The present status and future prospects of Computational linguistics in India[*]

Pradeep Kumar Das[**]

Abstract:

The present paper is divided into four parts. Part one tries to introduce the notion of Computational Linguistics (CL henceforth). It also highlights goals of the discipline. This also examines the problem of cooperation between the language engineers and computer software engineers. The second part introduces and explains term 'artificial intelligence' which is considered as the backbone of CL. It also presents a summary of branches of 'artificial intelligence'. The third part of the paper examines the domains or applications of CL. Having talked about the applications of the CL, the fourth part of the paper examines a variety of ongoing research work in Computational Linguistics in India. Though, it is not intended with any evaluative measures, but the paper also questions the methodology that is used in the ongoing research in the area and raises an unanswered question on the utility of the output of such research. The paper also suggests some measures at the end that could not only shape up the discipline but also help in ascertaining a more fruitful result by a better mutual cooperation of the two dynamic disciplines i.e. Linguistics and Computer Science.

Key words: Computational Linguistics, Language engineers, Computer, Software engineers, Artificial Intelligence, methodology.

1. Introduction

The time is the witness that every era that has made a landmark contribution and has altered the face of the civilization, has always found its due place in the history, and it is for this reason that the history has abundant examples of such epochs each of which has been marked by their distinctive characteristics and they helped mankind to reckon from a fixed line of thinking by paving the path of progress and development. The sparkling inventions and discoveries of 'Stone, Fire, Wheel etc.' in different periods of human civilization have had ever lasting impact in the overall development of mankind in the world. These face-altering periods are known in the history book as 'Stone-age, Copper-age, Bronze-age, Iron-age and so forth. In this connection, while keeping the cyclic movement of the development of human history in pace, I believe that the time has come when we can also vouch for and say that the remarkable contribution from our era (21st Century) is the 'Computer'.

2. Computational Linguistics in layman's term

Computer has indeed proven to be a remarkable possession of human being in modern time so much so that it has changed the very perspective of human thinking altogether. If we recall that the computer came to many developing country as a foe, but very soon we realized that it is a good friend to have around. And no matter where we live today in the modern world, the computer has become almost an inseparable part of our life. We really cannot think our day and night as the normal ones without the help and support of the computer. I would never forget the incident which I witnessed other day at HUFS itself when a professor of Drama and Theatre could not place himself properly in the audience because his personal computer (notebook) malfunctioned that day. So, in a word, computer has sneaked in our life so much that one day we might not be able to think our life without the computer.

The so called first stage of the use of the computer in human society has been very successful and it definitely has benefited us in every possible way. However, as we know that the human civilization is like any dynamic organism. It cannot stop on the way of its progression and development. It has to keep changing in order to accommodate the development of science and technology and introduce new facets of development to the human society for the betterment of the people who live in it.

Something quite similar happened in the field of computer technology and engineering. The computer scientists, having reached to a comfortable level of development in computing process by using the mathematical symbols to create several languages for computer, gradually found their interested in the exploration of a new possible avenue in the field of computer science. They wanted to establish a link between arbitrarily created computer language and languages of human beings. They named this new branch of computer science as 'Natural Language Processing' (NLP hereafter) or 'Artificial Intelligence' (AI hereafter). However, it was necessary for the computer scientists to work together with the linguists for such venture. This cooperation was required because the computer software engineers had to deal with the natural language phenomenon. The knowledge that an individual learns about his/her language is not sufficient or very helpful in order to use the language for the purpose of NLP or AI. There are issues and problems that are found in every language and we can not claim to explain those problems despite the fact that we have been using the language for so many years. It was, therefore, necessary for the computer software engineers to take the help from what I prefer to call the language engineers, the linguists, in order to bringing down the structure and forms of human language almost to the level of 'artificial language' of the

computer programming. I guess the advent of Noam Chomsky, the mathematician turned into linguist, in linguistics was not without any reason. He had foreseen or fore-visioned this prospect of linguistics or it must be just a coincidental that what he stated as 'Generative Enterprise' as early as in 1957 to break down the structures of human language to the level of mathematically viable symbols by introducing the 'concept' of TREE DIAGRAM with its 'generative power' in linguistics, has proven almost as a boon for computational linguistics.

The linguists, particularly working in the 'generative paradigms' [2] by this time were also in search of a domain where they wanted to prove that their understanding about the human language has reached to a level where any new or challenging theorization is possible and they can make use of this theorization and apply it even to the computer to analyse, explain and make the computer understand the human language. Therefore, they agreed to collaborate with the computer scientists in carrying out the research for NLP or AI. They, however, wanted to clarify a basic fact to the researchers working in CL with regard to the degree of logical parsing in human languages. It is quite often quoted in linguistics that human languages are logical but they do not behave logically in all circumstances. Whatever this statement could mean to the people working outside the discipline, but it makes perfect sense for a linguist. The statement simply implies that we, the linguists, can very well describe and explain as to what are the components or the constituents of a language and how they function at a given point of time in certain context. However, we can not attempt to answer any such question which demands why this or that happens in human language. We believe that such question can not be placed properly at any level of linguistic analysis of human language. The AI in the form of NLP thus was started as a new area of research with the abovementioned collaborative efforts in the United States in the 1950s to use computers to automatically translate texts from foreign languages into English.

Soon after the discipline launched the research in this new area, it started getting a good momentum and many other sub-disciplines like Psychology, Mathematics and Statistics etc. showed interests in their work and were ready to provide the probable inputs to strengthen the discipline. It was wonderful to witness that what started once as a mere exercise of translating one language into another, later tried to draw the attention of researchers from other sub-disciplines. With this collaborative effort, the computer software engineers got interested in

---

[1] The reference of the term 'generative paradigm' does not necessarily mean 'Chomskyan paradigm' because there are many other schools of linguistic theorizations which show the procedures of 'generating something new from the existing one by applying some rules' but not all of them follow Chomsky's 'generative model'.

understanding the grammar of both languages, including morphology and syntax. In order to understand syntax, one had to also understand the semantics and the lexicon, and also the pragmatics of language use. In a very short period of time, all these core areas of linguistics became the new field of studies and the new research work started demanding an underlined computing probability. In brief, CL gave a new direction to linguistics and provided the possibilities of new researches.

3. The artificial intelligence as the basis for CL

The artificial intelligence is the science and engineering of making intelligent machine, especially an intelligent computer program. It is thus important to know what Bolshakov and Gelbukh[3] say '....neither increasing the speed of computers, nor refinement of programming tools, nor further development of numerous toy systems for language 'understanding' in tiny domains, will suffice to solve one of the most challenging problems of modern science - automatic text understanding'. The quote by these computational linguists hints at a meaningful collaboration of linguists and engineers for achieving a tough goal in modern time. It means using computers to understand human intelligence; however, it does not have to incorporate the methods that are biologically viable in terms of human intelligence. In general, intelligence for computers involves mechanisms. If the mechanics of doing a task is logical in following its steps, we often get the prize of being intelligent. The AI tries to make computers follow and imbibe the mechanics of carrying out a given task in various specific domains. The computer programs embedded with this AI promises to give impressive performances and results.

The mathematical algorithm is the soul of AI and the accuracy of logical steps in any algorithm decides the success of a computer program. The great mathematical logicians Kurt Godel and Alan Turing established by their experiments in 1930s that there is no guarantee that algorithms can solve any problem in a mathematical domain. For example, whether a sentence of first-order logic is a theorem or not is one issue, and whether a polynomial equation in several variables has integer solution or not is another issue and they depend on the given domains of AI. Humans have solved problems in these domains all the time, and this has been offered as an argument that computers are intrinsically incapable of doing what humans do so easily. This is also the reason that we have to make the computer program very intelligent and we can infuse this intelligence into computer program if we carefully plan its components on the basis of mathematical logic.

---

[3] For more detail see Bolshakov and Gelbukh (2004) Computational Linguistics: Models, Resources and Applications.

It was in 1960s when the two computer scientists Steve Cook and Richard Karp who developed the theory of NP-complete problem domains[4]. The full form of this hypothesis is known as Nondeterministic Polynomial time domain. It was observed by the abovementioned scientists that the computers could solve the problems in these domains; however, it was found that it took the time-exponential in the size of the problem. We, the humans, often solve problems in NP-complete domains in much shorter time than what is given by the algorithms. It is, therefore, important to have algorithms in AI quite capable like human at solving the problems. It is also important to identify the sub-domains for which good algorithms exist because there are AI based computer programs in which problem solving modules are not associated with readily identifiable sub-domains. The theory by which such problems are identified is called the *computational complexity*. The theory has not interacted so far with AI as much it should have done in order to generate more fruitful results. Success in problem-solving by humans and by AI programs seems to rely on properties of problems and problem-solving methods and these have not been identified precisely by the computer scientists working in the domains of AI.

However, before we talk about the help that linguistics has extended to computer scientists, it is important to have a fleeting view of the branches or the integral parts of AI. These branches of AI have been discussed by McCarthy [5] in his various writings and there are lots of interpretations available on the web regarding these branches of AI. The following write-up is an attempt to present these branches in the simplest form to the readers without distorting the main theme from what McCarthy has explained.

a) Logical AI

Something that a computer knows about the world in general is about a specific situation in which it has to act and its goals are all stored in some mathematical or logical language. The program decides what action to take in which situation and what is appropriate for achieving its goal by referring these stored data. In practice this is what happens in a normal PC - the machine depending on the request, takes an action depending on the program. But the complexity, ambiguity and volume of data associated with AI in this context is far more ahead of a normal PC running any normal user friendly software.

---

[4] For further details on the topic please refer to http://en.wikipedia.org/wiki/NP-complete
[5] McCarthy 1990, 2000 and elsewhere, an useful link is http://www-formal.stanford.edu/jmc/mcc59.html

b) Search

Any computer program that is based on AI often has to examine a large number of possibilities e.g. moves in a chess game or inferences in a theorem proving program etc. The continuous path changing techniques and new discoveries are made about how to do this more efficiently in various other domains. The AI researchers name it as 'Search'.

c) Pattern recognition

When a computer program makes an observation of some kind, it is often expected from the program to compare what it sees in terms of a pattern. For example, a vision program may try to match a pattern of eyes and a nose in order to find a face. More complex patterns can also be targeted and studied in this program such as speech recognition of patters in natural language text, in chess positions, and in the events of the history etc. These complex patterns require quite different methods than what the simple patterns do which have been studied the most. The above mentioned voice recognition systems and vision recognition systems have reasonably made their advent into the commercial world and with the help of new research more complex task-based techniques which require more complex logic and vigorous programming are in the making now.

d) Representation

Facts about the world have to be represented in some way for the formulation of the concepts in computer programs. Fortunately and thankfully natural languages of mathematical logic are very helpful and useful to execute this job in AI. However, it is quite often said that the AI in terms of representing the facts for the purpose of building knowledge is still at its early stages and requires lot of development. Since knowledge is available in many forms, the method of representation must be such that it can be easily used in reasoning, it can be easily examined and updated, it can be easily judged as relevant or irrelevant to when applied to particular problems. The scientists have not been able to do justice to the issue of representation of facts, and thus remains an obscure one with regard to the development of AI. The reason that is put forward for this shortcoming is the data or the facts that are voluminous and are constantly changing. Some current methods include constraints, frames, logic, relational databases, scripts, semantic nets and summary charts which are hardly satisfactory for the effective implementation of AI.

e) Inference

The method of 'inference' is a great device in logic. Inference is the process by which explicit representations of knowledge can be created from implicit ones. In AI it is regarded as a device that is responsible for the creation of knowledge. Deductive inference uses a set

of axioms to produce new statements. Inductive inference starts with a set of facts or observations and produces generalizations, descriptions and laws. Inferencing teaches the computer program that from some facts others can be inferred. Mathematical logical deduction is adequate for such purposes, but new methods of *non-monotonic* inference have been added to logical inferencing in the early 1970s. The simplest form of non-monotonic reasoning is the default reasoning in which a conclusion is to be inferred by default, but the conclusion can be withdrawn if the evidence goes contrary to its targeted inference. This has helped the AI to save the time in terms of taking decisions for the correct indexing. For example, when we hear of a bird, we infer that it can fly, but this conclusion can be withdrawn when we hear that it is a penguin. It is the possibility that a conclusion may have to be withdrawn that constitutes the non-monotonic character of the reasoning.

f) Common sense (knowledge and reasoning)

This is one area that might help people who argue against intelligent computers. This is where AI is farthest from human-level. While there has been considerable progress, e.g. in developing systems of *non-monotonic reasoning* and theories of action etc., yet many more new ideas are needed to make the system function better. The Cyc-system[6] which has been added to the system contains a large but spotty collection of common sense facts.

g) Learning from experience

This step in natural language processing is very important. The approaches to AI based on *connectionism* and *neural networking* have to specialize in this process of learning from experience. There is also learning of laws expressed in logic. Programs can only learn what facts or behaviours their formalisms can represent, and unfortunately learning systems are almost all based on very limited abilities to represent information. However, a recent effort is being put in order to generate a log-reference in the programs of AI which could record the steps if they go wrong in computing the desired value and this log-reference can be used later as the prevention of such wrong moves in computation.

h) Planning

The step of planning in AI programs starts with a general fact about the world especially the fact about the effects of actions. It also concerns about the fact of any particular situation and the statement of a goal. From these, they generate a strategy for achieving the goal. In the most common cases, the strategy is just a sequence of actions.

---

[6] http://www.cyc.com/cyc/technology/whitepapers_dir/Cyc_Architecture_and_API.pdf

i) Epistemology

In AI the epistemology is a study of knowledge that is required for solving problems in the computation process.

j) Ontology

As we are aware that ontology is the study of the things that exist. In AI, the programs and sentences deal with various kinds of objects, and we study what these kinds are and what their basic properties are. Emphasis on ontology began in the early phase of computation in 1990s and it is an integral part of any computation process in present time, especially in the case of NLP.

k) Heuristics

A heuristic approach is a way of trying to discover something that is imbedded in a program. The term is used in a variety of ways in AI. *Heuristic functions* are used in some approaches to search to measure how far a node in a search tree can be located from a goal. *Heuristic predicates* that compare two nodes in a search tree to see if one is better than the other.

l) Genetic programming

Genetic programming is a technique in AI that helps the computer programs to solve a task by mapping random samples in the LISP programs. It also helps in selecting the best suitable node in millions of other nodes that are generated.

If we try to put all these branches together and try to see what they are capable of in terms of bringing out or teaching the computer with regard to the intelligence, we find in figure-1 as the evidence of what computer can generate out of given numbers as input:
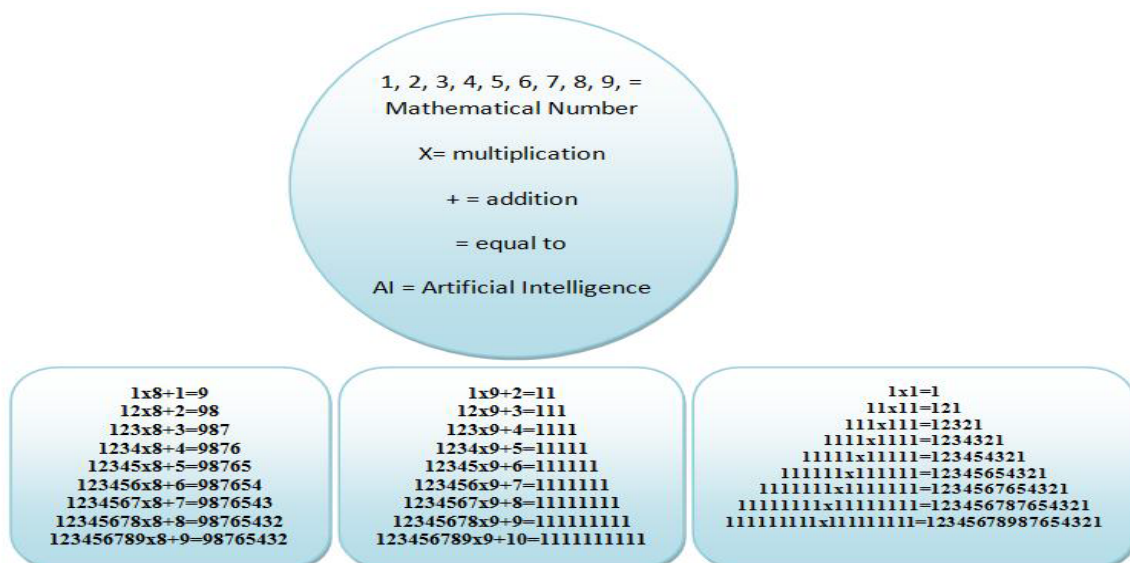

Figure-1

4. The domains of Computational Linguistics

We will now examine various domains in which the use of computational linguistics that itself is based on the abovementioned principles (branches) of AI and have been all around us and made our life comfortable.

a) Search engines: Google et al.

The search engines on internet are the first important applied field of computational linguistics. The data of various kinds are digitized and put on the web hosted protocols. The information is mostly in the free floating digital data format. They are indexed and digitized in natural languages. The search engines are compatible with the AI logic and they are efficient to tag the key-words or the sentences that are put in the search engines and successfully match the data that have been stored in 'World Wide Web' (www henceforth) and mine the target result. The gradual development in the accuracy and the advancement in the search procedure have been possible due the continuous growth of computational linguistics.

b) Web dictionaries

The online dictionaries are available all around the globe. This is facilitated with the hard work put together by the computer programmers and the language experts. With the help of these online dictionaries, the restoration of classical languages like Greek, Hebrew, Sanskrit and Old Tamil has been possible. There are materials and online tutorials available for these classical languages. This could not have been possible without the joint venture of computer programmers and language experts. The exciting and challenging research of computational linguistics has provided us with the facility to connect ourselves with our traditional and cultural heritage by making these classical languages revive again. Most of the texts of ancient time of many countries are easily available on web due to this process of digitization of the text-materials.

c) Web translation/transliteration

The new facility offered by 'Google Indic Transliteration' has proven almost like a boon for the academia as well to the people who are engaged with the work in Indian languages. We can get the targeted texts in Indian languages like Bengali, Gujarati, Hindi, Kannada, Malayalam, Marathi, Nepali, Tamil, Telugu etc. in the easiest way by using 'Romanised version of these Indian languages'. This is, to my mind, the awesome gift from the computational linguistics to the society. There are many portals on the web where we can translate one language to another without much problem. This too is being made possible due the hard work of the computational linguistics. It uses the AI based techniques to tag the

attempted strings from one language to another and brings almost the accurate results. The choice of the languages at the moment, however, remains limited and there is a need to do more research and enrich the data bank by adding other languages in the list and efforts are being made in this direction.

d) Text Messages of Cell phones

The cellular phones are getting smarter day by day in every part of the world, and this has been possible due to the new research that is being carried out in the field of computational linguistics. The text messages in language(s) of the place/country have a great impact at least on the youngsters in the society. The cellular companies like Nokia, Vodafone, Sony, LG, Samsung and others have now established their own labs for IT professionals and the computational linguists. The predictive text that one can type so fast on the cellular phone has been made possible due the research work in computational linguistics. In the predictive text messages procedure, the three or four alphabets of a language are put into one key-button of a phone. Now, imagine that we have to type a word 'before' on the screen for the text message; there are two ways of doing it. First, without using the smart-messaging technique, we can press the buttons containing '*abc*', '*def*', '*mno*', '*pqrs*' and '*def*' many times (15 times) to get a simple word like 'before'. However, if we use the smart predictive text-messaging, we can type this word just by pressing the above mentioned buttons five times. This is the simplest application of CL.

e) Voice and speech recognition system

Microsoft office in 1999-2000 started this additional tool to make it possible for the *OFFICE* users to allow them to type the documents in the computer hands-free i.e. without using the key-board. It is a wonderful tool, however, it did not pick up so much of the attention of the computer users because every individual has to train the PC with his/her voice recognition process before one can actually use this tool. However, it was a big achievement for computational linguistics. We must put lots of effort and make contribution in this area of computation. This will be a great help for the upcoming generation of human society.

f) Braille computer

Computational linguistics has given a marvellous gift to the physically challenged people in the society in the form of Braille computers. The joint venture of software engineers and language engineers (linguists) will always be appreciated for this contribution to the society. Lots of new research is being carried out every day in this area to make the machine better and user friendly.

g) GPS with voice activation

The so called Global Positioning System that we have got in our cellular phones and an add-on device in our vehicle has proven very useful ever since it has the voice activation system in it. This helps us to locate a destination and reach there without any hassle when we are driving our vehicle and our hands are not free to keep the map unfolding. This too has been made possible due to the current research in computational linguistics.

h) Automated bank and phone machines

There are many banks in the world which are totally voice activated and work very efficiently and they do not have any security threat because the voice activated machine are very advanced in terms of the security system. It has been practically demonstrated that even when one does the mimicry, the acoustic-token shown on the spectrogram can never be mapped similar between two individuals. The computational linguistics has made this possible again to have such a safe banking system. The so called '*Pin no*' which is mostly four-digit code of security is created in every bank through the voice-activation system. This too is a case of computational linguistics.

i) Mining of the legal documents

The legal documents in Europe and America have all been digitized and put in the public domain for cross reference and use of these documents for any kind of litigation case. The computational linguistics is used in the retrieval of these documents and it helps the jury to save time and come to their decision by examining the relevant cases in the past.

j) Mining of medical records

The medical records in many countries have also been digitized with the help of computational linguistics. These documents are made available in the form of data archive for the consultation of the medical practitioners. There are user-friendly softwares for homeopathy as well as allopathic and doctors can use this software to find out the history of medical symptoms and they can use this knowledge for the treatment of the patients.

k) Market statistics for product promotion

This is a very new venture in computational linguistics. This is also known as 'sentiment-analysis' in the field of B to B enterprise. There are softwares that gather data from internet web-pages and blogs for various kinds of products in the market. The companies use this statistical tool for the promotion of their products. Recently, a new tool has been added by the computational linguists in the list. It is known as 'EssayRater'. This tool is very popular amongst the academicians. This software runs the key-words of any literary write-up and mines the information whether any such research work has been already carried out by the

researchers in the world or not. This tool helps the editors and the publishers in the publishing industry to find out the cases of plagiarism. We must be grateful to the effort and hard-work put by the computational linguists to provide such a useful tool.

5. The scenario of Computational Linguistics in India

It is a statistical fact that about 30 million IT professionals work in US. They are treated as the backbone of the IT industry in US. There can be zillions of reasons behind this brain drain. This, in fact, should not bother the academicians. There are better opportunities and handsome salaries and so called better standard of living in US and elsewhere, which might be the driving force for these professionals to go abroad. However, what bothers me is the fact that barely half of these Indian IT professionals in US and elsewhere are no better than the English-speaking clerks that was created by Macaulay during the British Raj in India. If a grocery-store is automated in terms of various statistical details of purchase and sell, the so called foreign IT professionals are hired to make the software for the grocery-store. This is just one example, but there are hundreds of other examples like 'company making credit card reader, beer selling company, all big malls and even a company that has a chain of fast-food outlets and for the automation of these shops foreign software engineers are hired to make the software programs. These IT professions never get tired of boasting that they are software engineers. No one says that they are not software engineers, but what about their contribution to the society! Many of these professionals are trained in the premier institutions like IIT, RIT and BIT where government of India spends millions of dollars to train these professions for nothing but brain drain. This, of course, makes us to pity the poor Indian IT professionals working abroad.

In course of writing this paper, I wanted to know as to how many Indian academicians are there in the Computer Science Department in US universities. I randomly browsed through ten university web-sites for the faculty list and I am shocked to know that there are either no Indian faculty in many cases or in some cases the ratio is too low. I personally think that this a loss on the part of the great Indian tradition where learning and teaching was part and parcel of life and knowledge was considered the highest goal of life. Something seriously has gone wrong here.

This background information will help us re-locate the status of computational linguistics in India. It is still a better state of affair that there are many institutions and centres of learning where computational linguistics is taught and researched. All IITs in India, IIIT, Hyderabad, CIIL, Mysore, HP Lab, Bangalore, Hyderabad Central University, Centre for Computer Science and Sanskrit Centre in JNU, New Delhi, Linguistics Department and Computer

Science Department in Delhi University and NIC of India, CDAC, India and NCST, Mumbai are the main centres where computational linguistics is taught and researched. But if talk about the utility of the research or teaching is done in the above mentioned places, I am sorry to say that in many of these institution, the CL is just being detected in the foetus and this is also because the foreign companies are ready to invest and giving the guidelines for the research and bringing the project to these institutions. I purposely used the term 'foetus' for the discipline because it has not even taken birth in real sense. The best that we have been able to do in this regard is 'the process of digitization of the texts' that can be stored on web-portals. Of course, this is the first good step for the eventual growth of the discipline, but I believe there is a dire need the process of acceleration in this field. And this can be possible only if we do lots of research in our indigenous languages. The core meaning and success of the discipline lies in the mining of the knowledge that is hidden in thousand of the local languages of the country. If we learn this indigenous knowledge and make use of them and make it reach to the rest of the world through the World Wide Web, we can also make our own identity and the world will surely recognize us for our true knowledge. The following are the research works being carried out in India by the computational linguists:

a) **Speech recognition system**

The HP Lab in Bangalore is engaged in carrying out lots of research on speech recognition system on various Indian languages. The work is mostly done for corporate sectors like cellular phone companies and banking services. However, some sporadic research work is also done by the professionals for the development of knowledge based system in Indian languages.

b) **Search engines**

The computational linguists in JNU and IIT, Hyderabad are doing some useful work in the field of computational linguistics. The work at JNU especially focuses at digitizing the old texts of Indian tradition like Mahabharata and others. The 'Google India' has contributed a lot in the field computational linguistics by digitizing the texts and posting the famous 'Indic transliteration' facility on web which transliterates the Romanized version of texts in Bengali, Gujarati, Hindi, Kannada, Malayalam, Marathi, Nepali, Tamil, Telugu very successfully. This is great step towards the development and advancement of Indian languages; however, this facility is limited in a sense that it can only do the transliteration online. The facility of internet and being online 24x7 is itself a privilege that is not available for major chunk of the society.

c) Web dictionaries

There are few web dictionaries available for some Indian languages. There is just one website, 'Shabdkosh[7]' that I have found useful for Hindi and even this does not include many words that we often need to know.

d) Digitization of public documents

The online applications for various purposes that are in natural languages have started coming on the web-portals but there is still a long way to go for the standardization process for these documents. The great contribution of the chief minister of Karnataka, Mr. Chandrababu Naidu, must find a mention here. It is said that he got all the documents related to land possession digitized for the farmers in his state. In another words, the farmers can see their land and property online. Many other remarkable developments took place in the field of digitization of the documents in his regime.

CIIL, Mysore claims to have a huge digital archive in many Indian languages. Since the institute is a government supported institute, there are restrictions in terms of the use and distribution of the archive. However, this is a good news for the computational linguistics and its growth in India.

e) Machine translation

The first attempt of such research was done by the computational linguists in IIT Kanpur, India. The computational linguist, Rajeev Sangal is the one who have done a lot of research work in this area. Other people from various Institutions joined him in this work of machine translation and it is their hard work that has made this field very successful in India. Translating Indian languages in English and vice-versa has been a successful attempt in many cases. The recent process of standardization of fonts used for Indian languages into UNICODE has also helped the computational linguists to achieve a greater degree of success in machine translation. It is expected that it will soon be possible to translate the texts of Indian languages into each other by using the tools of machine translation. If this is made possible, the rich old tradition of Indian culture and local ethnic information will soon be available for the mass usage.

f) Braille computer

There is an urgent need for the computational linguists to do some research work and help the weaker section of the society by providing probably what they deserve in this modern age. I have no idea if there is even one good OCR (Optical character recognition) for the purpose of

---

[7] For further reference, please see http://www.shabdkosh.com/

Braille readers. The Braille computer in Indian languages is something that arrests our immediate attention and the computational linguists must do something about it very soon.

6. Conclusion

As the concluding remark, I must say that I do not quite agree with those who think that it is already too late for India to take some strong steps for the development of Indian languages. However, I do understand that the blind marathon for English language has forced us to ignore the development of Indian languages. It is an old saying that if a nation has to win the race for the development and advancement, it cannot do by adopting a foreign tongue. It is, therefore, very important for India and Indians to think about Indian languages and put a bit more than the maximum of our effort to promote and facilitate each and every possible thing for the development of our languages, and computational linguistics should be made the epicentre of this development. The country has marked its presence on the score board of development of the world, and if we recall, the IT sector of India is the one which is recognised worldwide. It is for this reason too that we must not let our recognition go a waste at any cost. Linguistics, if we trace back its root, can be claimed the gift to the world by India too. Panini's era might be debated for 6th century BC or 4th century BC, however what is remarkable about the era is that we had a fully developed method to analyse a language and a scientific explanation to every of the '*matras*' that was important either for sound, syllable, word or sentences. The Paninian grammar has been attested worldwide to possess the system of 'generative procedures'. There are rules and applications of those rules that work in more or less the same fashion as the modern 'generative rules' work under different approaches of grammar. This is also why we find lots of people using Paninian model of 'generative grammar' for their research works in computational linguistics as well. There is, however, an urgent need to do more and more research work in this area and make our presence on the world map and let the world accept India and Indian knowledge was not only great in the past, but it has the capability to sustain this fame and retain our dignity in the modern time too.

References:

Bharati, A., V. Chaitanya and R. Sangal. 2008, *Computation Linguistics in India: An overview*, Paper presented at 6th International Conference on Natural Language Processing ICON, Hyderabad.

Bharati, A., V. Chaitanya and R. Sangal. 1995, *Natural Language Processing: A Paninian Perspective*, Prentice-Hall of India, New Delhi.

Bolshakov I. and Gelbukh A. 2004, *Computational Linguistics: Models, Resources, Applications*. Instituto Politecnico, Direccion de Publicaciones, Maxico. (http://www.publicaciones.ipn.mx/PDF/1341.pdf)

McCarthy, J. 1990. Generality in artificial intelligence. In Lifschitz, V., ed., *Formalizing Common Sense*. Ablex. 226-236.

Mitkov R. (ed.) (2003) *The Oxford Handbook of Computational Linguistics*. Oxford University Press, Oxford.

Ralph, G. 1994, *Computational Linguistics: An Introduction*, The Press Syndicate of the Cambridge University, USA.

Web-references:

CDAC URL: http://www.cdac.org.in

Corpora URL: http://www.iiit.net

Dictionary URL: http://www.iiit.net

Machine translation URL: http://www.iiit.net / http://www.tdil.gov.in

Artificial Intelligence http://www-formal.stanford.edu/jmc/whatisai/node2.html

Cye-systme http://www.cyc.com/cyc/technology/whitepapers_dir/Cyc_Architecture_and_API.pdf