RESEARCH ARTICLE

PROBLEMS OF MACHINE TRANSLATION IN *PONNEELAN'S NITHIYAMANATHU* FROM TAMIL INTO ENGLISH

T. ANAND
Ph.D Research Scholar,
Department of Translation, Tamil University, Thanjavur-10

T. ANAND

ABSTRACT

Machine Translation (MT) is still a huge challenge for both IT developers and users. From the beginning of machine translation, problems at the syntactic and semantic levels have been faced. Today despite progress in the development of MT, its systems still fail to recognize which synonym, collocation or word meaning should be used. Although mobile apps are very popular among users, errors in their translation output create misunderstandings. The paper deals with the analysis of machine translation of general everyday language in Tamil to English. The results of the analysis shows that more than two thirds of all the sentences were translated incorrectly, which means that there is a relatively small possibility that a mobile app will translate sentences correctly. The results are disappointing, because even after almost 70 years of MT research and improvement, researchers still cannot offer a system that would be able to translate with at least 50% correctness.

Keywords: Machine Translation, syntactic and semantic sentences, Tamil and English language   pairs, words and phrases.

INTRODUCTION

The translation of natural languages by machine, first dreamt of in the seventeenth century, has become a reality in the late twentieth. Computer programs are producing translations - not perfect translations, for that is an ideal to which no human translator can aspire; nor translations of literary texts, for the subtleties and nuances of poetry are beyond computational analysis; but translations of technical manuals, scientific documents, commercial prospectuses, administrative memoranda, medical reports. Machine translation is not primarily an area of abstract intellectual inquiry but the application of computer and language sciences to the development of systems answering practical needs.

After an outline of basic features, the history of machine translation is traced from the pioneers and early systems of the 1950s and 1960s, the impact of the ALPAC report in the mid-1960s, the revival in the 1970s, the appearance of commercial and operational systems in the 1980s, research during the 1980s, new developments in research in the 1990s, and the growing use of systems in the past decade. This brief history can mention only the major and most significant systems and projects, and for more details readers are referred to the publications listed.

The technology is reaching new heights, right from conception of ideas up to the practical implementation. It is important, that equal emphasis is put to remove the language divide which causes

communication gap among different sections of societies. Natural Language Processing (NLP) is the field that strives to fill this gap. Machine Translation mainly deals with transformation of one language to another. Coming to the machine translation scenarios in India, it has enormous scope due to many regional languages of India. It is pertinent that majority of the population in India are fluent in regional languages such as Hindi, Punjabi *etc.*. Given such a scenario, machine translation can be used to provide an interface of regional language. This chapter aims to survey machine translation systems in India and aboard along with a brief history of machine translation.

Machine Translation

Machine translation is one of the major, oldest and the most active area in natural language processing. The word 'translation' refers to transformation of one language into other. Machine Translation is the process of using computers to automate some or all of the process of translation from one language to another. It is an area of applied research that draws ideas and techniques from linguistics, computer science, artificial intelligence, translation theory, and statistics. It is a focused field of research in linguistic concepts of syntax, semantics, pragmatics and discourse, computational-linguistic approaches such as parsing algorithms, semantic and pragmatic clarification and text generation, descriptive linguistics that deals with lexicon and language rules for particular languages and modeling human knowledge representation and manipulation. Research began in this field as early as in the late 1940s, and numerous methods some based on extensive linguistic theories and some ad-hoc have been tried over the past five decades.

Machine translation can also be defined as, the application of computers to the task of translating texts from one natural language to another. Today a number of systems are available that are capable of producing translations which, even though not perfect, is of sufficient quality to use in a number of specific domains. In the process of translation, which either carried out manually or automated through machines, the context of the text in the source language when translated must convey the exact context in the target language. While seeing from the surface, this seems straightforward, but it is far more difficult. Translation is not a just a word level replacement. A translator, either a machine or human, must interpret and analyze all the elements in the text. Also, he should be familiar with all the issues during the translation process and must know how to handle it. This requires widespread knowledge in grammar, sentence structure, meanings, etc., in the source and target languages, also with understanding with each language 's culture in order to handle idioms and phrases which gets originated from different culture and becomes an important issue that affect the accuracy of the translation.

It will be a great challenge for human to face various challenges in the designing a machine translation system, proficient of translating sentences by taking into consideration all the required information to perform translation. Even though, no two individual human translators can generate similar translations of the same text in the same language pair and it may take several revisions to make the translation perfect. Hence it will be a greater challenge for humans to design a fully automated machine translation system to produce quality translations.

Classification of Machine Translation Errors

Machine Translation receives a lot of criticism (Labutis, 2005; Rimkutė; Kovalevskaitė, 2007, 2007a). However, cost, speed, and size of dictionaries are the factors that largely determine the quality of machine-generated translations. Flanagan (1994) identifies a few reasons why machine translation quality is difficult to evaluate: a text can have more than one correct translation; errors can involve not only single words but also phrases, discontinuous expressions, word order or relationships across sentence boundaries, and as noted by Vilar et al. (2006) it makes simple counting of the number of wrong words in the translation pointless; one error can lead to another; and the cause of errors in MT output is not always clear, as the evaluator usually does not have ability to trace the tests and actions in the software, making it difficult to identify what went wrong in the text translation.

THEME OF NITHIYAMANATHU SHORT STORIES

The short stories Nithiyamanathu by Ponneelan deals with fully centered about the atrocities and injustice done to the women. Even the hard-hearted readers will become soft and merciful. The seed sown will

sprout in them and try to do some good deeds to women to lead their lives peacefully. It also expresses the women's sacrifice to her family. Each story is set in a way that reveals the difficulties of women of all ages. This collection of short stories emphasizes in every condition from childhood to old age. All the fourteen stories are very sensitive and heart melted stories which express about the women's feelings and emotions.

It is a collection of fourteen short stories. All the fourteen stories are based on external world and experiences. It speaks about family and women in particular. In the Indian scenario, women are given secondary treatment. All the fourteen stories of Ponneelan voice out the need for advancement and progress of women. It also speaks about the necessity to give due space for women in all the spheres of life – family and society – on par with men both economically and psychologically. Ponneelan is one of the writers who voice out for the empowerment of women.

Structures of Tamil and English Languages

In Tamil, word order is more flexible, as grammatical relations are signaled by inflections. English relies on word order as a means of expressing grammatical relationships within constructions. In generative linguistics, English with fixed word order is called configurational language and Tamil with fairly free word order is called non-configurational language. The core of the configurationally issue is about the question of special grammatical relation of subject and a different one of object, whatever these relations correspond to different positions in the hierarchy of the sentence. In Tamil, there is little or no evidence for a hierarchy as given below, but very often Tamil differentiates subjects and objects in crucial ways.

Many interesting points will be revealed for the purpose of transferring English language structure into Tamil, if we look at the correlating features of the two languages from the point of view of their typological characteristics as SOV and SVO languages respectively.

1.   Syntactically, English and Tamil are perhaps most saliently different in the basic word order of verb, subject and object in simple declarative clauses. English is an SVO language, meaning that the verb tends to come between the subject and object and Tamil is an SOV language, meaning that the verb tends to come at the end of basic clauses. So the two languages differ in their ordering of certain functional units. For example, English being an SVO language has prepositions, whereas Tamil being SOV language has postpositions.

2.   The affirmative sentence in English which are in SVO order becomes aux + SVO to form interrogative sentences which is a discontinuous order. In Tamil, the interrogation does not change the word order.

3.   English is a highly consistent SVO language. The government constructions observe SVO patterns, as do the nominal modifying constructions – with the exception of descriptive and limiting adjectives in an archaic order. As a consistent language, English exemplifies characteristic features of SVO languages, such as the many patterns that have been developed in the verbal modifying constructions, the wide use of substitutes, and the grammatical processes used to highlight elements of sentences. The verbal patterns make heavy use of auxiliaries, which are also involved as substitutes and in interrogative and negative constructions, differentiating English in this way from (S) OV languages like Tamil. The grammatical process involves function words, again in distinctive constructions like clef ting. Tamil is a typical (S) OV language in which the verb occurs at the final position of a sentence. Word order in the sentence is relatively free, as long as the sentence ends with a main verb. For example, the sentence *Ramu introduced Kala to Arun* in Tamil can have the following word- order- variants.

1. *ramu kalaavai arunukku arimukappatuttinaan*.

2. *ramu arunukku kalaavai arimukappatuttinaan.*

3. *kalaavai arunukku ramu arimukappatuttinaan.*

4. *arunukku kalaavai ramu arimukappatuttinaan.*

5. *arunukku ramu kalaavai arimkappatuttinaan.*

6. *kalaavai ramu arunukku arimukappatuttinaan.*

ai and *kku* are accusative and dative case markers and nominative is unmarked in Tamil. The above sentences are identical in logical content, but are different in discourse presupposition in a very subtle way. Ordinarily, constituents that represent older information precede those that represent newer information. The subject-initial sentence pattern is the most common among the various word order patterns. In declarative sentence with nominal subject and object, the dominant order is almost always one in which the subject precedes the object.

4.   English does not permit any order other than the above in unmarked sentences occurring as single utterances. This constraints applies also in subordination, as in the following sentences:

   Ramu shouted while Kala folded her hands.

5.   In English the verbal qualifiers must precede verbs. This position conflicts with the optimum position for subjects. To express negation, for example, the negative element might be prefixed to the verb.

   In Tamil, the negative element follows the verb:

   kala tan kaikalai matikkavillai

   'Kala did not folded her hands'

   *kala tan kaikalai illai matittaal

    Kala does not fold her hand

   *Kala folded not her hand

6.   English has been characterized by functional syntactic as a language in which the initial segment, or theme, often using old material, sets the scene for the new material, or rhyme.

   Kala folded her hands.

   The subject Kala is one of the important elements of the preceding discourse,   while the predicate folded her hands introduces a new action. SVO order provides a convenient basis for such organization of sentences. The same can be said for Tamil too.

   kala tan kaikalai matittaal. -'Kala folded her hands'

7.   For the basic sentential structures identified for Tamil, the corresponding English

   structures are given.

S V	S V
paravaigal paranthana.	Birds fly
S O V	S V O
somu pandtai erindtaan	Somu threw the ball
S C V	S V C
bala puticaali aavaan	Bala is clever
S A V	S V A
rani palliyil irukkiraal	Rani is in the school
S O A V	S V O A
raja pandtai meecai meel vaittan	Raja kept the ball on the table
S O C V	S V O C
ravi aval tavaru enru nirupittaan	Ravi has proved her wrong
S IO DO V	S V IO DO
muthu avalukku icai karpitaan	Muthu taught her music

(Here in this context A = Adjunct, C = Complement, I O = Indirect Object, DO = Direct Object, S = Subject, V=Verb)

SENTENCES IN TAMIL

In Tamil language, sentence is a grammatical unit consisting of a word or syntactically related group of words which expresses a complete thought.

- *aval arvǫmudan parthal*   She looked eagerly.

- *vanka akka vanka endral*   She called her to come in.

TYPES OF SENTENCES IN TAMIL

Sentences in Tamil may be classified into two types. They are major sentences and minor sentences. A major sentence consists of a subject and a predicate, minor sentence consists of a noun with or without a particle of an address.

MAJOR SENTENCES

Sentences which are constituted of a subject and a predicate are called major sentences. Major sentences may further be classified into three types. They are simple, complex and compound sentences.

SIMPLE SENTENCE

A simple sentence must have a single clause (a single verb) which is independent, and it cannot take another clause.

I always wanted to become a writer. (One clause – one verb)

- *priya, kulanthaiyilirunthu than parvaiyai vilakavillai*

  Priya didn't take her eyes off the child.

- *thai aluvaratha paathu avalukum aluga varuthu*

  On seeing her mother crying, she also started crying.

COMPLEX SENTENCE

A *complex sentence* also has more than one clause but of one them must be an independent clause and the other/others must be a dependent clause (es). There are also some particular connectors for the clauses of a complex sentence to be connected.

- I know that you always wanted to be a writer.

Here, a dependent clause is followed by a connector and an independent clause. The other way around is also possible.

- *anbukunu oru aala adaiyalam kaatanumna kannamava kaatalaam*

If one wants to show the symbol for love, Kannama is the best example.

- *erinji pona ava mugathil oru velicham*

Her face becomes bright though it was burned.

Compound Sentences

A *compound sentence* must have more than one independent clause with no dependent clauses. Some specific conjunctions, punctuation, or both are used to join together these clauses. The main clauses in compound sentences are connected with co-ordinating conjunctions such as anal, atanal, enave, etc. The structure of the main clause in a compound sentence is like that of a simple sentence. Compound sentences in Tamil may also frequently be constructed without any co-ordinating conjunction.

- I like coffee. Mary likes tea. → I like coffee but Mary likes tea.

T. ANAND

Two independent clauses – two verbs.

- *kuninji athuka pinji nethiyila menmaya mutham vachen*

I bent down and pressed a kiss gently on the forehead of the soft face

- *kilavi poi sollura aanalum ayyavoda viral pal illatha vaayala kavvathaan seyyura.*

She has lied but she grips dad's finger with her toothless mouth.

IMPERATIVE SENTENCE

An Imperative sentence expresses an order, request, command, and suggestion. We use this type of sentence on our daily basis. In an imperative sentence, the subject is always the second person that is you, and that's why we can easily recognize them.

An imperative sentence can mainly be found in stories, novels, biographies/autobiographies, literature.

- Imperative sentences can be used to make ads. While writing these sentences, we should know points-

- For normal command, the sentence must end with a full stop.

- For forceful command, the sentence should end with an exclamation sign (!)

  ➢ *thayavu seithu inga va*   Please come here.

  ➢ *kathaatha*   Don't shout.

  ➢ *maligai kadaiku po*   Go to the grocery shop.

VOCATIVE SENTENCE

A vocative is a word or phrase used to address a reader or listener directly, usually in the form of a personal name, title, or term of endearment (Bob and Doctor respectively). The person's name or term of address is set off in the sentence with vocative commas**.** It has a noun in the vocative form with or without a proceeding particle of address.

  ➢ ei, meena   'Hey, Meena'

  ➢ yo, raivarey   'Hey, Driver'

DECLARATIVE SENTENCE

An assertive sentence is also called a declarative sentence, as these sentences assert, state, and declares also. This sentence relates to facts, thoughts, opinions, and beliefs, ending with a full stop. These are the most usable sentences, among others, which can be of any length that contains a subject and an object.

- David loves traveling with friends.

- The moon is the satellite of the earth.

  ➢ *en ilaya magal paatiya kindal pandra*

  My younger daughter mocks at her grandmother'

  ➢ *aval idathu kaiyil vaala ilaiyai vaithirunthaal*

  She has banana leaf in her left hand

INTERROGATIVE SENTENCE

The word interrogative means to ask something. Similarly, an Interrogative sentence means a sentence that asks a question, and this sentence always ends with a question mark sign (?). These sentences always get accordingly

in the form of a question. Occasionally, the interrogative sentence's answers are only yes/no, but sometimes it needs a brief explanation. Wherever we need to collect some information, we use the interrogative sentences.

- When did he will come to my house?

- Did you complete your assignment, which I had given yesterday?

  ➢ *vela pesarathuku munna yen edutha?*

  Why do you take it before fixing the price?

  ➢ *inga vanthu perakkanumnu evvalavu naalak kaathiruntha?*

  How long have you been waiting to be born here?

EXCLAMATORY SENTENCE

A Sentence that expresses some strong, unexpected, and extraordinary emotion and surprise or sudden changing thoughts is called an exclamatory sentence. An Exclamatory sentence can easily be identified, which always ends with an exclamation sign (!).

➢ What the hell, man!

➢ Wow, what a beautiful flower it is!

- *ithu enna vipareetham!*

  What a disaster this is!

- *adeyappa, naatla ippadi panam peruthu pochey!*

  Ahoy! People become too wealthy!

TAMIL SENTENCE CORRESPONDING ENGLISH SENTENCE

Affirmative or assertive sentences

➢ *avan kovilukku ponaan*

➢ He went to temple

Question or interrogative sentences

➢ *avan kovilukku pookiraana?*

➢ Is he going to temple?

Negative sentences

➢ *avan kovilukku pookavillai*

➢ He is not going to temple

Imperative or command Sentences

➢ *kovilukku poo*

➢ Go to Temple

Exclamatory sentences

➢ *aa! evvalavu azhakaana kattitam itu!*

➢ How beautiful the building is!

This functional distinction of sentences is also crucial to our venture in finding translation equivalents in Tamil for English sentences. It should also be noted that the word order plays a crucial part in converting affirmative

sentences into interrogative sentences in English. In Tamil, word order does not play a crucial role while transforming an affirmative into an interrogative sentence, it makes use of clitics.

CONCLUSION

The Tamil to English language, machine translation refers to the use of computers for the task of translating automatically from one language to another. The differences between languages and especially the inherent ambiguity of language make machine translation a very difficult problem. Traditional approaches to machine translation have relied on humans supplying linguistic knowledge in the form of rules to transform text in one language to another. Given the vastness of language, this is a highly knowledge intensive task. Statistical machine translation is a radically different approach that automatically acquires knowledge from large amounts of training data. This knowledge, which is typically in the form of probabilities of various language features, is used to guide the translation process. Statistical machine translation (machine translation) treats the translation of natural language as a machine learning problem. by examining many samples of human-produced translation, machine translation algorithms automatically learn how to translate. Machine translation has made tremendous strides in less than two decades, and new ideas are constantly introduced.

One of the reasons for taking up the statistical machine translation approach is ambiguity. Word sense ambiguity and structural ambiguity create great amount of problem in building rule-based machine translation systems. Words and phrases in one language often map to multiple words in another language. For example, in the sentence, i went to the bank, it is not clear whether the "mound of sand" (karai in tamil) sense or the "financial institution" (vangki in tamil) sense is being used. this will usually be clear from the context, but this kind of disambiguation is generally non-trivial [nancy and veronis,1998]. Also, English and Tamil have their own idiomatic usages which are difficult to identify from a sentence. For example, India and Pakistan have broken the ice finally.

Machine Translation can be defined as the use of computers to automate some or all of the process of translating from one language to another. Machine translation is an area of applied research that draws ideas and techniques from linguistics, computer science, Artificial Intelligence (AI), translation theory, and statistics. Work began in this field as early as in the late 1940s, and various approaches — some ad hoc, others based on elaborate theories have been tried over the past five decades. This report discusses the statistical approach to MACHINE TRANSLATION, which was first suggested by Warren Weaver in 1949 [Weaver,1949], but has found practical relevance only in the last decade or so. This approach has been made feasible by the vast advances in computer technology, in terms of speed and storage capacity, and the availability of large quantities of text data.

A major drawback with the statistical model is that it presupposes the existence of a sentence-aligned parallel corpus. For the translation model to work well, the corpus has to be large enough that the model can derive reliable probabilities from it, and representative enough of the domain or sub-domain (weather forecasts, match reports, etc.) it is intended to work for. Another issue is that most evaluation of statistical machine translation has been with training documents that are very rigid translations of each other (parliamentary proceedings have been widely used). News articles and books, for example, are generally rather loosely translated — one sentence in the source language is often split into multiple sentences, multiple sentences are clubbed into one, and the same idea is conveyed in words that are not really exact translations of each other. In such situations, sentence-alignment itself might be a big challenge, let alone word-alignment. Statistical machine translation techniques have not so far been widely explored for Indian languages. It would be interesting to find out to what extent these models can contribute to the huge ongoing machine translation efforts in the country.

Some of statistical models entirely devoid of linguistic knowledge, but similar (non-linguistic) models have achieved encouraging results. Researchers believe that introducing linguistic knowledge can further strengthen the statistical model. Such knowledge may be in the form of morphological rules, rules about word-order, idiomatic usages, known word correspondences and so on. Intuitively, for translation between English and Tamil (or any other Indian language) such linguistic knowledge might be crucial because of the vast structural and lexical differences between the two languages.

Since statistical machine translation is in some sense word alignment (with probabilities). it can be used for lexicon acquisition also, apart from the larger goal of machine translation. The present system is only in its initial stage. Augmentation by increasing the corpus size and adding linguistic information can enhance the accuracy of the system.

REFERENCES

[1].   Aarts, J and Meijs, W (eds.) 1990. "Theory and practice in corpus linguistics". Rodopi, Amsterdam.

[2].   Annamalai, E. "Corpora Development in Indian Languages", in Agarawal and Pani(eds.)   Information Technology Applications in Language, Script and Speech, New Delhi: BPB Publication

[3].   Arnold D. 2003. "Why translation is difficult for computers" in Somers, H. (ed.). 2003.

[4].   Anusaaraka: Machine Translation in stages. Vivek, A Quarterly in Artificial Intelligence, Vol. 10, No. 3. , NCST, Banglore. India, pp. 22-25.

[5].   Bongers, H. 1947. Historical and Principles of Vocabulary Control. Wocopi, Worden.

[6].   Brown, R. 1973. A first language: the early states. Harvard University press, Cambridge.

[7].   Dash N.S. 2005. Corpus linguistics and Language Technology. Mittal Publications, New Delhi.

[8].   Durai Pandi. 2002. "English-Tamil Machine Translation System", in Kalyansundaram K,(ed.) Tamil Internet 2002: Conference Papers, Chennai: Asian Printers, page 86.

[9].   Hutchins, W.J. 1986. *Machine translation: past, present, future*. Chichester (UK):Ellis Horwood; New York: Wiley.

[10].   Hutchins, W. J. and Somers, H. L. 1992. An Introduction to Machine Translation.Academic Press, London, UK.

[11].   John Hutchins. Machine translation: History and general principles. The encyclopedia of   languages and linguistics, 5:2322–2332, 1994.

[12].   Labov, W. 1969.'The logic of non-standard English', Georgetown Monographs on Language and Linguistics 22.

[13].   Lehmann, T. 1993. A grammar of modern Tamil. Pondicherry Institute of Linguistics and   Culture, Pondicherry.

[14].   MD Okpor. Machine translation approaches: issues and challenges. International Journal of   Computer Science Issues (IJCSI), 11(5):159,2014.

[15].   Rajendran, and Vasugi.G, 2019, English to Tamil Machine Translation System Using Parallel Corpus, Journal, Coimbatore.

[16].   Roberts A.H. Zarechnak, 1994. "Machine Translation", Currents Trends in Linguistics,12 pp 2825-2870.

[17].   Thorndike, E. 1921. A teacher's wordbook. Columbia Teachers College, New York.