

Word Sense Disambiguation Using Semantic Web for Tamil to English Statistical Machine Translation

Santosh Kumar T.S.

Bharathiar University, India.

Type of Reviewed: Peer Reviewed.

DOI: <http://dx.doi.org/10.21013/jte.v5.n2.pl>

How to cite this paper:

T.S., Santosh Kumar (2016). Word Sense Disambiguation Using Semantic Web for Tamil to English Statistical Machine Translation. *IRA-International Journal of Technology & Engineering* (ISSN 2455-4480), 5(2), 22-31. doi:<http://dx.doi.org/10.21013/jte.v5.n2.pl>

© Institute of Research Advances



This work is licensed under a [Creative Commons Attribution-Non Commercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/) subject to proper citation to the publication source of the work.

Disclaimer: The scholarly papers as reviewed and published by the Institute of Research Advances (IRA) are the views and opinions of their respective authors and are not the views or opinions of the IRA. The IRA disclaims of any harm or loss caused due to the published content to any party.

ABSTRACT

Machine Translation has been an area of linguistic research for almost more than two decades now. But it still remains a very challenging task for devising an automated system which will deliver accurate translations of the natural languages. However, great strides have been made in this field with more success owing to the development of technologies of the web and off late there is a renewed interest in this area of research.

Technological advancements in the preceding two decades have influenced Machine Translation in a considerable way. Several MT approaches including Statistical Machine Translation greatly benefitted from these advancements, basically making use of the availability of extensive corpora. Web technology web3.0 uses the semantic web technology which represents any object or resource in the web both syntactically and semantically. This type of representation is very much useful for the computing systems to search any content on the internet similar to lexical search and improve the internet based translations making it more effective and efficient.

In this paper we propose a technique to improve existing statistical Machine Translation methods by making use of semantic web technology. Our focus will be on Tamil and Tamil to English MT. The proposed method could successfully integrate a semantic web technique in the process of WSD which forms part of the MT system. The integration is accomplished by using the capabilities of RDFS and OWL into the WSD component of the MT model. The contribution of this work lies in showing that integrating a Semantic web technique in the WSD system significantly improves the performance of a statistical MT system for a translation from Tamil to English.

In this paper we assume the availability of large corpora in Tamil language and specific domain based ontologies with Tamil semantic web technology using web3.0. We are positive on the expansion and development of Tamil semantic web and subsequently infer that Tamil to English MT will greatly improve the disambiguation concept apart from other related benefits. This method could enable the enhancement of translation quality by improving on word sense disambiguation process while text is translated from Tamil to English language. This method can also be extended to other languages such as Hindi and Indian Languages.

Keywords: Machine Translation, Word Sense Disambiguation, Semantic Web, Ontology.

1. Introduction

MT refers to the use of computers to automate the tasks of translating between natural languages. Development of a full-fledged bilingual, bidirectional MT system for any two natural languages with limited electronic resources and tools is a challenging and demanding task. Many attempts are being made all over the world to develop MT systems for various languages using different approaches, particularly rule-based and statistical-based approaches. MT systems can be designed either specifically for two particular languages, called a bilingual system, or for more than a single pair of languages, called a multilingual system. MT methodologies are commonly categorized as direct, transfer, and Interlingua. The methodologies differ in the depth of analysis of the SL and the extent to which they attempt to reach a language independent representation of meaning or intent between the source and target languages. There are several barriers in reaching a good quality MT output and the one of the most prominent one is the concept if ambiguity in natural languages.

Although we have seen research in Machine Translation for more than two decades it still remains a very challenging task for devising an automated system which could give us accurate translations of natural languages. However, there has been great success in this field because of the development of technologies of the web and consequently there is a renewed interest in this area of research. MT is considered one among the top in emerging technologies which will change the world. The MT technology has definitely

matured in recent years mainly owing to the global government and business needs. It is predicted to become an indispensable tool in everyday work in the days to come.

In this paper, we are proposing a technique which can further improve existing statistical Machine Translation methods by making use of semantic web technologies. This method will enable the enhancement of translation quality by contributing an effective WSD technique while text is translated from Tamil to English language. The suggested method will also be useful in the translation of Hindi other Indian languages to English.

2.0 Statistical Machine Translation

Statistical machine translation (SMT) is a machine translation paradigm where translations are generated on the basis of statistical models whose parameters are derived from the analysis of bilingual text corpora.

The statistical approach comes under Empirical Machine Translation (EMT) systems, which rely on large parallel aligned corpora. Statistical machine translation is a data-oriented statistical framework for translating text from one natural language to another based on the knowledge and statistical models extracted from bilingual corpora. In statistical-based MT, bilingual or multilingual textual corpora of the source and target language or languages are required.

There are three different statistical approaches in MT, viz. Word-based Translation, Phrase-based Translation, and Hierarchical phrase based approaches. The idea behind SMT comes from information theory. A document is translated according to the probability distribution function indicated by $p(e|f)$, which is the Probability of translating a sentence f in the SL F (for example, Tamil) to a sentence e in the TL E (for example, English).

There are several mathematical models of the translation process which could be trained statistically from what is known as parallel corpora. Parallel corpora are document collections which have been manually translated sentence-by sentence from one language to another. One such effort is currently under development under the DeitY funded project on Indian Languages Corpora Initiative¹. Literally millions of sentences of electronic text are available as raw material and as annotated text (initially at POS and chunk levels of annotation) for machine translation in Modern Indian languages and English. For languages other than a few parallel corpora are difficult to come by, and the focus of much current research is in bootstrapping statistical machine translation in the absence of large resources (Al-Onaizan et. al. 2000). Such will be the case for the application of statistical machine translation to the Tamil language (as corpora in most of the major Indian languages are still under development). For a more detailed description of corpora development in Indian languages, see Jha (2010).

One of the leading users of SMT is Google and Google Translate does translations based on patterns found in large amounts of text. In this translation, instead of trying to teach the machine all the rules of a language (generally known as machine learning), SMT lets computers discover the rules for themselves. It works by analyzing millions of documents that have already been translated by humans (similar to Translation Workstations or Translation Memory systems). The major components of SMT are the translation dictionaries, patterns and rules that the program develops. Google admits this approach to translation inevitably depends on the amount of texts available in a particular language.

¹For details on ILCI Project, see <http://sanskrit.jnu.ac.in/projects/ilci.jsp?proj=ilci>

According to Knight (1999), “We want to automatically analyze existing human sentence translations, with an eye toward building general translation rules we will use these rules to translate new texts automatically.” These statements point to the knowledge included in vast amounts of texts available in digital form on the internet, partly in the form of human sentence translations. At the same time that MT started clearly moving into using the Web to search for machine readable texts and translations that could be used in the expanding SMT. Berners-Lee & Hendler (2001) defined the knowledge, that is included in the Web content, to expand the traditional WWW to become a Semantic Web. As we are looking at an expanded view of how to use the Web, and specifically the Semantic Web, for our approach of MT, we would like to draw parallels between what has been said so far about MT and the innovative possibilities that the Semantic Web provides for MT research.

3.0 Semantic Web

The World Wide Web (WWW) was once designed to be as simple, as decentralized and as interoperable as possible (Berners-Lee, 1999). The Web evolved and became a huge success, however, information was limited to humans. In order to make information available to machines, an extending and complementary set of technologies was introduced in the new millennium by the W3C, the Semantic Web (Berners-Lee & Hendler 2001). The base technology of the Semantic Web is the data format, i.e. Resource Description Framework (RDF). It defines a structure that is meant to “be a natural way to describe the vast majority of the data processed by machines” (Berners-Lee & Hendler 2001). RDF expresses meaning by encoding it in sets of triples, composed of subject, predicate and object, which are, in the N3-notation format, likewise written down as triples: :subject :predicate :object

Following the definition of Tim-Berners-Lee, “The Semantic Web will bring structure to the meaningful content of the web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users”(Berners-Lee, Hendler & Lasilla 1999), the content (or documents) on the web are machine readable but not machine understandable. The main aim of Semantic Web is to enrich documents with semantic information about the content and to develop powerful mechanisms capable of interpreting this information. These goals are achieved through implementation of models, standards as well as annotation of resources. For a detailed information on the layers and Layer Cake architecture see Berners-Lee, 2003).

We see strong connections between MT and the W3C Semantic Web. A lot of ideas exist on how to augment the Resource Description Framework (RDF) - the base format of the Semantic Web - with natural language. Since the beginning, RDF itself provided capacities for a “human-readable version of a resource’s name” (Guha, 2004). Also in the area of relational semantics, particularly in WordNets, which might be considered as a more traditional NLP domain, W3C Semantic Web technology plays a role, as approaches were developed to bridge the gap between natural language representations within these WordNets and the design principles of the Semantic Web (Graves & Gutierrez, 2005). The conversion of Princeton WordNet², for example, to RDF/OWL is covered by a W3C Working Draft or the GermaNet WordNet equivalent approach by adapting the ideas of the Princeton WordNet conversation. Similar efforts are being made in the construction of Indian language WordNets, i.e. IndoWordNet³.

3.1 Semantic Web Technology and MT

The semantic web provides us with an array of possibilities with efficient technologies such as XML,RDFS,OWL,SPARQL, etc.The semantic web builds upon RDF (Resource Description Framework)

²<http://wordnet.princeton.edu>

³<http://tdil-dc.in/indowordnet>

and RDFS (Resource Description Framework Schema). Web Ontology Language (OWL) defines the types of relationship that can be expressed in RDF to indicate the hierarchies and relationships between different resources in other words it can be referred as ontology. When RDF resource descriptions are associated with an ontology defined somewhere on the web, intranet, or internet, it becomes possible for machines to retrieve the semantic information associated with each resource.

Enhancing NLP or MT with Semantic Web technology is a novel idea, and in this new approach, it is suggested to base MT on a newly defined set of rules, which differ both from rules known from earlier MT approaches but also from any rules that are applied in SMT. These rules follow Tim Berners-Lees vision, in that knowledge, once defined and formalized, is accessible in arbitrary ways.

So we will take advantage of the powerful Semantic Web tool set with the goal to produce a semantically good translation for the given sentence. In this paper we will develop the method making use of these tools to demonstrate the improvisation of Word Sense Disambiguation (WSD) process. Here we are working with the inherent lexical ambiguity of the source text and proceed to implement the disambiguation process without extensive parsing of lexical resources such as WordNet, OntoLex and other online-dictionaries etc. This method will make use of the existing markup information in available Corpora and Semantic web technology stack in the disambiguation process.

4.0 Review of Literature

Although a number of works have been reported in MT making use of various methodologies, comparatively much less has been done or reported for Dravidian languages such as Tamil, Kannada, Telugu, Malayalam, etc. Recently, Google released the alpha version of the MT online services for Tamil, Kannada, and Telugu. Google uses the Statistical Machine Translation (SMT) approach. The quality of the translation is not bad for the simple and frequently used sentences. Mostly, the translation output is comprehensible even though the long sentences have issues with the word ordering and morphological generation of the Tamil words.

Renganathan (2002) demonstrated a functional English-Tamil Rule based MT system with limited set of words and rules. No further work had been reported after that. Germann (2001) reported a SMT system trained using 5000 sentence parallel corpora. Most of the research and development in Tamil NLP has been reported by AU-KBC research centre⁴ which also includes a Prototype English-Tamil MT. The English to Indian Language Machine Translation (Anuvadaksh EILMT) consortia of 8 academic institutions and 2 government organizations focuses on developing MT systems in the domains of Health and Tourism. for 6 language pairs including English to Tamil.

5.0 Methodology

In this paper, we attempt to devise a method and will demonstrate by taking a couple of Tamil sentences with lexical ambiguity and suggest the disambiguation process. This method will make use of the existing markup information in available Corpora and Semantic web technology stack in the lexical disambiguation process.

The Semantic web technology we suggest will come into play during the Lexical disambiguation in the statistical MT process.

⁴ www.au-kbc.org

The state of the art technology of the World Wide Web to express information, facts and relations for both humans and machines is RDF. So it is not unlikely that nowadays expert knowledge is encoded in that format, too. In order to demonstrate how our approach will enhance existing MT systems, we chose to try out Google Translate.

Let us consider this sample Tamil sentence: “Kalaingar Ammavidam Pesinaar” meaning The political personality of Tamil Nadu – Dr. Karunanidhi a.k.a “kalaingar” spoke to another political personality also from Tamil Nadu, Dr. Jayalalitha –a.k.a “Amma”. The meaning will be understood perfectly well especially in the geographical area of Tamilnadu, but since the contextual meaning is complicated and as such the normal Machine Translation results in total Irrelevance:

Google Translate:

கலைஞர்	அம்மாவிடம்	பேசினார்
Mom	Talked	to the artist

Here we can conclude that following sentence cannot be rendered correctly in the target language – English without the help of “expert knowledge” that the word “amma” although in literally means “mother” or “mom”, in this context explicitly refers to the named entity, the name of a politician for whom the pet name is associated with. The same is applicable to the word “kalaingar” who a political personality is bearing a pet name which literally means an ‘artist’.

The translation fails miserably because it does not take any semantic relations and context into consideration. This is a systematic issue in MT, demonstrating the necessity of including world knowledge in the computation of the target translation. Ambiguities are a common MT problem, let us try to resolve this issue below with the help of the semantic web technologies.

For this example we are considering a semantic web implementation called friend-of-a-friend *foaf* database. Using this tool one can use the input field to search through millions of interconnected persons, organizations and places in the semantic web. One needs to enter the name, e-mail, nick, homepage, openid, mbox-hash or URI of the person, organisation or place of the person one is searching for.

In this context a sample FOAF database is simulated. This database is in the RDF file format with details of the person, in the format of triples :subject :predicate :object.

```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:foaf="http://xmlns.com/foaf/0.1/"
xmlns:admin="http://webns.net/mvcb/">
<foaf:PersonalProfileDocumentrdf:about="">
<foaf:makerrdf:resource="#me"/>
<foaf:primaryTopicrdf:resource="#me"/>
<admin:generatorAgentrdf:resource="http://www.ldodds.com/foaf/foaf-a-matic"/>
<admin:errorReportsTordf:resource="mailto:leigh@ldodds.com"/>
</foaf:PersonalProfileDocument>
<foaf:Personrdf:ID="me">
<foaf:name>KarunanidhiMuthuvel</foaf:name>
<foaf:title>Dr</foaf:title>
```



```

<foaf:givenname>Karunanidhi</foaf:givenname>
<foaf:family_name>Muthuvel</foaf:family_name>
<foaf:nick>Kalaingar</foaf:nick>
<foaf:mbox_sha1sum>5ad8292ef9ba9c654bbebfc3c548c37218a3ac82</foaf:mbox_sha1sum>
<foaf:homepagerdf:resource="www.kalaingar_wsd.com"/>
<foaf:phonerdf:resource="tel:12345678"/>
<foaf:knows>
<foaf:Person>
<foaf:name>Jayalalitha</foaf:name>
<foaf:mbox_sha1sum>13f1f54dc5fba6b4eab4af8abc48691ab52fa885</foaf:mbox_sha1sum>
<rdfs:seeAlsordf:resource="amma"/></foaf:Person></foaf:knows>
<foaf:knows>
<foaf:Person>
<foaf:name>stalin</foaf:name>
<foaf:mbox_sha1sum>2c677e38d4a44ca4b4d2780738d06f83f7232c3a</foaf:mbox_sha1sum></foaf:Per
son></foaf:knows>
<foaf:knows>
<foaf:Person>
<foaf:name>kanimozhi</foaf:name>
<foaf:mbox_sha1sum>b212800cdbfca22911c1188ae32cda648a31c571</foaf:mbox_sha1sum>
<rdfs:seeAlsordf:resource="kani"/></foaf:Person></foaf:knows></foaf:Person>
</rdf:RDF>

```

In the above foaf database we have the details of Mr. Karunanidhi including first name, lastname, nickname, e-mailid, etc. We can make a simple query using the SPARQL language to get to the real contextual meaning of ‘Kalaingar’ which is ‘Karunanidhi’. This query will complement the traditional methodology in SMT which only calculate the frequency of occurrence of the word ‘Kalaingar’ which will lead to undesired results. We should keep in mind that here the task is not to disambiguate the literal meaning of the word ‘Kalaingar’ which means ‘artist’ in English, but on a larger scale to specifically arrive at the context of the word in question.

Similarly the contextual meaning of ‘amma’ here in this context can be queried using sparql from the RDF file below:

```

<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:foaf="http://xmlns.com/foaf/0.1"
xmlns:admin="http://webns.net/mvcb/">
<foaf:PersonalProfileDocumentrdf:about="">
<foaf:makerrdf:resource="#me"/>
<foaf:primaryTopicrdf:resource="#me"/>
<admin:generatorAgentrdf:resource="http://www.ldodds.com/foaf/foaf-a-matic"/>
<admin:errorReportsTordf:resource="mailto:leigh@ldodds.com"/>
</foaf:PersonalProfileDocument>
<foaf:Personrdf:ID="me">
<foaf:name>JayalalithaJayaram</foaf:name>
<foaf:title>Dr</foaf:title>
<foaf:givenname>Jayalalitha</foaf:givenname>
<foaf:family_name>Jayaram</foaf:family_name>

```

```

<foaf:nick>Amma</foaf:nick>
<foaf:mbox_sha1sum>3567ef36911826f84da736cbc534ef9685d01faa</foaf:mbox_sha1sum>
<foaf:homepagerdf:resource="www.jayalalitha_wsd.com"/>
<foaf:phonerdf:resource="tel:987654321"/>
<foaf:workplaceHomepagerdf:resource="www.tn.gov.in"/>
<foaf:workInfoHomepagerdf:resource="https://en.wikipedia.org/wiki/Jayalalithaa"/>
<foaf:schoolHomepagerdf:resource="www.bishopcottongirls.com"/>
<foaf:knows>
<foaf:Person>
<foaf:name>jayafriend1</foaf:name>
<foaf:mbox_sha1sum>4da61572b73adbf3c7ee3bc246dec5bb6b28f771</foaf:mbox_sha1sum></foaf:Person>
</foaf:knows>
<foaf:knows>
<foaf:Person>
<foaf:name>jayafriend2</foaf:name>
<foaf:mbox_sha1sum>6b5a1869a4060dea6e79210854e78dcb4bce40b6</foaf:mbox_sha1sum></foaf:Person>
</foaf:knows></foaf:Person>
</rdf:RDF>

```

In the above example it is the proper name/nick name mapping that is discussed, we can see another example:

Tamil: அவன் அம்மாவுடன் கற்பகம் கல்லூரிக்குச் சென்றான்.

It is translated as “Karpagam with his mother and went to college”.

அவன்	அம்மாவுடன்	கற்பகம்	கல்லூரிக்குச்	சென்றான்
his	with mother	Karpagam	to college	went

Here the word “Karpagam” is translated as a name of a person and the so the meaning of the complete sentence is changed.

The correct Translation should be “He went with his mother to Karpagam College”.

Traditional WSD techniques may fail to relate the word ‘karpagam’ to a place of learning or college simply because those methods basically look for the relationship between the nearby words ‘karpagam’, ‘College’ and ‘mother’, instead of treating ‘karpagam College’ as one entity. The occurrence of such strings may be rare in traditional dictionary databases and corpora. At the same time a simple google search on the internet for the word ‘karpagam college’ brings up about 377,000 matches and out of which the first 14 pages at least directly relate to the college named Karpagam. This clearly indicates that the current translation systems lack the ability to disambiguate with accuracy.

Specifically with translations, the understanding the context of the source text correctly is very important. Such ambiguities are sure to lead to incorrect translations, as the examples above demonstrate.

Currently disambiguation is done only based on simple statistical methods. Semantic annotation of the examples, as well as the input text would increase the translation accuracy. This makes sense especially for translation of on-line resources which are supposed to be subsequently annotated.

As per the Google translation process the SMT method used generates translations based on patterns found in large amounts of text. Instead of trying to teach the machine all the rules of a language, the system effectively lets computers discover the rules for themselves. It works by analyzing millions of documents that have already been translated by humans. But in the above mentioned case, the translation fails miserably because it does not take any semantic relations and context into consideration. This is a systematic issue in MT, demonstrating the necessity of including world knowledge in the computation of the target translation. Ambiguities are a common MT problem, let us try to resolve this issue below with the help of the semantic web technologies.

5. Summary and Future Directions

This paper portrays the concept of WSD using semantic web techniques in translating text from Tamil to English language. It presents the main principles of semantic Web along with the possibility of its contribution to the improvement of MT system. An attempt is made to plan a strategy for integrating WSD in a modular way into the SMT system, which in the fully developed stage can perform a complete phrasal multi-word disambiguation. A suggestion for incorporating semantic web component in the process of WSD is presented here. However the architecture supposes the existence of the repositories for syntactic and semantic annotations as well as the ontologies in Tamil and English languages. To make things work for MT with semantic web technologies such repositories have to be created for different languages and the texts have to be annotated and ontologies available for a broad spectrum of linguistic phenomena. These methods can also be successfully implemented in translations involving Hindi and other Indian languages.

REFERENCES

- Allemang, D. & J. Hendler. 2011. *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*. Elsevier.
- Al-Onaizan, Yaser, et. al. 2000. 'Translating with Scarce Resources', In: *Proceedings of the National Conference on Artificial Intelligence (AAAI)*.
- Bechhofer, S. 2009. OWL: Web ontology language. In *Encyclopedia of Database Systems*. New York: Springer, pp.2008-09.
- Beckett, D. & B. McBride. 2004. RDF/XML syntax specification (revised). *W3C recommendation, 10*.
- Berners-Lee, Tim & Mark Fischetti. 1999. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor*. San Francisco: Harper.
- Berners-Lee, Tim, James Hendler and OraLasilla. 2001. "The Semantic Web", *Scientific American*, pp. 29-37.
- Berners-Lee 2003, Foreword to "Spinning the Semantic Web-Bringing the World wide Web t Its Full Potential", in D. Fensel, D., J. Hendler., H.Lieberman, and W. Wahlster, (eds.), MIT Press, 2003
- Berners-Lee, and J. Hendler, and O. Lasilla, 1999, "The Semantic Web", *Scientific American*, 1999
- Brickley, D. & R. V.Guha. 2004. *RDF vocabulary description language 1.0: RDF schema*.
- Brown, P. et.al 1990. "A Statistical Machine Approach to Machine Translation," *Computational Linguistics*, 16(2):pp.79-85.
- Elita, N. and A. Birladeanu. 2005. 'A first step in integrating an EBMT into the Semantic Web'. In: *Proceedings of the Workshop on Semantic Web Technologies for Machine Translation*, Thailand: Phuket. URL: www.mt-archive.info/MTS-2005-Elita.pdf

- Fensel, D. 2005. *Spinning the Semantic Web: bringing the World Wide Web to its full potential*. Cambridge: The MIT Press.
- Gaspari, F. 2002 “Using free on-line services in MT teaching,” in *Proceedings of the 6th EAMT Workshop on Teaching Machine Translation*, November 14-15, 2002, Manchester, pp.145-153.
- Germann, Ulrich. 2001. “Building a Statistical Machine Translation System from Scratch: How much Bang for the Buck Can We Expect,” In: *Proceedings of the ACL 2001 Workshop on Data Driven Machine Translation*. WO1-1409.
- Graves, A. & Gutierrez, C. 2005, ‘Data representations for WordNet: A case for RDF’. URL: <http://www.dcc.uchile.cl/~cgutierr/papers/wordnet-rdf.pdf>
- Grigoris Antoniou and Frank van Harmelen. 2008. *A Semantic Web Primer (2nd edition)*. Cambridge: The MIT Press.
- Guha, R. 2004, ‘RDF Vocabulary Description Language 1.0: RDF Schema’. URL: http://moodletest.ncnu.edu.tw/file.php/9506/references2009/RDF_schema\1.pdf
- Harriehausen-Muhlbauer, Bettina.,&TimmHeuss. 2014. Semantic Web Based Machine Translation – A New Approach to Machine Translation, *International Journal of Advanced Computer Science*, 4:8, 393-400.
- Jha, GirishNath. 2010. The TDIL Program and the Indian Language Corpora Initiative (ILCI). In: Calzolari, Nicolai et.al. (eds). *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*. European Language Resources Association (ELRA).
- Karthikeyan, A. &D. P. Sengottuvelan.2010. An novel approach using semantic information retrieval for Tamil documents. *International Journal of Engineering Science and Technology*, 2(9), 4424-33.
- Knight, K and J. Graehl. 1998. "Machine Transliteration," *Computational Linguistics*, 24(4).
- Knight, K. &P. Koehn. 2000. ‘What’s New in Statistical Machine Translation’, Tutorial, HLT/NAACL pp. 1–89. URL: <http://www.auai.org/uai2003/Knight-UAI-03.pdf>
- Knight, K. 1999. *A Statistical MT Tutorial Workbook*. Ms.
- Koehn, P. 2010. *Statistical Machine Translation*. Cambridge: CUP.
- Lassila, O. &R. R. Swick. 1998. Resource Description Framework (RDF) model and syntax specification.
- NgongaNgomo, A. C., Bühmann, L., Unger, C., Lehmann, J., & Gerber, D. (2013, May). Sorry, i don't speak SPARQL: translating SPARQL queries into natural language. In *Proceedings of the 22nd international conference on World Wide Web*, ACM, pp. 977-88.
- Pandian, S. L., J. Devakumar&T.V. Geetha. 2008. Semantic information extraction from Tamil documents. *International Journal of Metadata, Semantics and Ontologies*, 3(3), pp.226-32.
- Pascal Hitzler, Markus Krötzsch& Sebastian Rudolph. 2009. *Foundations of Semantic Web Technologies*. CRC Press/Chapman and Hall.
- PéterSzeredi, GergelyLukácsy&TamásBenkő. 2014. *The Semantic Web Explained - The Technology and Mathematics Behind Web 3.0*. Cambridge: CUP.
- Renganathan, V. 2002. An Interactive Approach to Development of English-Tamil Machine Translation System on the Web. *Tamil Internet 2002*, California, USA. 68-73.
- RajendraAkerkar. 2009. *Foundations of the Semantic Web*. Oxford: Alpha Science.
- Trujillo, A. 1999. *Translation Engines: Techniques for Machine Translation*. Verlag: Springer.
- Weaver, Warren. 1949. “Translation”, In: William N. Locke & A. Donald Booth (eds.). 1955. *Machine Translation of Languages: Fourteen Essays*, Cambridge: The MIT Press, 15-23.