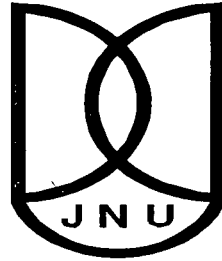


Challenges in Sanskrit-Hindi Noun Phrase Mapping

*Dissertation submitted to Jawaharlal Nehru University
in partial fulfillment of the requirements
for the award of the degree of*

MASTER OF PHILOSOPHY

KUMAR NRIPENDRA PATHAK



**Special Centre for Sanskrit Studies
Jawaharlal Nehru University
New Delhi-110067
INDIA**

2011




विशिष्ट संस्कृताध्ययनकेन्द्र
जवाहरलाल नेहरू विश्वविद्यालय
नई दिल्ली-११००६७


**SPECIAL CENTRE FOR SANSKRIT STUDIES
JAWAHARLAL NEHRU UNIVERSITY
NEW DELHI-110067**

21st July 2011

CERTIFICATE

This dissertation entitled “**Challenges in Sanskrit-Hindi Noun Phrase Mapping**” submitted by **Kumar Nripendra Pathak** to **Special Centre for Sanskrit Studies, Jawaharlal Nehru University, New Delhi-110067**, for the award of the degree of **Master of Philosophy**, is an original work and has not been submitted so far, in part or full, for any other degree or diploma of any University. This may be placed before the examiners for evaluation.


Dr. Vaishna Narang
(Chairperson)


Dr. Girish Nath Jha
(Supervisor)



विशिष्ट संस्कृताध्ययनकेन्द्र
जवाहरलाल नेहरू विश्वविद्यालय
नई दिल्ली-११००६७

**SPECIAL CENTRE FOR SANSKRIT STUDIES
JAWAHARLAL NEHRU UNIVERSITY
NEW DELHI-110067**

21st July 2011

DECLARATION

I declare that the dissertation entitled “**Challenges in Sanskrit-Hindi Noun Phrase Mapping**” submitted by me to **Special Centre for Sanskrit Studies, Jawaharlal Nehru University, New Delhi-110067**, for the award of the degree of **Master of Philosophy**, is an original work and has not been previously submitted, in part or full, for any other degree or diploma of any University/Institute.

Kumar Nripendra Pathak.
(Kumar Nripendra Pathak)

Devyarpanamastu

Acknowledgement

The effort made to bring out this dissertation is because of the inspiration of my supervisor Dr. Girish Nath Jha towards a good research in developing Sanskrit-Hindi Machine Translation (SHMT) tool kit. I was assigned the work of writing Transfer Grammar rules in the project titled 'Development of Sanskrit Computational Toolkit and Sanskrit-Hindi Machine Translation System' funded by DIT, Govt. of India. This opportunity helped me in conceptualizing the issues related to my topic. I am grateful to him for his tremendous support at every moment of my research.

I am feeling very fortunate that I received boundless blessings and valuable suggestions from Dr. Ashok Aklujār Sir. He promptly provided me with his paper titled '*Some theoretical observation on word order in Sanskrit*' through email which helped me a lot in my work. Dr. Madhav M Deshpandey Sir graced me with his valuable inputs and articles of Dr. S.D. Joshi from which I could easily understand different issues related to my topic. Brandon Gillon Sir, despite his busy schedule, provided me with his article on adjective through email. Prof. Steven Schaufele (Soochow University) helped me to understand the fact that 'free word-order' languages like Sanskrit have grammars that routinely generate continuous NPs -- but allow those NPs to be 'liberated'. It helped me in concentrating on the 2nd chapter as I could not find a single paper on Sanskrit Noun Phrase. Prof. V.N. Jha, Prof. P.K.S. Pandey, Dr. Malhar Kulkarni, Prof. Saroja Bhate, Prof. B. N. Patnaik and Dr. Ayesha Kidwai, despite their busy schedule, discussed the topic and graced me with their vast experience. Amba Kulkarni ma'am and Dipti Mishra Sharma ma'am helped me during my visit to Hyderabad for Transfer Grammar rules. Teachers from my centre, Prof. Shashiprabha Kumar, Dr. Hari Ram Mishra, Dr. Rajnish Kumar Mishra, Dr. Ram Nath Jha and Dr. Santosh Kumar Shukla were easily available to help me. I thank all of them with my deep sense of gratitude.

My all seniors, Office Staffs and Project Staffs were very much supportive. R. Chandrashekar ji took special interest in my work. Vibhuti Nath Jha, Diwakar Mishra,

Ritesh ji, Devananda Shukla, Avinesh and Piyush Arora also helped me in this research. I am grateful to them.

My friends: Shekhar, Jay, Devalina, Saurabh, Ritesh Jha, Rajneesh, Baldev, Esha, Avanti, Akansha, Resmi, Sheena, Usha Shank and Pinkey, deserve my special 'Thanks' for their help and care.

I thank my family (Dadi, Papa, Rajeev bhaiya & bhabhi, Onkar bhaiya & bhabhi, Himanshu bhaiya & bhabhi, Vira, Pranab, Ravi and Pari) for their consistent support, blessings and love. Whenever I need any support, specially, Onkar bhaiya is always there to help me in every condition. I am fortunate to find him as my bhaiya. Finally my thanks to Almighty for giving me courage and inspiration in life to move on.

Kumar Nripendra Pathak

Abbreviation

NP - Noun Phrase

MT- Machine Translation

SHMT- Sanskrit- Hindi Machine Translation

ILMT- Indian Language Machine Translation

TG- Transfer Grammar

SSF- Shakti Standard Format

IIIT- International Institute of Information Technology

TL- Target Language

SL- Source Language

NLP- Natural Language Processing

TDIL- Technology Development for Indian Languages

PLIL-Pseudo Lingua for Indian Languages

PG- Paninian Grammar

TAG-Tree Adjoining Grammar

LTAG - Lexicalized Tree Adjoining Grammar

C-DAC- Centre for Development of Advance Computing

GEB - Generalized Example-Base

REB - Raw Example-Base

EBMT – Example Based Machine Translation

KBCS - Knowledge Based Computer Systems

NCST - National Centre for Software Technology

CLIR - Cross Lingual Information Retrieval

UCSG - Universal Clause Structure Grammar

UNL - Universal Networking Language

MAT - Machine Aided Translation

OMA - Oriya Morphological Analyzer

OGC - Oriya Grammar Checker

OSC - Oriya Spell Checker

OSA - Oriya Semantic Analysis

OWP - Oriya Word Processor

RC-ILTS - Resource Centre for Indian Language Technology Solutions

SaHiT- Sanskrit-Hindi Translator

POS – Parts of Speech

R&D – Research and Development

SARS - *Sanskrit Anaphora Resolution System*

QUANT - Quantifiers

DET - Determiners

AP - Adjectival Phrase

REL - Relative clauses

N-COMP - Noun Complements

Contents

	<i>Page No.</i>
<i>Acknowledgement</i>	i-ii
<i>Abbreviation</i>	iii-iv
<i>Introduction</i>	1-4
Chapter 1: Machine Translation in India	5-25
1.1 Introduction	6
1.2 MT Systems for Indian Languages	7
1.2.1 Anglabharati	7
1.2.2 Anusaaraka by Indian Institute of Technology, Kanpur and University of Hyderabad	8
1.2.3 MANTRA by Centre for Development of Advanced Computing, Bangalore	9
1.2.4 AnglaBharti-II	10
1.2.5 AnglaHindi	11
1.2.6 Anubharti	11
1.2.7 Matra by Centre for Development of Advanced Computing, Mumbai (erstwhile NCST)	12
1.2.8 UCSG MAT by University of Hyderabad	12
1.2.9 UNL MT by Indian Institute of Technology, Mumbai	13
1.2.10 Tamil Anusaaraka by Anna University's K. B. Chandrasekhar Research Centre, Chennai	13
1.2.11 MAT by Jadavpur University	13
1.2.12 Anuvaadak by Super Infosoft	13
1.2.13 Anubaad hybrid machine translation system	14
1.2.14 Statistical MT by IBM	14
1.2.15 Oriya Machine Translation System (OMTrans) by Utkal University, Vanivihar	14
1.2.16 Shiva and Shakti machine translation	14
1.2.17 Tamil-Hindi machine aided translation system	15
1.2.18 English-Kannada machine-aided translation system	15

1.2.19	Hinglish machine translation system	15
1.2.20	English to (Hindi, Kannada, Tamil) and Kannada to Tamil language-pair example based machine translation system	16
1.2.21	Punjabi to Hindi machine translation system	16
1.2.22	Sampark	16
1.2.23	Hindi to Punjabi machine translation system	16
1.2.24	Development of Sanskrit Computational Toolkit and Sanskrit-Hindi Machine Translation System	17
1.2.25	SaHiT	17
1.2.26	Anusaarak Sanskrit Hindi MT System	17
1.3	Recent projects funded by DIT under the TDIL program	18
1.4	Work done in Sanskrit NP processing	20
1.4.1	Sanskrit POS Tagger	20
1.4.2	Sanskrit kāraka analyzer	20
1.4.3	Subanta analyzer	20
1.4.4	Kṛdanta analyzer	20
1.4.5	Gender recognition and analyzer of <i>sanjñā pada</i>	21
1.4.6	Śābdabodha System	21
1.4.7	DESIKA	21
1.4.8	Morphology of Sanskrit Case Affixes: A Computational Analysis	22
1.4.9	Sanskrit Morphological Analyzer of JNU	22
1.4.10	Sanskrit Morphological Analyzer	22
1.4.11	Vibhakti Divergence between Sanskrit and Hindi	22
1.4.12	Anaphora in Sanskrit	23
1.4.13	Discourse Anaphora and Resolution Techniques in Sanskrit	23
1.4.14	Sanskrit Anaphora Resolution System	23
1.4.15	Sanskrit Compound Processor	23
1.4.16	Knowledgebase for karma-kāraka	23
1.4.17	Lexical Ambiguities and their resolution in Sanskrit with Special Reference to Pañchatañtra	24
1.4.18	Named Entity Recognition for Sanskrit: a hybrid approach	24
1.5	Difference of the present work from the other available works	24

1.6.	Research methodology	25
Chapter-2: Sanskrit- Hindi Noun Phrase		26-43
2.1	Introduction	27
2.2	General Concept of Noun Phrase	27
2.3	Noun Phrase in Sanskrit	28
2.3.1	Pada in Sanskrit	31
2.3.2	Subanta pada	32
2.3.3	Nominal subanta	32
2.3.4	kṛt-pratyayānta subanta	32
2.3.5	Taddhita-pratyayānta subanta	33
2.3.6	Sāmāsika subanta	33
2.3.7	Indeclinable subanta	34
2.3.8	NP components in Vā. <i>ākhyātaṃ sāvyayakāra-kaviśeṣaṇaṃ vākyaṃ</i>	34
2.3.9	Usage of NP component	38
2.3.9.1	Simple sentence	38
2.3.9.2	Complex sentence	39
2.3.9.3	Compound sentence	39
2.4	Hindi Noun Phrase	40
2.4.1	Simple NP	40
2.4.2	Complex NP	41
2.4.2.1	Noun with Complement Clause	41
2.4.2.2	Noun Clause	42
2.4.2.3	Noun with Relative Clause	42
2.4.2.4	Noun with Participle Modifier	42
2.5	Basic difference between Sanskrit and Hindi NPs	42
Chapter 3: Challenges in Sanskrit-Hindi Noun Phrase Mapping		44-72
3.1	Introduction	45
3.2	Case markers difference between Sanskrit and Hindi	46
3.3	Different <i>vibhaktis</i> in a particular <i>kāraka</i> (case) in Sanskrit	47

3.3.1	<i>kartā</i>	47
3.3.2	<i>karma</i>	48
3.3.3	<i>karāṇa</i>	50
3.3.4	<i>sampradāna</i>	51
3.3.5	<i>apādāna</i>	51
3.3.6	<i>sambandha</i>	52
3.3.7	<i>adhikarāṇa</i>	52
3.4	Hindi cases and their different usage	53
3.4.1	Ergative in Hindi	53
3.4.2	<i>ko</i> case marker	54
3.4.3	<i>se</i> case marker	54
3.4.4	Genitive case markers	54
3.4.5	Locative case markers	55
3.4.6	Vocative case marker	55
3.5	<i>Vibhakti</i> Divergence and Challenges in NP- case mapping	55
3.6	Challenges in indeclinable mapping	61
3.7	Challenges in mapping compound words	63
3.8	Adjective and challenges in its mapping	64
3.8.1	Adjective	64
3.8.2	Challenges in adjective mapping	68
3.8.2.1	The identification of adjectival pada	68
3.8.2.2	Mapping of adjective	69
3.8.2.3	Challenges in mapping conjuncts	70
3.9	Challenges in mapping Gender and Number	70
3.10	Challenges in mapping word order	72
Chapter-4: Transfer Grammar for Sanskrit Hindi NP Mapping		73
4.1	Introduction	74
4.2	Task of Hindi generator	74
4.3	Transfer grammar	77
4.4	TG rules for case mapping	77
4.4.1	Division of <i>kartā kāraka</i> and TG rules	77

4.4.2	Division of <i>karma kāraka</i> and TG rules	78
4.4.3	Division of <i>karāṇa kāraka</i> and TG rules	81
4.4.4	Division of <i>sampradāna kāraka</i> and TG rules	81
4.4.5	Division of <i>apādāna kāraka</i> TG rules	82
4.4.6	Division of <i>adhikarāṇa kāraka</i> and TG rules	83
4.4.7	The non- <i>kāraka</i> relations and TG rules	83
4.4.8	Other relations and TG rules	85
4.5	Mapping nominal in compound syntax and conjunct <i>ca and vā</i>	86
4.6	Adjective mapping	87
4.6.1	Rules for identifying the adjectives	87
4.6.2	Mapping case marker in adjective	88
4.7	Dual number mapping	89
4.8	Word order mapping	89
	<i>Conclusion</i>	90-92
	<i>Bibliography</i>	93-100

INTRODUCTION

In the age of globalization, knowledge needs to be barrier-free and an effective Machine Translation (MT) system is supposed to do this with the help of emerging language technology which will be less time consuming as well as economic for the users. But this is a challenge for MT system developers to produce a reasonably good translation output to the common user so that it may be acceptable to them and there would not be a requirement of post editing in that output. Then only it may be claimed that MT can quickly provide barrier free knowledge in low cost and it will be bridging the gap between different language speakers. After that, it may be also useful for e-learning and e-governance as well. In order to provide comprehensible translation, an MT system should be capable of transferring the source language sentence into the correct syntactic structure of the target language. There is no doubt in the efficiency of machine as a quick performer but making a machine intelligent to produce a good output in Natural Language Processing (NLP), specially in MT, needs an understanding of language at syntactic and semantic levels. For this understanding, a comparative study of the syntax of Source Language (SL) and Target Language (TL) is required to find out the language contrast which is challenging to be mapped by a machine. Although languages within a family may seem easier to translate with the help of machine but a closer look into the structures of those languages provides some contrasts and difficulties. To handle the language contrasts between SL and TL at NP level, one has to look into the different aspects of each component in a meaningful sentence of SL and TL to know their functions in different contexts. While translating texts from SL to TL, a human translator may try to translate sentences as close to the meaning of source language and the structure of the target language as possible but the machines have a hard time translating these structures. A machine has to be taught the grammars of the two languages and the significant differences as well for the output to be as close to the input as possible. In this process, each type of component in a SL sentence needs to be listed at the level of NP. The contrasts found in the syntax can be mapped by writing the Transfer Grammar Rules to produce a reasonable output in TL. The present research is an attempt to understand the syntactic differences between Sanskrit (SL) and Hindi (TL) at Noun Phrase (NP) level

to evolve cross-linguistic model for Indian languages and write Transfer Grammar rules for Sanskrit-Hindi Machine Translation (SHMT).

There are 22 scheduled languages in India and MT activities are going on from mid 80s and early 90s between different language pairs. Some important MT systems are available for English to Indian Languages (IL) and IL to IL since more than two decades but SHMT is respectively new and it has to cover a long distance to produce a good output having less scope for post editing. This goal (SHMT) has a big challenge of handling NP and comparative study of NP between both languages is not done in theoretical linguistics yet, so this work is required to find out the issues and possible solution.

Sanskrit and Hindi are considered structurally close owing to genealogical relations. However, on a closer look, Hindi appears to have diverged significantly more in terms of structure than in lexical ingenuities. Gender, number distinctions, ergative, postposition, verb group, double causative, echo are some (among many) remarkable structural innovations that Hindi has gone through over the ages. While the structure of Sanskrit *vibhakti* was fairly organized, the same may not be true for Hindi. As we know that Sanskrit is an inflectional language and has respectively free word order, when Sanskrit is translated into Hindi by MT systems (being developed by a consortium of institutes) it has problems in Hindi structural transformation and in handling NP - mainly word order in output, case markers and modifier-modified identification. We have no exhaustive grammar in Hindi compared to Sanskrit to make the task easier. There is a need of a comparative study of Sanskrit and Hindi syntax. To get syntactically good Hindi translation from Sanskrit, it will help us in making Transfer Grammar rules as well.

This work contains the following four Chapters:

Chapter 1: Machine Translation in Indian Languages

In this chapter, MT work in Indian Languages is discussed. After giving a brief note on MT work on different IL pairs, Sanskrit-Hindi MT (SHMT) and work done in

SHMT is discussed. How the present research work is different from the previous works, is the last point of this chapter.

Chapter 2: Sanskrit-Hindi Noun Phrase

In this chapter, first the brief concept of Noun Phrase in modern linguistics is presented. Then the notion of Sanskrit NP is discussed. Here it is noticed that NP term is used by different scholars but there is a general paucity of literature in this area. So, here it is tried to present a hypothesis about the possible concept of NP in Sanskrit.

Chapter 3: Challenges in Sanskrit Hindi NP Mapping

In this chapter, direct transfer translation approach is adopted to identify the difference between Sanskrit-Hindi Syntax. The following topics are discussed in this Chapter:

- Challenges in case mapping
- Different vibhaktis in a particular case in Sanskrit
- Hindi cases and their different usage
- Vibhakti Divergence between Sanskrit and Hindi
- Challenges in indeclinable mapping
- Challenges in mapping compound nominals
- Challenges in mapping adjectives
- Challenges in mapping plural words used for a singular
- Challenges in Mapping Word order

Chapter 4: Transfer grammar for Sanskrit Hindi NP mapping:

In this chapter, transfer approach has been discussed for handling the challenges found in Sanskrit Hindi NP mapping. The challenges are divided into two levels: first the challenges that can be handled at generation level and second which can be handled at transfer level. Here the rules for handling challenges at transfer level are written. The challenges at generation level can be handled by improving the Hindi generator which is implemented in the Sampark system.

Chapter-1

MACHINE TRANSLATION IN INDIA

1.1 Introduction

“The mechanization of translation has been one of humanity’s oldest dreams. In the twentieth century it has become a reality, in the form of even computer programs capable of translating a wide variety of texts from one natural language into another. But, as ever, reality is not perfect. There are no ‘translating machines’ which, at the touch of a few buttons, can take any text in any language and produce a perfect translation in any other language without human intervention or assistance. That is an ideal for the distant future, if it is even achievable in principle, which many doubt”.¹

“MT is part of a wider sphere of ‘pure research’ in computer based Natural Language Processing in Computational Linguistics and Artificial Intelligence, which explore the basic mechanisms of language and mind by modeling and simulation in computer programs. Research on MT is closely related to these efforts, adopting and applying both theoretical perspectives and operational techniques to translation processes, and in turn offering insights and solutions from its particular problems. In addition, MT can provide a ‘test-bed’ on a larger scale for theories and techniques developed by small-scale experiments in computational linguistics and artificial intelligence”.²

“Most translation in the world is not of texts which have high literary and cultural status. The great majority of professional translators are employed to satisfy the huge and growing demand for translations of scientific and technical documents, commercial and business transactions, administrative memoranda, legal documentation, instruction manuals, agricultural and medical text books, industrial patents, publicity leaflets, newspaper reports, etc. Some of this work is challenging and difficult. But much of it is tedious and repetitive, while at the same time requiring accuracy and consistency. The demand for such translations is increasing at a rate far beyond the capacity of the translation profession. The assistance of a computer has clear and immediate attractions”.³

¹ <http://www.hutchinsweb.me.uk/IntroMT-1.pdf> p-1 web accessed on 7th June 2011 12:50 PM.

² <http://www.hutchinsweb.me.uk/IntroMT-1.pdf> p-2. web accessed on 7th June 2011 12:50 PM.

³ <http://www.hutchinsweb.me.uk/IntroMT-1.pdf> p-2. web accessed on 7th June 2011 12:50 PM.

In the age of Internet, the multiplicity of languages makes it even more necessary to have sophisticated machine translation systems. Many machine translation projects and related activities are going on in the country and abroad. Hence, it is essential to find out the current state of technology specially in Indian languages for this research work. R M K Sinha says that the MT work in India started in early eighties when they proposed using Sanskrit as Interlingua for translation to and from Indian languages.⁴ Salil Badodekar⁵ has presented a detailed survey of MT researches in India till dated 31/08/2003 which is a very useful source for knowing MT researches in India. Further Dwivedi and Sukhdeve⁶ have presented the latest development in this area. On the basis of these articles and the latest updates on TDIL website⁷, the MT research in India has been presented in this chapter which is divided into two parts. First part deals with the different MT Systems for Indian Languages (other than Sanskrit) and second part deals with the works done in the area of Sanskrit Computational Linguistics.

1.2. MT Systems for Indian Languages

1.2.1. Anglabharati⁸: The AnglaBharti project was launched at the Indian Institute of Technology; Kanpur in 1991 for Machine aided Translation from English to Indian languages. Sinha *et al.* (2001) has pioneered Machine Translation research in India. The approach and lexicon of the system is general-purpose with provision for domain customization. A machine-aided translation system specifically designed for translating English to Indian languages. English is a SVO language while Indian languages are SOV and are relatively of free word order. Instead of designing

⁴ "Computer processing of Indian languages and scripts - Potentialities and Problems", Jour. of Inst. Electron. & Telecom. Engrs., vol.30,no.6, 1984.

⁵ <http://www.cfilt.iitb.ac.in/Translation-survey/survey.pdf>

⁶ **Machine Translation System in Indian Perspectives**; Journal of Computer Science 6 (10): 1082-1087, 2010; ISSN 1549-3636 © 2010 Science Publications

⁷ <http://www.tdil-dc.in/>

⁸ <http://www.iitk.ac.in/>

<http://www.cse.iitk.ac.in/users/langtech/hist.htm>

<http://www.cse.iitk.ac.in/users/langtech/anglabharti.htm>

translators for English to each Indian language, AnglaBharti uses a (Dave *et al.*,2001) pseudo-interlingua approach. It analyses English only once and creates an intermediate structure called Pseudo Lingua for Indian Languages (PLIL). In AnglaBharti they use rule based system with context free grammar like structure for English, A set of rules obtained through corpus analysis which is used to distinguish conceivable constituents. Overall, the AnglaHindi (Sinha & Jain, 2003) system attempts to integrate example-based approach with rule-based and human engineered post-editing. AnglaBharti is a pattern directed rule based system with context free grammar (Sinha and Jain, 2003) like structure for English (source language) which generates a 'pseudo-target' (PLIL) applicable to a group of Indian languages (target languages). A set of rules obtained through corpus analysis is used to identify plausible constituents with respect to which movement rules for the PLIL is constructed. The idea of using PLIL is primarily to exploit structural similarity to obtain advantages similar to that of using Interlingua approach. The intermediate language structure has the word and word-group order as per the structure of the group of target languages. It also uses some example-base to identify noun and verb phrases and resolve their ambiguities. The system is a machine aided translation system for translation between English to Hindi, for the specific domain of Public Health Campaigns.

1.2.2. Anusaaraka by Indian Institute of Technology, Kanpur and University of Hyderabad⁹: Anusaaraka (1995) project which started at IIT Kanpur by Prof. Rajeev Sangal and his team including Prof Chaitanya and Amba Kulkarni It is now being continued at IIIT Hyderabad. It was started with the explicit aim of translation from one Indian language to another. The project was funded by Technology Development in Indian Languages (TDIL), Govt of India and Satyam Computers Private Limited. Anusaarakas have been built from Telugu, Kannada, Bengali, Punjabi and Marathi to Hindi. It is domain free but the system has been applied mainly for translating

⁹ <http://www.iiit.net/ltrc/>
http://www.iiit.net/ltrc/Anusaaraka/anu_home.html
http://www.iiit.net/ltrc/Publications/anu_brief.html

children's stories. Anusaaraka aims for perfect "information preservation". In fact, Anusaaraka output follows the grammar of the source language (where the grammar rules differ and cannot be applied with 100% confidence). For Example, a Bengali to Hindi Anusaaraka can take a Bengali text and produce output in Hindi which can be understood by the user but will not be grammatically perfect. An e-mail server been established for the Anusaaraka. To run the Anusaaraka on a given text, e-mail has to be sent with the name of the language in the subject line. For example, if 'Telugu' is put in the subject line, it involuntarily runs the Telugu to Hindi Anusaaraka. The focus in Anusaaraka is not mainly on machine translation, but on language access between Indian languages. Anusaaraka systems can be obtained from their website¹⁰. They are currently attempting an English-Hindi Anusaaraka machine translation system. Anusaaraka mainly focuses on language access between Indian languages, using principles of Paninian Grammar (PG) (Bharati *et al.*, 1995) and exploiting the close similarity of Indian languages. Badodkar¹¹ states in brief that the task of building an MT System is subdivided into two parts. The first module (called core anusaaraka) does language-based analysis: It takes all the information in the source text and resents it in its output, in an intermediate language that is quite close to the target language. The second module may do domain specific knowledge based processing, statistical processing, etc. in which it may utilize world knowledge, frequency information, concordances, etc. to produce output in the target language.

1.2.3. MANTRA by Centre for Development of Advanced Computing, Bangalore¹²: MACHiNE assisted TRANslation tool (MANTRA) (1999). It translates English text into Hindi in a precise domain of personal administration, specifically gazette notifications, office orders, office memorandums and circulars. Initially, the Mantra system was started with the translation of administrative document such as appointment letters, notification and circular issued in central government from English to Hindi. It is based on the Tree Adjoining Grammar (TAG) formalism from University of Pennsylvania. It uses Lexicalized Tree Adjoining Grammar (LTAG)

¹⁰ http://www.iiit.net/ltrc/Anusaaraka/anu_home.html

¹¹ <http://www.cfilt.iitb.ac.in/Translation-survey/survey.pdf>

¹² <http://www.cdacindia.com/html/about/success/mantra.asp>

(Bandyopadhyay, 2004) to represent the English as well as the Hindi grammar. Tree Adjoining Grammar (TAG) is used for parsing and generation. It is based on synchronous Tree Adjoining Grammar and uses tree transfer for translating from English to Hindi. The system is tailored to deal with its narrow subject domain. The Mantra has become part of “The 1999 Innovation Collection” on information technology at Smithsonian institution’s National museum of American history, Washington DC, USA. This system can be obtained from the C-DAC website (<http://cdac.in/html/aai/mantra.asp>) or by contacting Dr. Hemant Darbari. This project was funded by the Rajya Sabha Secretariat. The grammar is specially designed to accept, analyze and generate sentential constructions in “Officialese” domain. Similarly, the lexicon is suitably restricted to deal with meanings of English words as used in its subject domain. The system is ready for use in its domain. The system is developed for the Rajya Sabha Secretariat, the Upper House of Parliament of India. It translates the proceedings of parliament such as study to be laid on the table, Bulletin Part-I and Part-II. This system also works on other language pairs such as English-Bengali, English-Telgu, English-Gujarati and Hindi-English and also among Indian languages such as Hindi-Bengali and Hindi-Marathi. The Mantra approach is general, but the lexicon/grammar has been limited to the sub-language of the domain.

1.2.4. AnglaBharti-II: AnglaBharti-II (2004) (Sinha, 2004) addressed many of the shortcomings of the earlier architecture. It uses a Generalized Example-Base (GEB) for hybridization besides a Raw Example-Base (REB). During the development phase, when it was found that the modification in the rule-base was difficult and might result in unpredictable results, the example-base has been grown interactively by augmenting it. At the time of actual usage, the system first attempts a match in REB and GEB before invoking the rule-base. In AnglaBharti-II, provision were made for automated pre-editing and paraphrasing, The purpose of automatic pre-editing module is to transform/paraphrase the input sentence to a form which is more easily translatable. Automated pre-editing may even fragment an input sentence if the fragments are easily translatable and positioned in the final translation. Such fragmentation may be triggered in case of a failure of translation, by the ‘failure analysis’ module. The failure analysis consists of heuristics on speculating what might

have gone wrong. The entire system is pipelined with various sub modules. All these have contributed significantly to greater accuracy and robustness to the system.

1.2.5. AnglaHindi¹³: AnglaHindi is English to Hindi version of the AnglaBharti which uses example-base and statistics to obtain more acceptable and accurate translation for frequently encountered noun and verb phrases. Besides using all the modules of Anglabharti, AnglaHindi also makes use of an abstracted example-base. The example base is statistically derived from the corpus. Ambiguities in the meanings of the verb phrases are also resolved using an appropriate distance function in the example base. The system is claimed to generate approximately 90% acceptable translation in case of simple, compound and complex sentences up to a length of 20 words. It has provisions for augmenting its abstracted example-base specific to an application domain. This not only eliminates the alternative translations but also generates more accurate and acceptable translation. The AnglaHindi system attempts to integrate example-based approach with rule-base and human engineered post-editing¹⁴.

1.2.6. Anubharti¹⁵: Anubharti (2004) (Sinha, 2004) approach for machine-aided-translation is a hybridized example-based machine translation approach that is a combination of example-based, corpus-based approaches and some elementary grammatical analysis. The example-based approaches follow human-learning process for storing knowledge from past experiences to use it in future. In Anubharti, the traditional EBMT (Gupta and Chatterjee, 2003) approach has been modified to reduce the requirement of a large example-base. This is done primarily by is achieved by identifying the syntactic groups. Matching of the input sentence with abstracted examples is done based on the syntactic category and semantic tags of the source language structure. Both of these system architectures, AnglaBharti and AnuBharti, have undergone a considerable change from their initial conceptualization. In 2004 these systems named as AnglaBharti-II and AnuBharti-II. AnglaBharti-II uses a generalized example-base for hybridization besides a raw example-base, and the

¹³ <http://www.cse.iitk.ac.in/users/rmk/proj/proj.html>

¹⁴ Thanks to Pinki Nainwani for her brief presentation on this topic during group discussion.

¹⁵ http://www.iitk.ac.in/infocell/Archive/dirnov2/techno_machine.html

AnuBharti-II to cater to Hindi as source language for translation to any other language, though the generalization of the example-base is dependent upon the target language.

1.2.7. Matra by Centre for Development of Advanced Computing, Mumbai (erstwhile NCST)¹⁶: The MaTra system (2004), developed by the Natural Language group of the Knowledge Based Computer Systems (KBCS) division at the National Centre for Software Technology (NCST), Mumbai (currently CDAC, Mumbai) and supported under the TDIL Project is a tool for human aided machine translation from English to Hindi for news stories. It has a text categorization component at the front, which determines the type of news story (political, terrorism, economic and so on.) before operating on the given story. Depending on the type of news, it uses an appropriate dictionary. It requires considerable human assistance in analyzing the input. Another novel component of the system is that given a complex English sentence, it breaks it up into simpler sentences, which are then analyzed and used to generate in Hindi. They are using the translation system in a project on Cross Lingual Information Retrieval (CLIR) (Rao, 2001) that enables a person to query the web for documents related to health issues in Hindi.

1.2.8. UCSG MAT by University of Hyderabad¹⁷: MAT is a machine aided translation system for translating English texts into Kannada. It requires post editing. It works at sentence level. It parses an input sentence using the UCSG (Universal Clause Structure Grammar) parsing technology (developed by Dr. K. Narayana Murthy) and then translates it into Kannada using the English-Kannada bilingual dictionary, Kannada Morphological Generator and the translation rules.

¹⁶ <http://www.ncst.ernet.in/matra/>
<http://www.ncst.ernet.in/matra/about.shtml>

¹⁷ <http://www.uohyd.ernet.in/>

1.2.9. UNL MT by Indian Institute of Technology, Mumbai¹⁸: IIT, Mumbai is the Indian participant in Universal Networking Language (UNL) project. UNL is an international project of United Nations University. UNL is an interlingua for semantic representation. Input in the source language is converted into UNL and then de-converted from UNL to the target language. Currently, work on conversion and deconversion in English, Hindi and Marathi is going on.

1.2.10. Tamil Anusaaraka by Anna University's K. B. Chandrasekhar Research Centre, Chennai¹⁹: The aim is to build a Human Aided Machine Translation System for English-Tamil. The MT system has three major components, viz. morphological analyzer of source language, mapping unit and the target language generator. The Tamil-Hindi Machine Aided Translation (MAT) system has a performance in the range of 75%. The state-of-the-art Tamil Morphological analyzer can handle nearly 3.5 million word forms including compound words with more than 95% accuracy.

1.2.11. MAT by Jadavpur University²⁰: Jadavpur University at Kolkata has a rule-based English-Hindi MAT. It uses transfer approach. It works for news sentences.

1.2.12. Anuvaadak by Super Infosoft²¹: Anuvaadak 5.0 system has been developed by super Info soft private limited, Delhi under the supervision of Mrs. Anjali Rao Choudhury for a general purpose English-Hindi Machine Translation. For specific domains it has inbuilt dictionaries. It has specific domains like Official, formal, agriculture, linguistics, technical and administrative. The meaning of any English word is not available in Hindi in dictionary then there is facility of translation is provided. In the windows family this software runs on any Operating system. The Spell-checker is in both English and Hindi. It has an inbuilt thesaurus and grammar checker. Inbuilt grammar checker works in pre-translation and post-translation stages.

¹⁸ <http://www.cfilt.iitb.ac.in/>

¹⁹ <http://www.au-kbc.org/frameresearch.html>

²⁰ <http://www.jadavpur.edu/>

²¹ <http://www.mysmartschool.com/pls/portal/portal.MSSStatic.ProductAnuvaadak>

It has inbuilt *dictionaries for specific domains* e.g. official, formal, agriculture, linguistics, technical, and administrative. An English word processor is inbuilt. When Hindi meaning of the English word is not available in dictionary, facility of transliteration is provided. The software runs on any operating system in the Windows family.

1.2.13. Anubaad hybrid machine translation system: Anubaad a hybrid MT system is developed in the year 2004 for translating English news headlines to Bengali, developed by Bandyopadhyay (2000) at Jadavpur University Kolkata and. The current version of the system works at the sentence level.

1.2.14. Statistical MT by IBM²²: IBM India Research Lab at New Delhi has started work on statistical machine translation between English and Indian Languages. Their work is based on similar work at IBM for other languages.

1.2.15. Oriya Machine Translation System (OMTrans) by Utkal University, Vanivihar²³: In OMTrans, the source language is English and target language is Oriya. It does sense disambiguation using the N-gram model. It has a parser and Oriya Morphological Analyser (OMA), OGC (Oriya Grammar Checker), OSC (Oriya Spell Checker) and OSA (Oriya Semantic Analysis). These modules contribute to OWP (Oriya Word Processor) which facilitates multilingual editing.

1.2.16. Shiva and Shakti machine translation: The system Shiva is an Example-based and the system Shakti is working for three target languages like Hindi, Marathi and Telgu. Shiva and Shakti are the two Machine Translation systems from English to Hindi and are developed jointly by Carneige Mellon University USA, international institute of information technology, Hyderabad and Indian institute of science, Bangalore, India. The system is used for translating English sentences into the appropriate language. Shakti machine translation system (Bharati *et al.*, 2003) has been designed to produce machine translation systems for new languages rapidly.

²² <http://www.research.ibm.com/irl/projects/translation.html>

²³ <http://www.ilts-utkal.org/omt.htm>

Shakti system combines rule based approach with statistical approach whereas Shiva is Example-Based machine translation system. The rules are mostly linguistic in nature and the statistical approach tries to infer or use linguistic information. Some modules also use semantic formation. Currently system is working for three languages (Hindi, Marathi and Telugu).

1.2.17. Tamil-Hindi machine aided translation system²⁴: The system Tamil-Hindi Machine-Aided Translation system has been developed by Prof. C.N. Krishnan at Anna University at KB Chandrashekhar (AU-KBC) research centre, Chennai. The translation system is based on Anusaaraka Machine Translation System, the input text is in Tamil and the output can be seen in a Hindi text. It uses a lexical level translation and has 80-85% coverage. Stand-alone, API and Web-based on-line versions are developed. Tamil morphological analyser and Tamil-Hindi bilingual dictionary are the byproducts of this system. They also developed a prototype of English-Tamil Machine-Aided Translation system. It includes exhaustive syntactical analysis. It has limited vocabulary (100-150) and small set of transfer rules.

1.2.18. English-Kannada machine-aided translation system: English-Kannada MAT system is developed at Resource Centre for Indian Language Technology Solutions (RC-ILTS), University of Hyderabad by Dr. K. Narayana Murthy. The system is essentially a transfer-based approach and it has been applied to the domain of government circulars. English-Kannada machine translation system using Universal Clause Structure Grammar (UCSG) formalism. The system is funded by the Karnataka government.

1.2.19. Hinglish machine translation system: Hinglish a machine translation system for pure (standard) Hindi to pure English forms developed by Sinha and Thakur (2005) in the year 2004. It had been implemented by incorporating additional level to the existing English to Hindi translation (AnglaBharti-II) and Hindi to English translation (AnuBharti-II) systems developed by Sinha. The system claimed to be produced satisfactory acceptable results in more than 90% of the cases. Only in case

²⁴ <http://www.au-kbc.org/research-areas/nlp/demo/mat/>

of polysemous verbs, due to a very shallow grammatical analysis used in the process, the system is not capable to resolve their meaning.

1.2.20. English to (Hindi, Kannada, Tamil) and Kannada to Tamil language-pair example based machine translation system: English to {Hindi, Kannada and Tamil} and Kannada to Tamil language-pair example based machine translation system developed by Balajapally *et al.* (2006) in the year 2006. It is based on a bilingual dictionary comprising of sentence dictionary, phrases-dictionary, words-dictionary and phonetic-dictionary is used for the machine translation. Each of the above dictionaries contains parallel corpora of sentence, phrases and words and phonetic mappings of words in their respective files. Example Based Machine Translation (EBMT) has a set of 75,000 sentences most commonly spoken that are originally available in English. These sentences have been manually translated into three of the target Indian languages, namely Hindi, Kannada and Tamil.

1.2.21. Punjabi to Hindi machine translation system: Punjabi to Hindi machine translation system developed by Josan and Lehal at Punjabi University Patiala in the year 2007. This system is based on direct word-to- word translation approach. This system consists of modules like pre-processing, word-to-word translation using Punjabi-Hindi lexicon, morphological analysis, word sense disambiguation, transliteration and post processing. The system has reported 92.8% accuracy.

1.2.22. Sampark²⁵: Machine translation System among Indian language: Sampark: developed by the Consortium of institutions. Consortiums of institutions include IIT Hyderabad, University of Hyderabad, CDAC (Noida, Pune), Anna University, KBC, Chennai, IIT Kharagpur, IIT Kanpur, IISc Bangalore, IIT Alahabad, Tamil University, Jadavpur University in the year 2009. Currently these MT systems have two unidirectional (Hindi-Marathi and Hindi-Bangali) and five bidirectional (Panjabi-Hindi, Urdu-Hindi, Telugu-Hindi, Tamil-Hindi and Tamil-Telugu) systems

1.2.23. Hindi to Punjabi machine translation system: Hindi to Punjabi Machine translation System developed by Goyal and Lehal (2010) at Punjabi University Patiala

²⁵ <http://sampark.iit.ac.in/sampark/web/index.php/content>

in the year 2009. This system is based on direct word-to-word translation approach. This system consists of modules like pre-processing, word-to-word translation using Hindi-Punjabi lexicon, morphological analysis, word sense disambiguation, transliteration and post processing. The system has reported 95% accuracy. The overall conclusion of machine translation systems in Indian perspectives that from the year 1995 to 2009 the MT systems developed have achieved lots of success in translating languages. Still work has been carried out to achieve better than previous study.

1.2.24. Development of Sanskrit Computational Toolkit and Sanskrit-Hindi Machine Translation System: This project in the consortium mode under the leadership of Amba Kulkarni (University of Hyderabad) has been funded by DIT since April 2008 for building computational tools and Sanskrit-Hindi MT in the domain of children stories, and for building multimedia and e-learning content for kids.

1.2.25. SaHiT: Sanskrit-Hindi Translator is being developed by JNU under the supervision of Dr. Girish Nath Jha. Various tools for Sanskrit (SL) processing has been developed which is accessible at <http://sanskrit.jnu.ac.in> .

1.2.26. Anusaarak Sanskrit Hindi MT System: It has been developed by Dept of Sanskrit, University of Hyderabad, under the supervision of Amba Kulkarni. According to the information given on the website, it produces Hindi translation of a Sanskrit text and also provides intermediate outputs at various levels viz. *pada-viśleṣaṇa*, *kāraka- viśleṣaṇa*, *śabdārtha-nir dhāraṇa* etc.

1.3. Recent projects funded by DIT under the TDIL program ²⁶

Name of Ongoing Project	Started on	Output Target
Development of Sanskrit Computational Tool Kit and Sanskrit Hindi Machine Translation System , University of Hyderabad	28/3/2008	MT System & Language Tools and Resources for Sanskrit
Development of Major North Eastern Languages (Assamese, Bodo, Nepali, Manipuri) in Consortia mode C-DAC Pune	28/3/2008	Language Tools and resources for Major North Eastern Languages (Assamese, Bodo, Nepali, Manipuri)
Development of grammar checking system for Punjabi , Punjabi university, Patiala	28/3/2008	Punjabi Grammar checking system
Development of language tools and linguistics resource for Kashmiri, university of Kashmir	28/3/2008	Linguistics resource & Tools for Kashmiri Language
Resource Centre for Indian language solution 2nd phase Assamese and Bodo, IIT Guwahati	11/7/2007	Language Tools and Resources for Assamese and Bodo
Development of Manipuri Language Technology , Manipur University, Imphal	12/3/2007	Language Tools and Resources for Manipuri Language
Standardization of IPA-IL for 3 Indian Languages Hindi, Assamese, Bengali, CDAC Kolkata	19/3/2007	IPA-IL for 3 Indian Languages Hindi, Assamese, Bengali.
Development of adequate intonated TTS for Nepali, CDAC Kolkata	19/3/2007	TTS for Nepali
Development of phonetic engine for Indian languages Hindi, Telugu, IIIT Hyderabad	19/3/2007	Phonetic Engine for Telugu and Hindi

²⁶ <http://www.mit.gov.in/content/language-computing-group-vi>

Consortium for development of MAT System based on Angla Bharati Technology , IIT Kanpur (Consortia Leader)	22/3/2007	MAT System based on Angla Bharati Technology (Mission Mode); Six Languages pairs: English to Hindi/ Marathi/ Bengali/ Oriya/ Tamil/ Urdu.
Development of Robust document analysis and Recognition system for printed Indian scripts (Text OCR) system, IIT Delhi	5/9/2006	Text document analysis and Recognition system for printed Indian scripts system (Mission Mode) 10 Scripts: Bengali, Devanagari, Malayalam, Gujarati, Telugu, Tamil, Oriya, Tibetan,Nepali, Gurumukhi, Kannada
Development of online hand writing recognition system for Indian languages (OHWR), IISc Bangalore (Consortia Leader)	13/11/2006	Online hand writing recognition system for Indian languages (Mission Mode); 6 Scripts: Devanagari, Bengali, Tamil, Telugu, Kannada and Malayalam
Development of Indian Language-Indian Language Machine Aided Translation (EL-ILMAT) System, CDAC Pune (Consortia Leader)	30-8-2006	Machine Aided Translation system (Mission Mode) Six Languages pairs: English to Hindi/ Marathi/ Bengali/ Oriya/ Tamil/ Urdu.
Development of Indian Language to Indian Languages Machine Aided Translation (IL-IL MAT) System, IIIT Hyderabad	29/8/2006	Machine Aided Translation (Mission Mode) Nine Language pairs: Tamil-Hindi, Telugu-Hindi, Urdu-Hindi, Kannada-Hindi, Punjabi-Hindi, Marathi-Hindi, Bengali-Hindi, Tamil-Telugu, Malayalam-Tamil
Development of Cross Lingual Information Access (CLIA) system, IIIT Bombay (Consortia Leader)	29/8/2006	Cross Lingual Information Access (CLIA) system (Mission Mode) Six Languages: Bengali, Hindi, Marathi, Punjabi, Tamil and Telugu.
National Roll-Out Plan project C-DAC Pune	21/12/2005	Indian language technology Software tools and fonts

1.4. Work done in Sanskrit NP processing

Not a single work is directly related to Sanskrit NP processing but wherever nominal word is processed, it is assumed that they may partially contribute to NP processing. Some important analyzer tools which are somehow related to NP processing in SHMT are as follows:

1.4.1 Sanskrit POS Tagger is developed as a Ph.D. result by R. Chandrasekar under the guidance of Dr. Girish Nath Jha, Special Centre for Sanskrit Studies (SCSS), JNU. It performs as a primary tool in SHMT by tagging each string with its grammatical category.

1.4.2 Sanskrit *kāraka* analyzer²⁷ is developed as a Ph.D. result by Sudhir Kumar Mishra (2007) under the guidance of Dr. Girish Nath Jha, SCSS, JNU . This system analyses the *kāra*kas in a given input (i.e. Sanskrit prose text) on the basis of Paṇini and kātyāyaṇa *kāraka* formulation. *kāraka* identification in a NP is important for the handling of *vibhakti chinha* (case marker) in Hindi.

1.4.3 Subanta analyzer²⁸ is developed as a part of M.Phil. Dissertation by Subash (2006) under the guidance of Dr. Girish Nath Jha, SCSS, JNU which uses a hybrid approach of Paṇinian formalism and example-based technique and gives a comprehensive computational analysis of subanta pada. It also provides a brief information on Nominal Inflectional Morphology (subanta) of Paṇini.²⁹

1.4.4 Kṛdanta analyzer³⁰ is developed as a part of M.Phil. Dissertation by Surjit Kumar Singh (2008) under the guidance of Dr. Girish Nath Jha, SCSS, JNU. As kṛdanta also forms NPs in Sanskrit and a lot of children domain texts and dramas have kṛdanta form, it plays an important role in its analysis.

²⁷ <http://sanskrit.jnu.ac.in/karaka/analyzer.jsp> ; (21.5.2011; 3:45 PM).

²⁸ <http://sanskrit.jnu.ac.in/subanta/rsubanta.jsp> (21.5.2011; 3:45 PM).

²⁹ Machine Recognition and Morphological Analysis of Subanta-Padas, pp-35-37.

³⁰ <http://sanskrit.jnu.ac.in/kridanta/ktag.jsp> (21.5.2011; 3:47 PM).

1.4.5 Gender recognition and analyzer of *sanjñā pada*³¹ is developed as a part of M.Phil. research work by Manji Bhadra (2007) under the guidance of Dr. Girish Nath Jha, SCSS, JNU. It analyses the gender of a noun in SL input. In second chapter of the dissertation, It provides a description on the effect of gender in Hindi agreement and in third chapter, it provides information on simple Sanskrit NPs in brief as well.

1.4.6 Śābdabodha System developed by Academy of Sanskrit Research, Malkote, claims to process all types of sentences of Sanskrit and handling of generation and analysis of subantas. The website 'vedavid'³² has only one page which shows that Śābdabodha System was developed in 1995-96 and it contains Śemuṣī as an important component of the system. The website of Sanskrit Academy³³ has a brief description of Śemuṣī from which it is clear that the system basically functions as Subanta analyzer.

1.4.7 The Desika³⁴ system developed by Indian Heritage Group, C-DAC, Bangalore led by P. Ramanujan, claims to be a NLU system for generation and analysis for plain and accented written Sanskrit texts based on grammar rules of Pāṇini Aṣṭādhyāyī. In this system, *subanta* generation module is given which is downloadable from the website of TDIL. According to TDIL website, The *DESIKA* software helps in understanding a natural language input (typically an isolated sentence) through paraphrasing, voice change, query answering or summarizing, to develop a language-independent knowledge representation scheme based on ancient Indian Sciences, to develop tools for linguistic analysis and to assist in analysis and presentation of scriptural (accented text) knowledge, phonetic and language research, teaching etc., It was developed on DOS platform and has now been ported on Windows platform³⁵. According to C-DAC, Bangalore website³⁶ DESIKA is a comprehensive package for generating and analysing Sanskrit words which serves the need for an authentic Computer-based package for its study. It caters to different user

³¹ <http://sanskrit.jnu.ac.in/grass/analyze.jsp> (21.5.2011; 3:47 PM).

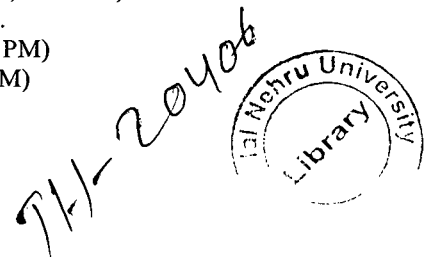
³² <http://www.vedavid.org/ASR/project.html> (21.5.2011; 3:47 PM).

³³ <http://www.sanskritacademy.org/Achievements1.htm#> (21.5.2011; 3:47 PM).

³⁴ <http://tdil.mit.gov.in/download/Desika.htm> (21.5.2011; 3:47 PM).

³⁵ <http://tdil.mit.gov.in/nlptools/ach-nlptools.htm> (22.5.2011, 10:45 PM)

³⁶ <http://www.cdac.in/html/ihg/pdf/desika.pdf> (22.5.2011, 10:55 PM)



communities like Academicians, Students, Researchers, Linguists, Computer Scientists etc.

1.4.8 Morphology of Sanskrit Case Affixes: A Computational Analysis is done by Dr. Girish Nath Jha (1993), JNU. This work takes into the account the Pāṇinian formalisms for inflectional morphology of case affixes and its explanation as provided by Bhaṭṭojidīkṣita's *Siddhāntakaumudī* and gives the paradigm of word forms for most of the words correctly. A small program for plural morphology of Hindi has also been given which generates plural forms of Hindi words (with or without postpositions).

1.4.9 Sanskrit Morphological Analyzer of JNU: The "Sanskrit Morphological Analyzer" is a collection of modules developed as a result of Computational Sanskrit R&D at Special Center of Sanskrit Studies, JNU under the supervision of Dr. Girish Nath Jha. It gives the morph analysis of Sanskrit input word.

1.4.10 Sanskrit Morphological Analyzer was developed by Vinish Jain, an M.Tech student of IIIT-Hyderabad, under the guidance of Amba P. Kularni, which was a further development of an earlier working Sanskrit Morphological Analyzers developed at ASR Melkote. Later with the collaboration of RSVP, Tirupati Sanskrit Hindi Machine Translation System (**Sampark**) is developed in HCU³⁷.

1.4.11 Vibhakti Divergence between Sanskrit and Hindi³⁸ Vibhakti divergence between the both languages is studied by Patel Preeti Khimji which presents the difference in case marking. She has classified the divergence into 7 categories³⁹ saying that the divergence discussed by Dorr [1994] is either of rare occurrences or does not pose much problem in MT. Her study is presenting the *kāraka prakaraṇa* of Aṣṭādhyāyī in a classified manner but the dependencies of Hindi case marker i.e. the cause of getting particular case marker in Hindi is not discussed during contrastive representation of both languages.

³⁷ <http://www.sanskrit.uohyd.ernet.in> 23.5.2011, 4:00PM

³⁸ M.Phil. Dissertation Submitted by Patel Preeti Khimji in Dept. of Sanskrit Studies, School of humanities, University of Hyderabad in June 2010.

³⁹ Optional, Exceptional, Differential, Alternative, Non-kāraka, Verbal and Complex-perdicate.

1.4.12 Anaphora in Sanskrit: Jha, Girish Nath (et al.)⁴⁰ has presented a comprehensive documentation and classification of Sanskrit anaphora and the resolution system is a part of Sanskrit Analysis System developed by the principal author at Computational Linguistics R&D, Special Centre for Sanskrit Studies, Jawaharlal Nehru University, New Delhi.

1.4.13 Discourse Anaphora and Resolution Techniques in Sanskrit: Jha, Girish Nath (et al.)⁴¹ has collected and classified cases of discourse anaphors in Sanskrit from a wide ranging sample from earliest times to the 18th century text of *Ambikā Dutta Vyāsa (Śivarājaviḥayam)* and then presented a computational model to handle such cases in Sanskrit.

1.4.14 Sanskrit Anaphora Resolution System: Madhav Gopal is currently working on Anaphora resolution system for Sanskrit called *Sanskrit Anaphora Resolution System (SARS)* in JNU. His work is primarily concerned with resolving pronominal anaphors in Sanskrit text *Pañcatantra*. He uses PNG constraints mainly for pairing anaphor and antecedent. He intends to provide a robust algorithm and a tool which is implemented in Java.

1.4.15 Sanskrit Compound Processor: This research work is being done by Mr. Anil Kumar in HCU under the supervision of Amba Kulkarni. Their paper *Sanskrit Compound Processor*⁴² presents the importance of compound processing to understand their meaning and decode the relations between its components and to produce their *vigrahavākya*. This work may also be seen as an important one for handling compound nominals in NP mapping in SHMT.

1.4.16 Knowledgebase for karma-kāraka: This research is being done by Ms. Manji Bharda⁴³ in JNU. In this research the most desired i.e. the karma of the given

⁴⁰ Johansson, C. (Ed.) *Proceedings of the Second Workshop on Anaphora Resolution (2008)*

⁴¹ Devi, Sobha Lalitha (ed al.) *Proceedings of 7th Discourse Anaphora and Anaphora Resolution Colloquium (2009)*

⁴² G.N.Jha (Ed.): *Sanskrit Computational Linguistics*, LNCS 6465, pp.57-69, 2010. © Springer-Verlag Berlin Heidelberg 2010.

⁴³ <http://sanskrit.jnu.ac.in/rstudents/phd.jsp>

sentence is being taken care. So this research will help us in handling karma during SHMT.

1.4.17 Lexical Ambiguities and their resolution in Sanskrit with Special Reference to *Pañchatantra*: This research is being done by Mr. Muktanand Agrawal⁴⁴ in JNU. In this research, lexical ambiguities are being resolved. So this research will help us in handling ambiguities during SHMT.

1.4.18 Named Entity Recognition for Sanskrit: a hybrid approach :

This research is being done by Mr. Sachin Kumar⁴⁵ in JNU. In this research Named Entity Recognition System is being developed. This research may help us in disambiguating Sanskrit names from adjectives.

1.5 Difference of the present work from the other available works

Except 1.2.11, all the above discussed researches deals with Sanskrit input texts mainly and does not map the NPs from SL to TL. Without NP mapping from SL to TL, appropriate case marking in TL (Hindi) is not possible and NP mapping between both language pairs needs a comparative study of the language pairs to find the contrastive properties in them. The work discussed in 1.2.11, although presents the divergence between Sanskrit and Hindi, but not gives any to handle these diversions. Where is *ko* getting dropped and why somewhere we need the *ko* marker? How do the animate-inanimate factor and transitive-intransitive factor affect the *ko* case marker? How to differentiate between *ko* and *ke liye*, *se* and *ke dvārā*, *mēn* and *para*? How to handle case markers with adjectives? Even how to differentiate the adjectives and the Substantives in Sanskrit? These things are completely untouched as nowhere these things are discussed. These will be discussed in the present study.

Further, if multiple case marker is possible in a particular sense then frequency of case markers should be checked and case marker with higher frequency should be

⁴⁴ <http://sanskrit.jnu.ac.in/rstudents/phd.jsp>

⁴⁵ <http://sanskrit.jnu.ac.in/rstudents/phd.jsp>

assigned. Case marker generally depends on gender of noun, animate-inanimate objects and verbs as well. So the dependency of case marker should be checked. Gerund in a noun phrase also needs to be tackled correctly for MT which depends on gender and number of the noun.

Change in gender of a noun (*dārā=patnī; agni=āga* etc), as well as handling of particles (i.e. *ca, vā, api, iva* etc.) is very important in NP mapping. Word order is another important issue with NP mapping as place of modifier of a noun has a fixed position in Hindi and contrary Sanskrit has respectively free word order. For example- if subject + modifier + auxiliary verb (i.e. *thā/hai/hogā* etc) sequence is there in a Hindi sentence, modifier may come before or after the subject. But if there is main verb in the sentence, modifier always comes before the subject in Hindi. In Sanskrit, adjective may take its anywhere in a NP. In this research, all these challenging aspects in NP mapping in SHMT will be discussed which will be giving a theoretical framework for correct output in SHMT. In modern linguistics specially, there is a lack of Sanskrit Hindi comparative study at NP level. So this research may fulfill that need and will help in evolving a cross-linguistic model for Indian languages and writing Transfer Grammar rules for SHMT.

1.6. Research methodology

For this kind of applied research work in Computational Linguistics, a theoretical base is required. Any exhaustive research work, directly related to the comparative study of Sanskrit-Hindi NP is not available as a theoretical base. As case level study of Sanskrit syntax and Hindi syntax is available separately, here it is obtained by the comparative study of both languages at the NP level which includes adjectives, determiners and case markers in it. On the basis of contrastive properties found by the comparative study of Sanskrit-Hindi Syntax, the transfer grammar rules will be written (theoretically) to map the NPs in SHMT.

Chapter-2

SANSKRIT-HINDI NOUN PHRASE

2.1 Introduction

Samuel R. Levin states two very important lines: a) the demonstration in 1786 by Sir William Jones that Sanskrit offered a close relationship to the Classical, Germanic, and Celtic languages ushered in the first major phase of what might be termed modern linguistic studies and b) The fact that Sanskrit, an extensively documented language, with a highly developed grammatical system and phonological and morphological sub-systems that were at once archaic and transparent, has provided the impetus for a century of intensive work in the comparative reconstruction of the parent Indo-European language.¹ These lines are presenting the important contribution of Sanskrit to modern linguistics and we find a deep philosophical discussion on syntactico-semantic relation in Sanskrit tradition. In this chapter, Sanskrit-Hindi noun phrases are compared to find the syntactic difference.

2.2 General Concept of Noun Phrase

Noun Phrases (NP) are phrases in which the head word is a noun, i.e. the noun is shown to be the head of the phrase by its primary stress, and syntactically whole group can be replaced by a simple noun. Noun phrases occupy the characteristic syntactic positions and case roles of nouns. These syntactic positions are most typically those of subjects, direct objects, indirect objects and nominal predicates. Pronouns and names make up the smallest noun phrase, since they typically come by themselves, with neither determiners nor any other modifiers. This is so because modifiers function, in various ways, to restrict the domain of possible reference of a noun; and both the pronouns and names refer to unique entities that require no further specification. A NP that is neither a name nor a pronoun is then made out of an obligatory head noun plus, optionally, some modifiers. The head noun is the core of the NP which determines its lexical-semantic type. Modifiers may indeed add various types of information to the head noun, but typically do not change its inherent lexical

¹ *Modern Linguistics: Its Development and Scope* ;The Journal of Higher Education, Vol. 36, No. 3 (Mar., 1965), pp. 137-146

type. The central role of the head noun in the NP may be expressed by the rule of semantic amalgamation of the NP i.e. whatever semantic features belong to the head noun also belong to the entire NP.

NP can be a pronoun, a name or a full noun phrase (i.e. grouped words headed by a noun). The general rule (in English) that orders the various optional modifiers relative to the head noun as well as vis-à-vis each other may be given as:

NP= (QUANT) (DET) (AP) N (REL/ POSS-NP/ N-COMP).

Here the modifiers that precede the head noun are, in order, Quantifiers (QUANT), Determiners (DET), and Adjectival Phrase (AP). Modifiers that follow the head noun are the Relative clauses (REL), Possessor NP (POSS NP) or Noun Complements (N-COMP).

2.3 Noun Phrase in Sanskrit

The concept of Noun Phrase (NP) in modern linguistics is well established and that approach is being applied for many languages including Hindi. In NP, there is a head Noun which may be further followed by an adjective and/or determiner (i.e. *ekah*, *kaścit* etc). Sanskrit (post Paninian), which is claimed as the most suitable natural language for Artificial Intelligence/Computational Linguistics², is studied from different perspectives by linguists. Spiejer and Tarapurwala have worked on Sanskrit Syntax and Brahamchari Surendra kumar has implemented Fillmore Case grammar on Sanskrit Syntax. Many scholars have used the term 'Noun Phrase' in their papers, but still there is scope for a discussion on the nature of Sanskrit and the possible concept of NP in Sanskrit. One needs an understanding of Sanskrit NP first then only one can move towards its comparison with Hindi NP.

It is generally claimed that Sanskrit, being inflectional language, does not possess a word order and that scrambling does not affect the meaning conveyed because of the

² Briggs, Rick (1985)

vivakṣā and that the kāraka relation (syntacto-symentic relation), word order becomes irrelevant. For example: a) *kamalaṃ vikasati* has Noun-Verb combination where the noun *kamalaṃ* alone is considered as NP and verb *vikasati* as VP. But in b) *nīlaṃ kamalaṃ sāyaṃ vikasati*, we clearly find the grouped words in NP as well as in VP. When c) *nīlaṃ kamalaṃ* (which is shown as NP in the above example) is presented as a sentence, can it be called NP? Or does it show NP-VP? Here *asti/bhavati/vartate* can be assumed to complete this sentence. But this kind of assumption may be called an extra effort to get the proper meaning and it is clearly a semantic processing of a given sentence. At a first look, it appears as a *samānādhikaraṇa* (or *viśeṣya-viśeṣaṇa*) and without assuming a verb, can we refer this as an NP? This type of problem gives a chance to raise question on the concept of Sanskrit NP. Further when we change the word order of (b) (i.e. *nīlaṃ kamalaṃ sāyaṃ vikasati*) and represent it like (d) *vikasati nīlaṃ sāyaṃ kamalaṃ*, we find this difficult to be grouped in a phrase. Although semantically this type of sentence consists of NP and VP but one can say that here free word order in the above example is restricting its representation as NP-VP. H.S. Ananthnarayana³ writes ‘The Noun Phrase can be manipulated to occupy any position in sentence without any modification of the meaning of the sentence although a certain order may be preferred to others.’ Here if we will prefer the general word order of (b) instead of (d), then we may have no difficulty in NP grouping.

Another example, *khādati* is a sentence⁴ but the agent (i.e. NP) is not expressed here. In this case how we can present the concept of NP in Sanskrit? Simply, this is the case of pro-drop parameter. There is no problem in identifying the first person and second person agents in Sanskrit sentences as verbal suffix (*si, thaḥ, tha, mi, vah, mah*) indicates its single possible agent. But verbal suffix *ti* (which takes third person singular number as its agent) may take different names as its agent or it may take pronominal words depending on the agent’s gender. So, only verbal suffix *ti* has problem in identifying its possible NP.

³ *The kāraka theory and case grammar*, Indian Linguistics, Vol.31, p-16.

⁴ *ekatiṅ vākyaṃ*

Further, ‘*dr̥ṣṭaḥ mṛgaḥ*’ can be used as a Noun Phrase if we assume *asti/bhavati/vartate* and in that condition *kṛdanta pada* ‘*dr̥ṣṭaḥ*’ functions as modifier. But when the word order of this sentence (*dr̥ṣṭaḥ mṛgaḥ*) is changed as *mṛgaḥ dr̥ṣṭaḥ*, *dr̥ṣṭaḥ* functions as verb instead of a modifier.

It is noticed that the change in word order sometimes doesn’t affect the meaning conveyed but sometimes clearly affect the meaning. So question is raised on the claim of being Sanskrit as a free word order language. This view is examined by scholars⁵ and it is established that there is a general word order which is usually found in simple prose texts. Ashok Aklujkar in his paper ‘*Some theoretical observation on word order in Sanskrit*’, points out the two views of J F Staal clearly “Anyone who wishes to find precise rules for word order in Sanskrit will constantly meet with similar problems [= problems similar to ones pointed out by Patañjali in Mbh.], provided the sentences he studies are not uncommonly short” (and hence that person will run into difficulties). He clears the conditional statement (provided the sentences he studies are not uncommonly short) and brings out the fact that the lesser the number of constituents in a sentence, the higher the probability of scrambling. When a long sentence is presented in Sanskrit, it becomes difficult to keep any constituent word anywhere frequently. There we can see the word order in Sanskrit. It can also be supported by the fact that the commentators first do the *anvaya* of any Sanskrit verse where he follows a general sequence by the following saying- ‘*viśeṣaṇaṃ puraskṛtya viśeṣyaṃ tadanantaram, karṭṛ-karma-kriyā-yuktaṃ etadvanvaya-lakṣaṇaṃ*’⁶. Here modifier (if available in a verse) is kept before the noun (modified) which is *karṭā* and the sentence pattern is SOV. Another saying regarding the word order is: ‘*ādau karṭṛ-padaṃ vācyam dvitīyādi padaṃ tataḥ, ktvā-tumun-lyap ca madhye tu kuryādante kṛyāpadaṃ*’⁷.

Thus we see that the word order is used generally in prose sentences and *anvaya*, basically a simplification of Sanskrit verse, is the first step in the processing of *Sanskrit vāṇmaya*. After establishing the fact that there is a logical word order in

⁵ Staal (1967), Aklujkar (1996), Gillon (1996).

⁶ Samasacakram, verse 5.

⁷ Aklujkar, Ashoka: *Some theoretical observation on word order in Sanskrit*

Sanskrit, we can present some sentences in a logical word order where we can try to find the NP in a Sanskrit sentence.

It is said that ‘one of the main differences between the Indian approach to language analysis and that of the most of the current linguistic theories is that the analysis of the sentence was not based on a noun-phrase model with its binary parsing technique but instead on a conception that viewed the sentence as springing from the semantic message that the speaker wished to convey.’⁸

Although, NP binary parsing is not necessary in Sanskrit Syntax, in a simple sentence, where agent and the verb is clearly mentioned, we can find the concept of phrase. The Indian approach of language analysis is semantic but the syntactic structure is also well discussed in *Aṣṭādhyāyī*. To understand the concept of Sanskrit NP, the understanding of *Sanskrit pada* and sentence (*pada samūha*) is needed.

2.3.1 Pada in Sanskrit: The concept of *nāman*, *ākhyāta*, *upasarga* and *nipāta* is well discussed in *Nirukta*. Here *nāman*⁹ and *ākhyāta*¹⁰ denotes subject and predicate respectively. The *sutra* ‘*prādayaḥ* (P.1.4.56)’ tells about 22 *prādi-particles* in Sanskrit and the next *sutra* ‘*upasargāḥ kriyāyoge* (P. 1.4.57)’ describes its use with the verbs in the sentence. In *nipātas*, some are used in predicate as *kriyā*. For example- *svāhā*, *svadhā śrausaṭ*, *vauṣaṭ* etc and some in subject. Therefore, it can be said that *upasarga* goes with a root (*ākhāta* or *kṛdanta*) and *nipāta* with *nāman* (with some exceptions broadly) and the traditional concept of *nāman*, *ākhāta*, *upasarga* and *nipāta* has two main categories *subanta* and *tiñanta* only. Necessity of *sup* and *tiñ* can be explained by the statements of grammarians. Bhartṛhari says that the *Śabda*, generally called ‘word’ in English, is the medium of knowledge:

na so ’sti pratyayo loke yaḥ śabdānugamādṛte|
*anuviddhamiva jñānaṃ sarvaṃ śabdena bhāsate||*¹¹

Its vast range of usage is presented precisely in Mahābhāṣya:

⁸ Briggs, Rick (1985)

⁹ *satvapradhānāni nāmāni; (Nirukta).*

¹⁰ *bhāvapradhānamākhyātam; (Nirukta).*

¹¹ *Vākyapadīya* (1/123), ed. Sharma, Pt. Raghunath (1988).

mahān śabdasya prayogaviśayaḥ. saptadvīpā vasumatī, trayo lokāḥ, catvāro vedāḥ sāṅgāḥ sarahasyā bahudhā bhinnāḥ- ekaśatamadhvaryuśākhāḥ, sahasravartamā sāmavedaḥ, ekaviṃśātidhā bahvṛcām, navadhā atharvaṇo vedaḥ, vākovākyaṃ, itihāsaḥ, purāṇaṃ, vaidyakamityetāvān śabdasya prayogaviśayaḥ.

Thanks to Pāṇini for a systematic grammar of language. He states that *padas* (*subanta* and *tiñanta*) are essential component in the formation of a sentence in Sanskrit¹². The *subanta* has *prātipadika* (base word derived from root) and *sup* suffix and *tiñanta* has root and *tiñ* suffix. Every verb (*tiñanta* as well as *kṛdanta*) requires an agent and a *subanta pada* comes as an agent as well as its modifier in a sentence. A *pada* may also be called *vākya* in Sanskrit¹³.

2.3.2 Subanta pada: There are five types of *subanta pada* in Sanskrit:- *Nominal subanta* (including *stri-pratyayānta subanta*), *Kṛt-pratyayānta subanta*, *Taddhit-pratyayānta subanta*, *Samāsānta subanta* and *Indeclinable subanta*.

2.3.3. Nominal subanta: From the sutra- ‘*arthavadadhaturapratyayaḥ prātipadikaṃ* (P.1.2.45)’ nominal *subanta* has a *prātipadika* (a meaningful base word) and *sup pratyaya* (suffix) is added to form a usable *pada* in a Sanskrit sentence. As - *rāmaḥ gacchati*. In this example, *rāmaḥ* (*nominal subanti*) has base word *rāma* and suffix *su* is added to make it usable in a sentence. Here *stri-pratyayānta subanta* is also included in nominal *subanta* because of *paribhāṣā- ‘prātipadika-grahaṇe liṅgā-viśiṣṭasyāpi-grahaṇaṃ’*. There are 8 types of *stri-pratyayas- ṭāp, ḍāp, cāp, ṇīp, ṇīṣ, ṇīn, ūñ* and *ti*, which are used with *prātipadikas* to form *stri-pratyayānta subanta*. These *stri-pratyayānta subanta* functions same as nominal.

2.3.4 kṛt-pratyayānta subanta (primary derived nominals): It is well discussed in Nirukta and Mahābhāṣya that all nominal forms are made of roots in Sanskrit¹⁴. Nominal *subanta rāma* may also be called *kṛt-pratyayānta* because it has $\sqrt{ram+ghañ}$

¹² *suptiñantaṃ padaṃ* (P.1.4.14).

¹³ *Padyate artho ‘vagamyate anena iti padaṃ nāma vākyaṃ.*

¹⁴ *nāma ca dhātujamāha nirukte vyākaraṇe śakaṭsya ca tokaṃ- in Mahā. on ‘uñādayo bahulaṃ’* (P. 3. 2. 1).

pratyaya. ($\sqrt{\text{ram}}+\text{ghañ} > \sqrt{\text{ram}}+\text{a} > \text{rāma}+\text{su} > \text{rāmaḥ}$)¹⁵. But *kr̥t-pratyayānta subanta* is mentioned separately because these are used as an agent/adjective as well as verb in Sanskrit¹⁶. as - *śraddhāvantaḥ paṇḍitāḥ na labhyante*. In this example, *śraddhāvantaḥ* is *kr̥t-pratyayānta subanta* which is used as an adjective of ‘*paṇḍitāḥ*’. Participle forms, which are also a kind of noun, are formed with *kr̥t-pratyaya*. *saḥ gataṅ* has ‘*gataṅ*’ *kr̥t-pratyayānta subanta pada* as a participle which has *ktavat* suffix in $\sqrt{\text{gam}}$. Same as- *tasyāḥ rodanaṃ srutvā janāḥ āgatvantaḥ*. *lyuṭ*, *ghañ*, *ktin* and the suffixes mentioned in the *adhikāra* of *bhāve* sutra are used in Sanskrit to form gerunds such as- *paṭhanaṃ* ($\sqrt{\text{paṭh}}+\text{lyuṭ}$) *pāṭhaḥ* ($\sqrt{\text{paṭh}}+\text{ghañ}$), *matīḥ* ($\sqrt{\text{man}}+\text{ktin}$) etc. When *tādarthyā* is expressed in the sentence, then the gerund is a part of VP (*svāsthyalābhāya bhramanaṃ āvasyakaṃ*). When it is unexpressed, for example, in this sentence-‘*bhramanaṃ svāsthyapadaṃ asti*’, gerund (‘*bhramnaṃ*’ in this case) becomes NP. In modern uses, verbs are changed into its *kr̥danta* forms for simplicity. For example- *bhramati*= *bhramanaṃ karoti*. So *bhraman* becomes noun after the split of verb *bhramati*.

2.3.5 Taddhita-pratyayānta subanta (secondary derived nominals): *Taddhita prakaraṇa* is the biggest section of *Aṣṭādhyāyī*. it is divided in many sub-prakaraṇas in which specific types of *taddhitānta* word formation is discussed. *Taddhita pratyayas* are added with *prātipadikas* and *kr̥dantas* mainly.¹⁷ It is used to derive genitive words in Sanskrit. For example- *garga+ yañ* =*gārgya*, *gārgya+phak*= *gārgyāyaṇa*. It functions as an agent as well as an adjective in Sanskrit NP. Ex- *māgadhaḥ nirṇayati* and *māgadhaḥ candr̥guptaḥ nirṇayati*. In the above examples, *māgadhaḥ* is an agent as well as an adjective respectively.

2.3.6 Sāmāsika subanta: Two or more *padas* are combined in a word to form a compound nominal. Sup suffixes are generally added into the *uttara-pada* of

¹⁵ *vivaḥśādhīnā hi śabdavyutpattiḥ* (*Aṣṭādhyāyī*-2.2.8) i.e. the analysis of language depends on speaker’s intention.

¹⁶ *kr̥dabhihito bhāvo dravyavadapi kriyāvadapi* (*Mahābhāṣya*-5.4.1)

¹⁷ Some *taddhit* suffixes (i.e. *ak*, *tamāṃ*, *tarāṃ kalpa* etc) are used with verbs. For ex- *pacati+ak*= *pacataki*.

compound *prātipadika* (made with the combination of two or more *padas*).¹⁸ For example- *rāmaśca lakṣmaṇaśca bharataśca śatrughnaśca = rāmalakṣmaṇa-bharataśatrughnāḥ*. *Samāsa* is well discussed in *Aṣṭādhyāyī*. *Samāsānta subanta* is mainly used as a subject or an adjective or an adverb in a sentence. As- *nīlāmbaraḥ sete* has *nīlāmbaraḥ* (a compound word) as an agent. When one more nominal *pada* is used with *nīlāmbaraḥ*, it can be an adjective of that nominal *pada*. In *tena tat vṛttāntaṃ yathāyathaṃ kathitaṃ, yathāyathaṃ* is an adverb which is compound nominal.

2.3.7 Indeclinable subanta: Indeclinable *subantas* are those words which have same form in all *liṅga* (genders), *vibhakti* (cases) and *vacana* (numbers).¹⁹ These are described in the following sutras in *Aṣṭādhyāyī* - *svarādinipātamavyayam, taddhitāścāsarvavibhaktiḥ, kṛnmejantaḥ, ktvātosunkasunaḥ, and avyayībhāvaśca* (P.1.1.38-41) and the result i.e. *vibhakti lopa* is said by the sutra *avyayādāpsupaḥ* (P.2.4.82). In *Amarakośa* and other lexicons also, the indeclinable words are listed. These are used for exclamations, conjuncts, pronouns (*kutra* etc), adverbs and adjectives frequently. It is also used as an agent but are few in number. For example, *trimuni, unmattagaṅgaṃ* (an example of *avyayībhāva* compound in Sanskrit) are an Indeclinable nominals which can be used as agents in Sanskrit sentence. As- *trimuninā proktaṃ*.

These are the five types of *subanta pada* in Sanskrit which is used in sentence. Here the *tiṅanta* is not discussed as it is not related with NP. Now the definition of a sentence will be discussed which helps in conceptualizing NP in Sanskrit and describes its components and possible usage in various ways. In the definition of a sentence, *ākhyāta* is also discussed which itself clears its importance in a sentence.

2.3.8 NP components in vā. ākhyātaṃ sāvyayakāravīṣeṣaṇaṃ vākyaṃ: *Aṣṭādhyāyī* is the base of Sanskrit grammar which governs the syntax by its precise rules. Here *Paṇini* didn't give the definition of *vākya* (sentence) but he mentions the

¹⁸ Some examples of 'devatādvandva' are its exceptions.

¹⁹ *sadṛśaṃ triṣu liṅgeṣu sarvaṣu ca vibhaktiṣu, vacaneṣu ca sarveṣu yanna vyeti tadavyayam.*

word *vākya* twice.²⁰ Kātyāyana wrote *vārtika* on Aṣṭādhyāyī and defines *vākya*: *ākhyātaṃ sāvyaayakāarakaviśeṣaṇaṃ vākyaṃ* which is explained in Mahābhāṣya²¹. It means that a sentence has *ākhyāta* which may be followed by *ayaya*, *kāra* and *kāra* (i.e. modified *kāra*). From its further explanation, we get different structures of a Sentence. In the *vārtika*, *sāvyaayakāarakaviśeṣaṇa* part is divided in *sāvyaaya*, *kāra*, and *kāra* and gives three types of sentence structure:

1) *ākhyātaṃ sāvyaayaṃ vākyaṃ*: It means that *ākhyāta* with indeclinable is a *vākya*. For ex- *uccaiḥ paṭhati*.

2) *ākhyātaṃ sakāraṃ vākyaṃ* i.e. *ākhyāta pada* which has *kāra*, is a *vākya*. Ex- *bālakaḥ khādati* or *phalaṃ khādati* or *bālakaḥ phalaṃ khādati* etc.

3) *ākhyātaṃ sakāra* i.e. *ākhyāta* with modified *kāra*, is a *vākya*. Ex- *sundaraḥ bālakaḥ madhuraṃ phalaṃ khādati*

The next *vārtika* is *sakriyāviśeṣaṇaṃ ca* which adds adverbs to *ākhyāta pada*. Ex- *mandaṃ calati*.

In all the above sentence structures, *ākhyāta* is present as a compulsory component of a sentence. If it is alone, it can also be a sentence by *ekatiṅ vākyaṃ*. In this context, Bhartṛhari's view of *vākya* can be discussed:

*ākhyātaśabdaḥ saṅghāto jātiḥ saṅghātavartinī,
eko 'navayavaḥ śabdaḥ kramo budhyanusamhṛtiḥ||1||
padamādyam prthaksarvam padam sākāṅkṣamityapi,
vākyaṃ prati matirbhinnā bahudhā nyāyavādinam||2||*

Here eight views are presented regarding *vākya*. They are: a) *ākhyātaśabdaḥ*, b) *saṅghātaḥ*, c) *saṅghātavartinī jātiḥ*, d) *eko 'navayavaḥ śabdaḥ*, e) *krama* f) *budhyanusamhṛtiḥ*, g) *ādyam padam* and h) *prthaksarvam padam sākāṅkṣam*. These are classified into two parts: *akhaṇḍa* and *sakhaṇḍa vākya*. Three views (c) (d) and

²⁰ *vākya* *āderāman* *tritasūyā* *saṃmati-kopa-kutsana-bhartsaneṣu* (8.1.8); *vākya* *ṣṭh* *pluta udāttaḥ* (8.2.82)

²¹ Mahābhāṣya (2.2.1.)

(f) are related to *akhaṇḍa vākya* and rest five are related to *sakhaṇḍa vākya*. In *sakhaṇḍa vākya* also, *saṃghāta* and *krama* are related with *abhihitānvayavāda* and rest three to *anvitābhīdhānavāda*. The three views of *akhaṇḍavākya* and *ādyaṃ padaṃ* of *sakhaṇḍa vākya* is related with the understanding of meaning. Four views - *ākhyātaśabdaḥ*, *saṃghātaḥ*, *kramaḥ* and *prthaksarvaṃ padaṃ sākāñkṣaṃ* are directly related to the syntax.

First view regarding *vākya* is *ākhyātaśabda* which states that a *ākhyāta pada* i.e. verb may be a sentence. It is cleared by the commentators that *ākhyātaṃ vākyaṃ* never states that only *ākhyāta* is a sentence but it states that where *kāraḥ* are not used and only verb is present in a sentence which gives a meaning, can also be a sentence. (*kvacid ākhyāta śabdo'pi vākyaṃ, yatra kāraḥ-śabda-prayogaṃ vinā kevalākhyātaśabdaprayoge'pi vākyaṃarthāvagatiḥ*). As - *pidhehi*.

Second view, *saṃghāta* states that the verb gets grouped with its *adhyāhrta* (assumed) agent to give its meaning. For example- *pidhehi* verb assumes *tvaṃ pada* as its agent and gives the meaning (*tuma*) *banda karo* in its Hindi. Karma may also be assumed as *dwāram pidhehi*.

Third view, *krama* says that the *padas* are when expressed in a sequence then they become a *vākya*. In the *ambākarṭka* comentory on *Vākyapadīya* (2/49), importance of *pada*-sequence is written as- “*na kramavyatiriktaṃ vākyaṃ nāma vākyaṃarthasyābhīdhāyakam*”²². The meaning of a sentence will be changed by the alteration in sequence of *padas* in a sentence. As - *ahaṃ eva gacchāmi* → *ahaṃ gacchāmi eva* (i.e. *maiṃ hī jātā huṃ* and *maiṃ jātā hī huṃ* in Hindi).

The definition, *prthaksarvaṃ padaṃ sākāñkṣaṃ* states that all the *padas* which have *sānnidhya* because of *ākāñkṣā*, are collectively a sentence. This definition is similar to the definition of *jaimini*- “*arthaikatvādekaṃ vākyaṃ sākāñkṣaṃ ced vibhāge*

²² Sharma, Pt. Raghunath (1980), *Vakyapadiyam-2/49*, p-95.

syār”²³. Both the definitions are same as the definition of *sāhityadarpaṇakāra*:-
“*vākyam syādyogyatākāṅkṣāsaktiyuktaḥ padoccayaḥ*”²⁴.

Thus it is clear from the above discussion that *ākhyāta* (verb) is the soul of any syntax²⁵. A verb may alone be used in a Sanskrit sentence. If a verb is alone, it will assume its agent because of its inflectional nature. As the definition of a sentence in Sanskrit is well discussed in grammar, philosophy and poetics, the different possible structures of sentences is basically based on the concept of *nāma* and *ākhyāta*, well known as nominal and verbal *padas*. In Sanskrit, when the minimum *ākāṅkṣā* is fulfilled to express a meaning, no other word is required. Any verb can be taken for its example because, in an inflectional language, suffix makes it easy to identify the agent of a verb. For example- *paṭhāmi* is a verb in Sanskrit which is a sentence (when used alone) and its agent is unexpressed. Here ‘āmi’ suffix of the verb indicates that ahaṃ is the only possible agent of this verb. There is no problem in identification of first person and second person agents in a Sanskrit sentence because they are pronouns. In third person, the agent can be identified by anaphora resolution. Similarly, where verb phrase is not clearly indicated or intentionally omitted, there is also a sentence. As *-rākṣaso nāma amātyaḥ* is a sentence where the verb (*asasti/babhūva* etc) is understood. In another example, *rāmaḥ gataḥ*, *gataḥ* is verb which is not a *tiñanta*, it is *ḥanta* and *ḥanta* is nominal in Sanskrit. It means that the *ākhyāta* is not necessarily a *tiñanta*, it may be *ḥanta* which is a nominal *pada*.

We can also see that nominal (*ḥanta*) is also used as *kriyā* (absence of *tiñanta pada* doesn’t mean that there is no verb). So while, *ḥanta* and *tiñanta* are two types of verb in Sanskrit which becomes VP (when *tiñanta* comes alone or comes with *ḥanta* (as an adverb) and *subanta*, *ḥanta* and *taddhitānta pada* becomes NP (either as an agent or as an adjective).

²³ Jaimini Sutra-2.1.46.

²⁴ *sāhityadarpaṇa*-2/1.

²⁵ The two contrastive views are debated on this issue *vyāpāramukhyaviśeṣyaka* and *pramātārthamukhyaviśeṣyaka śābdabodha*. In *devadattaḥ grāmaṃ gacchati*, *Vaiyākaraṇa* says that *gamanānukūla vyāpāra* (*kriyā/action*) is main (*pradhāna*) and *Naiyāyika* says that the *vyāpāravān devadatta* (agent) is main. *Mīmāṃsaka* also agrees with the first view.

In *rāmaḥ gacchati*, *rāmaḥ* is the NP. As there is no adjective or an article with agent 'rāma', it is clear that *rāma* is the only possible head of NP in this example. If more than one *subanta* is present in a simple Sanskrit NP, for example- *sundaraḥ rāmaḥ gacchati*, one will be the agent and another will be an adjective of that agent and both, jointly will be the NP of a sentence.

In the ex- *śraddhāvantaḥ na labhyante*, *śraddhāvantaḥ* is the NP which is a *kṛdanta pada* and in the example- *śraddhāvantaḥ paṇḍitāḥ na labhyante*, *śraddhāvantaḥ paṇḍitāḥ* is the NP.

In *jānakī āgatā*, here *jānakī* is a *taddhitānta pada* which is NP.

2.3.9 Usage of NP component:

The NP components are used in different types of Sanskrit sentences. There are three types of sentences which are classified on the basis of their syntax: simple, complex and compound. Study of these sentences is required to know the usage of different nominal *padas*.

2.3.9.1 Simple sentence- Generally a subject (a simple nominal/a compound nominal/a pronominal) and a predicate i.e. the agent and the verb is used in a simple sentence. Example - *bālakaḥ paśyati*. There may be a verb used alone in a simple sentence as well.²⁶ In this condition, of course, there is no scope of NP study. Leaving those sentences, we will look for the sentences which have NP. In these type of sentences, an adjective may be there to modify the agent. Example – [{*uttamaḥ* (adj.) *putraḥ*} {*dhyānena* (adv.) *paśyati*.}] The length of subject (*uddeśya*) can also be increased by using the genitives with it. As - *putraḥ> sarvagūṇasampannaḥ putraḥ> daśarathasya sarvagūṇasampannaḥ putraḥ> raghuvaṃśasya daśarathasya sarvagūṇa-sampannaḥ putraḥ*. When a compound nominal is used as a subject (specially *dvandva samāsa*), its different types of representation is possible. For example- *rāmaḥ lakṣmanaḥ ca/ rāmaśca lakṣmanaśca/ rāmalakṣmanau paṭhataḥ*.

²⁶ 'eka tiṅ vākyam'

2.3.9.2 Complex sentence-There is one main subject (*uddeśya*) and two or more infinite (*samāpikā*) verbs. As - *yasyārthāḥ tasya mitrāṇi*. In these sentences, there is one main clause and rest is dependent clause. There are three types of dependent clause: nominal clause, adjective clause and adverb clause which expand noun, adjective and adverb respectively. Nominal clause is used as an agent or an object, adjective clause always becomes an adjective of the agent or the object and adverbial clause is used in VP. Indeclinable is used at the start or end of these three types of dependent clauses. 'yathā'/'yad' is used at the beginning and 'iti' at the end of nominal clause. (ex- '*ayaṃ devaḥ*' *iti jñāmi*). Sometimes 'iti' is not used in nominal clause of a complex sentence. As – *kathaya 'satsaṅgatiḥ puṃsām kiṃ na karoti'*. Adjective clauses have the forms of relative pronoun 'yad-tad' (*yāvat-tāvat/yādrś-tādrś* etc.). Ex- '*yadāloke sūkṣmaṃ*' *vrajati tadvipulatām*. adverbial clause is denoted by *kṛdantas* and adjectival compound words (mainly *bahuvrīhi* and *tatpuruṣa*).²⁷ *yadā, yāvat, yāvanna, yadā yadā, yatra, yatra-yatra, iva, yathā* (with *tathā*), *yathaiva* (with *tathaiva*), *yathā yathā, yataḥ, yataḥ-tataḥ hi, yadi-tarhi, yena-tena, kena* etc are used in complex sentences.

2.3.9.3 Compound sentence- Two or more simple/complex sentences are added by conjunct and these are *samānādhikaraṇa* of each other. The indeclinable *ca, tathā, api* etc are used to add two or more statements. In a contrastive statement, *vā, tu, punaḥ, parantu* etc are used in a compound sentence. *ataḥ tasmāt* etc are used to express the conclusion of a statement. It is noticed that conjunct 'ca' and 'tathā' are used after the first word of the second sentence in a compound sentence. As- *rāmaḥ paṭhati syāmaḥ ca hasati*.

General Template for Sanskrit NP:

Following the condition of SOV word order, general template of Sanskrit NP can be presented as:

$$\text{NP} = \text{GEN} + \text{DET}_{(\text{ekaḥ/kaścit})} + \text{N}_{(\text{AP})} + \text{N} + (\text{REL})$$

²⁷ Apte, V.S. (1993), *Sanskrit Syntax* (Hindi translation by Dr. Umesh Chandra Pandey), pp-255-277.

Where NP is Noun Phrase, GEN is genitive, DET is determiner (*ekaḥ* or *kaścit*), AP is Adjectival Phrase, REL is relative clause and N is Nominal word (i.e. an agent) and there is a possibility of multiple APs in an NP.

Ex - *raghuvamśasya daśarathasya ekaḥ sarvaguṇa-sampannaḥ putraḥ, yaḥ vanavāsitamasti, dhyānena kamalaṃ paśyati.*

2.4 Hindi Noun Phrase

In Hindi, determiners and adjectives precede the head noun in the NP and relative and compliment clause do not precede the head noun. The NP may function as Subject, Object or compliment of verb. There are three types of sentences: Simple, Compound and complex. Depending on that an NP has been divided into two categories: Simple NP and Complex NP.

2.4.1 Simple NP: A noun by itself or a noun preceded by a determiner constitutes a simple NP. Hindi does not have a category of articles; instead, it uses other determiners, such as the numeral *eka* etc or the indefinite pronoun *koi* as indefinite determiners. Demonstrative *yaha/ye* are also used to make definite specific noun with added meanings of proximity and *vaha/ve* for remoteness of the referents.

In addition to a determiner, a noun may be produced by various quantifiers such as numerals which are further classified as Cardinals (*ek, do* etc), Ordinals (*pahalā, dūsarā..*), Fractionals (*ādhā, chauthāyi* etc), Multiplicatives (numeral added by suffix *gunā* i.e. *dogunā* etc), Aggregative (numeral added by suffix -o i.e. *dono* etc) and other quantifiers (i.e. *hara, kucha, lagabhaga karība* etc).

Adjectives and limiter Particles (quantifiers) may also precede a noun in a simple NP. The adjective which ends with *-ā* have a feminine form *ī* and a direct plural and oblique singular form ending in *-e* (as- *acchā, acchī, acche*). When it ends in consonants or in vowels other than *-ā*, have the same form in all genders, numbers and cases. As *lāla phūla* has adj. *lāla* and the singular object (noun) *phūla*, adjective

is unchanged when the object *phūla* is changed as plural (i.e. *lāla phūlō*), Comparative and superlative adjectives are forms having suffixes *-tara* and *-tama* borrowed from Sanskrit (*as- śreṣṭhatara* and *śreṣṭhatama*). All adjectives participate in periphrastic comparative and superlative constructions. The comparative construction is generally represented as: [compared entity + standard of comparison + *se* + adj. + linking verb]. For example: *vaha tumase adhika mehanatī hai*. The superlative construction has the following forms: [compared entity + *saba* + *se* + adj. + linking verb]. As - *gulāba sabase sundara phūla hai*. Kachru (1984) states that on the basis of specific syntactic positions i.e. whether a particular adjective can precede a noun or function as a compliment to a linking verb, or occur in both positions, adjectives can be classified as:

- a. Attributive (precedes the head noun) and Predicative (compliment as post-modifiers).
- b. Basic (non-derivational i.e. red, old etc.), Derived (Hindi uses much wider range of participial adjectives), and Complex Adjectives (preceding sub-modifiers as *ek bahuta acchā*).
- c. Pronominal (pronoun used as adjectives).
- d. Semantic Subclasses (i.e. gradable (*bahuta acchā* etc) and non-gradable (*jīvita, mṛta* etc), inherent (characterizes nouns directly; as - *purānī kitāba*) and non-inherent (do not characterizes nouns directly; as - *purānā dostā*), stative (to show the status as – *vaha vīra hai*) and dynamic (the status in progress as – *vaha lambā ho raha hai*)).

2.4.2 Complex NP: Complex NPs are of three types: a) Nouns with complement clause b) Noun clause²⁸ and c) Nouns with relative clauses and participle modifiers.

2.4.2.1 Noun with Complement Clause (CC) is the statement apposition to the noun head of the sentence which is added by the conjunct *ki*.

²⁸ Yamuna Kachru (1984), p-135.

Ex - [NP {N-(yaha kahanā) CC- (ki rājū jhūṭha bola rahā hai)}].

2.4.2.2 Noun Clause has a genitive form of nominal or pronominal word followed by gerund. It functions as the NP.

As- NP {(billī kā) (ronā)}.

2.4.2.3 Noun with Relative Clause: If the nominal/pronominal head is kept at the beginning of the sentence, the relative clause comes after the nominal/pronominal head.

As- {NP (vaha) (jo bāhara hai)}

or

{NP (vaha) (jisaki yaha kitāba hai)}

If the relative clause is kept at the beginning, nominal/pronominal head is kept just after the relative clause.

2.4.2.4 Noun with Participle Modifier: When a participle form is used as the modifier of any nominal or pronominal head, forms a NP.

As- NP {(rotā huā) (bacchā)}

Template for Hindi NP:

The template may be given as:

GEN + DET +N(AP) +N + (REL)

Ex- NP (gāva kā eka acchā laḍakā jo dillī meṅ paḍhatā hai)

2.5 Basic difference between Sanskrit and Hindi NPs

Sanskrit Template: NP = GEN + DET (*ekaḥ/kaścit*) + N_(AP) + N + (REL)

Hindi Template: NP = GEN + DET (*eka/koī*) + N_(AP) + N + (REL)

On the basis of two similar template in both the languages, one can say that both languages have same syntax which is generally presented in SOV word order. In that case, for sure, NP components can be classified as Noun head, case marker and modifier. Noun head may consist *ek, koi* etc as articles in Hindi and *ekaḥ, kaścī* etc in Sanskrit. But Sanskrit differs from Hindi at case marker and modifier levels. Sanskrit being inflectional language does not show any postposition while Hindi has postposition because of its derivational nature. At the level of modifier, two things can be noticed in Sanskrit, 1) case marker is inflected with modifier²⁹ and 2) modifier can take any place because of respectively free word order. In Hindi, modifier has no case marker or it takes oblique case marker. The conjunct *ca* can be used at the end of all nominals or with each nominal as it is used in *rāmaḥ lakṣmanaḥ ca* and in *rāmaśca lakṣmanaśca*. In Sanskrit *kṛdanta* nominal functions as modifier but when there is no verb phrase in the given sentence, *kṛdanta* functions as verb.

If we follow the *anvaya* theory, Simple and complex NPs have almost the same structure. The only difference is caused by the *ananvita* Sanskrit text where adjective can come anywhere. Here complex NP has a VP included in relative clause. As this research is not covering the VP, complex NP (which includes compound and complex sentence) will not be here.

These differences can be listed below:

Sanskrit NP	Hindi NP
<i>ekaḥ, / kaścī</i> are used as an article	<i>ek, koi</i> etc are used as an article
No fixed place for adjective	Fixed place for adjective
Case marker inflected	Case marker not inflected
Adjective has case marker same as substantive	Adjective drops case marker except oblique case

²⁹ *yā viśeṣyeṣu dṛṣyante liṅasaṃkhyāvibhaktayaḥ, prāyastā eva kartavyā samānārthe viśeṣaṇe.*

Chapter 3

***CHALLENGES IN SANSKRIT-HINDI NOUN
PHRASE MAPPING***

3.1 Introduction

This chapter discusses the challenges in NP mapping between SL (Sanskrit) and TL (Hindi). Because the nature of both languages is different, contrastive linguistic properties between Sanskrit and Hindi are given below:

Sanskrit	Hindi
Inflectional	Post positional
Three numbers	two numbers: singular & plural
Three genders: masculine, feminine & neuter.	Two genders: masculine and feminine.
Relatively free word order	Less free word order
No ergative	Ergative as special feature
No Subject-Verb agreement for gender in tinanta.	Have gender agreement at the level of verbs
Two <i>subvibhaktis</i> cannot come together	Two <i>vibhaktis</i> can come together (<i>usme se</i>)
<i>karma-pravacanīya</i> can come after or before a word	<i>karma-pravacanīya</i> can come only after a word.
Many <i>vibhaktis</i> (<i>case markers</i>) with <i>upapada</i>	Only genitive case marker with <i>upapada</i> (<i>ram ke sath</i>)
Conjunct <i>ca</i> at the end	Conjunct in between words
Adjectives and nouns both have same case markers	Only noun has case marker (except oblique cases)
Accusative form is used in multiple <i>kāraṅkas</i>	Accusative form (<i>ko</i> case) marker is used only in accusative and dative.

Table 1: contrastive property of Sanskrit and Hindi language.

These differences are capable of giving a clear picture about the challenges faced in SHMT. Challenges in NP Mapping can be discussed under the following subtopics:

- Case markers difference between Sanskrit and Hindi

- Different *vibhaktis* in a particular case in Sanskrit
- Hindi cases and their different usage
- *Vibhakti* divergence and challenges in case mapping
- Challenges in indeclinable mapping
- Challenges in mapping compound nominal
- Challenges in mapping adjectives
- Challenges in Mapping gender and number
- Challenges in Mapping Word order

3.2 Case marker differences between Sanskrit and Hindi

Because Sanskrit is inflectional language and Hindi is derivational, the case system must be looked into. We generally find the comparison between both languages in which *sup* vibhakti is kept in parallel with the Hindi case markers. As *kāraka* and *vibhakti* are different in Sanskrit, here *kāraka* is termed as case and *vibhakti* is termed as case marker. This general hypothesis states that the *kartā kāraka* takes *su/au/jas* *vibhakti* in Sanskrit and \emptyset , *ne* *vibhakti* (case marker) in Hindi. Following table shows the case marking difference:

Case	Sanskrit Case Marker			Hindi Case Marker
<i>kartā (Agentive)</i>	<i>su</i>	<i>au</i>	<i>jas</i>	\emptyset , <i>ne</i>
<i>Karma (Accusative)</i>	<i>aṃ</i>	<i>auṭ</i>	<i>śas</i>	\emptyset , <i>ko</i>
<i>karana (Instrumental)</i>	<i>tā</i>	<i>bhyāṃ</i>	<i>bhis</i>	<i>se, ke dvārā</i>
<i>sampradāna (Dative)</i>	<i>ṅe</i>	<i>bhyāṃ</i>	<i>bhyas</i>	<i>ko, ke lié</i>
<i>apādāna (Ablative)</i>	<i>ṅasi</i>	<i>bhyāṃ</i>	<i>bhyas</i>	<i>Se</i>
<i>Sambandha (Genitive)</i>	<i>ṅas</i>	<i>os</i>	<i>āṃ</i>	<i>kā, ke, kī</i>
<i>adhikaraṇa (Locative)</i>	<i>ṅi</i>	<i>os</i>	<i>sup</i>	<i>me, para</i>
<i>Sambodhana (Vocative)</i>	<i>su</i>	<i>au</i>	<i>jas</i>	<i>he, ho, re, are</i>

Table 2: General difference between Sanskrit Hindi case marker

Here first thing one can notice easily is that Sanskrit case markers are different in three numbers while Hindi case markers are not differentiated on the basis of number. Case marker *bhyām* in three cases and *bhyas* in two cases in Sanskrit presents challenge in its translation.

Basically, translation is not simply a map from lexical item to lexical item. There is some kind of syntactic-semantic relationship which is presented differently in different languages. In Sanskrit-Hindi NP mapping, it can be noticed at different levels: at the level of *kāraka* adjectives, genders, numerals as well as word order. These issues will be discussed here to find out the challenges in getting comparatively correct output in Hindi (TL). Let us, first, see the different usage of Sanskrit case marker to understand the possible case mapping in TL for SHMT.

3.3 Different *vibhaktis* in a particular *kāraka* (case) in Sanskrit

The statement *vivakṣātaḥ kārakāṇi bhavanti* makes it clear that the *kāra*kas may take any *vibhakti* (case marker) in Sanskrit depending on the desire of the speaker but the meaning should not be changed. The *kāraka prakaraṇa* presents a clear picture of the usage of different *vibhaktis* in a particular *kāraka* (case) in Sanskrit which is presented here.

3.3.1 kartā: Except *pañchami* (of *apādāna*), all *vibhaktis* are found in *kartā* in Sanskrit sentence.

3.3.1.1 Kartari prathamā- *prātipadikārtha-liṅga-parimāṇa-vacanamātre prathamā* (P.2.3.46) says that *prathamā* is introduced when nothing but nominal stem notion or nominal stem notion with gender, number or measure is to be denoted. In *rāmaḥ gacchati*, *rāmaḥ*, the agent, has *prathamā vibhakti* which denotes the gender and number.

3.3.1.2 Kartari dviṭyā- *gati-buddhi-pratyavasānārtha-śabdakarmakāṇāmapi kartā sa nau* (P.1.4.52) says that a *kāraka* which serves as agent of a non-*ṇic* root that either

has the signification *gati* (movement), *buddi* (perception), *pratyavasāna* (conjuncture) or has *sabda* (sound) as its object, or has no object at all, is termed as *karma* when the same root terminates in *ṇic*. In example- *śatrūn agamayāt svargaṃ, gamayate* verb has *ṇic* form of *gam* root. When the agent is marked as *karma* from the above said *sutra*, *karmanī dviṭyā* (P.2.3.2) is applied.

3.3.1.3 Kartari tṛṭyā - *karṭṛkaraṇayostrṭyā* (P.2.3.18) introduces *tṛṭyā* to denote the *karṭṛ* where sentence is in passive voice. In *rāmeṇa rāvaṇaḥ hataḥ*, agent is in *tṛṭyā vibhakti*.

3.3.1.4 Kartari caturthī- *pratyāñbhyāṃ śruvaḥ pūrvasya kartā* (P.1.4.40) says that a *kāraka* that serves as the agent of a prior act of requesting is termed *sampradāna* when root *śru* is used with the prefix *prati* and *āñ* in Sanskrit sentence. In *viprāya gāṃ pratiśṛṇoti*, *vipra* is the agent of the sentence to whom the pledge is made upon request. Here *caturthī sampradāne* (P.2.3.13) *sutra* is applied with *kartā*. In *hotre anugṛhṇāti/pratigṛhṇāti, anupratigṛhṇasca* (P.1.4.41) is applied which says that a *kāraka* that serves as the agent of a prior act of requesting is termed *sampradāna* when root *gr* (of *kryādi gaṇa*) is used with the prefix *anu* and *prati* in Sanskrit sentence. Here also, *caturthī sampradāne* (P.2.3.13) is applied with the agent.

3.3.1.5 kartari ṣaṣṭhī- According to *karṭṛkarmanoh kṛti* (P.2.3.65), *ṣaṣṭhī vibhakti* is used with *kartā* (and *karma* as well) when *kṛt* is there. As- *kṛṣṇasya kṛtiḥ*. The *karṭṛ* or *karman*, if expressed with the genitive, must be related with the action denoted by the root underlying the item which ends in *kṛt*.

3.3.1.6 kartari saptami- *arhāṇāṃ karṭṛtve 'narhāṇāṃ-akarṭṛtve tadvaiparītye ca* (vā. 2.3.37) says that *saptami vibhakti* is used to indicate the *karṭṛtva* of capable and *akarṭṛtva* of anable. As- *asatsu tiṣṭhatsu santaḥ taranti*. In this example *asatsu* is the *kartā* of verb *tiṣṭh* and it has *saptami vibhakti*.

3.3.2 karma: All 7 *vibhaktis* are found in *karma*.

3.3.2.1 karmanī prathamā- by the *sutra* 'prātipadikārtha liṅga parimāṇa vacanamātre prathamā (P.2.3.46)' *prathamā vibhakti* is used with *ukta karma*, the

karma which is already expressed (*ukte karmaṇī prathamā syāt*). As- *granthaḥ paṭhyate*.

3.3.2.2 *karmaṇī dvitīyā*- by the *sutra* '*karmaṇī dvitīyā* (P.2.3.2)' *anukta karma* has *dvitīyā vibhakti*. As- *grāmaṃ gacchati*. Here *karma* (*grāmaṃ*) is *anukta* so it has *dvitīyā vibhakti*.

3.3.2.3 *karmaṇī tṛtīyā*- *divaḥ karmaḥ ca* (P.1.4.43) says that a *kāraka* which serves as means, more than anything else, is termed as *karma*. As- *akṣaiḥ dīvyate. kariṅkaraṇayostrīyā* (P.2.3.18) says that a *tṛtīyā* occurs in *karta* and *karaṇa* where *kartā* is *anukta* and *karma* is *ukta*. *Sutra* '*sañjo 'nyatarasyāṃ karmaṇī* (P.2.3.22)' says that a *tṛtīyā* optionally occurs to express the object of *jñā* used with the suffix *sam*. For example- *pitṛā samjānīte*. The *vārtika* '*yajeḥ karmaṇaḥ karaṇasañjyā....* (vā 2.3.13)' says that the *karma* of root *yaj* occurs *karaṇa samjñā*. that's why *paśunā* has *tṛtīyā* in the sentence *paśunā rudraṃ yajet*.

3.3.2.4 *karmaṇī caturthī*- *kriyārthopapadasya ca karmaṇī sthāninaḥ* (P.2.3.14) says that *chaturthi* occurs to express the object of a verb which denotes an action for which another action is intended and *tumun* suffix is not used. Ex- *phalebhyo yāti*. *Sutra* *manyakarmanyānādare vibhāṣā 'prāṇiṣu'* (P.2.3.17) says that *karma* occurs *caturthī* optionally with the object of root *man* when disrespect is expressed. Ex- *na tvāṃ tṛṇaṃ/tṛṇāya manye*. The *sutra* '*gatyarthakarmaṇī dvitīyācaturthau ceṣṭāyāmanadhvani* (P.2.3.12)' says that either *dvitīyā* or *caturthī* occurs to express the object of a verb of movement (*gati*), provided that the object is not expressed otherwise, that it is not *adhvan* and that the action involves actual movement (*cheṣṭā*). Ex- *grāmaṃ/grāmāya gacchati*.

3.3.2.5 *karmaṇī pañcami*- *lyablope karmanyadhikaraṇe ca* (vā. 1.4.31) says that when the *lyab-pratyayānta* words are not used then the *karma* and *adhikaraṇa kāraka* have *pañcami*. As- *prāsādāt prekṣate, āsanāt prekṣate*. Here *āruhya* and *upviśya* is unexpressed in the examples and that's why *pañcami* is used instead of *dvitīyā*.

3.3.2.6 *karmanī śaṣṭhī* - The sutra *śaṣṭhī śeṣe* (P.2.3.50) says that *svasvāmibhāvādisambandha* is called *śeṣa* which is not considered as *kāraka* and there the *śaṣṭhī vibhakti* is used. In *mātuḥ smarati*, *mātuḥ* has *śaṣṭhī vibhakti*. *adhīgarthadayeṣāṃ karmaṇi* (P.2.3.52) says that the *śaṣṭhī vibhakti* is occurred after a stem to express, as a reminder, the karma of verbs which denote the sense of *adhika dayā* and *īś* roots. Ex- *sarpiṣo dayanaṃ (ghee kā dāna)*. (*kṛñjaḥ pratiyatne* (P.2.3.53) says that when *pratiyatna* is expressed then the object of *kṛ* root has *śaṣṭhī vibhakti*. For example- *edhodakasyopaskurute (īndhana jala me uṣṭā utpanna karatā hai)*. *kartṛkarmaṇoḥ kṛtī* (P.2.3.65) says that *śaṣṭhī vibhakti* is used with *kartā* and *karma* when *kṛt* is there. Ex- *dugdhasya pānam*. Here *dugdha* has *śaṣṭhī vibhakti* and *pānam* has *kṛt* suffix.

3.3.2.7 *karmaṇi saptamī- ktasyenviṣayasya karmanyupsamkhyānam* (vā.2.3.36) sates that the *karma* of the words formed by adding *in* suffix to the *ka-pratyayānta* word has *saptamī*. As- *adhītī vyākaraṇe (jisane vyākaraṇa padha liyā hai)*. *sādhasādhuprayoge ca* (vā.2.3.36) states that the *saptamī* case is used with the *karma* of *sādhu* and *asādhu*. As- *sādhuḥ mātari, asādhuḥ mātule* etc. *nimittāt karmayoge* (vā.2.3.36) states that the *saptamī vibhakti* is used with the purpose of the verb, if the purpose of the verb is associated with the object of the verb. Ex- *carmanī dvīpinaṃ hanti*. Here getting *carma* (skin) of tiger is the purpose of the verb. So it has *saptamī vibhakti*.

3.3.3 **karāṇa:** 4 vibhaktis are used in *karāṇa* in a Sanskrit sentence:

3.3.3.1 *karāṇe tṛtītā- kartṛkarāṇayostrītyā* (P.2.3.18) introduces *tṛtīyā* to denote the *kartṛ* as well as *karāṇa*. Ex- *bāṇena hataḥ*. Here *bāṇena* has *tṛtītā* case marker.

3.3.3.2 *karāṇe caturthī-* the sutra- *parikrayaṇe sampradānam-anyatarasyāṃ* (P.1.4.44) states that the means (i.e. instrument) which is helpful in hiring any person or get a thing, takes the *caturthī vibhakti*. As- *śatāya/śatena parikrītaḥ*.

3.3.3.3 *karāṇe pañcamī-* The sutra- *karāṇe ca stokālpa-kṛcchr-katipayasyāsatva-vacanasya* (P.2.3.33) states that a *tṛtīyā* optionally occurs to *pañcamī* to express

karaṇa after *stika* ‘a little’, *alpa* ‘a little’, *kṛcchra* ‘difficult’, or *katipaya* ‘a few’ provided they do not denote the *sattva* ‘substance’, and *karaṇa* is not expressed. As-
stoken stokād vā muktiḥ.

3.3.3.4 *karaṇe ṣaṣṭhī*- The sutra *jñō vidarthasya karaṇe* (P.2.3.51) states that a *karaṇe ṣaṣṭhī* occurs after a stem to express the *karaṇa* of *jñā* when not used in the sense of knowledge. As- *sarpiṣaḥ jñānam*. Here root *jñā* has the sense of *pravṛtti*. Thus the *ṣaṣṭhī vibhakti* is used in *sarpiṣa*.

3.3.4 *sampradāna*: 4 *vibhaktis* are found in *sampradāna* in a Sanskrit sentence.

3.3.4.1 *sampradāne prathamā* - In *dānīyo vipraḥ, dānīyaḥ* has *prathamā* by the sutra *prātipadikārtha līṅga parimāṇa vacanamātre prathamā* (P.2.3.46) where $\sqrt{dā}+anīyara$ (by *kṛtyaluṭo bahulaṃ* (P.3.3.113), is in dative which is ukta.

3.3.4.2 *sampradāne dvitīyā* - In *mānavakaṃ dharmam brūte/śāsti, mānavakaṃ* has *dvitīyā* by the sutra- *akathitaṃ ca* (P.1.4.51) where it is said that the *apradhāna karma* has also *dvitīyā vibhakti*. It is optional as here dative case can also be used. The *vārtika*- *yajeh karmanah karaṇasañjā sampradānasya ca karmasañjā* (vā. 2.3.13) states the *karma-sañjā* of *sampradāna*. As- *paśunā rudraṃ yajate* has *dvitīyā* in *rudraṃ* where *rudra* is *sampradāna-bodhaka*.

3.3.4.3 *sampradāne tṛtīyā* - The *vārtika*- *aśiṣṭavyavahāre dānaḥ proge caturthiarthe tṛtīyā* (vā. 2.3.23) states that when the donation is for the purpose of an immoral act, then the dative takes *tṛtīyā*. As- *dāsyā sañyacchate kāmukaḥ*. Here *dāsī* is *sampradāna* but has *tṛtīyā vibhakti*.

3.3.4.4 *sampradāne caturthī* - The sutra- *caturthī sampradāne* (P.2.3.13) says that *sampradāna* has *caturthī vibhakti*. As- *viprāya gāṃ dadāti*. Here *vipra* is the *sampradāna* and it has *caturthī vibhakti*.

3.3.5 *apādāna*: 3 *vibhaktis* are found in *apādāna* (locative).

3.3.5.1 *apādāne dviṭīyā-* The *sutra- akathitaṃ ca* (P.1.4.51) states that the *apradhāna karma* has also *dviṭīyā vibhakti*. Here in *baḷiṃ yācate vasudhāṃ*, *baḷi* is *apradhāna karma* and it has *dviṭīyā* in the sense of *apādāna* (ablative).

3.3.5.2 *apādāne ṭṛṭīyā - vibhāṣā guṇe'strīyāṃ* (P.2.3.25) says that a *pañcami* optionally occurs after a non-feminine stem which denotes *hetu* as well as *guṇa* 'quality'. *jāḍyāt jāḍyena vā baddhaḥ*. If the option of *pañcami* is not taken, *ṭṛṭīyā* can be used.

3.3.5.3 *apādāne pañcamī-* The *sutra- apādāne pāñcamī* (P.2.3.28) says that a *pañcamī* occurs after a nominal stem when *apādāna* is not expressed otherwise. *As-grāmāt āyāti*.

3.3.6 *sambandha:* Only one *vibhakti* is used in *sambandha* i.e. *ṣaṣṭhī* by the *sutra ṣaṣṭhī śeṣe* (P.2.3.50).

3.3.7 *adhikaraṇa:* 4 *vibhaktis* are used in *adhikaraṇa*.

3.3.7.1 *adhikaraṇe dviṭīyā-* In *vrajaṃ avaruṇaddhi gāṃ*, *vrajaṃ* is *apradhāna karma* and used as locus. Hence it has *dviṭīyā* by the *sutra- akathitaṃ ca* (P.1.4.51). The *vārtika- akarmakadhātubhīryoge deśaḥ kālo bhāvo gantavyo'dhvā ca karmasamjñakaṃ iti vācyam* (vā. P.1.4.51) says that a *karma-samjñā* occurs when an intransitive root is used to indicate the place, time, action and path. *As- kurun svapiti*. The *sutra- adhiśīṅsthāsām karma* (P. (P.1.4.46) says that a *kāraka* which serves as locus is termed as *karma* 'object' when *sīṅ*, *sthā* and *ās* root is used with the prefix *adhi*. *As- adhiśete/adhiṣṭhati/ adhyāste vaikuṇṭhaṃ hariḥ*. The *sutra - abhiniviśaśca* (P.1.4.47) states that a *kāraka* which serves as locus is termed as *karma* 'object' when *viś* is used with prefix *abhi+ni*. *As- sanmārgam abhiniviśate*. The *sutra-upānvadhyāñvasaḥ* (P.1.4.48) says that a *kāraka* which serves as locus is termed as *karma* 'object' when *vasa* is used with *upa*, *anu adhi* and *āñ*. *As-upavasati/anuvasati/ adhivasati/āvasati vaikuṇṭhaṃ hariḥ*.

3.3.7.2 *adhikaraṇe pañcamī-* The *vārtik- lyablope karmaṇyadhikaraṇe ca* (vā. 1.4.31) says that when the *lyab-pratyayānta* words are not used then the *karma* and

adhikaraṇa kāraka have pañcami. As- āsanāt prekṣate. Here *āsana* is the locus and *pañcami* is used with *āsana*.

3.3.7.3 *adhikaraṇe śaṣṭhī-* The *sutra- kṛtvorthaprayoge kāle 'dhikaraṇe ca* (P.2.3.64) says that a *śaṣṭhī* occurs to express locus after a stem which denotes time and is used in conjunction with a word which ends in an affix having a meaning of *kṛtvasuc*. *As-idameṣāmāsitaṃ gataṃ bhuktaṃ vā.*

3.3.7.4 *adhikaraṇe saptamī-* *saptamyadhikaraṇe ca* (P.2.3.36) simply says that the locus takes *saptamī vibhakti*. *As- kaṭe āste.*

3.4 Hindi cases and their different usage

In 3.3 it is observed that *kāraka* takes different *vibhaktis* in Sanskrit. Here it will be observed that a case marker in Hindi is used in different ways with different *kārakas*.

There are two cases in Hindi: direct and oblique. The direct case is used for nouns and is not followed by any postposition. The oblique case is used for nouns followed by a postposition. Adjectives qualifying nouns in the oblique case in the same phrase will inflect in the same manner. Case-suffixes and postpositions are used to express syntactico-semantic roles. Case suffixes are bound morphemes and are added only to the noun phrases. The NPs in Hindi takes oblique cases, when followed by postpositions. The vocative address forms may be preceded by the vocative morphemes *o/he/are*. Except for ergative, dative and passive subject, the default case marker (with agent) is null.

3.4.1 Ergative in Hindi: Ergative marker 'ne' is used with subject in perfect tense with transitive verbs. For example: *Gāndhiji ne kahā hai, Gāndhiji ne kahā hogā, Nehru ne kahā, usne kahā thā* etc. Here, when *ne* follows a noun, it is written separately and when 'ne' follows a pronoun, it is written as one word. Examples above show that *kahā hai. kahā and kahā thā* are used in perfect tense while *kahā hogā* is showing the assumption. It is clear that *ne* cannot be used in future tense.

3.4.2 ko case marker: It is used in a larger context in Hindi. It is generally assumed as *karma kāraka* (accusative) but it may be used in many cases. For example, *rāma ko ghara jānā hai*, *mohana ko patra likhanā hai* etc where *rāma* and *mohana* are the agents. In the sentence *pitā putra ko dekha rahā hai*, *putra* is object and the case marker *ko* is accusative. In another sentence, *mohana rāma ko pustaka detā hai*, *ko* case marker is used with dative *rāma*. *ko* casemarker can be used with locative. For example: *somavār ko padhāi hogī*. There must be ‘*ko*’ case marker in imperative sentence. For example – *bālaka ko padhanā cāhiye*.

3.4.3 se case marker: It is used with Instrumental and Ablative case frequently but it is also used in other cases. For example - *rāma se aba uthā nahī jātā hai*, *bacce se darda sahā nahī jātā*. In these sentences, *se* is used with agentive. In *mohana rāma se kahatā hai*, *se* is used as accusative. In *rāma kalama se likhatā hai*, *se* is used as instrumental. In *himālaya se gangā nikalatī hai*, *se* is used as ablative. In *peḍha se bandara laṭakatā hai*, *se* is used as locative. In *śera se sabhī ḍarate hain*, *se* is used in the sense of fear. *se* can be used in the sense of ‘with’ as *sonu caṭanī se roṭī khātā hai*, *acāra se parāṭhā khātā hai* etc. The sentence *tū mat padha, merī balā se* has the sense of negligence. *se* can be used with direction to denote the place. For example: *ayodhyā se mithilā gayā*. It is also used to show the time. For example- *tumhen kitane samaya se ḍhūṇḍha rahā hūn*. Case marker *se* is also used to show the *bhāva* of work. For example- *mana se paḍhatā hai*, *dhyāna se suntā hai*, *sthira se baithā hai*, *sthira se kāma karo* etc.

3.4.4 Genitive case markers: *kā* case marker is used with possessed noun-masculine. *āpkā bhātātā hai*, *āpakā svāsthya acchā hai* etc. *ke* case marker is used when the direction is right or left or nearby. *śyāma ke bāī or kursī hai*, *ūsake dāhinī ora baccā hai*, *pitā ke bagala me putra hai* etc. *ke* case marker is used with possessed noun followed by verb or adjective or other nouns. As- *latā ke ghara kī sundaratā dekhī*, *āpke makāna kā kāma pūrā huā*, *rāma ke vidyālaya me śikṣaka hain?* etc. Case marker *kī* is used with possessed noun- feminine. For example- *āpakī bahan jātī hai*, *āpkī tabiyata acchī hai*. It is used when the direction is followed by the word *diśā/ or*. As- *pūrva kī diśā me ghara hai*, *āge kī or baḍha gayā*. case marker *nā*, *nī* & *ne* are

always used with pronoun and the form becomes - *apnā*, *apanī* & *apne* respectively. As- *apnā glāsa lāo*, *apni kalam uthāo*, *apne bhāi ko bulāo* etc. Other case markers *rā*, *rī* & *re* are coming with *main* and *tum* in singular and plural both and gives *merā/mere/merī*, *hamārā/ hamāre/ hamārī* and *terā/ tere/ terī* respectively. As- *hamārā/merā bhāi āyā*, *hamāre /mere guru ne kahā*, *hamārī/merī mā viduṣī hai* etc.

3.4.5 Locative case markers: *meṅ* and *para* are used as case markers in locative. The case marker *meṅ* generally indicates the inner location of any object as- *pātra me jala hai*. The case marker *para* is used as an outer location as - *vrkṣa para caḍhatā hai*. But *maidāna me khelatā hai* has *me* case marker for outer location. So it is used vice-versa and hence use of this case marker is dependent on the type of location. For example, if the location is *ghaḍā* (i.e. *ghaṭa*), then the case marker is depending on the nature of the object/matter (*ādheya*). For example- if the object/matter is *jala*, locative case marker *me* will be used (*ghaḍe me jala hai*). If the object is *phala* then both the case marker may be used depending on the intention of the speaker. If the fruit is smaller, it may be kept inside the *ghaṭa* and if it is bigger, it will be kept on the *ghaṭa*. Hence *me* and *para* both the case markers are suitable.

3.4.6 Vocative case marker: It is generally used before the noun to address someone. It can also be used with adjective or the adjective which is used as name.

3.5 *Vibhakti* Divergence and Challenges in NP- case mapping

Sanskrit sentences are of three types- simple complex and compound. A single verb can be a simple sentence in Sanskrit (*ēka-tiṅ vākyaṃ*), where NP could not be found in absence of any nominal *pada*. Generally a simple sentence has an NP and VP i.e. *subanta* and *tiṅanta*. In Complex sentence, there is one main subject (*uddeśya*) and two or more *samāpikā* verbs. As - *yasyārthāḥ tasya mitrāṇi*. In these sentences, there is one main clause and rest is dependent clause. There are three types of dependent clause: nominal clause, adjective clause and adverb clause which expend noun, adjective and adverbial respectively. Nominal clause is used as an agent or an object, adjective clause always becomes an adjective of the agent or the object and adverbial

clause is used in VP. In a compound sentence, two or more simple/complex sentences are added by conjunct and these are *samānādhikaraṇa* of each other. The *indiclinable ca, tathā, api* etc are used to add two or more statements.

In these three types of sentence, simple sentence has the construction as NP and VP. NP may have modifiers and agent may take all the six case markers discussed with *kartā* in section 3.3.1 as well as genitive case. So it must be the *kartā* (agentive case). The rest cases will come under the VP which functions as object, instrument etc.

In complex sentence, NP has noun head and an adjectival clause. An adjectival clause is basically a complete statement about the agent and it functions as an adjective and comes under the NP category. Since it is just like a complete sentence, it has also two parts: NP and VP. It may be a very long compound sentence in which simple or complex sentence is concatenated with a complementizer. In a compound sentence, two or more simple/complex sentence is concatenated with a conjunct. So there are two or more complete sentences and each of them has NP-VP structure. The sentence pattern can be represented like this:

Simple sentence: [{NP (Adjective + Noun)} {VP((Adverb)+Verb)}]

Complex sentence: [{NP (Noun+ Adjectival clause)} {VP ((Adverb) +Verb)}]

Compound sentence: [{NP (Adjective + Noun)} {VP ((Adverb) + Verb)}] + conjunct
+[{NP (Adjective + Noun)} {VP((Adverb)+Verb)}] or

[{NP (Noun+ Adjectival clause)} {VP ((Adverb) + Verb)}] + conjunct +[{NP (Adjective + Noun)} {VP((Adverb)+Verb)}].

In this representation, we can again go for the subdivision of VP which consists of NP and verb. An NP which is coming under VP, may consist any case marker in Sanskrit. Here the challenges coming in those NP-case mapping will be discussed but the adjectival clause which is functioning as an adjective in a complex sentence, will not be discussed as they are basically a complete sentence which requires the study of VP for its complete mapping. The discussion of adjectival clause needs the study of NP

and VP. As this study is focused on NP only, the adjectival clause will not be discussed here.

To get an idea of divergences at the case level, comparison may be done in following similar sentences having same case marker in Sanskrit but differently translated in Hindi.

1. In the following two sentences, *vedaḥ adhīyate* and *dharmāḥ caryate*, *vedaḥ* and *dharmāḥ* has *prathamā vibhakti* in SL but their translation in Hindi (TL) is different. In *veda paḍhā jātā hai*, the case marker is dropped while *dharmā kā pālana kiyā jātā hai* has *kā* case marker. *rāmāḥ agacchat* (*rām gayā*) has no case marker in its translation while *rāmāḥ akathayat* (*rām ne kahā*) has ergative. Here *rāmāḥ* is the subject but *vedaḥ* and *dharmāḥ* are object in the above discussed examples.
2. The following examples: *āmraṃ khādati* (*āma khātā hai*), *vṛkṣaṃ ārohati* (*vṛkṣa para caḍhatā hai*), *bālakaṃ pāṭhayati* (*bālaka ko paḍhātā hai*), *sītāṃ pariṇināya* (*sītā se vivāha kiyā*), *phalaṃ icchati* (*phala kī icchā karatā hai*), *krośaṃ gacchati* (*kośa bhara jātā hai*), *dharmāṃ anugacchati* (*dharmā para calatā hai*), *rāmaṃ anugacchati* (*rāma ke pīche jātā hai*) etc are the examples where there is *dvitīyā vibhakti* in Sanskrit but their Hindi translation has different postpositions.
3. The following sentences have *ṭṛtīyā* but its translation have different postpositions in Hindi: *mayā kathitaṃ* (*maine kahā*), *gunaiḥ ātmasaḍrśūṃ kanyāṃ uḍavahe* (*guṇa meṅ apne samāna kanyā se tū vivāha kara*), *anena prakāreṇa kathaṃ likhasi?* (*isa prakāra kyōṅ likhate ho?*). The first sentence has ergative, second has the locative case marker while the third has no case marker.
4. These following sentences have *catuṛthī vibhakti* in Sanskrit but takes different postposition in Hindi Translation: *bālakāya phalaṃ dadāti* (*bālaka ko phala detā hai*), *vṛtrāya śāstraṃ praharati* (*vṛtra para śāstra pheṅkatā hai*), *nārī nāryaiḥ īrṣyati* (*nārī nārī se īrṣyā karatī hai*).
5. *akṣān/akṣaiḥ dīvyati* has been translated as – *pāson se khelatā hai*. Here it may also be translated as- *pāsā khelatā hai* where the case marker will be dropped.

Carudev Shastri ¹says that when the *vyavahāra artha* is taken for *√dīva*, translation will be *pāsā phenkatā hai*. Here we see that case marker is dropped depending on the intension of the speaker.

6. In *dhanikaḥ dhanāya/dhanena trīn sevakān dve varṣe yāvat parikrīnāti*, *dhanāya* may be translated as *dhana se* in Hindi but the other example shows the other possibilities in its case marking: *pratimāsaṃ dānīyebhyaḥ śatāya rupyakebhyaḥ* = *sau rupaye māśika deya par yaha naukara rakhā gayā hai*. Or it may be translated as – *prati-māsa deya sau rupaye par yaha naukara rakhā gayā hai*. Here *para* case marker is used for *dānīyebhyaḥ* and *sau rupaye* (*śatāya rupyakebhyaḥ*) has oblique form of *rupayā*. In another example, *viṃśati rupyakebhyaḥ parikrītamidam geham asmābhiḥ* = *hamāre dvārā yaha ghara bīsa rupaye me kirāye para liyā gayā*. Here *me* case marker is used instead of *se* or oblique case or locative case marker. So the same types of sentences have different possible translation and hence it becomes challenging for machine to handle.
7. Depending on the sentence pattern of *saḥ vidyālayāya gacchati* (*vaha vidhyālaya jātā hai*), if *saḥ grāmāya chalati* is translated as- *vaha gāvā chalatā hai*, it will not be a good translation. Here *vaha gāvā ke liye chalatā hai* or *vaha gāvā ki or chalatā hai* will be more appropriate. So the changing verb is affecting the postposition in Hindi.
8. The example- *mayā hariḥ sevyah²* can also be translated as *mujhako hari ki sevā karanī cāhiye* where *mujhe* is the oblique form of *main* with accusative case marker (i.e. *main + ko = mujhe/mujhko*). So here accusative *ko* case marker may also be mentioned in conclusion. It may also be translated as *mujhase hari sevita hai* or *mere dvārā hari kī sevā kī jānī chāhiye*. So *mere dvārā* has instrumental case marker while *mujhe/mujhako* has accusative case in Hindi.
9. The sentence *gām payaḥ dogdhi* is translated as *gāya se doodah duhatā hai* to indicate the ablative. In Hindi speaking areas, we can easily find *kā* instead of *se*

¹ vākvyavahārārtha, p-14.

² SK-p.794

in this example (*gāya kā doodah duhatā hai*). So *se* and *kā* are the two possible case markers in this example.

10. If the sentence *balim vasudhām yācate* is translated in two ways: *bali se vasudhā māṅgatā hai* and *bali se vasudhā ki yācanā kartā hai*, we can find that *vasudhā* is dropping its case marker when the verb *māṅganā* is there and in another translation, it is taking *ki* case marker with the verb *yācanā karanā*. So the effect of verb selection during translation is one very important factor in handling Hindi postposition.
11. In *gargān śataṃ daṇḍayati*, *garga* can take *ko* as well as *para* case marker as-
gargoṃ para sau rupayā daṇḍa lagātā hai and *gargoṃ ko sau rupayā daṇḍa detā hai*. Basically when we take *daṇḍa lagātā hai*, then *para* is coming and when *daṇḍa detā hai* is used, it takes *ko* case marker. Again it is the case of verb selection.
12. In the example of $\sqrt{\text{muṣ}}$, where *devadattaṃ śataṃ muṣṇāti* is translated as *devadatta se sau rupaye curātā hai*, *kā* case marker may also be used i.e. *devadatta kā sau rupayā curātā hai*.
13. In the example of $\sqrt{\text{nī}}$, *grāmaṃ ajāṃ nayati* is translated as *gāva me bakarī ko le jātā hai*. After scrambling we may get *ajāṃ grāmaṃ nayati* in Sanskrit. Then *bakarī ko gāva le jātā hai* is more suitable than *bakarī ko gāva me le jātā hai*. Case marker should not be affected by scrambling.
14. In the example *māsaṃ āste*, *māsaṃ* has accusative because it indicates the time and an intransitive verb is used with it. In its Hindi translation (*mahīne bhara rahatā hai*), the word *bhara* is added. In another example- *kalyaṃ uttiṣṭhati*, *kalyaṃ* also indicates the time but no additional word is inserted in its translation (*prātaḥ uṭhatā hai*), although it may take the locative case marker *me*.
15. The example *śatasya dīvyati vaṇik* and *sahasrasya dīvyati dhūrtaḥ* have *ṣaṣṭhī* instead of *dvitīyā* by the *sutra* - *divastadarthasya* in the sense of gamble and barter (exchange of goods). The first has *kā* case marker in the translation of

śatasya – baniya sai rupayā kā vyāpāra karatā hai, while the other has no case marker in the translation of *sahasrasya – juārī sau rupayā lagātā hai*³.

16. In *duṣṭāya krudhyati*, *duṣṭāya* has dative case marker because of √ *krudh*. It can be translated in two ways: *duṣṭa para krodha karatā hai* or *duṣṭa se gussātā hai*. So two different case marker is used to express the same meaning, depending of the verb selection.
17. In *durbuddheḥ daityasya mṛdhe pratikārayishyati*, *daityasya* has *ṣaṣṭhī* case marker but, in its translation- *mūḍha daitya se yuddha me badalā legā, se* case marker is there⁴.
18. There are two possible translation of the sentence *devadattāya dīrghamāyusyaṃ bhūyāt/devadattasya dīrghamāyusyaṃ bhūyāt* (i.e. *devadatta dīrghāyu ho* or *devadatta kī āyu lambī ho*). First example is dropping the case marker *ke liye* while *devadattasya* is not dropping the case (although the first translation can also be used for the second sentence).
19. Same as the above discussed example, *sarvebhyo sukhaṃ bhavatu* may have two possible translation as- *saba ko sukha ho* or *sabhī sukhī hoveṇ*. Here also one translation has case marker while other drops.
20. In *annasya hetoḥ vasati*, *anna ke hetu se rahatā hai* and *anna ke liye rahatā hai* are the two possible translations. The other example may be- *kena hetoḥ vasati*.

The above discussed points are clearly showing the divergences at case level. These divergences are causing errors in the output because case markers are getting changed in TL while the input sentence pattern is same. Animate-inanimate object, transitive-intransitive verb, effect of the verb selection (of TL) for translating into TL and singularity-plurality of the object are the reasons for the arbitrary change in case markers which is actually depending on the intention. The rules may be written to handle these challenges coming in case mapping on keeping the above reasons in mind.

³ vākvyavahārārtha, p-31.

⁴ vākvyavahārārtha, p-59.

3.6 Challenges in indeclinable mapping

Indeclinable is used in NP to express different meaning in different context. Mapping the indeclinable (used in NP) is thus important as well as challenging in Machine translation. These are described in the following sutras in Aṣṭādhyāyī - *svārādinipātamavyayam, taddhitāścāsarvavibhaktiḥ, kṛnmejantaḥ, ktvātosunkasunaḥ,* and *avyayībhāvaśca* (P.1.1.38-41). The sutra ‘*svārādinipātamavyayam*’ states that the *svārādi-avyaya* and *cādi-avyaya* which are the words of *ākṛtigaṇa*, are indeclinable.⁵ Next sutra ‘*taddhitāścāsarvavibhaktiḥ*’ states that the *taddhita-pratyayānta* words which are not found in all *vibhakti*, are also indeclinable. *kṛnmejantaḥ* states that *makārānta kṛdanta* and *ejanta kṛdanta* are indeclinable. *makārānta kṛdanta* such as *smāraṃ-smāraṃ* ($\sqrt{\text{smṛ}}+\text{ṇamul}$), *paṭhituṃ* ($\sqrt{\text{paṭh}}+\text{tumun}$) etc. The *ejanta kṛdantas* are mostly found in veda. As-jīvase, pibadhyai etc. *ktvātosunkasunaḥ* states that words ending with *ktvā, tosun* and *kasun* are also indeclinable. Example- *kṛtvā, udetoḥ* and *viśṛpaḥ* respectively. The result i.e. *vibhakti lopa* is said by the sutra *avyayādāpsupaḥ* (P.2.4.82). In Amarkoṣa (ṭṛtīya kāṇḍa) *avyaya* words are classified according to their meaning⁶.

Different types of usage of some important indeclinable words are discussed by V S Apte⁷. These may be shown as examples to understand the challenges indeclinable mapping in SHMT:

Indeclinable *aṅga* is used independently as *sambodhana*. Example - *aṅga kvacitkuśalī tāta*. Sometimes it is used for respect. As- *aṅga vidvan māṇavakaṃ adhyāpaya = he paṇḍit! māṇavaka ko paḍhāen*. Sometimes it is used with *kiṃ* to give the sense of *kiṃ punaḥ* or *kimuta*. As- (*trṇena kāryaṃ bhavatiśvarāṇām*) *kimaṅga vāgghastavatā nareṇa* (*pancatantra* 1/1) = *vāṇī aura hāthon se yukta manuṣya kī bāta hī kyā?*

⁵ List given in *siddhānta-kaumudī* is given separately in apendix-I.

⁶ Indeclinable classification is given in apendix-II

⁷ See, VS Apte (1993) Sanskrit Syntax: the students guide to Sanskrit Composition (Hindi translation by Dr. Umeshchandra Pandey, pp-165-126.)

Similarly different meaning of *atha* is found in sanskrit texts. It is used in the following sense: *aba/isake bāda, taba/usake bāda, praśnārthe, samuccayārthe, saṁśayārthe, saṁpūrṇārthe, yadiyarthe* in Hindi. As:

athāto brahmajijñāsā= aba/isake bāda brahmajijñāsā (ārambha hotā hai),

atha prajāñāmadhipaḥ prabhāte vanāya dhenuṃ mumoca = taba prajā ke svāmī ne prātaḥkāla gāya ko vana jāne ke liye choḍā,

atha kim= isase kyā?

bhīmo 'thārjunaḥ=bhima aura arjuna,

śabdo nityo'thānitya= śabda nitya hai yā anitya hai?,

atha maraṇamavaśyameva jantoḥ= yadi jīvon kā maraṇā niścita hai,

atha is used with *vā* in the sense of *vikalpa (alternate)*.

These are the different meaning of one indeclinable. The other indeclinable words are also used in different sense.

smṛtikārainna etāvat eva viramyate (smṛtikāra itane meṇ hī nahīn rukate) and *smṛtikārainna etāvat eva kathyate (smṛtikāra itana hī nahīn kahate)* shows the different translation because of *hi* insertion in first one.

The variation in meaning of indeclinable shows the challenges in its mapping. As these words are used in NPs, the task of NP mapping becomes challenging. Making any rule for handling different meanings of these indeclinable words can only be done after VP study as the example of *eva* shows the difference in translation because of change in verb.

3.7 Challenges in mapping compound words

In compound nominal mapping, first we have to process the compound to get the *vigrahavākya*. This *vigrahavākya* or *samāsa vighraha* can be translated to clarify the meaning of the given Sanskrit text. In Sanskrit, there are six types of *samāsa*⁸.

avyayibhāva: In *avyayibhāva*, compound nominal is formed with the help of indeclinables i.e. the indeclinable pada is *pūrvapada* and it is *pradhāna* (head which governs the compound) in *avyayibhāva*.

tatpuruṣa: In *tatpuruṣa*, *ūtṭarapada* is *pradhāna* and *pūrvapada* has different case markers. After getting *vigrahavākya*, here also case marker handling process is adopted.

karmadhāraya: In *karmadhāraya* compound, generally, *pūrvapada* is an adjective and *ūtṭarapada* is a nominal head. Here adjective should be identified first, then the case marking of substantive should be handled. (Handling adjective will be discussed in section 3.8).

bahuvrīhi: In this compound section, the meaning of compound word should be taken either by translating the *vighraha-vākya* or by its *rūḍha* meaning. Here, in the *vighraha-vākya* of any given *bahuvrīhi*, mostly the relative clause will be found. (If there will be any verb in the *vighraha-vākya*, that can be handled after the Verb Phrase (VP) study. That portion is leaved aside as present study is not handling the VP.)

dvandva: Compound nominal, *dvandva*, is used in NP which is sometimes added by indeclinable *ca*, *tathā* or sometimes dual/plural number is used to form a compound Nominal. For example, *rāmaḥ śyāmaḥ ca*, *rāmaśca śyāmaśca*, *rāmaśyāmau* etc are the different ways to write a compound nominal in a NP. Here single and double *ca* in the *vighraha-vākya* will require extra case as the conjunct in Hindi is neither kept at the end of the combining words nor used twice to combine the two given words. *bhrātarau*, *pitarau* are some compound words which are used to denote a male-

⁸ *dvandvosmi dvigurapi cāsmi, gṛhe ca me satatamavyibhāvashḥ, tatpuruṣa karmadhāraya yenāhaṃ syāṃ bahuvrīhiḥ.*

female pair but it looks like the dual form of *bhrāṭṛ* and *pitṛ*. So there is also the chance of ambiguity in its translation.

***dvigu*:** In *dvigu* compound, the *pūrvapada* is a numeral value which we get in *vigrahavākya* after processing the compound. Keeping numeral adjectives aside, the case marking of nominal pada may be done similarly as discussed earlier in the case handling section. Although numeral modifier in *vigrahavākya* generally drops the case marker, being the adjective of nominal head, it will also be discussed under adjectival mapping.

3.8 Adjective and challenges in its mapping

3.8.1 Adjective: Dash and Gillon (1995)⁹ say that there are two ways to categorize words: on the basis of properties intrinsic to words and on the basis of relations which words bear to be other words, or even extra grammatical entities. First way leads to the categories of noun, adjective, verb, preposition etc.; while the second way leads to such categories as predicate and modifiers. Sanskrit, being an inflectional language, requires essentially morphological criteria for determining membership in lexical categories. While noun and adjectives, both take declensional inflection, noun have gender intrinsically, while adjectives do not.

One element modifies another when the two elements are co-constituents in a constituent of which the later element is the head. It turns out that, in general, nouns, adjectives and prepositions can be modifiers in compounds. In every compound in Sanskrit which is not a co-ordinate, the non-head element modifies the head element. Even in phrase structure both adjectives and nouns can be modifiers. Adjectives agree with the nouns they modify. There is of course no agreement between nouns. Structural conditions for agreement of adjectives with nouns are quite precise: an adjective agrees with a noun in case, number and gender when the adjective is the head of an adjective phrase which modifies the noun (where the relation of

⁹ The Adyar Library Bulletin, pp. 285-294, 1995

modification is taken in its unextended sense). Relation of predication is the relation which the verb phrase of a clause bears to the subject noun phrase of the same clause. In *rāmaḥ sītāṃ paśyati*, *sītāṃ paśyati* is the predicate of the subject *rāmaḥ*. But in *strī sundarī*, the adjective *sundarī* is a predicate of the noun *strī* and there is no verb, in this condition, assuming null copula in absence of VP, we can say that the adjective *sundarī* is a predicate of the noun *strī*. Dash and Gillon (1995) present empirical support to the assumption of a phonetically null copula and shows that the adjective agrees with a noun of which it is a predicate or modifier (in phrase structure).

Adjectives (modifier) seem to be noun in absence of modified noun. In *dīnaṃ pālaya* *dīnaṃ* is performing the role of noun as this sentence contains a phonetically null noun/pronoun. Null noun/pronoun serves as a head noun in a NP in which the bare adjective is modifier. Thus Dash and Gillon (1995) conclude that a noun is provided for the adjective to modify, thereby supplying a noun phrase to be the subject of the sentence. They further give a template for this type of construction saying that these may be considered as compound. In this template, [N {Adj.} {Nominal}], head of such compound is a noun, so the compound itself has the status of the head.

Dr S D Joshi did the classification of nouns into adjectives and substantives in his paper titled *Adjectives and Substantives as a single class in the "Parts of Speech"*¹⁰. The structure of Sanskrit is such that it makes a clear distinction between verbs and nouns which can be defined semantically as well as morphologically. A verb denotes activities in process which consists of definite sequence of beginning middle and end. Nouns, on the other hand, do not denote a process but frozen actions in the form of a substance...Particles serve to reveal the relation existing between two different words, whereas the prepositions serve to specify the meaning of the verb. Grammatically *upsargas* are always connected to verbs while *nipātas* shows relation between the different words. The distinction between *upsarga* and *nipātas* is not structural or morphological but it is functional. This functional classification corresponds to the grammatical fact that prepositions are always connected with verbs in the sentence and bring out inherent signification of verbs whereas the *nipātas* are regarded as

¹⁰ Journal of the University of Poona, vol.25, pp.19-30, 1966.

word-connectives or sentence connectives....this division of words is based on the combined aspect of form, meaning and function, and it is suitable to the structure of Sanskrit.

According to Pānini, prepositions, particles, adverbs, adjectives fall under the category *nāman* or *prātipadika*. In his descriptive categorization, Panini includes *upsarga* and *nipāta* under the single category 'noun'. As mentioned earlier, noun presents the static notion and verb presents the process of happening, thus all adjectives, pronouns, conjunctions and indeclinables are noun. The adverbs are grouped under the category of nouns, because structurally they have similar forms with nouns. To Bhartṛhari, pronouns might be a subclass of nouns or adjective. It is either restrictive adjective or stands for the things in general. Thus they function either like adjectives or like noun, and when they restrict the sense of substantive, they are adjectives.

The adjectives and substantives have the same inflection in Sanskrit. The chief difference one may note is that the former class varies in three genders, as it is shown by their agreement with substantives. The substantives have definite gender¹¹. But there are some exceptions pointed out in *vyutpattivāda* where *vedāḥ pramānaṃ, śataṃ brāhmaṇāḥ* etc are shown as examples. Here we find the disagreement between adjectives and nouns in either their gender or number, or their gender and number.

Joshi (1996) states that Panini does not define the terms *viśeṣaṇa* and *viśeṣya* semantically or structurally but used the term in the rule II.1.57 (which means: a case inflected word standing for a qualifier is compounded with a case-inflected word standing for a qualificand). For example - *nīlotpala, raktotpala* etc. Here *nīla* and *rakta* are used as modifier which differentiate the blue and red lotuses from the white lotus etc). Further says that Patañjali also find it difficult to explain the term *viśeṣaṇa* and *viśeṣya*. The difference between *viśeṣaṇa* and *viśeṣya* lies in the point of view which we put forth. When the notion of qualifier and qualified is purely subjective with the reference to the wish of the speaker, adjectival notion can be easily turned

¹¹ *yalliṅgaṃ yadvacanaṃ yā ca vibhaktirviśeṣyasya, talliṅgaṃ tadvacanaṃ sā ca vibhaktirviśeṣaṇasyāpi. (samāsacakraṃ)*

into the substantive notion. Therefore Patañjali gives suitable terminology to point out the intrinsic difference between the adjectives and substantives. He calls adjectives *guṇavacanāḥ*: ‘denotative of qualities’, that are found in substances and terms substantives (*viśeṣya*) *dravyavacanāḥ*: ‘denotative of substances’. Panini has also used the term *guṇavacana* in sense of qualifying attribute¹². The grammarian’s term *guṇa* stands for the attribute or qualities separable or inseparable from the substances. Here the term *guṇavacana* means any attributive word which serves to distinguish one from the other which is cleared by the earlier example of *nīlotpala*, *raktotpala* etc. but this concept of *guṇavacāna* and *dravyavacana* is not useful for the *karmadhāraya* compound prescribed for the *viśeṣya-viśeṣaṇa* words because *karmadhāraya* compound is formed with the two *dravyavacana* as well as two *guṇavacana*. For example, in *āmraṅgāḥ*, we find two *dravya* *āmra* and *ṅga* respectively. Similarly *śuklakṛṣṇaḥ* has two *guṇavacana* *śukla* and *kṛṣṇa* respectively. The terms *guṇavacana* and *dravyavacana* have definite meaning and they cannot be normally interchangeable. When the notion of *viśeṣya-viśeṣaṇa-bhāva* is purely dependent upon *vivakṣā*, a classification of adjectival words into *guṇavacana* and *dravyavacana* can handle the non-compound construction in language processing. The grammatical fact of agreement or concord between the adjectives and substantives is well brought by Patañjali by classifying them logically as denotative of qualities and substances.

Joshi (1996) also presents the views of Bhartṛhari and Helārāja from *Vākyapadīya*. The *guṇa* cannot be conceived without *guṇī*. Here the *guṇa* is modifier and thus, Helārāja says: *bhedakatvaṃ cātra mukhyaṃ guṇalakṣaṇam*. Bhartṛhari, while discussing the term *viśeṣya* and *viśeṣaṇa*, pointed out that *viśeṣya-viśeṣaṇa* is a syntactic category and not a morphological or grammatical one. The designatives *viśeṣya* and *viśeṣaṇa* refer to a word as a member of combination and not as an isolated individual. If *nīla* and *ghaṭa* are not put together in combination but used separately, they only denote the nominal notion. But when it is used in combination, *nīla* becomes the modifier of the *ghaṭa*.

¹² Panini II.1.30;VI.2.115.

It is also presented by Joshi (1996) that the modifier is used not only for differentiation but also for the identification. In *prameyo ghaṭaḥ*, the universal attribute *prameya*, 'knowledge', cannot eliminate anything because there is no unknowable worldly object. In such case adjectives do not indicate the difference but simply identify the object with its qualifier. The qualifying word differentiates the individuals of the same class and not individuals of the other classes. In *nīlo ghaṭaḥ*, *ghaṭa* is representing the whole class of *ghaṭa* and *nīla* is differentiating it from the other possible *colours*.

So we conclude that 1) the main-difference between adjectives and substantives is that the substantives have a fixed gender while the adjectives vary in gender and number following the substantive character. And 2) the adjective is subordinate while the substantives are primary.

3.8.2 Challenges in adjective mapping: After a conceptual discussion, we can move towards few examples which are related to the identification of adjectival *pada*, the problem of case handling in adjectives and gender aspect affecting the adjective.

3.8.2.1 The identification of adjectival pada: In Sanskrit, except some compound nominals, the place of adjective is not fixed. In compounding, the adjective and substantive is observed by the commentators on the basis of *vācakatva*. If *jātivācaka/samjñāvācaka* word is kept with the *guṇavācaka* and *kriyāvācaka* then *guṇavācaka* and *kriyāvācaka* will be the adjectives in those compounds. In *nilotpalaḥ*, *nilaḥ* is qualitative and hence this is the adjective of *utpala* which is *jātivācaka*. In *pācakabrāhmaṇaḥ*, *kriyāvācaka pada pāchakaḥ* is the adjective of *brahmaṇa* (*jāti/samjñā*). (when *kriyāvācaka pada* will be the adjective, then Hindi will take the *wālā* construction). But when two qualitative words or the *kriyāvācaka padas* come together, there is no certain position for adjective.

Three types of nominal can function as an adjective: *subanta*, *kṛdanta* and *taddhita*. If two *subanta padas* are coming together as *samānādhikaraṇa* then anyone may be adjective. Same as, if two *kṛdanta padas* are coming together then also the place of adjective *pada* is uncertain. If *kṛdanta* and *subanta pada* comes together, then the

kṛdanta pada is adjective and *subanta* is substantive. If *kṛdanta/subanta* and *taddhita* is coming together, then *taddhita* is the adjective and the *kṛdanta/subanta* is substantive. As *taddhita* never becomes the adjective of a pronoun, when *taddhita* is used with a pronoun, pronoun will be the adjective and *taddhita* will be the substantive.

3.8.2.2 Mapping of adjective: There are two issues in adjective mapping: the case marking and the word order. The first difference between Sanskrit and Hindi adjective is that an adjective and a substantive in Sanskrit have the same case marker while adjective in Hindi does not show the case marker except oblique forms in pronominal adjectives and the adjectives with ā-ending. As- *tena bālakena kathitaṃ= usa bālaka ne kahā*. In *madhuraṃ phalaṃ dadāti*, adjective and substantive has the same case marker in Sanskrit, but in *mīthā phala detā hai*, adjective *mīthā* has no case marker. But the plural of *madhuraṃ phalaṃ*, *madhurānī phalānī (dadāti)*, (*mīthe phaloṃ ko detā hai*), has oblique form with the adjective in Hindi. In *sundaraḥ bālakaḥ asti*, adjective and substantive has the same case marker in Sanskrit, but in *sundara bālaka hai*, adjective *sundar* has no case maker. So here we see that the case marker is dropped. The NP *sushilā bālikā* can be translated as *sushilā bālikā* as well as *sushila laḍakī*. Here it can be noticed that when *bālikā* word is not changed with its synonyms, we can keep *sushilā* as an adjective of *bālikā*. But *sushilā laḍakī* is not used in Hindi, instead of *sushila laḍakī*. In Hindi, *acchī laḍakī* cannot be pluralized as *acchiyān laḍakīyān* instead of *acchī laḍakīyān*. So the claim of same number with adjective and substantive cannot be true in Hindi while this is applied in Sanskrit. We can see another example: *teja laḍakā* and *teja laḍakī*, where the gender of adjective word is not affected by the gender of substantive. As it is mentioned earlier that the adjectives and substantives have the same inflection in Sanskrit, the chief difference one may note is that the former class varies in three genders, as it is shown by their agreement with substantives. The substantives have definite gender¹³. But there are some exceptions pointed out in *vyutpattivāda* where *vedāḥ pramānaṃ*, *śataṃ brāhmaṇāḥ* etc are shown as examples. Here we find the disagreement between

¹³ *yalliṅgaṃ yadvacanaṃ yā ca vibhaktirviśeṣyasya, talliṅgaṃ tadvacanaṃ sā ca vibhaktirviśeṣanasyāpi. (samāsacakram)*

adjectives and nouns in either their gender or number, or their gender and number. So these challenges need to be address to get the correct output.

The next issue in mapping adjective is the place of adjective. In Sanskrit, adjective has no fixed place. In Sanskrit, it can be kept either before or after the substantive. In Hindi, adjective is kept before the substantive (i.e. Adjective + Noun) but the adjectival clause has no fixed place. So leaving the adjectival clause aside, we should focus on the adjective (non-adjectival clause) where we have, first, to identify the adjective in SL which will be arranged according to the Hindi output. One more thing can be notices that *bālakadvaya* types of constructions have numeral adjective where adjectives are attached with the substantive and making a single word. Here we have to separate the adjective from the substantive to handle the output.

3.8.2.3 Challenges in mapping conjuncts: In dvandva compound, the conjunct ca has no fixed number. Sometimes one ca can come at the end of substantives; sometimes multiple ca can be used i.e. ca with each substantive (depending on the number of substantive) or sometimes compound can take dual or plural number. For example-

- a. *rāmaḥ lakṣmanaḥ ca paṭhataḥ.*
- b. *rāmaśca lakṣmanaśca paṭhataḥ.*
- c. *rāmālakṣmanau paṭhataḥ.*

In their translation, Hindi takes only one conjunct at a fixed place i.e. just before the last word of the compound. This difference needs to be mapped in MT.

3.9 Challenges in Mapping Gender and Number

As Durga Singh stated on the rule *lokopacārāt grahaṇasiddhiḥ- śabdānāmekārthopi liṅgavacanabhedāḥ dārā kalatram bhāryeti*. According to Mahābhāṣya, the gender

has been described as *stanakeśavatī strī syāllomaśaḥ puruṣaḥ smṛtaḥ*¹⁴. Here the characters of male and female are indicated to differentiate the masculine and feminine gender. The problem increases in the identification of any non-human object where grammar is unable to explain the specific gender of a given object and the linguist tries to find out the solution from the cultural identity of the given object. In computational processing, when the gender of a given word (input) is changing during the translation in TL, the case marker gets changed. It specially affects the genitive case marker while it produces the output in Hindi (TL).the word *dārāḥ* is masculine plural, but having the meaning of wife, it is always applied to female singular. Here this is to be noticed that Panini says that the concord of gender and number of primitives derivative nouns, and of the attributes and substantives, need not to be taught or approved because it has the authority of conventional term or idiom. If the gender of the given object is changed in the output will also affect the *kr̥danta* verb-form in Sanskrit and thus this will lead to the problem of mapping VP as well.

As it has shown in the section 3.2, Sanskrit has three numbers but Hindi dropped its dual number. The dual form is generally changed into plural in Hindi. So handling the dual in Hindi needs the rules from the perspective of computational processing. It is noticed that, sometimes, Singular and plural, both gives the same meaning. As- *ahaṃ bravāmi* and *vayaṃ brūmaḥ* both are translated as 'I speak' by the *sutra asmado dvayośca* (P.I.2.59) where the plural of the pronoun is used optionally, though the sense requires a singular. The expressions like *dvayaṃ*, *trayaṃ* etc (when attached with a noun) having the meaning of dual and plural numbers take the form of singular in Sanskrit but are plural in Hindi. The *dvigu* compound has the singular form in Sanskrit. *dvandva* compound too is singular in Sanskrit. Even the meaning of *ahinakulaṃ*, *govyāgrāmaṃ* and *kākulokaṃ* is plural but they also take the singular number. Thus handling of the number is also an important issue in SHMT.

¹⁴ Mahābhāṣya, vol-II. P-196.

3.10 Challenges in Mapping Word Order

It is already discussed in second chapter that word order is an important issue in getting the meaning correctly. In the *ambākarṭṛka* comentory on Vākyapadīya (2/49), importance of *pada*-sequence is written as- “*na kramavyatiriktaṃ vākyam nāma vākyārthasyābhidhāyakam*”¹⁵. The meaning of a sentence will be changed by the alteration in sequence of *padas* in a sentence. As - *ahaṃ eva gacchāmi* (i.e. *main hī jātā huṃ*) → *ahaṃ gacchāmi eva* (*main jātā hī huṃ*). But this is also true that Sanskrit enjoys the liberty of being inflectional language. *vidyā buddhiṃ apekṣate* and *buddhiṃ vidyā apekṣate* has the same meaning. Despite of being free word order language, one characteristic in Sanskrit is clearly shown by Dr. Ashoka Aklujkar in his paper¹⁶ that when the number of constituents in a sentence is lesser, the probability of scrambling is the higher. Thus we can see the word order in Sanskrit in a longer sentence where word order cannot be changed arbitrary. It proves that Sanskrit has also the SOV word order which can be obtained by doing *anvaya* of given Sanskrit text. The theory of *anvaya* is given as- ‘*viśeṣaṇaṃ puraskṛtya viśeṣyaṃ tadanantaraṃ, karṭṛ-karma-kriyā-yuktaṃ etadanvaya-lakṣaṇaṃ*’¹⁷. It is clear that an adjective should be kept before the substantive and further the sequence of Subject, object and verb should be followed in a general statement. The same thing is cleared in these sayings: ‘*ādau karṭṛ-padaṃ vācyam dvitīyādi padaṃ tataḥ, ktvā-tumun-lyap ca madhye tu kuryādante kṛyāpadaṃ*’¹⁸. When the input text will be changed in this sequence (SOV), the output will be more accurate in TL.

¹⁵ Sharma, Pt. Raghunath (1980), Vakyapadiyam-2/49, p-95.

¹⁶ Aklujkar, Ashoka: Some theoretical Observation on Word Order in Sanskrit.

¹⁷ Samasacakram, verse 5.

¹⁸ Aklujkar, Ashoka: Some theoretical Observation on Word Order in Sanskrit.

Chapter-4

***TRANSFER GRAMMAR FOR SANSKRIT-
HINDI NP MAPPING***

4.1 Introduction

In this chapter, transfer approach has been discussed to handle the challenges found in Sanskrit-Hindi NP mapping in Machine Translation (MT). The challenges are divided into two levels: first the challenges which can be handled at generation level and second which can be handled at transfer level for example, supplying ergative 'ne', drop of 'ko' in certain cases in *karma vibhakti* because of animate inanimate effect on object, agreement between the adjective and a noun (because of drop of the *vibhakti* of an adjective in Hindi) are all taken care of by the Hindi generator. The challenges at generation level can be handled by improving the Hindi generator as generator part of SHMT is the same as that of the Sampark because it follows Sampark (ILMT) architecture which is based on analyze-transfer-generate paradigm. First, analysis of the source language text is done, then a transfer of vocabulary and analyzed structure to target language is carried out, and finally the target language text is generated. The Sampark pipeline has a simple parser, however, for SHMT a deeper level parser is being developed¹. Handling of *vibhakti* divergences, indeclinable mapping and handling the word order difference between Sanskrit and Hindi are the main problems in the TG. Here the task of Hindi generator and rules for handling challenges at transfer level (written as theoretical findings) are discussed. TG rules can be presented in SSF format of IIIT-Hyderabad to improve the output in Hindi.

4.2 Task of Hindi generator

As the challenges in Sanskrit-Hindi NP mapping are discussed in the third chapter, the following points will be examined briefly to decide what falls under the generation task and where TG rules are required:

- *Vibhakti* divergence and challenges in case mapping
- Challenges in indeclinable mapping

¹ http://sanskrit.uohyd.ernet.in/~scl/Reports/SHMT_Jan2011/shmt-report-jan2011.pdf (accessed on dated 7/7/2011 at 4:00PM.)

- Challenges in mapping compound nominal
- Challenges in mapping adjectives
- Challenges in mapping gender and number
- Challenges in mapping word-order

In *Vibhakti* divergence and challenges in case mapping, following task may fall under Hindi generation: 1) ergative insertion, 2) handling of *ko* case marker, 3) handling of instrumental *se* and *ke dvārā*, 4) handling of genitive case markers in output and 5) handling of Locative case markers *se* and *menī*. The challenges in indeclinable mapping will not fall under Hindi generation because some indeclinables are getting dropped in Hindi translation as they are *pādapūrṇārthaka*. So it will be handled by TG rules. The Challenges in mapping compound nominal will also be handled by TG rules as conjunct *ca* handling as well as adjective handling requires TG rules as compound nominal possess conjunct and adjectives. Gender mapping should be handled by Hindi generator because the gender of a given word is changing in TL (i.e. the gender of *dārā* or *agni* is changing in Hindi), so Hindi generator should function according to the (changed) gender of the TL word (there Hindi Generator must check the gender of the TL word from the lexicon). As Hindi has no dual number, the number mapping will be handled by TG rules. Lastly the word-order will fall under the TG rules as the SL syntax may not be presented in SOV order.

Let us start with the issues of Hindi generator. In ergative handling, the transitivity of the verb should be checked because intransitive verb does not allow an ergative ‘*ne*’. For handling *ko* case marker, animate-inanimate object should be checked because singular inanimate generally does not take *ko* case marker accept in-the case of emphasis. These tasks can be done by improving Hindi generator which will check the conditions of transitive-intransitive verbs and animate-inanimate objects.

In instrumental case handling, challenges can be found in mapping double causative construction such as- (1) *rāmaḥ devadattena gṛhakāryaṃ kārayati* (*rāma devadatta se gṛhakārya karawātā hai*) and (2) *rāmaḥ śyāmena mohanena putraṃ pāṭhayati* (*rāma śyāma ke dvārā mohana se putra ko paḍhawātā hai*). The sentence (1) can be

translated in Hindi by using post-position *se* or *ke dvārā* with *devadatta* while (2) indirect object (IO) 1 will take *ke dvārā* and the IO2 will take *se*. If (2) is presented as *anvita* text (i.e. presented in a proper word order), then *mohana* is the IO2 and it should take *se* and, then, *shyāma* (IO1) will take the case marker ‘*ke dvārā*’ because the second sentence cannot have *se* case marker twice because the sentence (3) **rāma śyāma se mohana se putra ko padhawātā hai* will be odd and may not be understood

As the genitive case marking depends on the gender aspect in Hindi, it is found that the Sampark output is presenting relations correctly². But in handling locative case marker, the difficulties are felt. As- *vṛkṣeṣu pakṣiṇaḥ santi* (*vṛkṣa para pakṣī haiṅ*) and *kṣetreṣu bālakāḥ krīḍanti* (*maidāna me bālaka khelate haiṅ*) have the outer locations but the case markers are different. It is not fixed that the case marker *meṅ* and *para* will always be used to indicate the inner and outer locations respectively. *deveṣu bhāramasti* (*devoṅ para bhāra hai*) and *deveṣu sāmartyamasti* (*devoṅ meṅ sāmartya hai*) have locative case marker with *deva* but their Hindi translations have *para* and *meṅ* case markers respectively. So it depends on the nature of the objects also. To handle the locative case markers in Hindi output, Hindi generator should be improved.

In gender mapping, the genitive case marking issue needs to be handled at NP level. At the VP level, *tiṅanta* form of verb in Sanskrit is not dependent on gender but the *krīḍanta* is. So the output at the VP level will also be improved. As- *dārāḥ gataḥ* is the Sanskrit Sentence which has *gataḥ* verb (which is used with masculine gender). When *dārāḥ* is translated as *patnī*, *gataḥ* must be translated according to the gender of the (TL) agent. So improving the Hindi generator will be helpful in producing correct output.

² Only some *kriyāmūla* verbs (which have the verb + *karanā/honā* form in Hindi) have problems in handling case markers which depends on verb selection in Hindi. For example - *devaṃ pūjati* may be translated as *deva kī pūjā karatā hai* or *deva ko pūjatā hai*. At this stage, these types of problems will not be handled as this work is not covering VP issues.

4.3 Transfer grammar

In machine translation, a set of Transfer Grammar (TG) Rules are required for mapping syntactic representations of a source language into the target language representations. The TG engine is developed at IIIT, Hyderabad which deals with transferring the sentence structures of one Indian language (IL-1) to another Indian language (IL-2). The module uses a set of rules which give corresponding structures of IL-1 – IL-2. In that module, necessary TG rules for Sanskrit-Hindi NP mapping can be added to get the desired output in SHMT. At the level of TG, basically we will be dealing with the syntactic differences. The very first and most important task is to handle the case marker which is related with *vibhakti* divergence and the challenges coming in case mapping. Mapping compound nominal and conjunct *ca* handling, adjective handling, dual number handling and word-order handling are the task of TG at NP level.

4.4 TG rules for case mapping

Sub-classification of *kāraḥ* may give us the idea about the possible case marking in a particular condition. On the basis of sub-classified *kāraḥ* which is well arranged by K V Ramkrishnamacharyulu in his paper titled *Annotating Sanskrit Texts Based on śābdabodha Systems*³, TG rules for case handling can theoretically be presented here.

4.4.1 Division of *kartā kāraḥ* and TG rules:

- 1) Default *kartā*: as- **devadattaḥ** *pacati* (*devadatta pakātā hai*) *anubhavī kartā*: as- **ghaṭaḥ** *naśyati* (*ghaṭa naṣṭa hotā hai*).
- 2) *amūrta kartā*: as- **krodhaḥ** *āgacchati* (*gussa ātā hai*)
- 3) *prayojaka kartā*: as- **devadattaḥ** *viṣṇumitreṇa pāchayati* (**devadatta** *viṣṇumitra se pakawātā hai*)

³ A. kulkarni and G. Hute (Eds.): Sanskrit Computational Linguistics, LNCS 5406, pp.26-39, 2009 @Springer-Verlag Berlin Heidelberg 2009.

- 4) *prayojya kartā*: as- *devadattaḥ viṣṇumitreṇa pāchayati* (*devadatta viṣṇumitra se pakawātā hai*)
- 5) *madhyastha kartā*: as- *devadattaḥ yajñadattena viṣṇumitreṇa pāchayati* (*devadatta yajñadatta ke dvārā viṣṇumitra se pakawātā hai*)
- 6) *abhipreraka kartā*: as- *modakaḥ rocate* (*modaka acchā lagatā hai*).
- 7) *karma karṭr*: as- *kāṣṭhaḥ bhidyate* (*svayameva*) (*lakaḍī khuda hī phaṭatā hai*)
- 8) *karaṇa- karṭr*: as- *asiḥ chinatti* (*kulhāḍī kāṭatī hai*)
- 9) *ṣaṣṭhī-kartā*: as- *ācāryasya anuśāsanam* (*ācārya kā anuśāsana*)

TG rules:

1. In *kartā kāraka*, default *kartā* and *prayojaka kartā*, drop the case marker except past perfect sentence where it takes ergative *ne*.
2. In *anubhavī*, *amūrta*, *utpreraka kartā*, *karma-kartā* and *karaṇa-kartā*, case marker is dropped.
3. *Prayojya kartā* takes *se* case marker.
4. *madhyastha kartā* takes *ke dvārā* case marker.
5. *ṣaṣṭhī-kartā* takes the genitive case marker- *kā*, *ke* and *kī* depending on the gender of substantives (i.e. *kā* with masculine, *ke* with honor and with masculine plural and *kī* with feminine. In Hindi, neuter gender is generally changed in masculine or feminine.)

4.4.2 Division of *karma kāraka* and TG rules:

1. Default *karma*: as- *śatrūn jayati* (*śatrū ko jītatā hai*), *odanaṃ bhunkte* (*bhāta khātā hai*)
2. *utpādyā karma*: as- *odanaṃ pacati* (*bhāta pakātā hai*), *putraṃ prasūte* (*putra ko janma detī hai*)
3. *vikārya karma*: as- *svarnaṃ kuṇḍalaṃ karoti* (*sona se kuṇḍala banātā hai*)
4. *karma* in *prayojya kartā*: as- *bālaṃ kṣīraṃ pāyayati* (*bālaka ko dūdha pilātā hai*)
5. *karma* in locus: as- *gṛhaṃ adhiśete* (*ghara meṅ sotā hai*)
6. *karma* in *deśavācī* word: as- *kurūn svapiti* (*kurū deśa meṅ sotā hai*)
7. *karma* in time: as- *māsaṃ āste* (*māha bhara rahatā hai*)

8. *karma* in *bhāva*: as- *godohaṃ āste* (*godohanakāla taka rahatā hai*)
9. *karma* in *mārga* (road measurement): as- *krośaṃ āste* (*kośa bhara (dūra) hai*)
10. *karma* in *sampradāna*: as- *paśunā rudraṃ yajate* (*paśu se yajña kī pūjā karatā hai*)
11. *karma* in unintended: as- *grāmaṃ gacchan tṛṇaṃ sprśati* (*gāva jāte hue tṛṇa chūtā hai*)
12. *karma* in *akathita*: as- *gopaḥ gāṃ dogdhi payaḥ* (*gvālā gāya kā dūdha dūhatā hai*)
13. *karma* in *gati-karma*: as- *grāmaṃ gacchati* (*gāva jātā hai*)
14. *karma* in instruments of playing: as- *akṣān dīvyati* (*pāsoṅ se khelatā hai*)
15. *karma* in anger: as- *krūraṃ abhikrudhyati* (*krūra para krodha karatā hai*)
16. *karma* in disrespect: as- *na tvāṃ tṛṇaṃ manye* (*tumako/tumheṅ tinkā bhī nahīn samajhatā hun*)
17. *ṣaṣṭhī karma*: as- *śabdānām anuśāsanam* (*śabdoṅ kā anuśāsana*)

TG rules:

- 1) In default *karma*, singular and plural animate objects and plural inanimate object takes *ko* case marker in Hindi. Singular inanimate object drops the case marker except the condition of emphasis.
- 2) In *utpādyā karma* also, inanimate has no case marker and animate has *ko* case marker.
- 3) In *vikāryā karma*, *se* case marker comes in Hindi.
- 4) When *karma* has been used with *prayojya kartā*, it takes *ko* case marker.
- 5) When *karma* is used with locus, generally it takes *meṅ* case marker for inner locus and *para* case marker with outer locus. Sometimes it is used as vice versa.
- 6) When *karma* is used with *deshvācī śabda*, *meṅ* case marker is used in Hindi.
- 7) When *karma* is used with *kālavācī śabda*, *taka* case marker is used in Hindi.
- 8) When *karma* is used with *bhāva* (activity), *taka* case marker is used in Hindi.
- 9) When *karma* is used with *mārgavācī śabda*, *bhara* is used in Hindi.
- 10) When *karma* is used with *sampradāna*, *se* case marker is used in Hindi.

- 11) When object is not intimated then animate and plural inanimate object takes *ko* case marker and singular inanimate drops the case marker in Hindi.
- 12) The *akathita karmas* take different case markers:
- a) *gām gogdhi payaḥ* has *akathita karma* 'gām' which is taking *se* case marker as it may be translated as:- *gāya se dūdhā dūhatā hai*.
 - b) Root *yāc* and *pac* takes *se* case marker. As- *balim yāchate vasudhām* takes *se* case marker in Hindi (as- *bali se vasudhā kī yācanā karatā hai*) and *taṇḍulān odanaṃ pacati* takes *se* case marker (as *cāwala se bhāta pakātā hai*).
 - c) Root *daṇḍa* takes *ko* as well as *para* in Hindi translation as *gargān śataṃ daṇḍayati* is translated as *garga ko sau rupaye kā daṇḍa detā hai / garga para sau rupaye kā daṇḍa lagātā hai*. Here we can see that the case marker is depending on the verb selection. When *detā hai* is used as verb, *ko* comes as case marker and when *lagātā hai* is used then *para* case marker comes in Hindi translation.
 - d) Root *rudh* takes *ko* case marker in Hindi (*gām vrajaṃ avarunaddhi = gāya ko bāḍe meṃ rokatā hai*).
 - e) Root *pracch* takes *se* case marker in Hindi as *māṇavakaṃ panthānaṃ pṛcchati* is translated as *bālaka se rāstā pūchatā hai*.
 - f) Root *ci* takes *se* case marker as *vṛkṣaṃ phalāni avacinoti* is translated as *vṛkṣa se phala chunatā hai*.
 - g) Root *brū* takes *se* case marker as *māṇavakaṃ dharmāṃ brūte* is translated as *bālaka se dharmā kahatā hai*. It may also be translated as *bālaka ko dharmā kahatā hai* or *bālaka ke liye dharmā kahatā hai*. Here we are taking only one case marker for TG rules which is used more frequently in Hindi translation.
 - h) Root *śās* takes *ko* case marker as *māṇavakaṃ dharmāṃ śāsti* is translated as *bālaka ko dharmā sikkhalātā hai*.
 - i) Root *ji*, *math* and *muṣa* takes *se* case marker as *devadattaṃ śataṃ jayati* is translated as *devadatta se sau rupaye jītātā hai*, *kṣīranidhiṃ sudhāṃ mathnāti* is translated as *kṣīrasāgara se amṛta mathatā hai* and *devadattaṃ śataṃ muṣnāti* is translated as *devadatta se sau rupaye curātā hai*.

- j) Root *nī* and *vah* take *meñ* case marker as *grāmaṃ ajāṃ nayati* is translated as *gāva meñ bakarī ko le jātā hai* and *grāmam ajāṃ vahati* is translated as *gāva meñ bakarī ko pahuicātā hai*.
- k) Root *hr* takes *se* case marker as *grāmaṃ ajāṃ harati* is translated as *gāva se bakarī kā haraṇa karatā hai*.
- l) Root *kṛṣ* takes *kī or* case marker as *grāmaṃ ajāṃ karṣati* is translated as *gāva kī or bakarī ko le jātā hai*.
- 13) The *gati-karma* drops the case marker as- *grāmaṃ gacchati* is translated as *gāva jātā hai*.
- 14) *Karma* of *div* root takes *se* case marker as- *akṣān divyati* is translated as *pāson se khelatā hai*. It may also drop the case marker (as in *pāsā khelatā hai*) but that is not taken in TG rule to avoid the complication of divergence issues.
- 15) *Karma* in anger takes *para* case marker as *krūram abhikrudhyati* is translated as *krūra para krodha karatā hai* in Hindi.
- 16) The *karma* in disrespect takes *ko* case marker as *na tvāṃ tṛṇaṃ manye* is translated as *tumako tinkā bhī nahīn samajhatā huñ*.
- 17) The *ṣaṣṭhī karma* takes genitive case markers as *śabdānām anuśāsanam* is translated as *śabdonī kā anuśāsana*.

4.4.3 Division of *karaṇa kāraka* and TG rules:

- 1) Default *karaṇa*: as- *bālakah kuñcikayā tālam udghāṭayati* (*bālaka cābhī se tālā kholatā hai*).
- 2) *karaṇa* in *karma*: as- *paśunā rudraṃ yajate* (*paśu se yajña kī pūjā karatā hai*).
- 3) *karaṇa* in *parikrayaṇa*: *śatena parikrīnāti* (*sau rupaye meñ parikrayaṇa karatā hai*).

TG rule: The *karaṇa kāraka* is taking only *se* case marker in these sense.

4.4.4 Division of *sampradāna kāraka* and TG rules:

- 1) *satvāśraya* (recipient with ownership): as- *devadattaḥ brāhmaṇāya gāṃ dadāti* (*devadatta brāhmaṇa ko gāya detā hai*)

- 2) *svīkartā* (recipient without ownership): as- *devadattaḥ rajakāya vastraṃ dadāti* (*devadatta rajaka ko vastra detā hai*)
- 3) *kriyayā abhipretaḥ* (intended to relate with activity): as- *patye śete* (*pati ke liye sotī hai*)
- 4) *jñīpsyamānaḥ* (addressed through price etc): as- *kṛṣṇāya ślāghate* (*kṛṣṇa kī praśānsā karatā hai*)
- 5) *uttamarṇa* (a creditor): as- *devadattāya śataṃ dhārayati* (*devadatta kā sau rupayā dhāratā hai*)
- 6) *īpsitaṃ* (desirēd): as- *puṣpebhyaḥ sprhayati* (*puṣpoṅ kī icchā karatā hai*)
- 7) *yaṃ prati kopaḥ saḥ*: as- *haraye krudhyati* (*hari para krodha karatā hai*)
- 8) *prīyamānaḥ* (location of desired): as- *devadattāya modakaḥ rocate* (*devadatta ko modaka acchā lagatā hai*)
- 9) *yasya vipraśnaḥ* (enquiry about): as- *kṛṣṇāya rādhyati* (*kṛṣṇa ke śubhāśubha kā vicāra karatā hai*)
- 10) *parikrayaṇa* (money in bounded labour): as- *sevakaḥ śatāya parikrītaḥ* (*sevaka sau rupaye meṅ rakhā gayā*).

TG rules:

1. *satvāśraya*, *svīkartā* and *prīyamānaḥ* take *ko* case marker.
2. *kriyayā abhipretaḥ* takes *ke liye* case marker.
3. *jñīpsyamānaḥ* and *īpsitaṃ* take *ki* case marker.
4. *uttamarṇa* takes *kā* case marker.
5. *yaṃ prati kopaḥ* takes *para* case marker.
6. *yasya vipraśnaḥ* takes *ke* case marker.
7. *parikrayaṇa* takes *meṅ* case marker.

4.4.5 Division of *apādāna kāraṇa* TG rules:

- 1) default *apādāna* (point of separation): as- *vṛkṣāt paṇaṃ patati* (*vṛkṣa se patte girate haiṅ*)

- 2) *bhaya hetuḥ* (cause of fear): as- *gr̥hasthaḥ caurāta vibheti* (*gr̥hastha chora se daratā hai*)
- 3) *ākhyāta-upayoge* (teacher): as- *chātraḥ upādhyāyāt adhīte* (*chātra upādhyāya se padhatā hai*)
- 4) *yasmāt vāraṇam* (point for obstruction): as- *kūpāt andhaṃ vārayati* (*kūen se andhe ko nikālatā hai*)
- 5) *yasya adarśanam iṣṭam* (person intended not to be seen): as- *mātuḥ nilīyate kṛṣṇaḥ* (*mātā se kṛṣṇa chipatā hai*)
- 6) *prakṛtiḥ* (raw material): as- *mṛdaḥ ghaṭaḥ jāyāte* (*mittī se ghaḍā banatā hai*)
- 7) *prabhavaḥ* (place of first appearance): as- *himavataḥ gaṅgā prabhavati* (*himālaya se gaṅgā nikalatī hai*)
- 8) *parājayah* (defeat from activity): as- *adhyayanāt parājayate* (*padhāi se hāratā hai*).

TG rule: *apādāna kāraka* is taking only *se* case marker.

4.4.6 Division of *adhikaraṇa kāraka* and TG rules:

- 1) *kālah* (time): as- *tretāyuge rāmaḥ āsīt* (*tretāyuga meṅ rāma the*)
- 2) *deśaḥ* (place): as- *rāmaḥ ayodhyāyām āsīt* (*rāma ayodhyā meṅ the*)
- 3) *viśayaḥ* (other than above): as- *mokṣe icchā asti* (*mokṣa meṅ ruche hai*)
- 4) *samayāvadhi* (time duration): as- *janavarītaḥ virāmaḥ* (*janavarī se virāma hai*)
- 5) *antarāla deśaḥ* (place in between): as- *tirupatitaḥ candragiriparyantaṃ bhavanāni santi* (*tirupati se candragiri taka bhavana haiṅ*)

TG rules:

1. *kāla*, *deśa* and *viśaya* are taking *meṅ* case marker.
2. *samayāvadhi* and *antarāla deśa* are taking *se* case marker.

4.4.7 The non-*kāraka* relations and TG rules:

- 1) *sambodhana* (addressed): as- *bho rāma! māṃ uddhara* (*he rāma! merā uddhāra karo*)

- 2) *prasajya-pratiśedha* (uncompounded relation): as- *rāmaḥ grhaṃ na gacchati* (*rāma ghara nahīn jātā hai*)
- 3) *sāmyam* (similarity): as- *brāhmaṇavat adhīte* (*brāhmaṇa kī bhānti paḍhatā hai*)
- 4) *kriyā-āvr̥tṭyantarālasamayaḥ* (time duration between the repetition of the same activity): as- *adya bhuktvā dinadvayāt bhoktā* (*āja khākara do dina ke bāda khāne wālā*)
- 5) *tādar̥thya* (purpose): as- *chātraḥ adhyayanāya gurukulaṃ gacchati* (*chātra adhyayana ke liye gurukula jātā hai*)
- 6) *hetuḥ* (cause): as- *vidyār̥thī adhyayanena vidyālayaṃ vasati* (*vidyār̥thī adhyayana ke liye vidyālaya meṃ rahatā hai*)
- 7) *vīpsā* (repetition): as- *śakuntalā āśrame prativr̥kṣaṃ siñcati* (*śakuntalā āśrama meṃ pratyeka vr̥kṣa ko siñcatī hai*)
- 8) *kriyā-āvr̥tṭi-gaṇanā* (counting of repetition): as- *bālakaḥ pāṭhaṃ pañcabāraṃ paṭhati* (*bālaka pāṭha ko pāñca bāra paḍhatā hai*)
- 9) *kriyāviśeṣaṇam* (manner adverb): as- *mṛgaḥ vegena dhāvati* (*mṛga vega se daudatā hai*)
- 10) *atyanta-sambaddhaḥ kālaḥ* (complete relation with time): as- *bālakaḥ gurukule māsam adhītaḥ* (*bālaka gurukula meṃ mahine bhara paḍhā*)
- 11) *atyanta-sambaddhaḥ mārgaḥ* (complete relation with road): as- *pāṭhaḥ krośaṃ adhītaḥ* (*kośa bhara pāṭha paḍhā*)
- 12) *atyanta-sambaddhaḥ kālaḥ saphalaḥ* (complete relation with time with result): as- *bālakaḥ māsenā anuvākaḥ adhītaḥ* (*bālaka mahine bhara meṃ anuvāka paḍha liyā*)
- 13) *atyanta-sambaddhaḥ mārgaḥ saphalaḥ* (complete relation with road with result): as- *bālakaḥ krosenā anuvākaḥ adhītaḥ* (*bālaka kośa bhara meṃ anuvāka paḍha liyā*)

TG rules⁴:

1. The *tādarthya* and *hetu* are taking *ke liye*,
2. *vīpsā* takes *ko* case marker,
3. *kriyā-āvṛtti-gaṇanā* (counting of repetition) takes *ko*,
4. *kriyāviśeṣaṇam* (manner adverb) takes *se*,
5. *atyanta-sambaddhaḥ kālaḥ* (complete relation with time) and *atyanta-sambaddhaḥ mārgaḥ* (complete relation with road) take *bhara*, *atyanta-sambaddhaḥ kālaḥ saphalaḥ* (complete relation with time with result)
6. *atyanta-sambaddhaḥ mārgaḥ saphalaḥ* (complete relation with road with result) take *bhara meṁ* as postpositions in Hindi.

4.4.8 Other relations and TG rules:

- 1) *śaṣṭhī* relation: as- *adhyāpakasya pustakaṁ chātrāḥ paṭhanti* (*adhyāpaka kī pustaka chātra paṭhate haiṁ*)
- 2) *ārambhasamayaḥ māpane* (starting point of time): as- *kārtikyāḥ āgrahāyaṇī māse* (*kārtika se lekara agahana māsa taka*)
- 3) *ārambhadeśaḥ māpane* (starting point of place): as- *tirupatitaḥ candragiriḥ krośe* (*tirupati se candragiri kośa bhar dūra hai*)
- 4) *lakṣaṇa* (point of direction): as- *vṛkṣaṁ prati viśyotate vidyut* (*vṛkṣa kī or vidyut camakatī hai*)
- 5) *tādarthya* (purpose): as- *bālakāya pustakaṁ krīṇāti* (*bālaka ke liye pustaka kharīdatā hai*)
- 6) *hetuḥ*: as- *kumbhakāraḥ daṇḍena ghaṭaṁ karoti* (*kumbhakāra daṇḍe se ghaṭā banātā hai*)
- 7) *saha sambandhaḥ* (association): as- *putreṇa saha pitā gacchati* (*putra ke sātha pitā jāte haiṁ*)

⁴ Here *sambodhana*, *prasajya-pratiṣedha* and *kriyā-āvṛtтыantarālasamayaḥ* have no case marking issue as *bho* is mapped as *he*, *na* is mapped as negation *naiṁ* and *kriyā-āvṛtтыantarālasamaya* has *tvā* suffix for *pūrvakālikatva* representation.

- 8) *vinā* (non-association): as- *dharmeṇa vinā jīvanam nāsti* (*dharma ke binā jīvana nahīn hai*)
- 9) *vibhaktaḥ* (comparison between two): as- *māthurāḥ pāṭaliputrebhyaḥ ādyatarāḥ* (*mathurāvāsī pāṭaliputrāvāsī se adhika sampanna haiṅ*)
- 10) *nirdhāraṇam* (Isolation one from a group – in the superlative degree context): as- *gavāṃ/goṣu kṛṣṇā bahukṣīrā* (*gāyon meṅ kālī (gāya) bahuta dūdha denewālī hai*)

TG rules:

1. *ṣaṣṭhī* relation takes *kā/kī/ke* depending on the gender of the substantives.
2. *ārambhasamayaḥ māpane* takes the postposition *se_ lekara* in Hindi,
3. *ārambhadeśaḥ māpane* takes *se*,
4. *lakṣaṇa* takes *kī or*,
5. *tādarthya* takes *ke liye*,
6. *hetuḥ* takes *se*,
7. *saha sambandhaḥ* and *vinā upapada* have *ke* case marker,
8. *vibhaktaḥ* has *se* case marker
9. *nirdhāraṇa* has *meṅ* case marker.

Thus all the possible rules for case handling can be written on the basis of above discussed facts. We need *kāraka* tagger tool to sub-classify the *kāraḥ*. Thereafter these rules may be presented in SSF format which can be applied in the TG engine of IIIT, Hyderabad.

4.5 Mapping nominal in compound syntax and conjunct *ca* and *vā*

In 3.7, it is discussed that *rāmaḥ śyāmaḥ ca*, *rāmaśca śyāmaśca*, *rāmaśyāmau* etc are the different ways to add the nominals in a compound syntax which is, actually, not an NP but considered. Here single and double/multiple *ca* in the *vigraha-vākya* require extra case as the conjunct in Hindi is neither kept at the end of the combining words nor used twice to combine the two given words. Following rules may handle these nominals:

- 1) If single *ca* is used between nominals in a compound syntax, the translation of *ca* (i.e. *aura* in Hindi) will be kept before the last substantive in TL. As- *rāmaḥ śyāmaḥ ca* = *rāma aura śyāma*.
- 2) If double/multiple *ca* is used between nominals in a compound, translation of last *ca* (i.e. *aura* in Hindi) will be kept before the last substantive in TL and the rest *ca* will be dropped from the SL input. As- *rāmaśca śyāmaśca* = *rāma aura śyāma*.
- 3) In *rāmaśyāmau* type of constructions, it will be checked that how many substantive words are there in the compound nominal, thereafter translation of *ca* conjunct should be inserted before the last substantive. As- *rāmaśyāmau* = *rāma aura śyāma*.

bhrātarau, pitarau etc are the compound nominals which are used to denote a male-female pair but it looks like the dual form of *bhrātr* and *pitṛ*. These types of compound nominal, first, should be handled under word sense disambiguation and then these can be translated on the basis of compound analysis. There also, we have to tackle the conjunct.

Similarly, when *vā* is used in any compound syntax, it should also be handled in the same manner. When single *vā* is used, then it's translation will be kept before the last substantive and when multiple *vā* is used, the last *vā* will be kept before the substantive and the rest *vā* will be dropped from the SL. As- *rāmaḥ śyāmaḥ vā, rāmaḥ vā śyāmaḥ vā* and *rāmaḥ vā śyāmaḥ*, these all will be translated as *rāma yā śyāma*.

4.6 Adjective mapping

In 3.8.2, we have discussed about the identification of the adjectival *pada* in SL and discussed the issues in mapping the adjective. There are two issues in adjective mapping: the order word and the case marking. On the basis of that discussion, the rules can be made for identifying the adjective.

4.6.1 Rules for identifying the adjectives:

1. If two *subanta padas* are coming together as *samānādhikaraṇa* then anyone may be adjective. Same as, if two *kṛdanta padas* are coming together then also the place of adjective *pada* is uncertain. Here it is generally assumed that the first *pada* is an adjective.
2. If *kṛdanta* and *subanta pada* comes together, then the *kṛdanta pada* is adjective and *subanta* is substantive.
3. If *kṛdanta/subanta* and *taddhita* is coming together, then *taddhita* is the adjective and the *kṛdanta/subanta* is substantive.
4. As *taddhita* never becomes the adjective of a pronoun, when *taddhita* is used with a pronoun, pronoun will be the adjective and *taddhita* will be the substantive.

4.6.2 Mapping case marker in adjective: The first difference between Sanskrit and Hindi adjective is that an adjective and a substantive in Sanskrit have the same case marker while adjective in Hindi does not show the case marker except oblique forms in adjectives with *ā*-ending. It is assumed that oblique form may be handled by the Hindi generator. So no TG rule is required for oblique case marking.

It is discussed earlier in 3.8.2.2 that the NP having feminine substantive takes feminine form of adjective in SL i.e. *suśīlā bālikā* can be translated as *suśīlā bālikā* as well as *suśīla laḍakī*. Here the adjective is changing because of change in taken synonym of the substantive in TL. In Hindi, *acchī laḍakī* cannot be pluralized as *acchiyān laḍakīyān* instead of *acchī laḍakīyān*. So the claim of same number with adjective and substantive cannot be true in Hindi while this is applied in Sanskrit. We can see another example: *teja laḍakā* and *teja laḍakī*, where the gender of adjective word is not affected by the gender of substantive in Hindi.

On the basis of the above observation, we can say that:

1. The *-ā* ending feminine adjectives are depending on the selection of the substantive.
2. The *-ā* ending feminine adjectives do not change their form in plural in Hindi.
3. Some adjectives (i.e. *teja*, *mūḍha* etc) have the same form in all gender.

For handling these differences, we can make a list of the adjectives under these three categories which will be helpful in making rules for correct output.

4.7 Dual number mapping

The following rules can be given to map the dual number:

1. When a substantive (without adding *dvau* before the substantive) is used in dual number, numeral *do* will be added before that substantive in Hindi translation.
2. When *dvau* is added with a dual number substantive, the numeral *dvau* will be dropped to avoid the repetition of *do* (i.e. *do do* in Hindi output) in TL.
3. When a substantive has *dvayam* as numeral adjective (substantive + *dvayam*), the meaning of *dvayam* will be kept before the substantive in the TL.

4.8 Word order mapping

After identifying the adjective in a Sanskrit sentence, the general word order in NP of the input sentence can be converted as per the template of NP.

NP= {DET (*ekaḥ/kaścit*)} / {(DET (*ekaḥ/kaścit*) + N_(AP)} + N + (REL)

Where DET is determiner (*ekaḥ* or *kaścit*), AP is Adjectival Phrase and N is Nominal word (i.e. an agent) and REL is Relative Clause.

In this way, the required TG rules can be written for mapping the structural differences between Sanskrit and Hindi to get the desired output in SHMT.

CONCLUSION

The goal of present research was to discuss the challenges in Sanskrit-Hindi NP mapping from the perspective of Machine Translation. Different tools have been made to achieve the goal of Sanskrit-Hindi Machine Translation (SHMT) but there are some challenges, such as- case mapping in substantives and adjectives, indeclinable mapping, conjunct and numeral mapping and word order handling, in SHMT which require the theoretical discussion first. On the basis of that discussion, the rules could be formulated (theoretically) to handle the challenges. These issues have been discussed at the NP level.

In first chapter, a survey has been presented to justify the importance of the present research. Second chapter titled: 'Sanskrit-Hindi Noun Phrase' presents the basic concept of NP in both languages. Challenges in mapping Sanskrit-Hindi NP have been discussed as third chapter to find out the issues which must be tackled by researchers to produce the desired output. On the basis of challenges identified, the 4th chapter deals with the problem tackling aspects. The task of handling all the discussed issues is divided into two parts: 1) task of Hindi Generator and 2) Transfer grammar (TG). Task of Hindi Generator shows the requirement of improving the generator as the Hindi generator is an independent tool which can be implemented to produce Hindi translation from any SL. In TG section, the rules are presented theoretically to map the syntactic differences. It can be implemented in any SHMT system to produce the reasonable output. To implement the rules of case marker mapping, which is thoroughly discussed, we need a Sanskrit *kāraka* tagger which will be tagging the *kāra*kas. The second requirement is a tagged dictionary in which substantives would be tagged as animate-inanimate and verb as transitive-intransitive. The animate-inanimate information will be helping in handling *ko* case marker and transitive-intransitive will be helping in handling the ergative *ne*. The POS tagger will help us in identifying the adjective in SL (Sanskrit) and the other issues can simply be handled by the TG rules.

This work has some limitations as well and that gives the scope for further research. In the present research, the relative clause is not handled because it needs the Verb Phrase (VP) study between both the language pairs as verb is affecting the case

marker in *kriyāmūlaka* roots (i.e. root + *karanā/honā*) in a relative clause as well as in a verb phrase. So VP level study would be the next level research which will be improving this research as well. Rules for indeclinable mapping could not be provided in this research due to a variation in meaning in different context which requires a statistical study from an exhaustive corpus.

BIBLIOGRAPHY

Books

1. Abhyankar, K.V. (1985): Fundamentals of Sanskrit Grammar. M.M.Vasudeo Shastri Abhyankar Publication Series (Sl.No.7), Sanskrit Vidya Parisamstha, Pune
2. Apte, V.S.(1890) The Students Guide to Sanskrit Composition, Chowkhamba Sanskrit Series Office.
3. Bahri, Hardev (1985): *Hindi Semantics*. Manasdharm, New Delhi.
4. Balchandran, Lakshmi Bai (1971), A case grammar of Hindi, Central Institute of Hindi, Agra, India.
5. Belvalkar, Krishna Shripad (2004): Systems of Sanskrit Grammar, Bhartiya Vidya Prakashan, New Delhi.
6. Borrow T. (2001), The Sanskrit Language, MLBD, New Delhi.
7. Chomsky N. (1957) Syntactic Structure, The Hague Mouton.
8. Chomsky N. (1965): Aspects of the theory of syntax, Cambridge Mass, MIT Press.
9. Deshpande, Madhav Murlidhar and Peter E Hook (2002), Indian Linguistic Studies: Festschrift in Honor of George Cardona, MLBD, New Delhi.
10. Dorr, Bonnie (1993): Machine translation: a view from the lexicon, The MIT Press, USA.
11. Goyal, Pawan and Sinha, R. Mahesh K. (2009): *Translation Divergence in English-Sanskrit-Hindi Language Pairs* in Amba Kulkarni (Et.al.), Sanskrit Computational Linguistics, Proceedings of Third International Symposium, Hyderabad. Springer-Verlag LNAI 5406 (2009).
12. Guru, Kamta Prasad (2010): Hindi Vyākaraṇa, Pancashil Prakashan, Jaypur.
13. Hock, Hans Henrich (1991), Studies in Sanskrit Syntax: A Volume in Honour of the centennial of Speijer's Sanskrit Syntax, MLBD, New Delhi.

14. Irach J.S. Taraporewala (1967), Sanskrit Syntax, Munsiram Manoharlal, New Delhi.
15. Iyer, K.A.Subramania (1983): The Vakyapadiya of Bhartṛhari, Vol.2. Motilal Banarsidass, Delhi.
16. Kachru, Yamuna (1966), An Introduction to Hindi Syntex, Dept. of linguistics, University of Illicins, Urbana.
17. Kachru, Yamuna (1973): Hindi rūpāntaraṅātṃmaka vyākaraṅa ke kucha prakaraṅa, kendriya Hindi Sansthan, Agra.
18. Kachru, Yamuna (2006): HINDI, John Benjamins Publication Company, Amstredam/Philadelphia.
19. Kale, M. R. (1894), A Higher Sanskrit Grammar, MLBD, New Delhi.
20. Kapoor Kapil (1967), *The ko-sentences in Hindi*, Paper presented at the Conference on Hindi Syntex at University of Illicins, Urbana.
21. Mimaṅsaka Y. (1979) (Ed.): Mahābhāṣya, Ramlal Kapur Trust, Bahgarh.
22. Mazumdar, Pradip Kumar (1985): The Meaning of Non-Denotative Words (A study on Indian Semantics), Sanskrit Pustak Bhandar, Calcutta.
23. Pandey, Dr. Umesh Chandra (2008), Sanskrit rachana (Hindi translation of V S Apte's Sanskrit Syntax), chaukhamba vidyabhavana, Varanasi.
24. Pandeya, Gopaladutta (2007): Vaiyakaraṅa Siddhanta Kaumudi of Sri Bhaṅḃjidikṅita ('Balamanorama'- dipika Sanskrit and Hindi commentary). Vol 1, Chaukhamba Surabharati Prakashan, Varanasi.
25. Pathak, Onkar Nath (Ed.) (2010): Paribhāṣāpradīpārciḃ of Udayankar Pathak, Sanskrit Pustak Bhandar, Kolkata.
26. Rajendran, C. (2002): Aspects of Pāṅinian Sementics, Sahitya academi, New Delhi.

27. Rishi, Dr. Umashankar Sharma (2008), Sanskrit vyākaraṇa meṅ kārakatattvānuśīlana, chaukhamba surbharti prakashan Varanasi.
28. Sahay, Caturbhuj (2007), Hindi Padavigyan, Kumar Publication, Agra.
29. Regmi, Acharya Sheshraj Sharma (Ed.) (2002): Sahityadarpan, Krishnadas Academy, Varanasi.
30. Shapiro, Michael C. (2000): *A Primer of Modern Standard Hindi*. Motilal Banarsidass, Delhi.
31. Sharma, Aryendra (1958): A Basic Grammar of Modern Hindi, Government of India, New Delhi.
32. Sharma, Raghunath (1988): Vākyapadīyam (with Ambakartṛka commentary), Sampurnanand Sanskrit Vishwavidyalaya, Varanasi.
33. Sharma, Rama Nath (2002): *The Aṣṭādhyāyī of Pāṇinī*, Vols 1-3. Munshiram Manoharlal, Delhi.
34. Shastri, Bheemsen, *Laghusiddhāntakaumudī*, Bhaimee Prakashan, 537, Lajapatrai Market, New Delhi-06
35. Shastri, Charudev (1969): Vyākaraṇa chandrodaya, Motilal Banarsidas, Delhi.
36. Shastri, Charudev (1986): Vākyavahārārtha, MLBD, Delhi
37. Singh, Harivansha (2003): kāraka vyākaraṇa kī dṛṣṭi se Hindi parasarga vicāra, kendriya Hindi Sansthan, Agra.
38. Sinha, B.K. (1986): Contrastive Analysis of English and Hindi Nominal Phrase, Bahari Publication (PVT.) LTD. New Delhi.
39. Speijer, J.S. (1886), (reprint-1988), Sanskrit Syntax, MLBD, New Delhi.
40. Staal, Frits. (1967). Word-Order in Sanskrit and Universal Grammar. Dordrecht: Reidel.

41. Tiwari, Bhola Nath (1986), Hindi ki roop sanracana, sahitya saahakar, Delhi.
42. Tripathi, Ramadeva (1977): Bhāṣāvijñāna kī bhāratīya paramparā aura Pāṇinī, Bihar Rashtrabhasha Parishad, Patna.
43. Vajpayee, Kishoridas (1998): Hindi śabdānuśāsana, Nāgarīpracāriṇī sabhā, Varanasi.
44. Vasu, Srisa Chandra (2003): The Siddhanta Kaumudi of Bhaṣṇoji Dikṛita. Vol I. Motilal Banarsidass, Delhi
45. Verma, M. K. (1971): The structure of the noun phrase in English and Hindi, MLBD, Delhi.

Research Papers

1. Aklujkar, Ashok (1996): *Some theoretical observations on word order in Sanskrit*, Festschrift Paul Thieme, Stll, (1996), S.1-25.
2. Anantanarayana, H.S. (1970), *The kārak theory and case grammar*, Indian linguistics Vol.31 No. 12. Poona.
3. Briggs, Rick (1985): Knowledge Representation in Sanskrit and Artificial Intelligence, AI Magazine 6 (1985), pp.32-39.
4. Dash, Siniruddha and Brenden Gillon, (1995): *Adjectives in Sanskrit*, The Adyar Library Bulletin- 1995.
5. Deshpande, Madhav M. (1987); *Paninian Syntax and the changing notion of sentence*, in Hock (1991), Studies in Sanskrit Syntax: A Volume in Honour of the centennial of Speijir's Sanskrit Syntax, MLBD, New Delhi.
6. Deshpande, Madhav M. (1990); *Semantics of Karakas in Panini: an exploration of philosophical and linguistic issues*, in Sanskrit and related studies:

- contemporary researches and reflections. Matilal (Et.al.) Sri Satguru Publications, 33-57.
7. Dwivedi and Sukhdeve (2010): Machine Translation System in Indian Perspectives; Journal of Computer Science 6 (10): 1082-1087, 2010; ISSN 1549-3636 © 2010 Science Publications.
 8. Jha, G. N., Sobha L, and Diwakar Mishra (2009): *Discourse Anaphora and Resolution Techniques in Sanskrit*, in Devi, Sobha Lalitha (ed al.) Proceedings of 7th Discourse Anaphora and Anaphora Resolution Colloquium (2009).
 9. Jha, Girish N (1995). *Proposing a computational system for Nominal Inflectional Morphology in Sanskrit*, Proceeding of national seminar on “Reorganization of Sanskrit Shastras with a view to prepare their computational database”.
 10. Jha, Girish N (2003). *A Prolog Analyzer/Generator for Sanskrit Subanta Padas*, in Language in India, volume 3.
 11. Jha, Girish N, March (2004). *Generating nominal inflectional morphology in Sanskrit*, SIMPLE 04, IIT-Kharagpur Lecture Compendium, Shyama Printing Works, Kharagpur, WB. Page no. 20-23.
 12. Jha, Girish N, Prof. Kapil Kapoor and Prof. G.V. Singh, (April 11-12th 1996). *Lexical conceptual structure and domain based machine translation system*, presented at a symposium on Machine Aids for translation & communication, JNU, New Delhi.
 13. Joshi., S.D. (1962): Verbs and Nouns in Sanskrit, Indian Linguistics, Vol-23. pp. 60-63.
 14. Joshi., S.D. (1966): Adjectives and Substantives as a single class in the “Parts of Speech”, Journal of the University of Pune, Vol. 25, pp.19-30.
 15. Joshi., S.D. (2001): *Synyactic and semantic device in Aṣṭādhyāyī of Pāṇini*, in journal of Indian Philosophy 29: 155-167, @ 2001 Kluwer Academic Publishers, Printed in the Netherlands.

16. Mishra, Sudhir K., Jha Girish N. (2004). *Identifying Verb Inflections in Sanskrit morphology*, in proc. of SIMPLE 04, IIT Kharagpur, Page no. 79-81.
17. Mishra, Sudhir K., Jha Girish N. (2004). *Sanskrit karaka analyzer for Machine Translation*, In SPLASH proc. of iSTRANS, Tata McGraw-Hill, New Delhi. Page no. 224-225.
18. P. Shukla, D. Shukl, A. Kulkarni (2010): *Vibhakti Divergence between Sanskrit and Hindi*, in G N Jha (Ed.) Sanskrit Computational Linguistics, LNCS 6465, pp-198-208, 2010. @ Springer-Verlag Berlin Heidelberg 2010.
19. Pathak, K. N. and G. N. Jha (2011): *Challenges in NP Case-Mapping in Sanskrit-Hindi Machine Translation*, in C. Singh (et al.) (Eds.): ICISIL 2011, CCIS 139, pp-289-293, 2011, @ Springer-Verlag Berlin Heidelberg 2011.
20. Ramkrishnamacharyulu, K. V. (2009): *Annotating Sanskrit Texts Based on Śābdabodha Systems*, in A. Kulkarni and G. Huet (Eds.): Sanskrit Computational Linguistics, LNCS 5406, pp. 26-39, @ Springer-Verlag Berlin Heidelberg 2009.
21. Samuel R. Levin (1965): *Modern Linguistics: Its Development and Scope*, in The Journal of Higher Education, Vol. 36, No. 3 (Mar., 1965), pp. 137-146
22. Sinha, R M K (1984): *Computer processing of Indian languages and scripts - Potentialities and Problems*, JIE&TE, Vol.30, no.6, 1984.
23. Staal, Frits and Kiparsky, Paul. (1969): *Syntactic and Semantic Relations in Paṣini*, Foundations of Language 5: 83-117.
24. Subash and Girish N. Jha (2005). *Morphological analysis of nominal inflections in Sanskrit*, Presented at Platinum Jubilee International Conference, L.S.I. at Hyderabad University, Hyderabad, pp-34.
25. Subash and Girish N. Jha (2006). *Sanskrit Subanta recognizer and analyzer for machine translation*, in the proceeding of 28th All India Conference of Linguists (28th AICL). Department of Linguistics, Banaras Hindu University (BHU), India.

26. Subash. (2006). *Computational identification and analysis of complicated Sanskrit noun phrases proceeding of 2nd International Conference on cognitive Science (ICCS-2006)*. Centre for Behavioral and Cognitive Sciences, University of Allahabad, Allahabad, U.P. India.

