Chapter 2
Survey of Literature

The Chapter discusses about the available Machine Translation Systems and other translation Support Tools for foreign as well as Indian Languages and the approach followed for its development.

2.1 Categories of Machine Translation Systems

The Automatic Systems can be categorized into following sub categories.

- **MT system as software for automatic translation**: The system where input units are full sentences of one natural language and the output units are corresponding full sentences of another language.

- **MT system for home use**: An automatic translation system designed by the producer for personal use by the general public, i.e. by persons normally with no experience or training in translation.

- **MT system for internet/web**: The system developed specifically for translating electronic documents on the Internet such as electronic mail, webpages, chat discussions, etc.

- **MT system for professional use**: The system designed for use by professional translators, usually working independently or for translation agencies.
• **MT system for enterprise**: The system designed for company intranets. For example, Client-Server Systems to support a team of professionally trained translators.

• **MT system for websites**: The system designed for company websites to provide online translation of webpages.

• **MT system for spoken language**: The system designed to translate speech directly.[49]

### 2.2 Machine Translation for Non Indian Languages

This section briefly discusses some of existing Machine Translation Systems being developed for Non-Indian languages from time to time.

**Georgetown Automatic Translation (GAT) System (1952)**, developed at Georgetown University, was used for translating Russian texts, mainly from Physics and Organic Chemistry to English. GAT became operational in 1964 with its delivery to the Atomic Energy Commission at Oak Ridge National Laboratory, and to Europe's corresponding research facility EURATOM in Ispra, Italy. The GAT strategy was direct and local, simple word for word replacement, followed by a limited amount of transposition of words to result in something vaguely resembling English. [47]

**CETA: Centre D’Études Pour La Traduction Automatique (1961)**, was started at Grenoble University in France and was used to translate Russian into French. Grenoble began the CETA project with a clear linguistic theory. The theoretical
basis of CETA was Interlingua at the grammatical level and Transfer at the lexical level. The CETA system was under development for ten years and during 1967-71 it was used to translate 400,000 words of Russian mathematics and physics texts into French. [48, 49]

**METAL : Mechanical Translation and Analysis of Languages (1961),** is a joint project of the Siemens and Linguistic Research Center at University of Texas. The system translates technical documents from German into English. It is based on transfer approach. Its modular design consists of purely monolingual lexicons, transfer lexicons, and an augmented phrase structure grammar. [50]

**TAUM-AVIATION (1965),** is a transfer approach based English to French Machine Translation System for weather forecasts. It was developed at University of Montreal. After short span of time, the domain for translation shifted to translating aviation manuals by adding semantic analysis module to the system. The translation is produced indirectly, by means of an analysis/transfer/synthesis scheme. The overall design of the system was based on the assumption that translation rules should not be applied directly to the input string, but rather to a formal object that represents a structural description of the content of this input. Thus, the source language (SL) text was mapped onto the representations of an intermediate language, prior to the application of any target language-dependent rule. The system was evaluated and the low accuracy of the translation by this
system forced the Canadian Government to cancel the funding and thus TAUM project in 1981. [51, 52]

**SYSTRAN (1968),** is a direct Machine Translation System developed by Hutchins and Somers. In 1970, SYSTRAN System was installed at United States Air Force (USAF) Foreign Technology Division (FTD) at Wright-Patterson Air Force Base, Ohio, and is still operational there. A large number of Russian scientific and technical documents are translated using this system. The quality of the translations, although only approximate, was usually adequate for understanding content. In 1974, NASA also selected SYSTRAN to translate materials relating to the Apollo-Soyuz collaboration, and in 1976, EURATOM replaced GAT with SYSTRAN. The Commission of the European Communities (CEC) purchased an English-French version of SYSTRAN for evaluation and potential use. Unlike the FTD, NASA, and EURATOM installations, where the goal was information acquisition, the intended use by CEC is for information dissemination - meaning that the output was to be carefully edited before human consumption. The quality for this purpose is not adequate but improved after adding lexicon entries specific to CEC related translation tasks. Also in 1976, General Motors of Canada acquired SYSTRAN for translation of various manuals for vehicle service, diesel locomotives, and highway transit coaches from English into French on an IBM mainframe. The system was originally built for English-Russian Language Pair,
but currently, SYSTRAN System is available for translating in 29 language pairs.

[21, 53]

**ALPS (1971),** is a direct approach based system, which is used to translate English into French, German, Portuguese and Spanish. It was started with an aim to develop fully automatic MT System but later in 1973, it became Machine Aided System. It is an Interactive Translation System that performs global analysis of sentences with human assistance, and then performs indirect transfer again with human assistance. But this project was not successful and hence not operational. [54]

**SuSy (1972),** The Russian-German prototype was starting point for the research on multilingual system known as SuSy. The languages involved included German, Russian, English and French. SuSy used transfer approach for translation. The emphasis was being laid on analysis and generation of German. The translation steps were broken into several sub-processes. The analysis and synthesis models were language specific.[27]

**The METEO (1977),** is a truly fully automatic MT System for Canadian Meterological Centre’s(CMC’s), nationwide weather communication networks. METEO scans the network traffic for English weather reports, translates them directly into French, and sends the translation back over to the communication network automatically. This system is based on the TAUM technology. It is probably the first MT system where translators had involved in all phases of the
design, development and refinement. Rather than relying on post-editors to discover and correct errors, METEO detects its own errors and passes the offending input to human editors and the output deemed correct by METEO is despatched without human intervention. The correctness of this translation system varies from 90-95%, shuttling the other 5-10% to the human CMC translators.[55, 56]

**English to Japanese Machine Translation System (1982),** is developed by Makoto Nagao et. al. The title sentences of scientific and engineering papers are analyzed by simple parsing strategies. Title sentences of physics and mathematics of some databases in English are translated into Japanese with their keywords, author names, journal names and so on by using fundamental structures. The translation accuracy for the specific areas of physics and mathematics from INSPEC database was about 93%.[57]

**PONS (1995),** is an experimental interlingua system for automatic translation of unrestricted text, developed by Helge Dyvik, Department of Linguistics and Phonetics, University of Bergen. PONS exploits the structural similarity between source and target language to make the shortcuts during the translation process. The system makes use of a lexicon and a set of syntactic rules. The system had been tested on translation of sentence sets and simple texts between the closely related languages Norwegian and Swedish, and between the more distantly related language i.e. English and Norwegian. [58]
Multi-Lingual Translation System (MLTS) (1997), is created by CIMOS, a French based company, that translates between Arabic, English and French. It incorporates lexicon entries and idioms saved for more than 15 years of conventional translation services. It is automated translation software conceived to help translators but not to replace them. It is capable of providing quickly the rough draft of an understandable and acceptable translation. MLTS produces a high quality draft translation at the rate of about 100,000 words per hour. The translator can then proceed to the proof-reading phase and adapt the translated text to the target environment. MLTS uses a semantic analyzer and universal grammar. Translation Memory capabilities are also included in MLTS. [59, 60]

English-to-Filipino MT system (2000), is a transfer based MT System that is designed and implemented using the lexical functional grammar (LFG) as its formalism. It involves morphological and syntactical analyses transfer and generation stages. The translation process involves only one sentence at a time. [61]

Turkish to English Machine Translation System (2000), is a hybrid Machine Translation System combined of two different approaches to MT. The hybrid approach transfers a Turkish sentence to all of its possible English translations, using a set of manually written transfer rules. Then, it uses a probabilistic language model to pick the most probable translation out of this set. The system
is evaluated on a test set of Turkish sentences, and compared the results to reference translations. The accuracy comes out to be about 75.6%.[62]

**Bulgarian-to-Polish Machine Translation System (2000),** has been developed by S. Marinov. This system has been developed based on the approach followed by PONS discussed above. The system needs a grammar comparison before the actual translation begins so that the necessary pointers between similar rules are created and system is able to determine where it can take a shortcut. The system has three modes, where mode 1 and 2 enable system to use the source language constructions and without making a deeper semantic analysis to translate to the target language construction. Mode 3 is the escape hatch, when the Polish sentences have to be generated from the semantic representation of the Bulgarian sentence. The accuracy of the system has been reported to be 81.4%.[63]

**Tatar (2001),** a Machine Translation System between Turkish and Crimean, developed by Altintas K. et al., used finite state techniques for the translation process. It is in general disambiguated word for word translation. The system takes a Turkish sentence, analyzes all the words morphologically, translates the grammatical and context dependent structures, translates the root words and finally morphologically generates the Crimean Tatar text. One-to-one translation of words is done using a bilingual dictionary between Turkish and Crimean Tatar.
The system accuracy can be improved by making word sense disambiguation module more robust. [64]

**Antonio M. Corbí-Bellot et. al. (2005)**, developed the open source shallow-transfer Machine Translation (MT) engine for the Romance languages of Spain (the main ones being Spanish, Catalan and Galician). The Machine Translation architecture uses finite-state transducers for lexical processing, hidden Markov models for part-of-speech tagging, and finite-state based chunking for structural transfer. The author claims that, for related languages such as Spanish, Catalan or Galician, a rudimentary word-for-word MT model may give an adequate translation for 75% of the text, the addition of homograph disambiguation, management of contiguous multi-word units, and local reordering and agreement rules may raise the fraction of adequately translated text above 90%. [65]

**Carme Armentano-oller et al (2005)**, extended the idea of A.M.Corbi-Bellot et. al. and developed an open source Machine Translation tool box which includes (a) the open-source engine itself, a modular shallow transfer Machine Translation engine suitable for related languages (b) extensive documentation specifying the XML format of all linguistic (dictionaries, rules) and document format management files, (c) compilers converting these data into the high speed format used by the engine, and (d) pilot linguistic data for Spanish—Catalan and Spanish—Galician and format management specifications for the HTML, RTF and plain text formats. They use the XML format for linguistic data used by the
system. They define five main types of formats for linguistic data i.e. dictionaries, tagger definition file, training corpora, structural transfer rule files and format management files. [66]

**Apertium (2005),** developed by Carme Armentano-oller et. al is an open-source shallow-transfer Machine Translation (MT) system for the [European] Portuguese ↔ Spanish language pair. This platform was developed with funding from the Spanish government and the government of Catalonia at the University of Alicante. It is a free software and released under the terms of the GNU General Public License. Apertium originated as one of the Machine Translation engines in the project OpenTrad and was originally designed to translate between closely related languages, although it has recently been expanded to treat more divergent language pairs (such as English–Catalan). Apertium uses finite-state transducers for all lexical processing operations (morphological analysis and generation, lexical transfer), hidden Markov models for part-of-speech tagging, and multi-stage finite-state based chunking for structural transfer. For Portuguese–Spanish language pair, promising results are obtained with the pilot open-source linguistic data released which may easily improve (down to error rates around 5%, and even lower for specialized texts), mainly through lexical contributions from the linguistic communities involved. [67]

**ga2gd (2006),** is a robust Machine Translation System, developed by Scannell K.P., between Irish and Scottish Gaelic despite the lack of full parsing technology.
or pre-existing bilingual lexical resources. It includes the modules Irish Standardization, POS Tagging, Stemming, Chunking, WSD, Syntactic Transfer, Lexical Transfer, and Scottish Post Processing. The accuracy has been reported to be 92.72%. [68]

**SisHiTra(2006),** is a hybrid Machine Translation System from Spanish to Catalan. It was developed by Gonzalez et. al. This project tried to combine knowledge-based and corpus-based techniques to produce a Spanish-to-Catalan Machine Translation System with no semantic constraints. Spanish and Catalan are languages belonging to the Romance language family and have a lot of characteristics in common. SisHiTra makes use of their similarities to simplify the translation process. A SisHiTra future perspective is the extension to other language pairs (Portuguese, French, Italian, etc.). The system is based on finite state machines. It has following modules: preprocessing module, generation module, disambiguation module and post-processing module. The word error rate is claimed to be 12.5% for SisHiTra system.[40]

**METIS-II: (2007),** is a hybrid Machine Translation System. This system makes use of monolingual corpora for undertaking translation. The system endeavors to make use of minimum resources and tools for both source and target languages. It is useful in such situations where other Machine Translation Systems are not available or are of insufficient quality. METIS-II works just on basic resources. METIS makes use of pattern matching techniques and target language retrieval
for free text translation system. It is used for translation between the European languages like Dutch, German, Spanish to English.[23]

**SYSTRAN Enterprise Server 7 (2009),** is a fully integrated translation server powered by SYSTRAN’s hybrid technology. Based on self-learning techniques, SYSTRAN Enterprise Server 7 can be trained on existing and validated bilingual texts to achieve cost-effective high quality translations. Hybrid technology combines the flexibility of statistical Machine Translation and the reliability of rule based linguistic approach. In February, 2011, the company has announced the availability of a prototype of its hybrid Machine Translation (MT) engines dedicated to the translation of European Union legal documents into 56 language pairs. This prototype is just one of the results of the ongoing cooperative research program between SYSTRAN and Philipp Koehn of Edinburgh University.[44, 45]

### 2.3 Machine Translation for Indian Languages

This section briefly describes some of the translators from English to Indian languages and vice versa and also translators developed for translation among Indian Languages.

**Anglabharati (1991):** Anglabharti is a rule-based Machine Aided Translation system with source language as English and uses a pseudo-interlingua to cater to all Indian languages. Ambiguity/Complexity is handled by post-editing. The system retains all possible ambiguous constructs, and the user has to select the
correct choices using a post-editing window to get the correct translation. The system’s approach and lexicon is general-purpose, but has been applied mainly in the domain of public health. The project is primarily based at IIT-Kanpur, in collaboration with ER&DCI, Noida, and has been funded by TDIL. This is the basic translation process translating the English source language to Pseudo Interlingua (PLIL) with most of the disambiguation having been performed. The PLIL structure is then converted to each Indian language through a process of text-generation. Since the effort in analyzing the English sentences and translating into intermediate is estimated to be about 70% and about 30% effort is needed for text-generation. Thus only with an additional 30% effort, a new English to Indian language translator can be built. In Anglabharti design 90% of the task is done by the machine and 10% is left to human. In the system a set of rules obtained through corpus analysis is used to identify plausible constituents with respect to which movement rules for the pseudo-interlingua is constructed. The idea of using PLIL is primarily to exploit structural similarity to obtain advantages similar to that of using Interlingua Approach. It also uses some example-base to identify noun and verb phrasals and resolve their ambiguities.[36,69]

**Anusaaraka(1995):** Anusaaraka is a machine aided translation system among Indian languages. It has been built with funding from TDIL project. The Anusaaraka system presents an image of the source text in a language close to
the target language. Anusaarakas has been built for five pairs of languages: Telugu, Kannada, Marathi, Bengali and Punjabi to Hindi. The work on the Anusaaraka project started at the Indian Institute of Technology, Kanpur. It is now being carried out at the Language Technologies Research Center, Indian Institute of Information Technology, Hyderabad with financial support from Satyam Computers Private Limited.[70, 71]

**Anubharati (1995)**, used EBMT paradigm for Hindi to English translation. The translation is obtained by matching the input sentences with the minimum distance example sentences. The system stores the examples in generalized form to contain the category/class information to a great extent. This has made the example-base smaller in size and its further processing partitioning reduces the search space. This approach works more efficiently for similar languages that are among various Indian languages. [72]

**Mantra System(1999):** The Mantra system translates appointment letters in government from English to Hindi. It is based on synchronous Tree Adjoining Grammar and uses tree transfer approach for translation. The grammar is specially designed to accept, analyze and generate sentential constructions in "officialese". Similarly, the lexicon is suitably restricted to deal with meanings of English words as used in its subject-domain. [53]
MAT (2002), is a machine assisted translation system for translating English texts into Kannada, and has been developed by K. Narayana Murthy at Resource Centre for Indian Language Technology Solutions, University of Hyderabad. Its approach is based on using the Universal Clause Structure Grammar (UCSG) formalism. The input sentence is parsed by UCSG parser and outputs the number, type and inter-relationships amongst various clauses in the sentence and the word groups, which take on various functional roles in clauses. Keeping this structure in mind, a suitable structure for the equivalent sentence in the target language is first developed. For each word, a suitable target language equivalent is obtained from the bilingual dictionary. The MAT System provides for incorporating syntactic and some simple kinds of semantic constraints in the bilingual dictionary. The MAT system includes morphological analyzer/generator for Kannada. Finally, the target language sentence is generated by placing the clauses and the word groups in appropriate linear order, according to the constraints of the target language grammar. Post Editing tool has been provided for editing the translated text. MAT System 1.0 had shown about 40-60% of fully automatic accurate translations. It has been applied to the domain of government circulars, and funded by the Kamataka government. [73]

VAASAANUBAADA (2002), an Automatic Machine Translation of Bilingual Bengali-Assamese News Texts using Example-Based Machine Translation technique, has been developed by Kommaluri Vijayanand et. al. It involves
Machine Translation of bilingual texts at sentence level. In addition, it also includes preprocessing and post-processing tasks. The bilingual corpus has been constructed and aligned manually by feeding the real examples using pseudo code. The longer input sentence is fragmented at punctuations, which results in high quality translation. Backtracking is used when the exact match is not found at the sentence/fragment level, leading to further fragmentation of the sentence. The results when tested by authors are fascinating with quality translation. [74]

AnglaHindi (2003), is a pseudo –interlingual rule-based English to Hindi Machine-Aided Translation System, developed by Sinha et. al. at IIIT, Kanpur. It is a derivative of AnglaBharti MT System for English to Indian languages. AnglaHindi besides using all the modules of AnglaBharti, also makes use of an abstracted example-base for translating frequently encountered noun phrases and verb phrasals. The system generates approximately 90% acceptable translation in case of simple, complex and compound sentences upto a length of 20 words. [75]

Anglabharti-II (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 is found that the modification in the rule-base is difficult and it may result in unpredictable results, the example-base is 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, provisions were made for automated pre-editing & paraphrasing, generalized & conditional multi-word expressions, recognition of named-entities. It incorporated an error-analysis module and statistical language-model for automated post-editing. 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 by 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. [37, 76]

Matra(2004), is a general purpose fully automatic system for English-Hindi translation. MaTra achieves this by using a judicious mix of corpus-based or statistical tools and techniques for shallow parsing, word-sense disambiguation, abbreviation handling, and transliteration, and rule-based techniques for lexical and structural transfer. The structural transfer component has at its core a relatively simple and intuitive intermediate representation called MSIR – MaTra Structured Intermediate Representation that can accommodate most types of sentences that are found in real-world texts. MaTra does not attempt either a deep parse or an elaborate semantic analysis of the English sentence. The
parsing strategy is hybrid – a combination of statistical and rule-based techniques. One of the primary goals of the design is graceful degradation from full sentence structures all the way down to word-by-word structures. Essentially, MaTra follows a structural and lexical transfer approach, using semantic information only when required. The domain being explored is news, but the approach is applicable to any domain. [77]

**ANUBHARTI-II (2004),** has been generalized to cater to Hindi as source language for translation to any other Indian language. The system uses hybrid Example-based Machine Translation approach which is a combination of example-based approach and traditional rule-based approach. The example-based approaches emulate human-learning process for storing knowledge from past experiences to use it in future. It also uses a shallow parsing of Hindi for chunking and phrasal analysis. The input Hindi sentence is converted into a standardized form to take care of word-order variations. The standardized Hindi sentences are matched with a top level standardized example-base. In case no match is found then a shallow chunker is used to fragment the input sentence into units that are then matched with a hierarchical example-base. The translated chunks are positioned by matching with sentence level example base. Human post-editing is performed primarily to introduce determiners that are either not present or difficult to estimate in Hindi. [76]
**Shakti (2004),** is a Machine Translation System from English to any Indian language currently being developed at Language Technologies Research Centre, IIIT-Hyderabad. It has already produced output from English to three different Indian languages – Hindi, Marathi, and Telugu. It combines rule based approach with statistical approach. The rules are mostly linguistic in nature and the statistical approach tries to infer or use linguistic information. Although the system accommodates multiple approaches, the backbone of the system is linguistic analysis. The system consists of 69 different modules. About 9 modules are used for analyzing the source language (English), 24 modules are used for performing bilingual tasks such as substituting target language roots and reordering etc., and the remaining modules are used for generating target language. The overall system architecture is kept extremely simple. All modules operate on a stream of data whose format is Shakti Standard Format (SSF). [78]

**English-Telugu Machine Translation System,** has been developed jointly at CALTS with IIIT, Hyderabad, Telugu University, Hyderabad and Osmania University, Hyderabad. This system uses English-Telugu lexicon consisting of 42,000 words. A word form synthesizer for Telugu is developed and incorporated in the system. It handles English sentences of a variety of complexity. [42]

**ANUBAAD(2004),** an example based Machine Translation System for translating news headlines from English to Bengali, has been developed by Sivaji Bandyopadhyay at Jadavpur University Kolkata. During translation, the input
headline is initially searched in the direct example base for an exact match. If a match is obtained, the Bengali headline from the example base is produced as an output. If there is no match, the headline is tagged and the tagged headline is searched in the Generalized Tagged Example base. If a match is obtained, the output Bengali headline is to be generated after appropriate synthesis. If a match is not found, the Phrasal example base will be used to generate the target translation. If the headline still cannot be translated, the heuristic translation strategy is applied i.e. translation of the individual words or terms in their order of appearance in the input headline will generate the translation of the input headline. Appropriate dictionaries have been consulted for translation of the news headline. [79]

**Hinglish (2004)**, is a Machine Translation System for pure (standard) Hindi to pure English forms developed by R. Mahesh K. Sinha and Anil Thakur. It had been implemented by incorporating additional layer to the existing English to Hindi translation (AnglaBharti-II) and Hindi to English translation (AnuBharti-II) systems developed by Sinha. The system claimed to produce satisfactory acceptable results in more than 90% of the cases. Only in case of polysemous verbs, due to a very shallow grammatical analysis used in the process, the system is unable to resolve their meaning. [1]

**INGIT**, is a Hindi-To-Indian Sign Language MT system that has been built for the railway reservation domain. The system takes input from the reservation clerk
and translates into ISL. The output of the system is an animated representation of the ISL-gloss strings. [80]

**English to Indian Sign Language Machine Translation System** can be used to disseminate information to more than 1.5 million deaf people in India. A framework for the syntactic transfer of English text to ISL is presented. Our approach uses transfer grammar rules for the ISL sentence generation. The prototype MT system developed is a unidirectional system performing a structural transfer between English and ISL. [80]

**IBM-English-Hindi Machine Translation System**, has been initially developed by IBM India Research Lab at New Delhi with EBMT approach. Now, the approach has been changed to statistical Machine Translation between English and Indian languages. [81-83]

**English to {Hindi, Kannada, Tamil} and Kannada to Tamil Language-Pair Example Based Machine Translation (2006)** has been developed by Prashanth Balajapally. It is based on a bilingual dictionary comprising of sentence-dictionary, phrases-dictionary, words-dictionary and phonetic-dictionary and is used for the Machine Translation. Each of the above dictionaries contains parallel corpora of sentences, phrases and words, and phonetic mappings of words in their respective files. Example Based Machine Translation (EBMT) has a set of 75000 most commonly spoken sentences that are originally available in English.
These sentences have been manually translated into three of the target Indian languages, namely Hindi, Kannada and Tamil. [81-83]

**Google Translator (2007)**, is based on Statistical Machine Translation approach, and more specifically, on research by Franz-Josef Och. Before using statistical approach, Google translator is using SYSTRAN for its translation till 2007. Currently, it is providing the facility of translation among 51 languages pairs. It includes only one Indian language Hindi. The accuracy of translation is good enough to understand the translated text. [Internet Source: http://translate.google.com/]

**Punjabi To Hindi Machine Translation System (2007)** is The Punjabi to Hindi Machine Translation System developed at Punjabi University, Patiala. It has been implemented with various research techniques based on Direct MT architecture and language corpus. The Direct MT system is based upon exploitation of syntactic similarities between more or less related natural languages. The system accepts a string in Punjabi language and returns its corresponding Hindi string. Various online applications are also created. An application enables the user to translate a webpage in Punjabi to Hindi on the fly. Another interface enables the users to write E-mail in Punjabi. The message is translated into Hindi and sent to the target email address. The receiver gets the mail in Hindi. The accuracy percentage for the system is found out to be 90.67% [84]
**Sampark (2009)** is an automated system for translating one Indian language to another. Sampark is a hybrid system consisting of traditional rules-based algorithms and dictionaries and newer statistical machine-learning techniques. It consists of three major parts and 13 modules arranged in a pipeline. It is developed by 11 institutions headed by International Institute of Information Technology Hyderabad (IIIT-H). [85]

**Bengali to Hindi Machine Translation System (2009)** is a hybrid Machine Translation System, developed at IIT Kharagpur. This system uses multi-engine Machine Translation approach. It is based on the unfactored Moses SMT system with Giza++ derived phrase table as a central element. This system uses dictionary consisting of 15,000 parallel sysnets, Gazeteer list consisting of 50,000 parallel name lists, monolingual corpus of 500K words both from source and target languages, suffix list of 100 Bengali linguistic suffixes. The BLUE score obtained during system evaluation is 0.2318. [41]

**A Hindi To Punjabi Machine Translation System (2009)** is the system for Hindi to Punjabi translation developed at Punjabi University, Patiala has been implemented with various research techniques based on Direct MT architecture and language corpus. A web site is created with an interface that enables a user to write his input sentence in Hindi and system will produce the output in Punjabi. The present system translates any complex sentence. The System accuracy measure is up to 95%. [86]
**English to Malayalam Translation (2008)(A Statistical Approach):** The system uses statistical models for undertaking translation. Monolingual corpus of Malayalam is used and bilingual corpus is used for English language. This system uses various pre-processing techniques for undertaking translation. The structural difference between English Malayalam pair is resolved applying order conversion rules. This system verifies the translation by using BLEU, F measure and WER evaluation metrics. [32]

**GB Theory Based Hindi To English Translation System(2009):** This Translation system makes use of Government and Binding (GB) theory for undertaking translation. GB emphases common phase structure for all languages. The system takes Hindi as the input language and English as the target language. It consists of a parsing module and generating module. The overall translation is undertaken by making use of an example.[30]

**English to Sanskrit Translation (EST)(2010):** English to Sanskrit Translation system makes use of rule based approach along with Artificial Neural Network (ANN). The author makes use of feed forward ANN to make selection of Sanskrit words such as nouns, verbs, objects and adjectives from English to Sanskrit User Data Vector (UDV). This system employs only morphological markings to identify Subject, Object, Verb, Preposition etc. [24]
Machine Translation at IBM India Research Lab. IBM’s India Research Lab have been working on SMT systems for Indian languages. IBM Research has developed several NLP tools which can be used to speed up the development of SMT systems and improve the quality of translation. They have developed a framework for SMT for Indian languages and have developed a Hindi-English SMT system and an English-Hindi SMT system. They have also developed a sentence-aligned Hindi-English parallel corpus, a monolingual English corpus, a monolingual Hindi corpus and also constructed a statistical dictionary for English-Hindi. [21]

EILMT System, aims to design and deploy a Machine Translation System from English to Indian Languages in Tourism and Healthcare Domains. The project is funded by Department of Information Technology, MCIT and Government of India. It uses statistical model and its primary objective is to initially build an English-Hindi translation system capable of translating free flow text and gradually adapt it to other Indian language pairs as well. [87]. The Consortium Members of EILMT system are C-DAC Mumbai, IISc Bangalore, IIIT Hyderabad, C-DAC Pune, IIT Mumbai, Jadavpur University Kolkata, Utkal University Bangalore, Amrita University Coimbatore and Banasthali Vidyapeeth, Banasthali.
2.4 Summary

The present chapter discusses the various language translation systems developed in the world following different approaches. These include English to French, Spanish to English, ALP (English into French, German, Portuguese and Spanish), GAT (English-Russian), Portuguese-Chinese Translation, ALP (Automated Language Processing) from English to Multiple Languages, SYSTRAN (English-Russian), English-Arabic and Hindi to Punjabi, Punjabi to Hindi, English to Sanskrit, Hindi to English, Telgu to Tamil, English to Telgu Machine Translation Systems. It is concluded that direct approach for Machine Translation is most suitable for closely related languages i.e the languages with similar structure. The indirect and statistical approach is suitable for languages with different structures. The best approach with highest accuracy is the hybrid approach which is a combination of rule based and statistical approach.