carries a sense of 'classification'. Different kinds of मामला are classified and kept separately on the basis of their nature and thus we get the senses of 'civil case', 'criminal case' etc.

Further consideration of the problem will make it clear that, in fact, मामला is also linked to the same *sutra*.

The *sutra* is clearly discernible in the way the senses are shown to unfold below:

   खोल -> पेटी[बक्सा]{विशेष} -> वर्गीकरण -> मामला

pillow-case..bangle-case..accusative or dative case..lower or upper case..case_statement{programming}..medical/legal_case
On the basis of the above, one way to formulate the following '*sutra*' would be:
"case","N","1. मामला[~~<खोल]"

2.3 'light'

Here is an attempt at formulating a '*sutra*' for the English word 'light'. Observe the analysis of the word 'light':

1.
"light","A","1.हलका-"

"light","N","1.प्रकाश"

"light","V","1.जलाना"
    Mohan lighted the match-stick.

2. सुलगाना
    Mohan lighted the cigarette.

3. प्रकाशित-करना
    The torch lighted the way for him.

4. चमक~आना
    His face lighted up when he heard the news.

Various uses of the word 'light' in the sample sentences point to a peculiarity of the English language. Very often in English the same word can be used as a verb or a noun. Is this also a mere coincidence?

It is a special feature of English that despite functioning as a noun, verb or adjective in different contexts, all the senses of 'light' are interconnected.

(i) In English, very often a noun can also be used as a verb. In order to change a noun into a verb, a "∅" (zero) affix is used which may be said to have the inherent meaning of "to do". Thus, on the basis of the example 'Mohan lighted the lamp.' we may infer the meaning of 'light' as 'cause to give light'.

(ii) Any 'action' has two inherent components in its meaning i) activity and ii) result. The extension of the core meaning of a verb (denoting an action) could be either related to the result or activity or some sub-activity etc. For example :

      Activity-related (व्यापार-साम्य): Mohan lighted the cigarette.

      Result-related (फल-साम्य): Mohan lighted his 4-celled torch before entering the corridor.

      Sub-activity-related (उप-व्यापार(करण का)): The torch lighted the way
                                    for him.

(iii) Metaphorical use: The glow on one's face due to enhanced circulation of blood caused by happiness / excitement may be poetically expressed as "His face lighted up when he heard the news."

(iv) Because 'light' has very little density (therefore, weight), the word may have acquired the meaning of 'having very little weight'.

On the basis of the discussion above, we may formulate the '*sutra*' for 'light' in the following manner:

    प्रकाश~[के~योग्य~करना]/हलका

Procedure for Formulating a Padasutra

From the analysis of the sample sentences for 'issue', 'case' and 'light', it is by now hopefully clear that although a word may appear to have multiple senses at the superficial level, these senses are linked at a deeper level.
Thus, we have to formulate a '*sutra*' to show the relation between the multiple senses of a word. We have tried to illustrate this procedure with the English word 'fire'.

i) To begin with, make a list of sentences in which the word is used in different senses.

"fire","N","1.आग"
The fire at the cinema hall resulted in many deaths.

"2.गोली~चलाना"
The police opened fire to hold back the violent mob.

"fire","V"

1. **"प्रहार~करना{बंदूक~या~तीर~से}"**

He fired at the robbers.

2. "निकाल~देना"
   He has recently been fired from his job.

3. "प्रोत्साहित~होना"
   Ram's words fired his imagination for developing a new application for
   this product.

4. "भट्टी~में पकाना"
   Bricks are fired in fire kilns to make them strong.

ii) The relationship between the different senses can be derived in the following manner.

To operate a cannon in the olden days, one had to first light it with fire. In this sense, it may have acquired the meaning of 'firing' a cannon or a gun. It was after it was 'lit with fire' that the cannon balls or bullets 'ejected' from the cannon/gun. This might have led to the sense of being 'thrown out' as in 'He had recently been fired from his job'.

The use of 'fire' in sentence (3) is interesting. It is somewhat related to the energy/heat generated by some excitement/anger 'आग-बबूला हो जाना', where anger may generate heat as well as redness in the face (like fire). In the sample sentence (3) though there is no physical manifestation of 'fire', it relates to the emotional excitement/energy/heat generated by 'Ram's words'.

'Baking in a kiln' in example 4 is the same as processing in the fire.

Iii) After the analysis of various sample examples of the word consideration, arrange its 'senses' in an order to indicate the distance from the core meaning.

आग~लगाना>तोप~में~पलीता~लगाना~>गोली~या~तीर~चलाना>निकाल~फेंकना)

fire in the cinema hall...fired at the enemy...fired at the robbers...fired from the job

iv) Formulate a **sutra** on the basis of the sample sentences.

"fire","आग~[>फेंकना]~लगाना"

v) Additional 'notation' can be used, if necessary, to formulate the sutra. For instance, the '>' sign indicates how we arrive at the sense of 'thrown out' from the core word 'fire'. Information regarding such notations are given in Appendix 5.1.

Appendix

Appendix 5.1 – Special notation for formulating Padasutras.