Light on Machine Translation Literature Survey- Part-1

Promila Bahadur
Asst. Lecturer
Rashtriya Sanskrit Sansthan
Gomti Nagar, Lucknow.

ABSTRACT

Translation of natural language has always attracted attention of scholars world-wide, be it manual or machine based. We can witness machine translation from, the mid of 20th century. Since, then we have seen progress of machine translation by leaps and bounds, as several agencies like government or private have been involved in this work. Machine Translation have been attempted on various languages across the globe be it Asian or European. It has also reported a good success rate of translation.

Keywords

Machine Translation, Generation, Indian language, Foreign Language, Devnagri.

1. INTRODUCTION

Machine Translation of natural language is a widely discussed and challenging topic. It has attracted linguistic as well as researchers, from, the mid of 20th century. This century witnessed birth and growth of machine translation. The progress can be recorded decade-wise, since 50's and prominent developments took place in each decade [2]. The era can be broadly categorized into five progressive generations [3,4,5].

First Generation- During the initial years, inputs from government and intelligence agencies, showed the need of machine translation. It was heavily funded by agencies due to its high potential, high speed quality translation being visualized. This concept grew with the time.

Second Generation- By its beginning, it was realized that automated translation is not achievable target as per the outcomes of the first generation. A report from National Academy of Sciences condemned the field as well as its workers. Though, the report was criticized as narrow and short sighted but the recommendations were adopted [2].

Third Generation- The situation worsened by its advancement. The government funded projects were functioning only and most of the projects elsewhere stopped [6]. Gradually, later in the era, private companies floated machine translation projects. Though, high expectations resulted in disrepute of machine translation.

Fourth Generation- Machine translation gained momentum and attracted attention from various walks of society e.g., government, business and industry. Now machine translation and Machine Aided Translation (MAT) system were in use [7]. Private and Government sectors started funding machine translation projects as the expectations with machine translation were more realistic. It came into light that machine translation has great potential.

This generation witnessed, the popularity of machine translation, among developing countries, like India, Japan etc. Researchers [1] suggested demand for technical translation. It was reported that worldwide work on machine translation is going on that includes array of projects. The research included various categories of translation like rough translation, full translation using machine translation, Value Added Network (VAN) service based on machine translation etc.

Fifth Generation- New dimensions of the machine translation were explored which included Statistical Machine Translation (SMT). It was introduced by IBM researchers in a workshop sponsored by the US National Science Foundation and Johns Hopkins University's Center for Language and Speech Processing [8, 9].

By this time machine translation had became a household name and many projects were undertaken to carry out translation from one language to another [10,11]. As an important issue, the percentage of accuracy of the translation persisted. Many international and national projects like SYSTRAN, METEO, LOGOS [12], Anusaarka, AksharBharati [86] gained momentum. Individual efforts also surfaced in the project named Transliteration [14].

Translation on spoken language projects such as C_Star, ART, Eutrans, TC-Star, PF-Star etc [3] were also attempted. The written language translation was also attempted. For researchers international languages like Spanish, Portuguese, Japanese, German, Chinese, Arabic, Hebrew, Sinhalese etc., became desirable languages for translation. In Indian-Sub continent widely attempted languages for translation were Bengali, Hindi, Gurumukhi, Telegu, Kannad, Oriya etc.

2. BASE OF MACHINE TRANSLATION

Studies, on machine translation, can be broadly divided on the basis of languages attempted for translation and techniques evolved for translation. Languages attempted for translation, can be categorized into Foreign Language and Indian Language, as shown in Fig. 1.

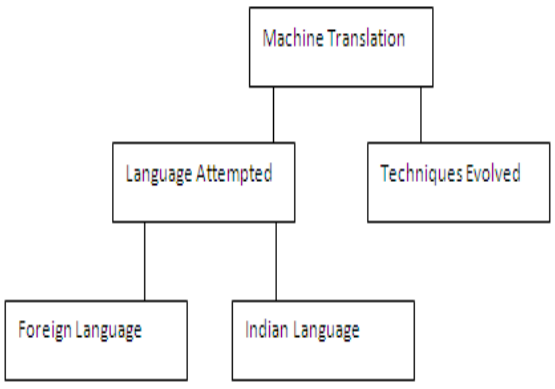

Fig. 1: Base of machine translation

3. LANGUAGES ATTEMPTED FOR TRANSLATION

i. Machine Translation in context of Foreign Language

Foreign languages are translated from one to another, depending on their popularity and population coverage. We can say that popularity is directly propotional to the choice of the language, more the popular language is, attracts the researchers. Projects on translation from Chinese to English, Japanese to English, Arabian to English etc [15, 16, 17] and vice versa had been widely taken up. English language remained a favorite language for translation as source or target language, in India or overseas; due to its reach, popularity and availability of literature across the nations. The detailed discussion is given under section 3.1.

ii. Machine Translation in context of Indian Language

Indian language translation, in itself is a marathon task, which is considered exclusively. We have almost twenty-two official languages in India. Each one of them has its own literature and cultural heritage which makes it desirable to be shared across the length and breadth. Various projects [18] have been taken in this direction to be discussed under section 3.2.

iii. Techniques Adopted for Machine Translation

Irrespective of the language used, many researchers [19, 20, 21] have discussed various techniques for machine translation, with the aim to provide more accurate, fast, un-ambiguous translation. These techniques would be discussed in later paper.

3.1 Machine Translation in Context Of Foreign Language

This section can be divided on the basis of Asian language and European language. Project on machine translation from English to Japanese, Chinese Simplified to English, Chinese Traditional to English, Dutch to French, German to Spanish, Greek to Italian, Japanese to Korean, etc and vice-versa [15,16,17] have been undertaken.

The work is abundant in terms of research papers, journals as well as in terms of availability of online translation engines offering text, e-document, website and e-mail translation. Many language pairs are taken for translation with variety of techniques [21, 23, 24]. The rate of accuracy reported is very encouraging.

3.1.1 Machine Translation in Context of Asian Languages

Projects on Asian Languages like Chinese, Arabic, Turkish, Korean, Czech, Japanese, Persian, Farsi are in abundance. It can be summarized as follows-

I. Chinese to English [17, 25, 26] machine translation has attracted many researchers. The language is also paired up with many international languages like Arabic, English etc. The Arabic-Chinese machine translation uses English as pivot language and does sentence pivoting and phrase table pivoting. It does translation successfully and capable of handling complex sentences [25].

II. Bi-lingual translation over Chinese and English web data in various forms to translate Chinese noun-noun (NN) compounds into English was reported [27]. Bi-lingual dictionaries bridged gap between the English language and the Chinese language. Their method is based on contextual similarity between the two languages, without assuming parallelism or comparability in corpus data. An impressive ninety-five percent success was reported over a dataset of one thousand instances.

III. Partial bi-lingual web pages in Japanese were reported [22], to extract translation pairs. The implementation was done by accessing web pages containing a given Japanese expression, and looking for the English expression which occurs most reliably in its immediate vicinity. This method scored an impressive accuracy, over a combination of simplex nouns and compound nominal.

IV. An Interlingua-Based Chinese-English machine translation system [17] describes translation from Chinese to English language. It discusses about method of parsing which comprises of syntactic parsing, semantic parsing and design of Interlingua.

V. Integration of algorithm for English-Chinese word segmentation and alignment is given in [28]. The algorithm takes bilingual word segmentation and alignment work synchronously and interactively. It avoids the errors of word segmentation from being transferred into word alignment. Experimental result shows that it distinctly improves accuracy of both word segmentation and alignment.

VI. Cross-Language Information Retrieval (CLIR) adopting English-Chinese query translation as the dominant strategy is reported in [29]. English queries are used as translation objects, and English-Chinese machine readable dictionary is utilized as the important knowledge source to acquire correct translations. Machine-aided translation (MAT) and machine translation (MT) to the translation of museum relic texts is outlined in [30].

VII. Interlingua-based English–Korean Two-way Speech Translation of Doctor–Patient Dialogues with CCLINC [31] proposed two-way speech translation system trained on doctor patient dialogues. Key features of the system includes language–independent meaning representation which preserves the hierarchical predicate–argument structure of an input utterance, providing a powerful mechanism for discourse understanding of utterances originating from different languages, word-sense disambiguation and generation of various word orders of many languages. It also discusses adoption of the DARPA Communicator architecture, a plug-and-play distributed system architecture which facilitates integration of component modules and system operation in real time. Automatic acquisition of grammar rules and lexicons for easy porting of the system to different languages and domains was also discussed.

VIII. English-Turkish interlingua-based machine translation system [32] uses KANT-knowledge-based machine translation system with integration of a Turkish generation system. These two independently constructed systems were integrated, through development of a module which maps KANT Interlingua expressions to Turkish syntactic structures. The combined system is able to translate completely and

correctly forty-four of fifty-two benchmark sentences in the domain of broadcast news captions.

IX. English-Turkish example based machine translation with Synchronous Structured String Tree Correspondence for the representation of the sentence in parallel corpora presents, translated sentences or parts of sentences before hand which are stored as corpus of translated sentences. ORHUN Example Based Machine Translation system is developed from which very encouraging results have been obtained [33].

X. English-Czech phrase based machine translation uses explicit modeling of morphology, using a separate morphological language model. It also catches verb modifier error. The system has reported seventy-five percent of success [75].

XI. Translation by machine of Complex Nominal discusses about a method for compositionally translating noun-noun (NN) compounds is given in [15]. A word-level bi-lingual dictionary is used. Syntactic templates and dictionary statistics are used for candidate and corpus generation. The system is evaluated over English-Japanese machine translation technique. The system developed is robust over low-frequency noun-noun (NN) compounds.

XII. English-Japanese machine translation for noun-noun (NN) compounds was presented [34] using a word-level bilingual dictionary and syntactic templates for candidate generation and corpus and dictionary statistics for selection was proposed. A support vector learning-based method employing target language corpus and bilingual dictionary data [35] is used. The evaluation is done over English-Japanese machine translation system. This technique registered sixty-five percent of success. It emphasizes on translation of noun phrase from source to target language. Word level translation for each word in NN compound to be translated is considered. The selected NN compound is executed through a set of constructional translational template to generate translation candidate cases.

XIII. Rule-based technique is used for the representation, acquisition, and application of translation knowledge on the Japanese-English machine translation [36]. The translation knowledge is learnt automatically from parallel corpus using structural matching between the parse trees of translation examples. A comfortable user interface is developed, which makes it possible to invoke the translation functionality

XIV. Neural network Assisted Japanese to English machine translation system combines strength of logic programming, procedural programming and Neural Networks. Over one-thousand sentences were taken into consideration and the success rate reported is near to ninety-eight percent [37]. Machine translation for sentences with Fixed Expressions in English to Japanese considers sentences which are difficult to translate and have shown accuracy up-to seventy percent [38]. MAJESTY (Morphological Analyzer for Japanese Text Analysis) is a morphological analyzer for syntactic, semantic parsing purposes. The system has been tested on news paper article and tagged parts of speech etc., and has shown ninety-eight percent results [17].

XV. Extended super function based Chinese-Japanese machine translation is divided into super functions for sentences and phrases. It pairs in Chinese Japanese business letters Encyclopedia which consists of over one hundred thousand sentences. The results obtained are nearly sixty percent correct [39].

XVI. Among multi word translation Korean- Chinese work has been presented. It combines domain tuned dictionary and target language corpus. A hybrid method has been suggested which combines statistical method with linguistic rules. Experimental results have shown ninety-one percent success in this dictionary based method [40].

XVII. Interlingua based English-Korean Two Way Speech Translation of Doctor Patient dialogue has integrated various techniques and tackles challenge of real time speech translation. The system is developed with real time common coalition language system at Lincoln Laboratory (CCLINC).The performance evaluation of the system is based on impact of discourse understanding and speech recognition accuracy on translation quality which showed better results [41]. Korean-English machine translation based on Idiom Recognition considers idiom based approach. This approach showed promising results as it give best translation for idiomatic expressions without complex rules [42].

XVIII. Negative Expression Translation in Japanese-Chinese machine translation explores rules for the translation. The system has taken One hundred fifty sentences in consideration and has accuracy is reported to about seventy-eight percent [43].

XIX. Word sense disambiguation using second long monolingual corpus is based on statistical model and constraint propagation algorithm which handles all disambiguates. The method is evaluated on Hebrew and German language. The system proves six-eight percent accurate for examples picked up from newspapers [44].

XX. Various approaches have been adopted for English-Arabic language pair translation e.g., Prototype English-Arabic Interlingua-based machine translation system [16] describes generation of Arabic sentences from Interlingua Representations (IRs). The mapping of sentential components is done and differences between English and Arabic are discussed, such as, agreement in number, this cannot be transferred exactly from the Interlingua representations of an English sentence [16]. Another technique [45] discusses about the architecture of the Arabic language and mapping system. The goal and result of mapping is a target language components whose contents reflect the contents of the Interlingua, expressed in terms of the syntactic and lexical properties of the target language

XXI. Mapping semantic features for each Interlingua concept to the appropriate syntactic features is proposed in [46]. The generation of properly inflected Arabic verbs and nouns is a concern of both the mapper and the generator for a partial integration of the Arabic Morphology system into the KANT system [47].

XXII. Hierarchical Phrase reordering model has taken two language pairs, Chinese-English and Arabic-English. The translation model is meant for phrase movements. It showed performance up to fifty-six percent [48].

XXIII. Automatic English to Persian text translator proposed in [49] translates simple English sentences into Persian, exploiting a combination of rule based and semantic approaches.

XXIV. The development of a Farsi-German and German-Farsi translation system is shown in [50]. The system has reported

considerable success on the BLEU1 scale.

XXV. Transliteration system for English to Sinhalese machine translation has used Finite State Automation (FST) to develop transducers. Two FST are constructed namely Type-1 and Type-2 for testing Wordnet and Sinhala Chat Box respectively [14]. A theoretical approach [51] was proposed for English-Sinhala machine translation through the concept of Varanagema (conjugation) in Sinhala Language.

3.1.2 Machine Translation in terms of European Languages

There is abundant work on machine translation in European Language which can be summarized as follows:

I. Open source machine translation for Portuguese to Spanish language and vice-versa [52] discusses about shallow-transfer machine translation toolbox. The error rates have been minimized to a reasonable limit and the contributions from the linguistic communities were involved.

II. English-German and English-Spanish language pair has also gained a lot of momentum. A web data to select English translations for compositional German and Spanish noun compounds is given in [53]. The system achieves an impressive accuracy of eighty-seven percent. The translation task is intrinsically simpler, however, in that it considers only those compounds which translate in noun-noun NN compounds into English.

III. Evaluation for Spanish-English Pronominal Anaphora Generation [54] presents the pronominal anaphora generation module of a complete Interlingua based machine translation approach. The approach named as AGIR (Anaphora Generation with an Interlingua Representation) allows the generation of anaphoric expressions into the target language from the Interlingua representation of source text. AGIR uses different kinds of knowledge (lexical, syntactic, morphological and semantic information) to solve the Natural Language Processing (NLP) problems of the source text. The paper presents the evaluation of the generation of English and Spanish (including zero pronouns) third person personal pronouns into the target language. The following results have been obtained: a precision of eighty percent and eighty-four percent in the generation of Spanish and English pronominal anaphora has been reported.

IV. Much of the Spanish-English translation is based on online tools like SYSTRAN, SDL, WordLingo, InterTran etc. The tools have reported the success of eighty-five percent [12].

V. English generation from Interlingua by example based method [25] discusses the generation of English language from Interlingua with the help of English dictionary and concept dictionary. Various verb selections have been made to test effectiveness of the system.

3.2 Machine Translation In Context Of Indian Language

Various premier research led institutes and government organizations [55, 56, 57, 58] have taken up the wide task of translation and achieved remarkable success. India has twenty-two officially recognized languages, such as, Assamese, Bengali, Bodo, Dogri, Gujrati, Hindi, Kannada, Kashmiri, Konkani, Maithili, Malyalam, Manipuri, Marathi, Nepali, Oriya, Punjabi, Sanskrit, Santahli, Sindhi, Tamil, Telugu and Urdu. It is necessary to have user- friendly machine translation systems to solve day to day problem. Many national and international projects have been taken up in this regard and fruitful results have been obtained [59].

Machine translation of Indian Language have been done both from one Indian language to another as well as in English Language also. Anuvadak, Anusaaraka, AnglaBharati, MANTRA, MAtRA, Shakti [59] etc. are various translation tools available for translation in different Indian languages. We can divide it in terms of North Indian and South Indian languages, as a lot of projects have been undertaken.

3.2.1 Machine Translation in terms of North Indian Languages

We section is discussed as follows:

I. A language oriented parsing algorithm for Urdu language sentences is given [55] by initiating tagging only for morphologically closed classes of words like postpositions, conjunctions, verb morphemes, etc. By utilizing linguistics features of these closed classes words are collected into chunks. The chunks are formed by applying grammar rules.

II. A rule based token mapping system presented [56] for translation between language pair Gujarati-Hindi. However, most appropriate translation in certain cases of post positions markers, pronouns, adjectives and adverbs, are under the influence of grammatical properties of other elements of the sentence cannot be obtained. Though, GH-MAP with special empirical rules, formed based on observed patterns have resolved the translation problem. These empirical rules help GH-MAP to obtain more appropriate translation.

III. The morphological analysis and generator tool for Hindi Language using paradigm approach for Windows platform having GUI is presented [58]. The Hindi morphological analyzer and generator stores all the commonly used word forms for all Hindi root words in its database. This approach prefers time and accuracy to memory space. The system is also applied on Punjabi language. A unique approach to develop machine translation system based on the insights of information dynamics from Paninian Grammar Formalism is presented in [57].

IV. Akshar Bharati group [13,76] have developed many NLP tools for Indian languages like Morphological Analyzers, Anusaraka, Shakti Machine translation System. The tool is developed in many Indian languages like Hindi, Telugu, Marathi, Kannada, Punjabi, Telugu etc. For a given word the Morphological Analyzers gives the root word with its feature information like gender, number, person, tense etc.

V. Anusaaraka [59] is a Language Accessor cum machine translation system based on the fundamental premise of sharing the load producing good enough results according to the needs of the reader. The system strives to provide faithful representation of the translated text and no loss of information while translating. The layered output provides an access to all the stages of translation making the whole process transparent.

[1] A software to test effectiveness of machine translation system

VI. Interlingua-based English–Hindi Machine Translation and Language Divergence given in [60], also discusses about UNL2 as intermediate language. The meaning conveyed by the source language is represented in Interlingua structure and it generates a target language structure.

VII. Assamese Machine Translation System is based on Machine Aided Translation System (AnglaBharati) [61,62]. The system developed basically a rule-based approach and relies on bilingual English to Assamese dictionary. Anubaad machine translation system [59] for news sentences using the transfer approach is developed at Jadavpur University on the basis of example-based English-Hindi language. This system translates short single program news items from English to Hindi language.

VIII. Oriya Machine Translation System (OMT) translating from English to Oriya is being developed by the Research Centre of the Department of Computer Science, Utkal University, Vanivihar [13, 63]. The architecture of the OMT is divided into six parts: Parser, Translator, OMT System, OMT Database, Disambiguator and the Software tools. The heart of the system is the OMT database (bilingual dictionary). [13,64] Anuvaadak 5.0 Machine Translation system from English to Hindi language was developed by Super Infosoft Pvt. Ltd. Some important features of this system are filling and printing facility, inbuilt grammar checker for both languages, over a hundred Hindi free fonts, user-friendly interface with pull-down menus and facility to add multiple Hindi meanings to the existing ones.

IX. Mantra project has been developed by C-DAC, Bangalore [13]. A sub-language English-Hindi Machine Translation system has been developed for the domain of gazette notifications pertaining to government appointment [13]. MaTra is developed by C-DAC, Mumbai, aims at machine-assisted translation from English to Hindi, essentially based on a transfer approach using a frame-like structured representation.

X. Anubharati is a project at IIT Kanpur, dealing with template-based machine translation from Hindi to English [13,65]. The project uses Hybrid Example-Based Model for Machine Translation (HEBMT), combining the strategies used in the pattern/rule-based approach and the example-based approach.

3.2.2. Machine Translation in terms of South Indian Languages

This is section is discussed as follows:

I. The translation of Sinhala-Tamil language pair using statistical machine translation was reported in [66] and the use of Statistical Machine Translation due to language pair similarity is argued

II. Similarly, a lot of work of the system translation has been done for Indian languages. UNL De-convertor for Tamil [26], discusses about generation of Tamil sentence from UNL structure by tackling at morphologicallevel and syntactical level. A generation of UNL structure from Tamil structure is outlined in [67]. The UNL is described as intermediate language which can be used both as en-converter and de-converter. UNL is said to have standard and global format.

English-Kannada Machine Translation system developed by the Computer Science department at the University of Hyderabad [13,61] uses Universal Clause Structure Grammar (UCSG) formalism. This is based on a transfer-based approach. It also requires post-editing and works at sentence level.

III. A Tamil-Hindi language accessor has been built based on the Anusaaraka formalism at the Anna University, Chandrasekhar Research Centre at Chennai [13]. Resource Centre for Indian Language Technology Solutions (RCILTS) is working various projects like ORIDIC (e-Dictionary Oriya - English -Hindi), OriNet (Oriya Word Net), Sanskrit WordNet, Trilingual Word Processor, Grammar, ORI-SPELL-CHECK (Oriya Spell Checker), Oriya Morphological Analyser (OMA) and also Oriya Machine Translation (OMT) System. A proto-type of Sanskrit WordNet using Navya-Nyaya and Paninian Grammar is developed. It has 300 Sanskrit words (250 Nominal words and 50 Verb forms) having synonym, antonym, hyponym, hypernym, holonym and meronym relations with their analogy, etymology, and definitions [86].

IV. Divyadrusti [65] system claims that printed Oriya documents can be recognized with up to ninety-seven percent accuracy. This system has integrated with Text-To-Speech (TTS) system.

V. Interlingua based translation having Data Oriented Parsing system (DOP) for target language generation mechanism was proposed. The system displayed considerable amount of success up to eighty-two percent [26].

3.2.3. Machine Translation in terms of Sanskrit

This section focuses on projects undertaken specifically for Sanskrit language. The language has the potential of being a pseudo-language for translation, among Indian languages as well as foreign, due to its scientific lineage.

I. A study of Example Based English to Sanskrit Machine Translation is given in [68]. Rashtriya Sanskrit Vidyapeetha (RSV) Tirupati has presented linguistic resources for Natural Language Processing in Sanskrit [13]. It segregates Sanskrit compound words into its components. It works on Sanskrit ISCII text in Linux environment and the system is also concentrating on kridanta and tiganta analyzers and also generators for subanta, tiganta and samasa.

II. Shabdabodha project developed at TDIL Academy of Sanskrit Research (ASR), Melkote is given in [13,77]. The project has an interactive application built to analyze the semantic and syntactic structure of Sanskrit sentences. It has two sections- user input and input file. User input provides the user to input any sentence and shows the corresponding syntactically compatible sentence and shows all the morphological details. Input file section provides the user to input any file name and shows the corresponding syntactic file. The system can process all types of sentences of Sanskrit. It can handle generation and analysis of subanta of more than twnty-six thousand stems, tiganta conjugations of roots, in two voices, ten lakaras and four modes viz. kevala, tiganta, kridanta and sananta. It also handles the generation, analysis and identification of case inflected forms of eleven types of kridanta of one-hundred and fifty roots. Apart from this it is said to have a database of six-hundred and ninety avyayas, twenty-six thousand nominal stems, six-hundred verbal roots, and kridanata forms of six-hundred verb roots [69]. Lexical Knowledge Bases like Wordnet, has been developed with tools like Morphological Analyzer and Parts Of

[2] Universal Networking Language

Speech (POS) taggers for Hindi and Marathi with Universal Networking Language (UNL) as Interlingua [70].

III. A tagset for Sanskrit has been given in [71]. The system developed has Parts of Speech (POS) tagger and has a tag set containing sixty-five word class tags, forty-three feature sub-tags, and twenty-five punctuation tags[72]. A Sanskrit subanta recognizer and Analyser System is being developed, the system has been developed according to Paninian formulation which accepts only sandhi-rahita Sanskrit text in Devanagari script (Unicode) and fully depends on both the rule base, example base and a database of other linguistic resources. The system claims to give an average accuracy of ninety-one percent accuracy.

IV. e-learning projects in Sanskrit have also been reported in [73]. Tiganta analyzer using reverse Paninian and database approach is also reported. Analyzer applying Paninian rules has also been reported in [74].

4. REFERENCES

[1] James M Lufkin Currentg Trands in Technical Translation 1989 IEEE.

[2] A Survey of machine translation –it's history, current status and future prospects, Jonathan Sloculn Microelectronics and Computer Technology Corporation Austin, Texas Computational Linguistics, Volume 11, Number 1, January-March 1985

[3] Integration of Speech Recognition and Machine

Translation in Computer-Assisted Translation Shahram Khadivi and Hermann Ney, Senior Member, IEEE, IEEE Transactions on Audio Speech and Language Processing, VOL. 16, NO. 8, NOVEMBER 2008

[4] An experiment on the machine translation of languages carried on BESM, I. S Mukhin, Digital Computer Techniques, Paper no 2144, Nov 1956.

[5] A New Level of knguuge Generafiion Technology: Cupubiliis und Possibiii#ies Eduard H. How, Information Sciences Institute of the University of Southern California, 0885/9000/92/0400-0012 $3.00 0 1992 IEEE.

[6] Computer translation —is it worth while? by W. L. PRICE, B.Eng., Ph.D. Electronics & Power September 1967

[7] Benefits of automating data translation, Sandra A Mamra , Julie Barnes, 07407459/93/370C/0082/ C IEEE July 1 9 9 3

[8] Statistical Language Approach Translates into Success Steven J. Vaughan-Nichols, Technology News, November 2003.

[9] An Online Relevant Set Algorithm for Statistical Machine Translation Christoph Tillmann and Tong Zhang, IEEE Transactions on Audio Spech and Language Processing, Vol. 16, NO. 7, September 2008.

[10] Toward Human Level Machine Intelligence—Is It Achievable? The Need for a Paradigm Shift Lotfi A. Zadeh University of California, USA, August 2008 | IEEE Computational Intelligence Magazine, Digital Object Identifier 10.1109/MCI.2008.926583

[11] Machine Translation Inching toward Human Quality Jan Krikke, 1541-1672/06/ © 2006 IEEE Intelligent Systems, Published by the IEEE Computer Society

[12] Milam W. Aiken, Zachary Wong, Spanish-to-English Translation Using the Web

[13] Sudhir Kumar Mishra, Sanskrit Karaka Analyzer for Machine Translation,PhD thesis submitted at Jawaharlal Nehru University 2007

[14] B.Hettige,Karunananda ,Transliteration System for English To Sinhala Machine Translation,Second International Conference on Industrial and Information Systems,ICIIS 2007,8-11 August 2007 ,Sri Lanka.

[15] Timothy Baldwin etal., Translation by Machine of Complex Nominals: Getting it Right In Proceedings of the ACL 2004 Workshop on Multiword Expressions: Integrating Processing, Barcelona, Spain.

[16] Abdelhadi Soudi etal. A Prototype English-to-Arabic Interlingua-based MT system

[17] QI Xuan. An Interlingua-Based Chinese-English MT System Volume 17 No.4 July 2002 Springer Press.

[18] QI Xuan. An Interlingua-Based Chinese-English MT System Volume 17 No.4 July 2002 Springer Press.

[19] Lemanh Hai,Phan Thi Tuoi ,2009 IEEE Three Algorithm for Word to Phrase MT

[20] Advath Siddharthan, Syntactic Simplifications & Text Cohesion, Research of Langauge Cohesion 2006 Springer

[21] Azza Abdel Monem · Khaled Shaalan,Ahmed Rafea · Hoda Baraka,"Generating Arabic text in multilingual speech-to-speech machine translation framework", Mach Translat (2008) 22:205–258

[22] Nagata et al. (2001)

[23]] Josep M. Crego · François Yvon,Factored bilingual n-gram language models for statistical machine translation Mach Translat (2010) 24:159–175.

[24] Chinese R&D in Natural Language Technology Chengqing Zong, Chinese Academy of Sciences Qingshi Gao, Beijing University of Science and Technology, 1541-1672/08/$25.00 © 2008IEEE Intelligent Systems, Published by the IEEE Computer Society.

[25] Eiji Komatsu etal. English generation from interlingua by example-based method, Proc. 15th conference on Computational linguistics - Volume 1, Kyoto, Pages: 363 - 368 ,1994

[26] T.Dhanabalan etal, Tamil to UNL En-Converter, Proc. of ICUKL 2002, November 2002, Goa, India

[27] Cao and Li (2002)

[28] Zhi-Ming Xu et al, Intergration Algorithn of English-Chinese Word Segmentation and Alignment, Proceedings of the Fifth International Conference on Machine Learning and Cybernetics, Dalian, 13-16 August 2006

[29] Zhi-Ming Xu et al, Intergration Algorithn of English-Chinese Word Segmentation and Alignment, Proceedings of the Fifth International Conference on Machine Learning and Cybernetics, Dalian, 13-16 August 2006

[30] Lu Mingyu Si Xianzhu,Application of Machine Translation to Chinese-English Translation of Relic Texts in Museum, 2010 International Conference on Intelligent System Design and Engineering Application.

[31] YOUNG-SUK LEE1, etalInterlingua-based English–Korean Two-way Speech Translation of Doctor–Patient Dialogues with CCLINC 2004 Kluwer Academic Publishers.

[32] Dilek Zeynep Hakkani4. etal English-Turkish interlingua using KANT knowledge-based machine translation system.

[33] N.deniz Alp,English to Turkish Example Based Machine Translation with Synchronous SSTC,Fifth International Conference on Information technology ,New Generation DOI 101109/ITNG 2008.64

[34] An Efficient Interlingua Translation System For Multilingual Document Production. Proceedings of the 3rd Machine Translation Summit.Nyberg, E.H. and T. Mitamura, 1992.

[35] Jim Breen. 1995. Building an electronic Japanese-English dictionary. Japanese Studies Association of Australia Conference

[36] Werner Winiwarter, Automatic Acquisition of Translation Knowledge Using Structural Matching Between Parse Trees, Proceedings of the First International Conference on the Digital Society (ICDS'07).

[37] Todd Law+ Hidenori Itoh Hirohisa Seki,,Gokiso-cho, Showa-ku, Nagoya A Neural Network- Assisted Japanese-English Machine Translation System, Proceedings of 1993 International Joint Conference on Neural Networks

[38] Naoto Katoh 1 Teruaki Aizawa,Machine Translation of Sentences with Fixed Expressions, NHK Science and Technical Research Laboratories,1-10-11

[39] Xiao Sun, Fuji Ren,Extended Super Function based Chinese Japanese Machine Translation,978-1-4244-4538-7/09/$25.00 ©2009 IEEE

[40] Byeong-Kwu Kang, Yi-Rong Chen, Bao-Bao Chang, Shi-Wen YU,Translating Multi Word Terms into Korean from Chinese Documents, cn0-7803-9361-9/05/$20.00 02005 IEEE Proceeding ofNLP-KE'05

[41] Young-Suk Lee, Daniel J. Sinder and Clifforrd J. Weinistin, Interlingua-based English–Korean Two-way Speech Translation of Doctor–Patient Dialogues with CCLINC, Machine Translation 17: 213–243, 2002. © 2004 Kluwer Academic Publishers. Printed in the Netherlands.213

[42] Hagyu Lee and Yung Taek Kim,Korean English machine translation based on idiom recognition, IEEE TENCON‘93 ,Beijing.

[43] Hong Zhang, Fuji REN, Negative Expression Translation in Japanese and Chinese Machine Translation,978-1-4244-2780-2/08/$25.00 ©2008 IEEE

[44] Ido Dagan,Word Sense Disambiguation Using a Second Language Monolingual Corpus © 1994 Association for Computational Linguistics.

[45] Arabic Morphology Generation Using a Concatenative Strategy". Proceedings of the North American Association For Computational Linguistics (NAACL), 2000, Seattle, United States.

[46] Timothy, A.B., 1990. Lexicographic Notation of Arabic Noun Pattern Morphemes and their Inflectional Features. Proceedings of the Second Cambridge Conference on Bilingual Computing in Arabic and English.

[47] The KANT System: Fast, Accurate, High Quality Translation in Practical Domains. Proceedings of COLING’92.

[48] Michel Galley, A Simple and Effective Hierarchical Phrase Reordering Model, Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, pages 848–856,Honolulu,October 2008. ©2008 Association for Computational Linguistics

[49] Yasaman Motazedi, Mehnoush Shamsfard, English to Persian Machine Translation

exploiting Semantic Word Sense Disambiguation, Proceedings of the 14th International CSI Computer Conference (CSICC'09).

[50] Somayeh Bakhshaei, Farsi – German Statistical Machine Translation Through Bridge Language, 2010 5th International Symposium on Telecommunications (IST'2010)

[51] Feiliang REN, Translate Chinese Organization Names Using Examples and Web, 978-1-4244-4538-7/09/ ©2009 IEEE

[52] Carme Armentano-Oller et.al , Open Source Portugese-Spanish Machine Translation, PROPER 2006, Lecture Notes in Artificial Intelligence, Volume 3960, pp 50-59, Springer, 2006

[53] Grefenstette (1999)

[54] Jesús Peral etal, IL MT System.Evaluation for Spanish-English Pronominal Anaphora Generation

[55] S. M. Jafar Rizvi, Language Oriented Parsing Through Morphologically Closed Word Classes in Urdu, 0-7803-8871-2/04/$20.00 ©2004 IEEE.

[56] Kalyani A Patel, Jyoti S Pareek, Rule Base to Resolve Translation Problems due to Differences in Gender Properties in Sibling Language Pair Gujarati-Hindi,International Conference on Computer &Communication Technology,ICCCT’2010.

[57] Sriram Chaudhury, Ankitha Rao, Dipti M Sharma, Anusaaraka: An Expert System based Machine Translation System, 978-1-4244-6899-7/10/ c2010 IEEE.

[58] Vishal Goyal, Gurpreet Singh Lehal, Hindi Morphological Analyzer and Generator, First International Conference on Emerging Trends in Engineering and Technology, First International Conference on Emerging Trends in Engineering and Technology, DOI 10.1109/ICETET.2008.11

[59] Jian LI and Bo WANG,Patterns and Matching Algorithm in an English-Chinese Machine Translation System, 0-7803-9361-9/05/ C2005 IEEE 839 Proceeding ofNLP-KE'O5

[60] Shachi Dave etal, Interlingua-based English–Hindi Machine Translation and Language Divergence, Machine Translation, Volume 16 , Issue 4, Pages: 251 – 304, 2001.

[61] Sudip Naskar and Sivaji Bandyopadhyay,Use of Machine Translation in India: Current Status,http://www.mt-archive.info/MTS-2005-Naskar-2.pdf (accessed on 7th April 2007).

[62] 'Progress Report of Resource Centre for Indian Languages Technology Solutions', 2004, Department of Computer Science and Engineering, IIT Guwahati, Assam, site: http://www.iitg.ernet.in/rcilts/download/progressreport_Aug04.pdf (accessed on 7th April 2007).

[63] Mohanty, S. and Balabantaray, R.C., 2004, 'Machine Translation System (Orya)', Proceedings of Symposium on Indian Morphology, Phonology and Language Engineering SIMPLE 2004, held in IIT Karagpur, March, 2004.

[64] Anuvaadak,http://tdil.mit.gov.in/TDIL-Jan-April-2004/super%20infosoft%20pvt%20ltd.pdf (accessed: 7April 2005).

[65] Divyadrusti, RCILTS, Utkal University, http://www.ilts-utkal.org/ocrpage.htm (accessed: 7 April 2005).

[66] S. Sripirakas, A. R. Weerasinghe, D. L. Herath, Statistical Machine Translation of Systems for Sinhala – Tamil, 2010 International Conference on Advances in ICT for Emerging Regions (ICTer)

[67] T.Dhanabalan, T.V.Geetha, UNL Deconverter for Tamil, International Conference on the Convergence of Knowledge, Culture, Language and Information Technologies, Convergences 2003, December 2 - 6, 2003, Alexandria, EGYPT.

[68] Study of Example Based English to Sanskrit Machine Translation,Vimal Mishra,R.B Mishra http:// www.gelbukh.com /polibits / 37_05. pdf

[69] Language Processing Tools: TDIL website, http://tdil.mit.gov.in/nlptools/ach-nlptools.htm (accessed: 7 April 2006)

[70] Resource Centre for Indian Language Technology Solutions (CFILT) - http://www.cfilt.iitb.ac.in/(accessed: 14.10.2006)

[71] R. Chandrashekar, 2006, 'Part-Of-Speech Tagging For Sanskrit', submitted for Ph.D. degree at SCSS,JNU, New Delhi.

[72] Chandra, Subash, 2006, 'Machine Recognition and Morphological Analysis of Subanta-padas', submitted for M.Phil degree at SCSS, JNU, New Delhi

[73] Bhowmik, Preeti & Jha, Girish Nath, 2006, 'Sanskrit Language Pedagogy: an e-2learning approach', In the Souvenir Abstracts of 28th AICL, BHU, Varanasi, p.150.

[74] Kumar, Sachin & Jha, Girish Nath, 2006, 'Issues in sandhi processing of Sanskrit', In the Souvenir Abstracts of 28th AICL, BHU, Varanasi, p.129.

[75] Ondˇrej Bojar, English-to-Czech Factored Machine Translation,Proceedings of the Second Workshop on Statistical Machine Translation, pages 232–239,Prague, June 2007. ©2007 Association for Computational Linguistics

[76] Akshar Bharati, is a personification of groups (@ IITK, IIIT-H, University of Hyderabad, etc.) working on NLP with special emphasis to Indian Languages giving due attention to Indian theories of grammar and language

[77] Shabdabodha, TDIL, Gov. of India, http://tdil.mit.gov.in/download/Shabdabodha.htm (accessed: 7 April 2005)