

**உலகத் தமிழ்
இணைய மாநாடு
INTERNATIONAL
TAMIL INTERNET
CONFERENCE
2 0 1 2**

Topic A

Internet

இணையம்

The Tamil Wikipedia: Criteria for Evaluation and Enrichment

N.Murugaiyan

Chief Resource Person, Central Institute of Classical Tamil

‘Wikipedia’, according to the founder Jimmy Donal Wales, an American financial-trader -turned internet-entrepreneur ‘is a freely licensed encyclopedia written by thousands of volunteers on the Internet in over a hundred languages and it’s one of the fastest growing Web sites’. Wikipedia had a predecessor in Nupedia, launched in March 2000 as the free encyclopedia, a peer-reviewed open-content encyclopedia, that lasted for only three years before it became defunct in September 2003. The Hawaiian word ‘wiki’ means ‘quick’ and the Wikipedia, founded in a collaboratively edited, multilingual free encyclopedia supported by the non-profit Wikimedia Foundation, established in Florida, USA and Headquartered in Sanfrancisco, California, USA, with more than 23 million articles in 285 languages, launched in 2001 has grown so rapidly that the English Wiki alone has more than 4.1million articles and most of the wikis such as Malayalam wiki, Urdu wiki, Arabic wiki etc have more than 1000 articles in each of them. The Wikipedia in its various versions has become the largest and the most popular general reference work on the Internet, having as many as 365 million readers approximately.

The Tamil Wikipedia

The Tamil Wikipedia is a free online encyclopedia like its counterparts such as English, Telugu, Hindi, Gujarati, Kannada, Malayalam etc encyclopedias run by the Wikimedia foundation. The Tamil Wikipedia established in September 2003 has more than 48000 articles by the second week of October 2012. This is the 59th Wikipedia of the world by article count. English wiki with its 4077000 articles is the largest Wikipedia of the world. Among the Indian languages Tamil ranks third, having Hindi and Telugu above it by article count.

Poor Turnout of Articles from Tamil Nadu

Even though the Govt. of Tamil Nadu in the year 2010 conducted a contest, being the first of its kind, for promoting among college students the habit of contributing to the Tamil Wikipedia, the number of articles produced by students and others remaining nothing but meager. In spite of the fact that the general awareness of and the interest in Tamil Studies is on the rise, the number of Wikipedia articles produced ever since the inception of the Tamil edition of the Wikipedia is deplorably low. The contribution of scholars from Tamil Nadu where Tamil is predominantly spoken and used widely in all walks of life is far from satisfactory. The scholars and professionals from Tamil Nadu rarely come forward to make entries to the Tamil Wikipedia while most of the development of Tamil wiki comes from Tamil Diaspora spread over a number of countries such as Malaysia, Singapore, south Africa, France, Germany, Australia, UAE, Switzerland, Canada, the USA, Fiji etc. The percentage of entries received from the former is only 7% while those received from Tamil Diaspora is 93%.

Reasons for the Meager Participation by Mainland Tamils

It has been seen that there is a poor turnout of articles that suffer not only quantitatively but also qualitatively. Besides the participation of the Tamil community in Tamil Nadu in this world class Tamil Wiki project is meager to say the least. The number of contributors from Tamil Nadu where Tamil is predominantly spoken, where Tamil is the official language of the state, where Tamil is used as a language of administration, education, culture etc. in a world class project of this kind is nothing but minimal. Even among those who contribute only a handful among them are either active or very active contributors while the vast majority among the contributors being desultory.

There is a disconnect or mismatch between computing skills needed for potential contributors and the subject knowledge or specialization required for giving a comprehensive account in an encyclopedia. Those who have subject knowledge or specialization may lack in computing skills such as technical ability required for serving either as a contributor or as a collaborative editor. There are many who have technical skills but they lack in knowledge relating to the field chosen for making a presentation in the form of an encyclopedic entry. With scanty or inadequate knowledge in the area chosen for an entry, the articles produced by them tend to become a stub lacking in details and citations. Rarely do we come across a situation in the Tamil mainland i.e., Tamil Nadu where these two skills go together.

Single-authorship and Multiple-authorship

Some of the contributors may be happy as a single-author contributor. They may be uncomfortable with collaborative editing or writing since While involved in collaborative writing or editing, one is expected to ensure the use of stylistic equivalences at various levels of language use. To preserve uniformity in style is a stupendous task with which some of the potential contributors may not be familiar with. This may be a possible reason for many not coming forward to make contributions to the Tamil Wikipedia.

Physical Facilities

One can identify other reasons connected with physical facilities such as non-availability of computer with internet connectivity in urban or semi-urban or rural areas on occasions in which they would like to spend time engaging themselves either in collaborative editing or in writing to the Tamil Wiki. This sort of situation prevails more often in remote or rural areas of Tamil Nadu than in its urban or semi-urban areas. This is particularly true in the case of Tamil Nadu which was in the forefront in getting classical language status to Tamil fighting for this privilege over a period of one hundred years and more.

Two Specimen Entries

Following are two versions of specimen entries relating to the classical Tamil work நான்மணிக்கடிகை (The salver of Four Gems as translated by G.U. Pope) . A careful evaluation of these two entries will focus on problems connected with making a satisfactory entry suitable for the Tamil Wikipedia.

Specimen 1

நான்மணிக்கடிகை பதினெண்கீழ்க்கணக்கு நூல்களுள், ஒன்று. இது ஒரு நீதி நூல் . விளம்பி நாகனார் என்னும் புலவரால் இயற்றப்பட்ட இந்நூல் நூற்றியொரு பாடல்களைத் தன்னகத்தே கொண்டுள்ளது . ஒவ்வொரு பாடலும் நான்கு அடிகளால் ஆனது. இந் நூற் பாடல்கள் ஒவ்வொன்றிலும் , நான்கு மணியான கருத்துக்கள் சொல்லப்படுகின்றன. இதனாலேயே இது நான்மணிக்கடிகை எனப்பெயர் பெற்றது .

எடுத்துக்காட்டு

கல்வியறிவு இல்லாதவர்களுக்கு அவர்கள் வாயிலிருந்து வரும் சொற்கள் இயமன் ஆகின்றன . வாழை மரம் தான் ஈனுகின்ற காயினாலேயே அழிந்துபோகி றது. செய்யத்தகாதவற்றைச் செய்பவர்களுக்கு அறமே இயமன். ஒரு குடும்பத்துக்கு, தீய ஒழுக்கம் கொண்ட பெண்ணே இயமனாவாள் என்ற பொருள்கொண்ட, நான்மணிக்கடிகைப் பாடலொன்றைக் கீழே காண்க .

கல்லா ஒருவர்க்குத் தம்வாயிற் சொற்கூற்றம்
மெல்லிலை வாழைக்குத் தானீன்ற காய்கூற்றம்
அல்லவை செய்வார்க்கு அறங்கூற்றம் கூற்றமே
இல்லத்துத் தீங்கொழுகு வான்

சங்க இலக்கியங்கள்

சங்க இலக்கியத்தை மேற்கணக்கு , கீழ்க்கணக்கு நூல்கள் என்று இரு பிரிவாகப் பிரிக்கலாம். மேற்கணக்கில் பத்துப்பாட்டு எட்டுத்தொகை ஆகிய பதினெட்டு நூல்கள் அடங்கும் . கீழ்க்கணக்கில் நாலடியார், நான்மணிக்கடிகை என்று தொடங்கும் பட்டியலிலும் பதினெட்டு நூல்கள் உண்டு . பத்துப்பாட்டு நூல்கள் ஒவ்வொன்றிலும் ஒரே கவிஞரால் எழுதப்பட்ட ஒரு நீண்ட கவிதை உள்ளது. எட்டுத்தொகை நூல்களில் பலரின் கவிதைகள் தொகுக்கப்பட்டிருக்கின்றன . கீழ்க்கணக்கு நூல்கள் பலராலோ அல்லது ஒரு சிலராலோ எழுதப்பட்ட கவிதைகளை கொண்டுள்ளன . மேற்கணக்கு நூல்களெல்லாம் ஆசிரியப்பாவிலோ , பரிபாடலிலோ அல்லது கலிப்பாவிலோ உள்ளன . ஆனால்

கீழ்க்கணக்கு நூல்கள் அனைத்தும் இரண்டு முதல் ஆறடி வரையுள்ள வெண்பாக்களினால் ஆனவையாகின்றன.

கீழ்க்கணக்குநூல்கள்

கீழ்வரும் தனிப்பாடல் கீழ்க்கணக்கு நூல்களை பட்டியலிட்டு காட்டுகின்றது . கீழ்வரும் வெண்பா எளிதாக அவற்றை நினைவுகூற பயன்படுகின்றது. எழுதியவர் யாரென்று தெரியாத பாடல்களில் இதுவும் ஒன்று.

நாலடி நான்மணி நானாற்பது ஐந்திணைமுப்
பால் கடுகங் கோவை பழமொழி -- மாமூலம்
இன்னிலைய வாங் காஞ்சியுடன் ஏலாதி என்பவே
கைநிலை யுமாம் கீழ்க்கணக்கு

Specimen 2

நான்மணிக்கடிகை

நான்மணிக்கடிகை பதினெண்கீழ்க்கணக்கு நூல்களுள், ஒன்று. இது ஒரு நீதி நூல் . விளம்பி நாகனார் என்னும் புலவரால் இயற்றப்பட்ட இந் நூல் நூற்றி யொறு பாடல்களைத் தன்னகத்தே கொண்டுள்ளது [1]. ஒவ்வொரு பாடலும் நான்கு அடிகளால் ஆனது[2] இந் நூற் பாடல்கள் ஒவ்வொன்றிலும், நான்கு மணியான கருத்துக்கள் சொல்லப்படுகின்றன . இதனாலேயே இது நான்குவகை மணிகளால் ஆன ஆபரணம் நான்மணிக்கடிகை என்று அழைக்கப்படுகிறது .

விளம்பி நாகனார்

நான்மணிக்கடிகையின் ஆசிரியர் விளம்பி என்ற ஊரில் பிறந்த நாகனார் என்ற பெயர் உடையவராக இருக்க வேண்டும் என்று கருதப்படுகிறது . அல்லது இவர் ஆற்றிய தொழில் காரணமாக இவர் விளம்பி நாகனார் என்று அழைக்கப்பட்டிருக்கலாம் என்று கருதுவாருமுள் . இந்நூலில் உள்ள கடவுள் வாழ்த்துப்பாடல்கள் இரண்டும் திருமாலைப்பற்றி இருப்பதால் இவர் வைணவ பிரிவை சேர்ந்தவராக இருக்கலாம் என்றும் இந்நூல் வைணவ இலக்கியம் என்றும் கூறுவாறுமுள் . இவர் பெயரை நாயனார் நயினார் என்று கூறி , இந்நூலில் பல இடங்களில் சமண சமயக் கருத்துக்களான பொய்யாமை , கொல்லாமை, புலால் உண்ணாமை ஆகியவை வலியுறுத்திக் கூறப்படுவதால் இதன் ஆசிரியர் விளம்பி நாகனார் ஒரு சமணர் என்றும் நான்மணிக்கடிகை சமணர்களின் இலக்கியம் என்று கூறுவாரும் உள் . இக்கருத்தை ஏற்பவர்கள் திருமாலைப்பற்றிய செய்யுட்கள் இரண்டும் இடைச்செறுகல் என்று வாதிடுவர் . கி. ஆ.பெ. விசுவநாதன் பதிப்பித்துள்ள மும்மணிகளும் நான்மணிகளும் [3] என்னும் நூலில் கடவுள் வாழ்த்து செய்யுட்கள் இரண்டும் நீங்கலான நூற்றுநான்கு பாடல்களே உள்ளன [4].

எடுத்துக்காட்டு

கல்வியறிவு இல்லாதவர்களுக்கு அவர்கள் வா யிலிருந்து வரும் சொற்கள் யமன் ஆகின்றன . வாழை மரம் தான் ஈனுகின்ற காயினாலேயே அழி ந்துபோகிறது. செய்யத்தகாதவற்றைச் செய்பவர்களுக்கு அறமே யமன். ஒரு குடும்பத்துக்கு, தீய ஒழுக்கம் கொண்ட பெண்ணே யமனாவாள் என்ற பொருள்கொண்ட , நான்மணிக்கடிகைப் பா டலொன்றைக் கீழே காண்க . இப்பாடல் டி .எஸ் பாலசுந்தரம் பிள்ளை பதிப்பித்துள்ள நான்மணிக்கடிகை நூலில் 85 வது பாடல்[5]

கல்லா ஒருவர்க்குத் தம்வாயிற் சொற்கூற்றம்
மெல்லிலை வாழைக்குத் தானீன்ற காய்கூற்றம்
அல்லவை செய்வார்க்கு அறங்கூற்றம் கூற்றமே
இல்லத்துத் தீங்கொழுகு வாள்

இவ்வெண்பாபில் முதல் மூன்று கருத்துக்களும் கூற்றம் (death) என்ற சொல்லில் முடியும் போது நான்காவது கருத்து கூற்றம் என்ற சொல்லுடன் தொடங்குகிறது . முதல் மூன்று கருத்துக்களும் ஒரு படிவத்தை (Pattern) கொண்டுள்ளபோது நான்காவது கருத்து படிவமாற்றம் (pattern variation) ஒன்றை பயன்பாட்டில் கொண்டுவருவதற்கான நோக்கம் என்னவாக இருக்கக்கூடும்? கூர்ந்து மேற்கண்ட செய்யுள் அமைப்பை பார்ப்போமானால் ஒரே படிவத்தையே பயன் படுத்தியிருந்தால் செய்யுளுக்கு இப்பொழுது கிடைக்கும் இனிமை கிடைத்திருக்காது . மற்றும் கடைசிக் கருத்துக்கு படிவமாற்றம் இன்றி அளிக்கப்பட்டிருந்தால் இப்பொழுது அதற்கு கிடைத்திருக்கும் அழுத்தம் (emphasis) படிவ மாற்றம் இல்லாதபோது கிடைத்திருக்காது .

மேற்கோள்களும் குறிப்புகளும்

- 1 ஒரு சில பதிப்புகளில் நூற்று ஆறு வெண்பாக்கள் கொடுக்கப்பட்டிருக்கின்றன
- 2 மூன்று வெண்பாக்கள் ஐந்தடிகளை கொண்டனவாக இருக்கின்றன
- 3 மும்மணிகள் என்ற தொடர் திரிகடுகத்தையும் நான்மணிகள் எனும் தொடர் நான்மணிக்கடிகையையும் குறிக்கின்றன
- 4 மும்மணிகளும் நான்மணிகளும், கி. ஆ. பெ விசுவநாதன், பாரி நிலையம், சென்னை 600 108, முதற் பதிப்பு 1954, 11ம் பதிப்பு 2007
- 5 பதினெண்கீழ்க்கணக்கு நூல்கள்: நான்மணிக்கடிகை, இளவழகனார் என்கிற டி.எஸ்.பாலசுந்தரம்பிள்ளையின் உரையுடன், திருநெல்வேலி சைவசித்தாந்த நூற்பதிப்புக்கழகம், முதற் பதிப்பு 1904, திருவரங்கனார் அச்சகம், சென்னை -600018, 1980

Criteria for Evaluation

A careful comparison of entries 1 and 2 given above relating to Nanmanikkatikai, a classical Tamil book, one of the eighteen minor works will bring out facts connected with how far entry 2 is an improvement over version 1 cited above. Listing of parts of the entry, an omission in 1 above is presented in 2. There is no information on Vilambi Naganar while a sub-section of 2 deals with it. There is a quatrain, though two different verses, quoted from the original work in both versions, the latter contains paraphrase of the quatrain as well as a stylistic interpretation of the verse while the former comprises only the paraphrase and no attempt at criticism is made in version 1 above. There is hardly any citation in the former while there are citations, references, bibliography etc in the latter. Further there is an effort made to place Nanmanikkatikai among the classical Tamil Literary works.

Criteria for Enrichment

Parts of the entry: A clear Indication

An encyclopedic article begins with an indication of the various parts of the text. It provides a thumb nail sketch of the various parts of the entry. It can also list the areas relating to the theme chosen for the entry.

Coverage

Particular areas connected with the theme chosen for treatment must be indicated with suitable captions. The principle behind the sequencing of the areas must be clearly understood. What are the areas connected with the theme remain uncovered in the present entry? There must be a statement indicating why certain areas are chosen in preference over the other areas.

Illustrations

Verbal or pictorial illustrations are provided in the entry wherever possible. Any theoretical statement must be followed by illustrative verbal statements. Use of sign- post words such as 'for instance', 'for example', 'by way of illustration', 'in

conclusion', 'to put it in a nutshell' etc can be used wherever necessary. The pictures or photographs presented must have a suitable caption and information on the picture also form part of the illustration.

Citations

Authenticity of the ideas presented can be ensured with appropriate brief citations in the body of the article. Details about the works cited or authors referred to can be reserved for the bibliographical citations presented at the end of the article. By referring to similar ideas expressed by other writers and their works as briefly as possible, the readers of the article are convinced about genuineness and relevance of the ideas presented. There can be references presented at the end of the article that provide additional information on the point or idea or author or work or website referred to. Many an article in the Tamil Wikipedia suffers from want of citations.

Collaborative Editing

There are certain issues connected with collaborative editing. To use it effectively at the beginning may be a challenging task to the new user. Besides if it is not effectively used, the readability of the text will suffer. Is there any way of overcoming problems arising as a result of collaborative editing in use? How can we bridge the gap between stylistic differences found in the language used by collaborative editors of the Tamil wikipedia? These are questions that need to be addressed by the team of editors of various versions of wikipedias. One way of overcoming this prolem is arranging for bringinging in uniformity of style by the deployment of editors with experience for this. As it is a difficult or challenging task a group of personnel may be specially trained for this purpose. A single author entry and a mulitiple author entry produced by collaborative editing may be presented to all indicating the differences involved and suggesting ways and means of bringing in uniformity in style with sample materials.

Conclusion

This article focuses on certain criteria relating to evaluation and enrichment of the Tamil Wiki using improved and unimproved versions of specimen materials

Bibliography

1. Tamil Wikipedia: A case Study, L . Balasundara Raman, presented at the Wikimania Conference, 2009
2. A Companion Site for the C-SPAN, Series airing Sunday Nights at 8 p.m., and 1 p.m., Q&A Interesting People . Informative Conversations Sep. 25, 2005, September 25, 2005, Jimmy Wales, Wikipedia Founder

WCAG Compliance of Tamilnadu Government Websites: Are We in the Right Track?

S Arulselvan

Associate Professor, Department of Electronic Media and Mass Communication, Pondicherry University
arulselvan.senthivel@gmail.com

Introduction

Government institutions have their presence on the internet, enabling citizens to engage in e-governance services (Pieterse, Ebbers, & van Dijk, 2007). Around the world, government portals have become one of the most important channels for public service delivery and for citizen–government interaction, and the performance of these portals reflects the effectiveness of a country's e-government platform (Liu Yuan 2012). This study aims to analyze the Tamil Nadu Government Websites' compliance to the WCAG 2.0 guidelines which largely helps to make the websites accessible by everyone, including the persons with disabilities.

Jiang and Xu (2009) studied the on-line structure of China's provincial government websites to find out how governments used the websites to improve their legitimacy and to control citizens' political participation. Holliday and Yep (2005) observed the progress of e-government development in China by conducting a content analysis of government websites (Liu Yuan 2012). A study conducted by Steven R Sligar 2008 confirms that web access remains problematic for persons with disabilities. Gant and Gant (2002) studied the functionality of 50 state government web portals, using four dimensions, customization, openness, transparency and usability, as a framework for analysis. Accessibility was evaluated as a part of the dimension of usability and they found that 34 states had accessible websites and 16 failed to provide reasonable access.

The United Nations adopted the Convention on the Rights of Persons with Disabilities (CRPD) in December 2006 which asserts a range of fundamental rights and freedoms that people with a disability enjoy as members of society. Article (4) (1) (g) of the Convention calls on parties to “promote access for persons with disabilities to new information and communications technologies and systems, including the Internet”. Article 9 of the Convention requires countries 'to identify and eliminate obstacles and barriers and ensure that persons with disabilities can access their environment, transportation, and public facilities and services, and information and communications technologies'. India is a signatory to the Convention and as a signatory to the UN Convention, the government of India and all its state governments are bound to provide a friendly services to the disabled on all forms of services including the websites and portals.

WCAG 2.0 Guidelines

The World Wide Web Consortium (W3C) is an international organization that develops standards for the Web and Web based technologies, including accessibility. W3C recommends 14 general guidelines and 66 associated specific checkpoints to evaluate a web site's level of accessibility (Steven R Sligar 2008). The Web Content Accessibility Guidelines (WCAG) 2.0 of the World Wide Web Consortium - Web Accessibility Initiative (W3C WAI), makes web content 'accessible and benefit users with temporary or long-term disabilities, including those who are blind or have low-vision, limited mobility, hearing disabilities, and those who require adjustments due to characteristics associated with aging or limiting environments'.

The WCAG 2.0 recommendations may be largely categorized into four primary domains: 1. Perceivable, 2. Operable, 3. Understandable, 4. Robust. The perceivable suggestions provide text alternatives for any non-text content so that it can be changed into other forms people need. The Operable features make all functionality available from a keyboard. The Understandable features make text content readable and understandable. And the Robust features maximize compatibility with current and future user agents, including assistive technologies.

Tamil Nadu Government Websites

The Tamil Nadu Government's website directory <http://www.tn.gov.in/usefullinks/links-state.htm> has listed 251 websites operated by the Government by its various departments, public sector undertakings, boards, societies, corporations, statutory organizations, district administrations, and educational institutions. Of the 251 websites, a sample of ten percent of the total websites on each category were randomly selected and a total of 25 websites were studied.

Sampling of Web sites on each category

TN Government Website Category	No. of Total websites	No. of websites selected on sampling
Departments	86	8
Undertakings / Boards / Corporations / Societies	65	6
Municipalities / Municipal Corporations	08	1
Statutory Organizations	08	1
Districts	33	3
Universities / Educational Institutions	45	5
Miscellaneous	06	1
Total	251	25

Results and Discussion

All the 25 sampled Tamil Nadu government websites were analyzed on their compliance of the WCAG 2.0 guidelines. Each website was reviewed on the following 12 criteria as stipulated by the WCAG 2.0 Guidelines, and in addition, availability of Tamil language version of the website content was also taken into account:

- 1.1 Text Alternatives availability: Providing text alternatives for any non-text content so that it can be changed into other forms people need, such as large print, braille, speech, symbols or simpler language.
- 1.2 Time-based Media: Providing alternatives for time-based media.
- 1.3 Adaptability: Creating content that can be presented in different ways without losing information or structure.
- 1.4 Distinguishable: Making it easier for users to see and hear content including separating foreground from background.
- 2.1 Keyboard Accessible: Making all functionality available from a keyboard.
- 2.2 Enough Time Available: Providing users enough time to read and use content.
- 2.3 Free from facilitating Seizures: Not designing content in a way that is known to cause seizures.
- 2.4 Navigable: Providing ways to help users navigate, find content, and determine where they are.
- 3.1 Readable: Making text content readable and understandable.

3.2 Predictable: Making Web pages appear and operate in predictable ways.

3.3 Input Assistance: Helping users avoid and correct mistakes.

4.1 Compatible: Maximising compatibility with current and future user agents, including assistive technologies.

5.1 Native Language: Facilitating the availability of Tamil version of the website.

Table 1: Tamil Nadu Government websites' compliance with WCAG 2.0 guidelines

Dept.	URL	1.1	1.2							2.4	3.1					TOTAL SCORE	5.1	Percentage of the Score
Civil Supplies and Consumer Protection Department	http://www.consumer.tn.gov.in	0	0	0	0	0	0	0	0	1	1	1	0	0	3	Y		33.33
Commercial Taxes Department	http://www.tnvat.gov.in/	0	0	0	0	0	0	0	0	1	1	1	0	0	3	Y		33.33
Archeology Department	http://www.tnarch.gov.in/	0	0	0	0	0	0	0	0	1	1	1	0	0	3	N		25
Highways Department	http://www.tnhighways.org/	0	0	0	0	0	0	0	0	1	1	1	0	0	3	N		25
Hindu Religious and Charitable Endowments Administration Department	http://www.tnhrce.org/	0	0	0	0	0	0	0	0	1	1	1	0	0	3	N		25
Information Technology	http://www.elcot.com/	0	0	0	0	0	0	0	0	1	1	1	0	0	3	N		25
Tamil Nadu Police	http://www.tnpolice.gov.in/	0	0	0	0	0	0	0	0	1	1	1	0	0	3	N		25
Pension Directorate	http://www.tn.gov.in/dop/default.htm	0	0	0	0	0	0	0	0	1	1	1	0	0	3	N		25
Dept. of Economics and Statistics	http://www.tnstat.gov.in/	0	0	0	0	0	0	0	0	1	1	1	0	0	3	N		25
Sarva Shiksha Abhiyan	http://www.ssa.tn.nic.in/	0	0	0	0	0	0	0	0	1	1	1	0	0	3	N		25
School Education	http://www.tn.gov.in/school_education	0	0	0	0	0	0	0	0	1	1	1	0	0	3	N		25
Tamil Nadu Legislative Assembly	http://www.assembly.tn.gov.in/	0	0	0	0	0	0	0	0	1	1	1	0	0	3	N		25
Tamilnadu State Transport Corporation Ltd.	http://www.tnstc.in/	0	0	0	0	0	0	0	0	1	1	1	0	0	3	N		25
Vazhndhu Kaatuvom Project	http://www.vazhndhukaatuvom.org/	0	0	0	0	0	0	0	0	1	1	1	0	0	3	N		25
Tiruchirappalli Municipal Corporation	http://www.trichycorporation.gov.in/	0	0	0	0	0	0	0	0	1	1	1	0	0	3	N		25

Tamil Nadu Pollution Control Board	http://www.tnpcb.gov.in/	0	0	0	0	0	0	0	0	1	1	1	0	0	3	N	25
Dharmapuri District	http://www.dharmapuri.tn.nic.in/	0	0	0	0	0	0	0	0	1	1	1	0	0	3	N	25
Karur District	http://www.karur.tn.nic.in/	0	0	0	0	0	0	0	0	1	1	1	0	0	3	N	25
The Nilgiris District	http://www.nilgiris.tn.gov.in/	0	0	0	0	0	0	0	0	1	1	1	0	0	3	N	25
Bharathiar University	http://www.b-u.ac.in/	0	0	0	0	0	0	0	0	1	1	1	0	0	3	N	25
Madras Medical College (MMC)	http://www.mmc.tn.gov.in/	0	0	0	0	0	0	0	0	1	1	1	0	0	3	N	25
Manonmaniam Sundaranar University	http://www.msuniv.ac.in/	0	0	0	0	0	0	0	0	1	1	1	0	0	3	N	25
Tamil Nadu Dr. Ambedkar Law University	http://www.tndalu.ac.in/	0	0	0	0	0	0	0	0	1	1	1	0	0	3	Y	33.33
Tamil Nadu Open University (TNOU)	http://www.tnou.ac.in/	0	0	0	0	0	0	0	0	1	1	1	0	0	3	N	25
TN School Books Online	http://www.textbookcorp.tn.nic.in/	0	0	0	0	0	0	0	0	1	1	1	0	0	3	N	25

1.1 Text Alternatives availability 1.2 Time-based Media 1.3 Adaptability 1.4 Distinguishable 2.1 Keyboard Accessible 2.2 Enough Time Available 2.3 Free from facilitating Seizures 2.4 Navigable 3.1 Readable 3.2 Predictable 3.3 Input Assistance 4.1 Compatible 5.1 Availability of Tamil version.

The above table presents the scores on the compliance of Tamil Nadu Government websites of the twelve features of the WCAG 2.0 guidelines. Scores have been allotted for each compliance unit, at the rate of one point per compliance unit. For the 12 compliance units, the range was set between the maximum at 12, where all the compliance units were fully complied, and the minimum at 0 where none of the compliance units honored in the website. Besides these compliance units, the availability of native language version of the web content in all the sample sites were also reviewed.

The table amply illustrates the status of the government websites on its poor compliance with the guidelines. Except on Navigability, Readability and Predictability the government websites have not complied with any of the guidelines of WCAG 2.0 . Out of 25 websites that were reviewed in the study, none of the websites have found to be considering any guideline of the WCAG 2.0, while designing their content, features and lay out of their website.

This shows that the Tamil Nadu Government has no policy on its presence cyberspace . The analysis further reveals that the government and quasi government institutions have not expressed any commitment towards the differently abled persons' right to access the internet. Surprisingly, except three websites, which is just 12% of the sample, none of the Tamil Nadu Government websites provide Tamil version of their web content. Being a signatory to the UN Convention on the Rights of Persons with Disabilities, the government of India and all its state governments have the bounden duty to facilitate access to cyberspace, friendly to the Persons with Disabilities. The WCAG 2.0 guidelines have provided adequate directions in implementing the UN Convention on the Rights of the Persons with Disabilities.

Conclusion

Evaluating governmental websites would obviously facilitate further improvement of their quality in service. This study is a preliminary evaluation that attempted to provide an overview of the status and pattern of web content accessibility features that have been facilitated by the Tamil Nadu Government on all its websites. The analysis has exposed the

inherent weakness of the Tamil Nadu Government Websites, which have no regard for the rights of persons with disabilities. Tamil Nadu Government should facilitate the Tamil version of all its content in all its websites. Only 3 websites, out of the 25 sampled websites, have offered the Tamil version of their web content. This study further confirms that the Tamil Nadu government has not paid adequate attention to address the needs of the persons with disabilities in accessing the web based content. Before moving to implement e-governance projects, the government of Tamil Nadu should address the issues related to implementing the WCAG 2.0 guidelines in word and spirit, without any further delay.

Bibliography

1. Abebe Rorissa, Dawit Demissie (2010) An Analysis of e-Government service websites, *Government Information Quarterly* 27, pp. 161-169
2. Gant J.B.; Gant D.B (2002) Web portal functionality and state government e-service. *Proceedings of the 35th Hawaii International Conference on System Sciences*, Waikoloa, HI 35, 1627-1636.
3. Hanne Sørsum, Kim Normann Andersen & Ravi Vatrpu (2012), Public websites and human– computer interaction: an empirical study of measurement of website quality and user satisfaction, *Behaviour and Information Technology* 31:7, 697-706
4. Holliday, I., & Yep, R. (2005) . E-government in China. *Public Administration and Development*, 25, 239–249.
5. Jiang, M., & Xu, H. (2009) . Exploring online structures of Chinese government portals: Citizen political participation and government legitimation. *Social Science Computer Review*, 27 (2), 174– 195.
6. Kevin Cullen, Lutz Kubitschke (2009) Study on “Web accessibility in European countries: level of compliance with latest international accessibility specifications, notably WCAG 2.0, and approaches or plans to implement those specifications”. *European Commission: Bonn*
7. Kulkarni R.B, Pooja R. Rajeshwarkar S.K Dixit (2012) Cultural Analysis of Indian Websites, *International Journal of Computer Applications*, Volume 38– No.8, January PP. 15-21
8. Liu Yuan, Chen Xi, Wang Xiaoyi (2012), evaluating the readiness of government portal websites in China to adopt contemporary public administration principles, *Government Information Quarterly* 29, 403–412
9. National Forum on Education Statistics (2011) *Forum Guide to Ensuring Equal Access to Education Websites – Introduction to Electronic Information Accessibility Standards*, NFES, Alexandria.
10. Norman E. Youngblood ☐, Jo Mackiewicz, (2012) a usability analysis of municipal government website home pages in Alabama, *Government Information Quarterly* 29 582–588
11. Pieterse, W., Ebbens, W., & van Dijk, J. (2007) . Personalization in the public sector: An inventory of organizational and user obstacles towards personalization of electronic services in the public sector. *Government Information Quarterly*, 24 (1), 148–164.
12. Rob Law, Shanshan Qi, Dimitrios Buhalis (2010), *Progress in tourism management: A review of website evaluation in tourism research*, *Tourism Management* 31, 297–313
13. Sanne Elling, Leo Lentz, Menno de Jong, Huub van den Bergh (2012), Measuring the quality of governmental websites in a controlled versus an online, setting with the ‘Website Evaluation Questionnaire’, *Government Information Quarterly* 29 pp. 383–393
14. SEOYOUNG HONG and JINWOO KIM (2004), Architectural criteria for website evaluation – conceptual framework and empirical validation, *BEHAVIOUR & INFORMATION TECHNOLOGY*, SEPTEMBER–OCTOBER VOL 23. NO. 5. PP.337-357,
15. Steevan R Sligar, Xiaoming Zeng (2008) Evaluation of Website Accessibility of State Vocational Rehabilitation Agencies, *Journal of Rehabilitation* Vol. 74 No.1 pp. 12-18
16. Wen-Chih Chiou, Chin-Chao Lin, Chyuan Perng (2010) a strategic framework for website evaluation based on a review of the literature from 1995–2006, *Information & Management* 47 PP. 282– 290
17. Wen-Chih Chiou, Chin-Chao Lin, Chyuan Perng (2011), A strategic website evaluation of online travel agencies *Tourism Management* 32 (2011)
18. Wen-Hsien Tsai, Wen-Chin Chou, Chien-Wen Lai (2010) an effective evaluation model and improvement analysis for national park, *Tourism Management* 31, 936–952

தமிழ் எழுத்துருக்களின் வளர்ச்சி சங்க காலம் முதல் இன்று வரை ஒரு மீள் பார்வை

பேரா. முனைவர் வெ.இராமன்

B.Sc., M.A., M.A. D.F.Tech., D.L.L., DACP., Dip.in Rus., CCP., CMM., CFA., M.Phil., PhD
raman600@gmail.com, Prof.Dr.V.Raman@gmail.com

கல் தோன்றி மண் தோன்றா காலத்து முன் தோன்றிய மூத்த தமிழானது முச்சங்ககாலத்தில் பெரு வளர்ச்சி கண்டு இன்றும் இளமை குன்றாது திகழ்கிறது . காலவளர்ச்சியின் ஊடே தமிழும் தனது பெருமையை விரித்து வளர்ந்தது . காலப்போக்கின் விளைவாக தமிழின் எழுத்துரு ஒரு புதிய பரிணாம வளர்ச்சியைக் கண்டது . மண்ணிலும், கல்லிலும் எழுதிய தமிழன் ஓலைகளில் தனது எழுத்துருவைப் பதிக்கத் துவங்கினான் . காலவளர்ச்சியின் வேகம் எழுத்துருவை மாற்றியதுடன் புதிய எல்லைகளையும் தொடத் துவங்கியது . அச்சுக்கலை புகுந்த உடன் எழுத்துருவானது அதீத வளர்ச்சி கண்டதுடன் உலகெங்கும் பரவத்துவங்கியது . தமிழானது இலக்கிய , இலக்கண எல்லைகளைத் தாண்டி இருபதாம் நூற்றாண்டின் இறுதியில் அறிவியல் எல்லைக்குள் பயணிக்கத் துவங்கியது . இதன் விளைவாக முத்தமிழானது நான்காம் தமிழான அறிவியல் தமிழை தன்னுள் உள்வாங்கிக் கொண்டது . எழுத்துருவானது அறிவியல் தமிழில் தனது எல்லைகளைப் பரப்பியதுடன் உலகெங்கும் பரவவும் , புதிய மாற்றங்களைப் பெறவும் செய்தது . முத்தமிழுடன் நான்காம் தமிழாகிய அறிவியல் தமிழும் அதன் உட்கூறாகிய கணினித்தமிழும் கணினியாலும் , இணையத்தாலும் இன்று பெருவளர்ச்சி கண்டுள்ளது . இவ்வாறு வளர்ந்த தமிழ் எழுத்துருக்களின் வளர்ச்சியினை சங்க காலம் தொட்டுக் காண்போம் .

தமிழ் எழுத்துக்களின் தொன்மை

எழுத்து என்ற சொல்லிற்கு 1. எழுதப்படுதலின் எழுத்து என்றும் 2. எழுப்பப்படுதலின் எழுத்து என்றும் இருவகைப் பொருள்களைத் தொல்காப்பியம் உணர்த்துகிறது . முதலாவது வரி வடிவத்தையும் , இரண்டாவது ஒலி வடிவத்தையும் குறிக்கும் . தமிழ் மொழிக்கென்று உருவாக்கப்பட்ட எழுத்து முறைகளைத் தொல்காப்பியம் , சங்க இலக்கியம், காப்பியம், கல்வெட்டுகள், ஓலைச் சுவடிகள் முதலியன தெளிவாகப் புலப்படுத்துகின்றன .

எழுத்து என்ற சொல்லிற்கு வரைதல் , செதுக்குதல், பொறித்தல் முதலிய பொருள்களையும் பழைய இலக்கியங்கள் குறிப்பிடுகின்றன . `கடவுள் எழுத ஓர் கல்தாரான் எனின் ` (சிலம்பு 25.130) என்று இளங்கோவடிகள் குறிப்பிடுகிறார். எழுத்து - ஓவியம், சிற்பம் என்ற பொருள்களிலும் உரைக்கப்பட்டுள்ளது. எழுத்துநிலை மண்டபம் (19.53), எழுது எழில் அம்பலம் (18.28) என்று பரிபாடல் உரைப்பது நோக்கத்தக்கது .

பாலை நிலங்களில் நாட்டப்பெற்றிருந்த நடுகல்லில் வீரனின் பெயரும் புகழும் தமிழில் பொறிக்கப்பட்டிருந்தன . இந்த தமிழ் எழுத்துக்களைக் கூரிய உளியைக் கொண்டு பொறித்துள்ளனர் . இவற்றைக் குயில் எழுத்து - குயின்ற எழுத்து (செதுக்கப்பெற்ற, பொறிக்கப் பெற்ற எழுத்து) என்று அகநானூற்றுப் புலவர் மதுரை மருதனிளநாகனார் இரு பாடல்களில் குறித்துள்ளார் .

இருங்கவின் இல்லாப் பெரும்புன் நாடிக்
கடுங்கண் மறவர் பகழி மாய்த்தென
மருங்குல் நுணுகிய பேஎழுதிர் நடுகல்
பெயர்பயம் படரத் தோன்றுகுயில் எழுத்து
இயைபுடன் நோக்கல் செல்லாது அசைவுடன்
ஆறுசெல் வம்பலர் விட்டனர் கழியும் அகம் 297, 5-10

இப்பாடலின்வழி குயில் எழுத்துகள் சிதைவுற்றதை பாடலாசிரியர் எடுத்துரைக்கிறார் .

மரம்கோள் உமண்மகன் பேரும் பருதிப்
புன்தலை சிதைத்த வன்தலை நடுகல்..
கூர்உளி குயின்ற கோடுமாய் எழுத் (து) அவ்
ஆறுசெல் வம்பலர் வேறுபயம் படுக்கும்அகம் 343, 4-8

இப்பாடலின்வழி கூருளியால் வெட்டப்பட்ட கோடு மாய்ந்த எழுத்துக்கள் புழங்கியதை உணரமுடிகிறது . வளைந்த கோடுகளையே மிகுதியாகக் கொண்டவை வட்டெழுத்துக்கள் என்பர் தி .நா.சுப்பிரமணியன். எனவே இப்பாடலின்வழி கோடுமாய் எழுத்து என்று சுட்டப்படுவதால் தமிழகத்தின் சங்ககாலத்தில் வட் டெழுத்து வழங்கி வந்தது உணரப்படுகிறது .

தொல்காப்பியத்தில் எழுத்து முறை

எழுத்து எனப்படுப
அகர முதல் னகர இறுவாய்
முப்பஃது என்பதொல் - எழுத்து

என்று தொல்காப்பியம் தெரிவிப்பதன் வழி தொல்காப்பியர் காலத்தில் எழுத்துக்கள் மிகவும் வளர்ச்சி பெற்ற நிலையிலும், வரையறுக்கப்பட்ட நிலையிலும் இருந்ததாலேயே இவ்வாறு இலக்கணம் வகுக்கப்பட்டுள்ளதை உணர முடிகிறது.

பதினெட்டு மெய்யும் அகரத்துடன் கூடி உயிர்மெய்யாக உருவம் கொள்ளும்பொழுது புள்ளியில்லாத தம் உருவத்துடன் ஒலிக்கப்படுவன . அவை ஏனைய பதினொரு உயிர்களுடன்கூடித் தம் உருவில் தி ரிபு பெற்று ஒலிக்கப்பெறுவன. திரிபு பெறுதலாவது , மேலும் கீழும் விலங்கு பெற்றும் , கோடு பெற்றும் , புள்ளி பெற்றும் , புள்ளியும் கோடும் உடன் பெற்றும் உயிர்த்தலாம் என்று முன்னைய உரைவழி நின்று நச்சினார்க்கினியர் விளக்கியுள்ளார். எழுத்துக்கள் வரிவடிவத்தில் பெற்று வந் துள்ள மாற்றங்களையும் தொல்காப்பியம் (17) சுட்டி நச்சினார்க்கினியர் விவரித்துள்ளார் . எனவே தொல்காப்பியர் காலம் (கி.மு 500) முதலாகவே தமிழ் மொழிக்கென்று எழுத்துமுறை வரையறுக்கப்பட்டிருந்ததும் , அம்முறை காலந்தோறும் திரிபு பெற்று வந்துள்ளதை உணரமுடிகிறது.

யாப்பருங்கல விருத்தியில் எழுத்துக்கள்

கி.பி பதினோராம் நூற்றாண்டில் தோன்றிய யாப்பருங்கலம் எனும் செய்யுள் இலக்கணம் தமிழ் எழுத்து வகைகள் அக்காலத்தில் விளங்கிவந்த வகைமை குறித்து தெளிவாக எடுத்துணர்த்துகிறது .

உருவே உணர்வே ஒலியே தன்மையென
இருவகை யெழுத்தும் ஈரிரண்டாகும்

என்பதன் மூலம் எழுத்தானது உருவம் , உணர்வு, ஒலி, தன்மை என நான்கு வகையாகப் பிரிக்கப்பட்டிருந்ததை அறியமுடிகிறது.

காணப் பட்ட உருவம் எல்லாம்
மாணக் காட்டும் வகைமை நாடி
வழுவில் ஓவியன் கைவினை போல
எழுதப்படுவது உருவெழுத் தாகும்

என்றதன்வழி உருவ எழுத்து எனப்படும் ஓவிய எழுத்துக்கள் அல்லது படவெழுத்துக்கள் (Pictorial Writing) எழுதப்பட்ட வகை புலப்படுகிறது.

கொண்டவோர் குறியாற் கொண்ட வதனை
உண்டென்று உணர்வது உணர்வெழுத் தாகும்

என்ற பாடலின்வழி மக்களிடையே உணரப்பட்ட குறியெழுத்தினை உணர்வெழுத்து என்று குறிப்பிட்டனர் .

இசைப்படு புள்ளின் எழாஅல் போலச்
செவிப்புல னாவது ஒலியெழுத் தாகும்

என்ற நாற்பாவின்வழி பறவையின் ஒலியைப் போலச் செவியிற் சென்றடைந்து பொருள்தரும் ஓசையினை ஒலியெழுத்து என்று உணர்த்தினார் .

முதற்கா ரணமும் துணைக்கா ரணமும்
துணைக்கா ரணத்தொடு தொடரிய வுணர்வும்
அவற்றொடு புணர்ந்த வகத்தெழு வளியின்
மிடற்றுப் பிறந்திசைப்பது தன்மை யெழுத்து

என்ற பாடலின்வழி உள்ளிருந்து வெளிப்படும் காற்று , மிடறு உள்ளிட்ட பேச்சுறுப்புக்களில் தொடர்புற்று ஒலிக்கப்படுவது தன்மை எழுத்து எனப்படும் என்பதை அறியமுடிகிறது .

இதன்மூலம் நூற்பாக்கள் யாவும் யாப் பருங்கலத்திற்கு மிகவும் முற்பட்டவையாக உரைமரபில் நெடுங்காலம் வழங்கி வந்தவை என்றுணரமுடிகிறது . தொன்றுதொட்டுத் தமிழகத்தில் வழங்கி வந்துள்ள எழுத்து வகைகள் பற்றிய ஒரு பழைய மரபினை இந்நூற்பாக்கள் சுட்டுவன ஆகும் .

கி.பி. எட்டாம் நூற்றாண்டில் உருவான தமிழ் நிகண்டான திவாகரம்,

பெயர்எழுத்து முடிவெழுத்து வடிவெழுத்துத் தன்மை
எழுத்தென வெழுத்தின் பெயர் இயல்பி னாரே

என்று குறிப்பிடுவதின்மூலம் பெயர் , முடிவு, வடிவு, தன்மை என்ற நால்வகை எழுத்துகள் புழங்கியதை அறியமுடிகிறது.

இரண்டாயிரம் ஆண்டுகளுக்கு முன்பிருந்து...

கிறித்து ஆண்டு தொடக்கத்திற்கு முன்னரும் சில நூற்றாண்டுகளுக்குப் பின்னரும் இங்கு வழக்கத்திலிருந்த எழுத்துக்களைப் பற்றி கி.மு. 3ஆம் நூற்றாண்டைச் சார்ந்த சமவயங்க சுத்த என்ற சமண நூலும், அதே காலத்தைச் சார்ந்த லலிதவிஸ்தர என்ற பௌத்த நூலும் குறிப்பிடுகின்றன . சமண நூ லில் 18 வகையான எழுத்துக்களும் , பௌத்த நூலில் 64 வகையான எழுத்துக்களும் தரப்பட்டுள்ளன . சமண நூல் பிராகிருத மொழியில் குறிப்பிடும் 'தாமிலி' என்பது தமிழ் எழுத்துக்களைக் குறிக்கும் சொல்லாகும் . அதன்பின்னர் யாப்பருங்கலிருத்தியில் இராசியெழுத்து, ஆயவெழுத்து, நாளெழுத்து, தன்மையெழுத்து உட்பட பதினைந்து வகையான எழுத்துக்கள் குறிக்கப்பட்டுள்ளன எனினும் இவைகளுக்கு இன்னமும் முழுமையான விளக்கம் யாராலும் தரப்படவில்லை . தி.நா.சுப்பிரமணியன் 1933ம் ஆண்டில் இவை குறித்து எழுதுகையில் தமிழ் எழுத்துக்களுக்கும் இவைகளுக்கும் உள்ள ஒற்றுமை குறித்து விவரித்துள்ளார் .

கி.மு 1000லிருந்து கி.மு 500 ஆண்டுகளுக்கு உட்பட்டதாகக் கருதப்படும் விழுப்புரம் மாவட்டம் திருக்கோயிலூர் வட்டம் கீழ்வாலையில் கண்டுபிடிக்கப்பட்ட உருவ எழுத்தே தமிழகத்தில் உணரப்பட்ட பழைய உருவெழுத்தாக அறியப்படுகிறது. தமிழகத்தின் பல பகு திகளில் கிடைக்கும் அகழ்வாய்வின்போது எடுக்கப்பட்ட கருப்பு சிவப்பு பாணைகளோடு சில உருவங்கள் வரையப் பெற்ற பாணை ஓடுகள் பெரும்பாலும் பெருங்கற்காலப் பொருட்களோடே கிடைத்து வருகின்றன . ஆதலால் இவைகளை பெருங்கற்கால மக்களே வரைந்திருக்கலாம்

எனலாம். இவை உருவ எழுத்தின் (Pictograph) தொடர்ச்சியே எனலாம். அக்கால கட்டத்தில் அவ்வெழுத்துக்கள் ஒலிவடிவம் (Phonograph) பெற்றிருக்கலாம். அனால் அவற்றை நாம் அறிய முடியவில்லை. பாணை ஓடு உருவங்களில் 80 சதவிகிதம் மொகஞ்சதாரோ, அரப்பா உருவ எழுத்துக்கள் போலவே இருப்பது இங்கு குறிப்பிடத்தக்கது.

இலங்கை யாழ்ப்பாணத்திற்கு அருகே அனைக்கோட்டை என்ற இடத்தில் நிகழ்ந்த அகழ்வாய்வில் பெருங்கற்காலச் சின்னத்தில் கண்டெடுக்கப்பட்ட ஒரு செப்பு முத்திரையில் ஒரு வரியில் உருவ எழுத்துக்களும், மற்றொரு வரியில் தமிழ் பிராமி எழுத்துக்களும் பொறிக்கப்பட்டிருந்தது. இவை ஒரே பொருளைக் குறிப்பவனாக இருக்கலாம் எனக்கருதப்படுகிறது. உருவ எழுத்துக்கள் தமிழகத்தில் கண்டெடுக்கப்பட்ட பெருங்கற்கால பாணை ஓடு உருவங்களைப் போன்றே உள்ளன. எனவே தமிழகப் பாணை ஓடு உருவங்கள் ஏதோ பொருளை உணர்த்தும் உருவ எழுத்தெனவும், கிறித்து ஆண்டின் தொடக்க காலத்து மக்கள் உருவ எழுத்தையும், தன்மை எழுத்தையும் (Standardized Script) அறிந்தவர்களாக இருந்திருக்கிறார்கள் என்பதை உணரமுடிகிறது. அதனால் பிந்தைய இலக்கிய ஆசிரியர்கள் உருவ எழுத்தைப் பற்றிய செய்தியை எடுத்தாண்டிருக்கிறார்கள் என்று அறியமுடிகிறது.

பண்டைத்தமிழ் எழுத்து

தமிழகத்தின் மிகப்பழமையான தமிழ் எழுத்தானது **பிராமி** எனவும், **தென்னிந்தியப் பிராமி** எனவும் **பண்டைத் தமிழ் எழுத்து** (Archaic Tamil Script) எனவும் வழங்கப்படுகிறது. இவ்வெழுத்துக்கள் கி.மு 3ஆம் நூற்றாண்டுக்கு முன்பிருந்தே தமிழகத்தில் வழங்கி வந்திருக்கலாம் என்று கருதப்படுகிறது. தமிழ்நாட்டு பண்டைக் கல்வெட்டுகளில் 10 உயிரெழுத்துக்கள் தான் எழுதப்பட்டுள்ளன. 'ஐ' காரமும், 'ஒ' காரமும் காணப்படவில்லை. கி.பி 3-4ஆம் நூற்றண்டளவில் இப்பண்டைத் தமிழ் எழுத்தின் வடிவம் மாற்றம் பெறத் துவங்குகிறது. கல்லில் உளியாலும், தூரிகை கொண்டு வண்ணத்திலும் எழுதி வந்த தமிழர்கள் ஓலையில் எழுத்தாணி கொண்டு எழுத்த் தொடங்கிய காரணத்தால் இவ்வடிவ மாற்றம் தொடங்கி இருக்கலாம் என்பது அறிஞர்தம் கருத்து. முழு வட்ட வடிவம் பெற்ற தமிழ் எழுத்துக்களை கி.பி 8-9ஆம் நூற்றாண்டுக்குப் பின்னர் காண முடிகிறது. இந்த வட்டெழுத்தே இன்றைய தமிழ் எழுத்துக்களுக்கு முன்னோடி எனலாம். இவ்வகையிலான எழுத்துருக்களின் வளர்ச்சியினை கீழ் காணும் படங்களின் வழி காணலாம்.

படம்-1 உயிர் எழுத்துகள்: நூற்றாண்டுகள்											
கி.மு. 3 முதல் கி.பி. 19 வரை											
சா. நூ. ண. டு. Century	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ
கி. மு. 3	X	X	.	+	L	E	P	P	L	L	L
கி. மு. 2	X	H	.	+	L	E	P	P	L	L	L
கி. மு. 1	X	H	.	+	L	E	P	P	L	L	L
கி. மு. 4	X	X	.	+	L	E	P	P	L	L	L
கி. மு. 5	E				2						
கி. மு. 6	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ
கி. மு. 7	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ
கி. மு. 8	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ
கி. மு. 9	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ
கி. மு. 10	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ
கி. மு. 11	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ
கி. மு. 12	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ
கி. மு. 13	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ
கி. மு. 14	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ
கி. மு. 15	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ
கி. மு. 16	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ
கி. மு. 17	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ
கி. மு. 18	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ
கி. மு. 19	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ

படம்-2 மெய் எழுத்துகள்: நூற்றாண்டுகள்											
கி.மு. 3 முதல் கி.பி. 19 வரை											
சா. நூ. ண. டு. Century	க	ங	ச	ட	த	ந	ப	ம	ய	ர	ல
கி. மு. 3	+	C	h	C	I	h	+	h	+	+	+
கி. மு. 2	+	C	h	C	I	h	+	h	+	+	+
கி. மு. 1	+	C	h	C	I	h	+	h	+	+	+
கி. மு. 4	+	C	h	C	I	h	+	h	+	+	+
கி. மு. 5	+	C	h	C	I	h	+	h	+	+	+
கி. மு. 6	+	C	h	C	I	h	+	h	+	+	+
கி. மு. 7	+	C	h	C	I	h	+	h	+	+	+
கி. மு. 8	+	C	h	C	I	h	+	h	+	+	+
கி. மு. 9	+	C	h	C	I	h	+	h	+	+	+
கி. மு. 10	+	C	h	C	I	h	+	h	+	+	+
கி. மு. 11	+	C	h	C	I	h	+	h	+	+	+
கி. மு. 12	+	C	h	C	I	h	+	h	+	+	+
கி. மு. 13	+	C	h	C	I	h	+	h	+	+	+
கி. மு. 14	+	C	h	C	I	h	+	h	+	+	+
கி. மு. 15	+	C	h	C	I	h	+	h	+	+	+
கி. மு. 16	+	C	h	C	I	h	+	h	+	+	+
கி. மு. 17	+	C	h	C	I	h	+	h	+	+	+
கி. மு. 18	+	C	h	C	I	h	+	h	+	+	+
கி. மு. 19	+	C	h	C	I	h	+	h	+	+	+

இவ்வகைத் தமிழ் எழுத்துக்கள் சோழர்கள் காலத்தில் பெரும் ஆதரவைப் பெற்று பிற்காலப் பாண்டியர்கள் , விசயநகரப் பேரரசர்களும் பயன்படுத்தியமையால் இங்கு வலுப்பெற்றுவிட்டது . பின்னர் வந்த நாயக்க மன்னர்களும் இவ்வெழுத்தையே பயன்படுத்தியமையால் இப்புதிய எழுத்துக்களே பொலிவோடும் , வலிவோடும் தமிழகத்தில் நிலைபெற்றுவிட்டது . அச்சுக்கலைக் காலத்திற்குப் பின்னர் இவ்வெழுத்து முறை அதிக வலுப்பெற்று ஆட்சிபுரியத் தொடங்கியது எனலாம்.

கணினிக் காலம்

அண்மைக்காலத்தில் கணினி மற்றும் இணையப் பயன்பாட்டில் ஏற்பட்ட வியத்தகு மாற்றங்களில் ஒன்று கணினித்தமிழ். பெரும்பான்மை கணினிப் பயன்பாட்டில் ஆங்கிலத்தையடுத்து , தமிழின் பயன்பாடு சிறப்பிடம் வகிக்கிறது. தொடக்க காலத்தில் கணினியில் தமிழைப் பயன்படுத்த மேற்கொண்ட முயற்சிகள் பல . அம் முயற்சிகளின் முதல்படியாகக் கணினியில் தமிழ் எழுத்துரு , குறியாக்கத்தின் வளர்ச்சி அமைந்தது . இணைய வளர்ச்சியானது எழுத்துருக்களுக்கு குறியீட்டு முறைகளைக் கொடுத்ததுடன் , புதிய குறியீட்டு முறைகளைப் பெற்று உள்ளங்கையில் உலகம் சுருங்கி நிற்க வழிவகுக்கும் இணையத்தில் , உலகெங்கும் பரவி நிற்க வழிவகை செய்துள்ளது. இச்சூழ்நிலையில் எழுத்துருக்கள் மற்றும் குறியீட்டு முறைகள் பற்றிய ஆய்வு தமிழுக்கு புதிய கருத்துப் பொருளை கொணர்ந்து சேர்ப்பதுடன் புதிய வழிவகை பற்றிய அறிவு பிறக்க ஏதுவானதாகவும் அமையும்.

தமிழ் எழுத்துரு குறியாக்கத்தின் தொடக்கம்

கணினியின் பயன்பாட்டில் தமிழானது தொடக்கத்தில் தமிழ் தட்டச்சுக் கருவியைத் தழுவி ‘பாமினி’ என்கிற எழுத்துரு அறிமுகமானது . இது ஆங்கிலக் குறியாக்கத்தில் அமைந்த எழுத்துருக்களுக்குப் பதிலாகத் தமிழை உட்புகுத்தியது. இதுபோன்றே பலரும் சொந்தத் தயாரிப்பில் பலவகை எழுத்துருக்களை உருவாக்கினர் . ஆனால், காலப்போக்கில் இணையத்தின் வரவால் பல்வேறு நடைமுறைச் சிக்கல்கள் இதில் உருவாகின . இக்காலத்தில் வேறுசில குறியாக்க மென்பொருள்களும் பயன்பாட்டிற்கு வந்தன . அதனால் இணையத்தில் செய்தி பரிமாற்றத்தில் பலவகைச் சிக்கல்கள் ஏற்பட்டன.

தகுதரம் (திஸ்கி) (TSCII)

இந்த ஏற்பாட்டில், இணையத்தின் வரவு புதிய நடைமுறைச்சிக்கல்களை உருவாக்கியது . இதே காலப் பகுதியில் வேறு பல நியமங்களும் உருவாகத் தொடங்கின . இதனால் கோப்புக்களைப் (File) பரிமாறுவதில் சிக்கல்கள் ஏற்பட்டன. தவிர்ந்து, பல தரவுத் தளங்களில் (Database) ஒரு எழுத்துருவை மாத்திரமே ஏற்றுக் கொள்வதால் தமிழையும் ஆங்கிலத்தையும் ஒன்று சேற்க இயலாமல் போனது . மேலும் இவ்வகையில் தமிழ் - ஆங்கிலம் என இரு மொழிகளிலும் ஆவணம் ஒன்றைத் தொகுப்பது கடினமானது . இந்நிலையில் பல மென்பொருள் தயாரிப்பாளர்கள் , தமிழ் அறிஞர்கள், தமிழைக் கணினியில் பயன்படுத்துவோர் , தமிழ் ஆர்வலர்கள் மற்றும் பலரின் முயற்சியால் புதிய தமிழ் நியமக் குறியாக்க முறை உருவாக்கப்பட்டது .

இம்முறையென்பது, “கணினி சார் தேவைகளுக்குத் தமிழ் எழுத்துக்களைப் பயன்படுத்துவதற்கென உருவாக்கப்பட்ட 8 பிட் அடிப்படையில் அமைந்த , தமிழ் - ஆங்கிலம் என இரு மொழிகளைக் கையாளத்தக்க ஒரு எழுத்துக் குறிமுறை நியமமாகும் . எனவே இவற்றைக் கருத்திற் கொண்டு தகவற் பரிமாற்றத்திற்கான தமிழ் நியமக் குறியீட்டு முறை (அதாவது, Tamil Standard Code for Information Interchange [TSCII]) உருவாகியது. தமிழில் திஸ்கி எனவும், தகுதரம் எனவும் (தமிழ் குறியீட்டுத் தராதரம்) வழங்கப்படும். இதில் முதல் 0-127 எழுத்துகள் தகவல் பரிமாற்றத்திற்கான அமெரிக்க முறையை (American Standard Code for Information Interchange [ASCII]) ஒத்தது. மிகுதியான 128-155ல் தமிழ் எழுத்துகள் நிரப்பப் பட்டன. விண்டோஸ் 3.1, 95, 98, Me ஆகிய பதிப்புக்களில் TSCII அதிகம் பயன்படுத்தப்பட்டது . இதுவே முதன்முதலில் உலகம் தழுவிய இணைய உரையாடல் மூலம் தரப்படுத்தப்பட்டு உருவாக்கப்பட்ட குறிமுறை நியமமாகும்” என்று தமிழ் விக்கிப்பீடியா உரைக்கிறது .

அதுவரை நிலவிவந்த பல்வேறு சிக்கல்களுக்கு , தமிழ் எழுத்துக் குறியாக்க வரலாற்றில் ஒருமொழி (ASCII) குறியாக்கத்தால் ஏற்பட்ட வெற்றிடத்தை நிறைவு செய்ய உருவாகிய அடுத்தகட்ட வளர்ச்சியே திஸ்கி . கணினியில்

ஆஸ்கி குறியாக்கத்திற்கான இடம் போக மீதமுள்ள (129 முதல் 256 வரையுள்ள கீற்றுகளில்) இடத்தில் தமிழ் எழுத்துக்களைப் பிரதியீடு செய்தலே திஸ்கி குறியாக்கத்தின் அடிப்படை . ஆங்கில எழுத்துகளின் குறியாக்கத்தில் தமிழைப் பிரதியீடு செய்த முறையினும் இது முன்னேற்றம் உடையதாக , இருமொழிப் பயன்பாட்டிற்கு எளிதாக இருந்தது.

தமிழ் 99 விசைப்பலகை

அரசால் ஏற்றுக் கொள்ளப்பட்டு வெளியிடப்பட்ட தமிழ் 99 விசைப்பலகை அச்சுமுறையில் தமிழ் எழுத்துருக்கள் TAM என்றும் TAB என்றும் இருவகையாகப் பிரிக்கப்பட்டன . TAM என்பது Tamil Monolingual. TAB என்பது Tamil Bi-lingual. TAM என்பது முழு வடிவிலான தமிழ் எழுத்துருக்களைக் கொண்டதாகும் . இம்முறையில் தட்டச்சு செய்யப்படும் எழுத்துருக்கள் முழு வடிவில் இருக்கும் . சான்றாக, 'நிலா' என்று தட்டச்சு செய்தால் ந , ி, ல, ா என்று தனித்தனி எழுத்துக்களாக இல்லாமல் நி , லா என முழு எழுத்துக்களாக இருக்கும் . அழகுணர்ச்சிக்காக முழு எழுத்துக் களாக இக்குறியீட்டு முறை உருவாக்கப்பட்டது .

TAB முறையில் தமிழ் எழுத்துருக் களைக் குறியாக்கம் செய்யக் கிடைத்த 128 கீற்றுகளில் 247 எழுத்துக்களைப் பொருத்த முடியாத காரணத்தால் எழுத்துக்களைக் கூறிட்டுப் பொருத்த வேண்டியதாயிற்று . 'கொடு அழகு மொழிச் சொல்' என்னும் வாக்கில் உள்ள கொ , டு, அ, ழ, கு, மொ, ழி, ச், சொ, ல் என்னும் பத்து எழுத்துக்களை இக்குறியாக்கத்தில் ெ , க, ா, டு, அ, ழ, கு, ெ, ம, ா, ழ, ி, ச், ெ, ச, ா, ல் எனப் பதினேழு கீற்றுகளாகக் குறியாக்கம் செய்தனர். இம்முறைக் குறியாக்கம் ஒரு குறையாகவே கருதப்படுகிறது.

ஒருங்குறி கூட்டமைப்பு (Unicode Consortium)

ஆங்கிலம் தவிர உலகளாவிய மொழிகளில் ஏற்பட்ட இக்குறியாக்கப் பிரச்சனையைப் போக்கவும் , உலகில் உள்ள அனைத்து மொழிகளையும் ஒன்றிணைக்கவும் ஒருங்குறி கூட்டமைப்பு (Unicode Consortium) என்ற அமைப்பு உருவாக்கப்பட்டது. இலாப நோக்கற்ற நிலையில் மிகப்பெரிய கணினி நிறுவனங்களும் , உலக நாடுகள் பலவும், கணினி ஆர்வலர் பலரும் உறுப்பினர்களாக உள்ள இவ்வமைப்பு உலக மொழிகள் அனைத்தையும் ஒரு குடையின் கீழ் கொண்டுவரவே நிறுவப்பட்டது .

இக்கூட்டமைப்பில் உறுப்பினர்களாக உள்ளவர்கள் அவரவர் மொழியை ஒருங்குறி மூலம் கணினியில் எப்படிக் கொண்டுவருவது என்றும் , அதில் ஏற் படும் சிக்கல்களுக்கு எவ்வாறு தீர்வு காண்பது என்றும் ஆராய்ந்து திட்ட அறிக்கைகளை இவ்வமைப்பிடம் வழங்குவர் . அதனை ஆராய்ந்து ஒவ்வொரு மொழிக்கும் எவ்வளவு இடம் ஒதுக்குவது என இவ்வமைப்பு முடிவு செய்யும் . இவ்வமைப்பில் இந்திய அரசும் , தமிழக அரசும் உறுப்பினர்களாக உள்ளன.

ஒருங்குறி குறியாக்க முறை (Unicode Encoding)

ஒருங்குறி குறியாக்க முறை (Universal Coding என்றும் கூறுவர்) என்பது உலகளாவிய குறியாக்க முறை . இதில் உலக மொழிகள் அனைத்திற்கும் இடம் உண்டு . இது 32 பிட் திட்டம் . இதில் 65000 கீற்றுகளில் பல்வேறு மொழிகளுக்கும் இடம் ஒதுக்கப்படுகிறது. சைனீஸ் போன்ற சில மொழிகள் தவிர ஒவ்வொரு மொழிக்கும் 128 பிட்கள் ஒதுக்கப்படும். இவ்வாறு இந்திய மொழிகளுக்கு ஒதுக்கப்பட்ட குறியாக்க முறை ISCII (Indian standard code for Information Interchange) என்று அழைக்கப்பட்டது . ஆனால், இது தேவநாகரி முறையில் குறியாக்கம் செய்யப்பட்டதால் தமிழ் மொழிக்குக் குறைந்த இடங்கள் ஒதுக்கப்பட்டது . இதனால் தமிழ் பயன்பாட்டில் சில சிக்கல்கள் ஏற்பட்டன . என்றாலும் ஒருங்குறி , நமது அசிரத்தையினால் இஸ்கி சார்ந்த குறியாக்க முறையை ஏற்றுக்கொண்டுவிட்டது என்பார்கள்.

ஒருங்குறி தமிழ் குறியாக்கமும் எழுத்துரு பயன்பாடும்

தற்போது ஒருங்குறியில் தமிழ் மொழிக்கும் 8 பிட் அடிப்படையில் குறியாக்க முறை வரையறை செய்யப்பட்டுள்ளது. இதனால் நம்மிடையே இருமொழிப் பயன்பாட்டிற்கென உள்ள TAB TSCII முறைகளை விட ஒருங்குறியில் பன்மொழி உள்ளடக்கிய எழுத்துருவைப் பயன்படுத்த வழி ஏற்படுகிறது . மேலும், TAM, TAB, TSCII

போன்ற குறியாக்க முறையில் குறியீடுகள் வேற்று மொழி எழுத்துருக்களைக் கொண்டதாக இருக்கும் . குறியாக்க முறையிலும் ஒரு குறியாக்க முறை 'அ' வை 140 ஆவது இடத்தில் புகுத்தி இருந்தால் , வேறொரு குறியாக்க முறை 'ன்' வைப் புகுத்தியிருக்கும். ஒருங்குறியில் மேற்கண்ட இடர்ப்பாடு களையப்படுவதுடன் , பிறமொழிக் குறியாக்கத்தில் தமிழைப் பிரதியீடு செய்த முறையும் மாற்றம் பெற்று தமிழ் மொழி எழுத்துருக்கள் உட்புகுத்தப்படுகிறது.

ஒருங்குறியில் எண் 2944 முதல் எண் 3071 ((U+DB80 – U+OBFF) வரை தமிழுக்காக இடம் ஒதுக்கப்பட்டுள்ளது. அந்நிலையில், எந்தமொழியைச் சேர்ந்தவராயினும் 2949 என்ற எண்ணை ஒருங்குறியில் எழுதினால் அது தமிழ் 'அ' வாகவே கணினியில் வெளிப்படும் . மேற்கூறியவாறு உலகம் முழுவதும் ஒரே குறியாக்கம் பயன்பட்டால் செய்திப் பரிமாற்றத்தில் குழப்பமேற்படாது.

இவ்வாறு ஆங்கில முறையான அஸ்கி (ASCII) முறையில் துவங்கிய கணினியின் எழுத்துருவானது நீண்ட பாதையில் பல்கிப் பெருகி மாற்றங்களடைந்து இன்று ஒருங்குறியில் (Unicode) பயணித்துக் கொண்டுள்ளது.

முடிவுரை

இதன்மூலம் தமிழ் எழுத்துருக்களின் வளர்ச்சி சங்ககாலத்திலிருந்து இன்றைய காலகட்டம் வரை எவ்வாறு மாற்றம் அடைந்துள்ளது என்பதை நம்மால் அறிய இயலுகிறது . இதன்வழி தமிழகத்தில் ஏற்பட்டுள்ள சமூக , அரசியல் மாற்றங்கள் தமிழ் எழுத்துருவை மாற்றியுள்ளதும் உணரமுடிகிறது . எப்படியிருப்பினும், இவ்வாறு சங்ககாலத்தில் தொடங்கிய தமிழ் எழுத்துருக்களின் வளர்ச்சியானது இன்றைய கணினிக்காலத்தில் உலகெங்கும் ஒரு குடையின் கீழ் பவனி வரும் எழுத்துரு ஆட்சிக்கட்டிலில் கொடிவீசி நிற்பதனால் தமிழர்களின் தொடர்பு எல்லை இப்புவி எங்கும் பரவி நிற்கிறது . இதனால் தமிழர்கள் உலகெங்கும் ஒன்றுபட்டு செயல்படும் சூழ்நிலையும் உருவாகியுள்ளது என்பது தமிழர்களாகிய நமக்கு மகிழ்ச்சி அளிக்கும் என்பதில் ஐயமில்லை .

மேற்கோள்கள்

1. அகநானூறு, சங்க இலக்கியங்கள் மூலமும் உரையும் , தொகுதி 10 & 11, வர்த்தமானன் பதிப்பகம் , சென்னை. 2012
2. உயர்கல்வியில் ஒருங்குறி குறியீட்டுமுறை செயலாக்குவதின் அவசியம், பேரா.முனைவர்.வெ.இராமன் – சென்னைப்பல்கலைக்கழக இணையக் கருத்தரங்கம், 2010
3. தமிழ் இலக்கிய வரலாறு - மு.வரதராசன், சாகித்திய அக்காதெமி - புது தில்லி, 1972
4. தமிழ் எழுத்தும் கிரந்த எழுத்தும் - சோ.ந.கந்தசாமி, தமிழியல் - 79&80 ஜூன் & டிசம்பர் 2011 உலகத்தமிழாராய்ச்சி நிறுவனம், சென்னை.
5. தொல்காப்பியம் - கழக வெளியீடு, சென்னை, 1973
6. யாப்பருங்கல விருத்தி, கழக வெளியீடு, சென்னை, 1973

முகநூல் கருத்தாடலில் உணர்ச்சிப் பரிமாற்றம்

மலர்விழி சின்னையா (kovaimalar@yahoo.com)

பரமசிவம் முத்துசாமி (param@fbmk.upm.edu.my)

Fakulti Bahasa Moden dan Komunikasi,
Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

ஆய்வுச் சாரம்

நேருக்கு நேர் மேற்கொள்ளப்படும் உரையாடலைப் போலவே சமூக வலைத்தளங்களில் மேற்கொள்ளப்படும் உரையாடல்களிலும் முகபாவனை, சைகைகள், தொனி, நிறுத்தம் போன்றவை பயன்படுத்தப்படுகின்றன. ஆங்கில மொழியில் பெரிய எழுத்துகளால் (capital letters) தட்டச்சு செய்வது, நேர்முக உரையாடலில் குரலை உயர்த்தி அல்லது கத்திப் பேசுவதற்குச் சமமானது. இது கோப உணர்ச்சியின் வெளிப்பாடு என்பதை இணையப் பயணர்கள் அறிவர். சிறிய எழுத்து-பெரிய எழுத்து (small letters-capital letters) எனும் வேற்றுமை இல்லாத தமிழ்மொழியில் இஃது எப்படிச் சாத்தியப்படுகிறது எனும் வினாவே இக்கட்டுரை உருவாக வழிகோலியது. முகநூல் பயணர்கள் மொழிசார்ந்த (verbal) உணர்ச்சிச் சொற்கள், மொழியியல் குறியீடுகள் (linguistic markers) போன்றவற்றோடு மொழிசாரா (non-verbal) நிறுத்தக்குறிகள், உணர்ச்சிக் குறியீடுகள் (emoticons) போன்றவற்றையும் பயன்படுத்தித் தங்கள் உணர்ச்சிகளைப் பரிமாறிக்கொள்கின்றனர் என்பது இவ்வாய்வில் கண்டறியப்பட்டுள்ளது. மேலும், இக்கட்டுரை முகநூல் கலந்துரையாடலில் பரிமாறப்படும் உணர்ச்சிகளைப் பகுத்தாய்ந்து விளக்கியுள்ளது.

முன்னுரை

உலகில் காணக்கிடைக்கின்ற அனைத்து வசதிகளையும் இன்றைய மாய உலகமாகத் திகழும் இணையத்தளத்தில் கொண்டுவருவதில் தொழில்நுட்ப வல்லுனர்கள் பெரும் வளர்ச்சியைக் கண்டுள்ளனர் ; கண்டும் வருகின்றனர் . அவ்வகையில் உலக வாழ்க்கை யில் மனிதர்கள் எப்படிச் சமுதாயமாகக் கூடி வாழ்கின்றனரோ அதே போல் இணையத்திலும் கூடிப் பழகவே சமூக வலைத்தளங்களைக் கண்டருளினர் . அப்படிக் கண்டறியப்பட்ட சமூக வலைத்தளங்களுள் முகநூலே (facebook) அதிகமான பயணர்களைக் கொண்டுள்ளது என்று உறுதியாகக் கூறலாம் . முகநூலில் தொடர்புகொள்கின்றவர்கள் கருத்துப் பரிமாற்றம் செய்து கொள்வதோடு அக்கருத்துகளின் ஊடே தங்களது உணர்ச்சிகளையும் பரிமாறிக்கொள்கின்றனர் . அவ்வகையில் முகநூலில் தமிழ்மொழி கருத்தாடலில் பரிமாறிக்கொள்ளப்படுகின்ற உணர்ச்சி வகைகளையும் உணர்ச்சிகளைப் பரிமாறிக்கொள்ள கையாளப்படு ம் உத்திகளையும் இக்கட்டுரை விவரிக்கிறது .

ஆய்வின் பிற்புலம்

ஹார்வர்ட்டு பல்கலைக்கழக முன்னாள் மாணவரான மார்க்கு சுக்கெர்பெர்க்கு (Mark Zuckerberg) அவருடைய நண்பர்களின் துணையோடு உருவாக்கியதுதான் முகநூல் எனும் சமூக வலைத்தளம் ஆகும் . தொடக்கத்தில் (2004இல்) ஹார்வர்ட்டு கல்லூரி மாணவர்களுக்காக மட்டுமே உருவாக்கப்பட்ட இந்த வலைத்தளத்தில் , தொடங்கப்பட்ட இரண்டே மாதங்களில் போஸ்டன் (Boston) வட்டாரத்திலுள்ள பிற கல்லூரி மாணவர்களும் ஒரே ஆண்டில் பல உயர்கல்விக்கூடங்களின் மாணவர்களும் உறுப்பினர்களாயினர் . 2006ஆம் ஆண்டு செப்டம்பர் திங்கள் 11ஆம் நாளிலிருந்து மின்னஞ்சல் முகவரியுள்ள எவரும் முகநூல் உறுப்பினராகப் பதிவு செய்துகொள்ளும் வாய்ப்பு வழங்கப்பட்டது . 2007ஆம் ஆண்டு ஜூலை திங்கள் வரையில் உலகெங்கிலுமிருந்து 3, 40, 00, 000 பேர் முகநூலில் உறுப்பினர்களாகப் பதிந்து கொண்டனர் . உருவாக்கப்பட்ட ஒரே ஆண்டில், அதிகமான பயணர்களைக் கொண்ட வலைத்தளங்களின் பட்டியலில் முகநூல் அறுபதாவது இடத்திலிருந்து ஏழாவது இடத்திற்கு உயர்ந்தது . முகநூலின் புகழ் நாளுக்கு நாள் வளர்ந்து கடந்த ஆண்டு (2011) இறுதிவரை 5, 70, 00, 000 உறுப்பினர்களைக்

கொண்டிருப்பதாக முகநூல் அறிவித்துள்ளது. 2007ஆம் ஆண்டு தொடங்கி ஒவ்வொரு நாளும் ஏறக்குறைய 2, 50, 000 பேர் முகநூலின் உறுப்பினர்களாகப் பதிவு செய்துகொள்வதோடு மட்டுமல்லாமல் முகநூல் உறுப்பினர்களின் மொத்த எண்ணிக்கை ஒவ்வொரு ஆறு மாத இடைவெளியிலும் இரு மடங்கு அதிகரித்திருப்பதாகவும் சொல்லப்படுகிறது (நூருல் ஹுஸ்னா, 2011). தற்பொழுது உலகில் அதிகமான முகநூல் பயணர்களைக் கொண்ட நாடுகளில் அமெரிக்கா முதலிடத்திலும் இந்தியா இரண்டாவது இடத்திலும் இருக்கின்றன. எவர் விரும்பினாலும் விரும்பாவிட்டாலும் முகநூல் பலரும் விரும்பி பயன்படுத்தும் சமூக வலைத்தளமாக நிலைத்திருக்கும் (இஸ்கண்டார், 2010).

இப்படி அதிகமானோரால் விரும்பிப் பயன்படுத்தப்படும் சமூக வலைத்தளத்தில் தமிழர்களின் பங்களிப்புக்கும் குறைவில்லை. ஆனால், தமிழ்மொழி பயன்பாட்டில்தான் சற்று குறையிருப்பதாகத் தெரிகிறது. தமிழ் எழுத-படிக்கத் தெரியாத தமிழர்கள் இணையத்தளங்களில் ஆங்கிலத்தையோ ரோமன் எழுத்துகளில் தமிழையோ பயன்படுத்துவது ஒரு குறையன்று. ஆனால் தமிழ் எழுத-படிக்கத் தெரிந்த தமிழர்களும் தமிழில் தட்டச்சு செய்யாமல் ஆங்கிலத்தையே பயன்படுத்துவதுதான் வருந்தத் தக்கது. இந்நிலைக்குக் காரணங்கள் பல இருக்கலாம். இருந்தாலும் அவற்றையெல்லாம் களைந்து, தமிழறிந்தவர்கள் அனைவரும் தமிழிலேயே உரையாடினால்தான் இக்கணினியுத்திலும் தமிழ் செழிக்கும். தமிழ்மொழியில் உரையாட முனையும்பொழுது உரையாடலுக்கேற்ற உணர்ச்சிகளையும் வெளிப்படுத்தினால்தான் அவ்வுரையாடல் உயிரோட்டமாக இருக்கும். அம்முயற்சியின் ஒரு படியாக இக்கட்டுரை தமிழர்களிடையிலான கருத்தாடலில் காணப்படும் தமிழ்மொழியில் அமைந்த உணர்ச்சிப் பரிமாற்றங்களையே ஆராய்கிறது.

ஆய்வுச் சிக்கல்

மனித உரையாடலில் மொழியும் உணர்ச்சியும் பிணைந்தே செயல்படுகின்றன. நேர்முக உரையாடலின் போது குரலில் வெளிப்படும் ஏற்றத்தாழ்வு, வேகம், தொனி, ஆகியனவற்றோடு தலையை ஆட்டுதல், முகத்தைச் சுளித்தல், கண்களைச் சிமிட்டுதல் முதலான உடல் அசைவுகளாலும் உணர்ச்சிகளைத் தெளிவாக வெளிப்படுத்தவும் சரியாகப் புரிந்து கொள்ளவும் இயல்கின்றன. ஆனால் வெறும் எழுத்துகளை மட்டுமே பயன்படுத்தி உணர்ச்சிகளை வெளிப்படுத்துவதும் புரிந்துகொள்வதும் அவ்வளவு எளிதன்று. தவறான உணர்ச்சி வெளிப்பாட்டாலும் தவறான புரிதலாலும் கருத்து முரண்பட்டு, மனவருத்தம் ஏற்படும். எனவே, முகநூலில் பரிமாறப்படும் உணர்ச்சி வகைகளையும் அவற்றைப் பரிமாறிக்கொள்ளும் உத்திகளையும் தெரிந்துகொள்வதாலும் அவற்றை ஒருமுகப்படுத்துவதாலும் இச்சிக்கலை ஓரளவு களைய முடியும்.

ஆய்வு முறை

தமிழ்மொழிக் கருத்தாடலில் இடம்பெறும் உணர்ச்சிப் பரிமாற்றங்களை விவரிப்பதே இக்கட்டுரையின் நோக்கம் ஆகும். அதற்கு முகநூல் எனும் சமூகவலைத்தளம் ஆய்வுக் களமாகத் தேர்ந்தெடுக்கப்பட்டது. தமிழ்மொழியிலுள்ள உணர்ச்சிகளைக் கண்டறிவதற்குத் தமிழ் எழுத-படிக்கத் தெரிந்தவர்களைத் தரவாளர்களாகத் தேர்ந்தெடுப்பதே பொருந்தும் என்பதால் தமிழ்மொழியில் அமைந்த பொது பார்வைக்கான முகநூல் சுவர்ப்படங்களில் (public view wall post) இடப்பட்ட கருத்துகளே இவ்வாய்வின் தரவுகளாகப் பயன்படுத்தப்பட்டுள்ளன. கடந்த அக்டோபர் திங்கள் தொடங்கி நவம்பர் திங்கள் வரை முகநூலில் இடப்பட்ட கருத்துகளை உற்றுநோக்கி அவற்றுள் ஆய்வுக்குப் பொருத்தமான கூற்றுகள் மட்டும் தேர்ந்தெடுக்கப்பட்டன. இடம், வயது எனும் கட்டுப்பாடுகளின்றி மொத்தம் 350 பயணர்கள், 20 சுவர்ப்படங்களில் பதிவு செய்த 475 கூற்றுகள் அடையாளங்கண்டு தேர்ந்தெடுக்கப்பட்டன. தேர்ந்தெடுத்த கூற்றுகளுக்குப் பனுவலின் அடிப்படையிலும் கூற்றுகளின் அடிப்படையிலும் ப1கூ1, ப1கூ2, ... என்று அடையாளங்களிட்டு அக்கூற்றுகள் ஆராயப்பட்டன.

முகநூலில் பரிமாறப்படும் உணர்ச்சி வகைகள்

மனித உணர்ச்சிகளை வரையறுப்பதிலும் வகைப்படுத்துவதிலும் பலரும் பலவித கருத்துகளைக் கொண்டுள்ளனர். இவ்வாய்வு தொல்காப்பிய மெய்ப்பாட்டியலில் சொல்லப்பட்டுள்ள

‘நகையே, அழுகை, இளிவரல், மருட்கை,
அச்சம், பெருமிதம், வெகுளி, உவகை, என்று
அப்பால் எட்டே மெய்ப்பாடு’ என்ப

எனும் எட்டு வகை உணர்ச்சிகளை அடிப்படையாகக் கொண்டு அமைக்கப்பட்டுள்ளது . மேற்சொன்னவற்றுள் நகை (சிரிப்பு), மருட்கை (வியப்பு), பெருமிதம் (வீரம்), உவகை (காதல்) ஆகியனவற்றை உடன்பாட்டு உணர்ச்சிகள் என்றும் அழுகை , இளிவரல் (அவமானம்), அச்சம், வெகுளி (சினம்) ஆகியனவற்றை எதிர்மறை உணர்ச்சிகள் என்றும் இரு வகையாகப் பிரித்துப் பார்க்கலாம் . முகநூல் உரையாடல்களில் இந்த எண்வகை உணர்ச்சி வெளிப்பாடுகளும் அடையாளங்காணப்பட்டன .

அட்டவணை 1: உணர்ச்சி வெளிப்பாடுகள்

உணர்ச்சி வகைகள்		வெளிப்பாடுகள்
உடன்பாட்டு உணர்ச்சிகள் (Positive Emotions)	சிரிப்பு	:), :-), ☺ ஜோக்கா சொன்னீங்க ஹஹஹா, கிகீஹிஹிஹ்ஹிஹ்
	வியப்பு	: (, வாவ் அடேயப்பா, அடேங்கப்பா எப்படிச் சொல்ல
	பெருமிதம்	^ அருமை, அற்புதம் வாழ்த்துகள், வாழ்க
	உவகை	♥ சந்தோசம், மகிழ்ச்சி விரும்புகிறேன் இனிமையான வரிகள்
எதிர்மறை	அழுகை	: (வருந்துகிறேன்

உணர்ச்சிகள் (Negative Emotions)		கவலையா இருக்கு
	அவமானம்	அடக்கடவுளே அடப்பாவி வடை போச்சே
	அச்சம்	ஐயோ!, ஐயையோ!, ஆஹா! லொல்லு தாங்கலையே ஆண்டவா! பயமா இருக்கு போச்சிடா..... பெரிய அக்கா வந்துட்டாங்க
	சினம்	:X விட்டுத் தொலை ஒத்துக்க மாட்டேன் நான் கோவக்காரனும் கூட அடிச்சிடுவேன்

அட்டவணை 1-இலிருந்து முகநூலில் பரிமாறிக்கொள்ளப்படும் உணர்ச்சி வகைகளையும் அவற்றின் வெளிப்பாடுகளையும் அறியலாம் . வெறும் எழுத்துகளையும் குறியீடுகளையும் கொண்டு நிகழ்த்தப்படுகின்ற முகநூல் உரையாடல்களிலும் நேர்முக உரையாடலில் வெளிப்படக்கூடியவை என்று வரையறுக்கப்பட்டுள்ள எட்டு வகை உணர்ச்சிகளையும் காணமுடிகின்றன . ஊடகம் எதுவான போதிலும் உரையாடல்களில் வெளிப்படுகின்ற உணர்ச்சிகளில் மாற்றமில்லை என்பது இங்குத் தெளிவாகிறது .

முகநூலில் உணர்ச்சிகளை வெளிப்படுத்த கையாளும் உத்திகள்

முகநூலில் வெளிப்படுகின்ற வெளிப்படுத்தப்படுகின்ற உணர்ச்சி வகைகளை அறிந்து கொண்ட நாம் அவற்றை முறையாக வெளிப்படுத்துகின்ற உத்திகளையும் அறிந்துகொண்டால்தான் உயிரோட்டமான உரையாடல்களை மேற்கொண்டு, தவறான கருத்தாடல்களைத் (miscommunications) தவிர்த்து விளைபயன்மிக்க வகையில் உறவுகளை வளர்க்க முடியும். இவ்வாய்வில் உணர்ச்சிப் பரிமாற்றத்திற்காக முகநூல் பயணர்கள் பயன்படுத்துகின்ற ஏழு வகை உத்திகள் அடையாளங்காணப்பட்டுள்ளன. இவ்வுத்திகளை மொழிசார் உத்திகள் மொழிசாரா உத்திகள் எனும் இருபிரிவுகளில் காணலாம்.

மொழிசார் உத்திகள்

மொழிசார் உத்திகள் என்பவை சொற்கள், தொடர்கள் ஆகியவற்றால் உணர்ச்சிகளை வெளிப்படுத்தும் முறைகள் ஆகும். முகநூல் பயணர்கள் தங்களது கருத்தாடலில் உணர்ச்சிச் சொற்கள் (கோபம், மகிழ்ச்சி, கவலை, ...), உணர்ச்சிகளை உணர்த்தும் ஒலிக்குறிப்புகள் (ஐயோ, ஐயையோ, ...) மொழியியல் குறியீடுகள் (linguistic markers), திரைப்பட வசனங்கள் என நான்கு வகை மொழிசார் உத்திகளைப் பயன்படுத்துகின்றனர்.

உணர்ச்சிச் சொற்கள் என்பவை அகராதிப் பொருளில் மனித உணர்வுகளைக் குறிக்கப் பயன்படுகின்ற மகிழ்ச்சி, காதல், பயம், வேதனை போன்றவையாகும்.

எ.கா: (i) ஓ... அண்ணா.....சந்தோசம், வாழ்த்துகள்:)) (ப13கூ1)

(ii) எனக்கு ரொம்ப பயமா இருக்கு (ப17கூ4)

சில வேளைகளில் நாம் உணர்ச்சிகளை வெளிப்படுத்த வெறும் ஒலிக்குறிப்புகளை மட்டும் உச்சரிப்பது உண்டு . இவற்றிற்குச் சொற்பொருள் என்பது கிடையாதென்றாலும் இவை சூழலுக்கேற்ற உணர்ச்சிப் பொருளை உணர்த்த வல்லன.

எ.கா: (i) ஐயோ! மாம்ஸ்.... ஆரம்பிச்சிட்டீங்களா!!! (ப14கூ17)

(ii) வாவ்! அண்ண ஒரே ரோமண்டிக் மூட்லெ இருக்கிற மாதிரி

தெரியுது (ப14கூ25)

உணர்ச்சிச் சொற்கள் , உணர்ச்சி ஒலிகள் ஆகியன தவிர்த்து சில தொடர்களின் வழி குறிப்பால் நாம் உணர்ச்சிகளை வெளிப்படுத்துவதுண்டு . " நான் இப்போது கோபமாக இருக்கிறேன் , " என்று சொல்வதை விட "தொலைச்சுப் புடுவேன் " என்ற கூற்றின் வழி கோபத்தை வெளிப்படுத்துவதுதான் இயல்பாகும் . இதனையே மொழியியல் குறியீடுகள் என்கிறோம் . பெரும்பாலும் இது போன்ற மொழியியல் குறியீடுகளின் வழிதான் ஒருவர் தம் உணர்ச்சிகளை வெளிப்படுத்துவார் . இதை விடுத்து "நான் கோபமாக இருக்கிறேன் " என்று கூறுவது இயல்பான உணர்ச்சி வெளிப்பாடாக இல்லாமல் நாடகத்தன்மை கொண்ட உரையாடலாகவே அமையும் .

எ.கா: (i) நீங்க ரொம்ப ராங்கி... கொஞ்சம் அடங்குங்க... இல்லேனா!!! (ப10கூ20)

(ii) சத்தியமா சொல்லுறேன் நீயெல்லாம் உருப்படவே மாட்ட (ப15கூ16)

தமிழ்மொழியில் கருத்தாடும் முகநூல் உறுப்பினர்கள் , புகழ்பெற்ற திரைப்பட வசனங்களின் துணையாலும் , சொல்ல விழைகின்ற கருத்துக்கேற்ற தொனியைக் கொடுத்துத் தங்கள் உணர்ச்சிகளை வெளிப்படுத்துகின்றனர் .

எ.கா: (i) யென்ன கொடுமை சார் இது! (சந்திரமுகி திரைப்பட வசனம்) (ப15கூ25)

(ii) ஆஹா! கவிதெ கவிதெ! (குணா திரைப்பட வசனம்) (ப19கூ6)

மொழிசாரா உத்திகள்

நிறுத்தக்குறிகள், உணர்ச்சிக் குறியீடுகள் (emiti-con), எழுத்துகளின் நிறங்கள் , எழுத்துகளின் அளவுகள் , உயிரெழுத்துகளின் நீட்சி (உயிரளபெடை) போன்றவற்றையே உணர்ச்சி வெளிப்பாட்டிற்கான மொழிசாரா உத்திகள் என்கிறோம் . தமிழில் அமைந்த முகநூல் உரையாடலில் உணர்ச்சிக் குறியீடுகள் , நிறுத்தக்குறிகள் என்ற இருவகை மொழிசாரா உத்திகள் மட்டுமே அடையாளங்காணப்பட்டன . மின்னஞ்சல், அலைபேசி போன்ற மின்னியல் தொடர்பு ஊடகங்களில் முக பாவனையைப் பிரதிபலிக்க பயன்படுத்தப்பட்ட உணர்ச்சிக் குறியீடுகள் முகநூல் உரையாடல்களிலும் இடம்பெற்றுள்ளன .

எ.கா: (i) அனைவருக்கும் நன்றி _/_ நன்றி _/_ நன்றி _/_ (ப8கூ12)

(ii) இப்படி ஒரு தோழன் எனக்கு இல்லையேனு வருந்துகிறேன் ((ப10கூ21)

உணர்ச்சிக் குறியீடுகளைப் போலவே நிறுத்தக்குறிகளும் அதிகமான முகநூல் பயணர்களால் உணர்ச்சி வெளிப்பாட்டிற்காகப் பயன்படுத்தப்படுகின்றன.

எ.கா: (i) ஹாஹா... மாம்ஸ் அக்காவே அப்படினா?? தங்கச்சிங்க இன்னும்

கொஞ்சம் டெர்ரரா இருப்போம்ல...!!! (ப14கூ14)

(ii) இது யெப்போங்க!!??? சூப்பர்!!! (ப18கூ5)

இங்குக் குறிப்பிட்டுள்ள உணர்ச்சி வெளிப்பாடுகள் யாவும் கிடைக்கப்பெற்ற தரவுகளின் அடிப்படையிலேயே ஆராயப்பட்டுள்ளன. அவ்வகையில் இவ்வாய்வு முகநூலில் தமிழ்மொழி கருத்தாடலில்

பரிமாறிக்கொள்ளப்படுகின்ற எட்டு வகை உணர்ச்சிகளையும் , அவ்வுணர்ச்சிகளைப் பரிமாறிக்கொள்ள கையாளப்படுகின்ற ஏழு உத்திகளையும் கண்டறிந்துள்ளது .

நிறைவு

பிறமொழி உரையாடல்களில் பயன்படுத்துவது போலவே தமிழ்மொழிக்கும் இணையப்பயணர்கள் தாமாகவே ஒரு சில வழிகளைக் கண்டறிந்து உணர்ச்சிகளைப் பரிமாறிக்கொள்கின்றனர் . இந்த உணர்ச்சிப் பரிமாற்றத்திற்குப் பயன்படுத்தும் சொற்கள் , தொடர்கள் , உணர்ச்சிக் குறியீடுகள் போன்றவற்றையெல்லாம் சேகரித்து அவற்றைத் தரப்படுத்தி (standardize) பயன்படுத்தினால் தமிழில் அமைந்த சமூக வலைத்தள கருத்தாடல்களில் ஏற்படக்கூடிய தவறான புரிதல்களைக் குறைக்க முடியும் .

மேற்கோள் நூல்கள்

1. சிங்காரவேலன், சொ. (2006) . *தொல்காப்பியம் எளிய உரை (பொருள்)* . சென்னை: தென்னிந்திய சைவசித்தாந்த நூற்பதிப்புக் கழகம் லிட்ட்.
2. திருவாசகம், க. (2010) . கணினித் தமிழ் வளர்ச்சி ; *உலகத் தமிழ் செம்மொழி மாநாட்டுச் சிறப்பு மலர்* . தமிழ்நாடு: கோவை.
3. Iskandar Ab. Rashid. (2010) . *Kembangkan Rangkaian Sosial Menggunakan Facebook*. Selangor: Venton Publishing.
4. Nurul Husna. (2011) . *Facebook Sebelah Syurga, Sebelah Neraka*. Selangor: Al-Hidayah House of Publishers Sdn. Bhd.

செல்பேசியில் தமிழ்க் கலைச்சொல்லாக்கம்

முனைவர் நா.ஜானகிராமன்

தமிழ்த்துறைத்தலைவர், பாரதிதாசன் பல்கலைக்கழக கல்லூரி, பெரம்பலூர்.

தமிழ்மொழியின் பயன்பாடு இன்று உலகளாவிய நிலையில் பல்வேறு தளங்களில் பரந்து விரிந்துள்ளது . அறிவியல் துறையில் நுழைந்த தமிழ் இன்று கணினித்துறையில் வெகுவாக முன்னேற்றம் கண்டு வருகின்றது. இதனைப்போல் செல்பேசியைப் பயன்படுத்தாதவர் யாரும் இல்லை என்று கூறும் அளவிற்கு செல்பேசியின் பயன்பாடு தவிர்க்க முடியாத ஒன்றாக உள்ளது . வலைத்தளங்களைக் கூட தமிழில் காண மென்பொருட்கள் வந்துவிட்டன. ஆனால் இன்று நாம் கடைகளில் வாங்கும் அனைத்து செல்பேசிகளில் நிறுவன செல்பேசிகளிலும் தமிழ்பயன்பாட்டுக்கு வந்துவிடவில்லை . நோக்கியா போன்ற நிறுவனமும் , சில சீனத்தயாரிப்புகளிலும் மட்டுமே தமிழில் தமிழில் இயங்குதளம் உள்ளது . ஏனையவற்றுக்கு இல்லை . தமிழர்கள் வாங்கும் அனைத்து செல்பேசிகளிலும் தமிழ்மொழி இடம்பெற்றால் நாம் தமிழின் பெருமையை உயர்த்தியவர்களாகிவிடுவோம் . அவ்வகையில் செல்பேசிகளில் அனைத்து கூறுகளிலும் தமிழ் இடம்பெற்றால் தமிழ் வளரும் .

கலைச்சொற்கள் இதற்கு அடிப்படையாகும் . சில செல்பேசிகளில் இன்பாக்ஸ் என்று ஒலிபெயர்த்து எழுதப்பெறுகின்றன. கருவியின் செயல்பாடு , அமைப்பு, கருத்து எனப் பல்வேறு கோணங்களில் நின்று கலைச்சொல்லினை உருவாக்கலாம் . ஏற்கனவே, பல்வேறு கலைச்சொற்கள் கணினி சார்ந்து உருவாக்கப்பெற்றுள்ளன .

இன்று பயன்பாட்டிலிருக்கும் சாம்சங் என்ற நிறுவனம் வெளியிட்ட செல்பேசியில் இடம்பெற்றுள்ள அறிவியல் கலைச்சொற்களுக்கு ஏற்ற தமிழ்ச்சொற்கள் படைக்கும் முயற்சில் இந்த ஆய்வு அமைகின்றது .

கலைச்சொற்கள் வரலாறும், நோக்கமும்

தமிழில் கலைச்சொற்களைப் படைத்துக்கொள்வதின் மூலம் தமிழ்மொழியைப் பல்வேறு தளங்களில் வளர்த்தெடுக்க முடியும் என்ற நோக்கில் கலைச்சொற்கள் படைக்கப்படுகின்றன . வினை, பண்பு, பெயர், தொழில், அடை என்ற கூறுகளில் கலைச்சொற்கள் தோற்றம் பெறுகின்றன . அறிவியலையும் தொழில்நுட்பத்தையும் தமிழில் கொண்டுவர தமிழகத்தாரும் புலம்பெயர்ந்தோரும் எண்ணற்ற பணிகளில் ஈடுபட்டுவருகின்றனர் . இவற்றில் தமிழ்நாட்டினரைவிடவும் புலம்பெயர்ந்தோர் பங்களிப்பு மகத்தானது . தமிழில் அறிவியலை வளர்க்கவும், அது தமிழில் நிலைபெறவும் தரமான கலைச்சொற்கள் மொழிக்குத் தேவையாகின்றன . எல்லோரும் கலைச்சொல்லாக்கத்துறையில் ஈடுபட்டு உழைத்தாலும் அவர்கள் உருவாக்குகின்ற சொற்கள் தரமுடன் , செந்தரத்துடன் இயங்குகின்றனவா ? என்பதிலும் அச்சொற்களை கையாள்கின்ற நபர்கள் எங்ஙனம் அதனைப் பயன்படுத்துகின்றனர் என்ற நோக்கிலும் கலைச்சொல்லியலாளர்கள் ஆராயவேண்டும் (இராதா செல்லப்பன் (2006:34) . அந்தவகையில் இன்றைய பயன்பாட்டில் கலைச்சொற்களின் பயன்பாடு பல்வேறு இடங்களில் இடம்பெறினும் செல்பேசியில் இடம்பெறல் என்பது மொழியினைப் பயன்படுத்துகின்றவர்கள் ள் மேலும் மொழியினை அறிவதற்கும் அறிவியலைத் தமிழில் தெரிந்துகொள்வதற்கும் வாய்ப்பாக இருக்கும் என்பதில் கருத்து வேற்றுமை இருக்க முடியாது.

அறிவியல் தமிழ் பயில் என்று சொன்ன பாரதிதாசனும் , அறிவியலை தகவல் தொடர்பிற்கு எவ்விதம் பயன்படுத்தவேண்டும் என்று அன்றே சொன்ன பா ரதியாரும் கலைச்சொல்லாக்கப் பாதைக்கு வழிகாட்டியவர்கள் எனலாம்.

கலைச்சொல்லாக்க நெறிகள்

கலைச்சொல்லாக்கத்திற்குப் பல்வேறு நெறிகள் பின்பற்றப்படுகின்றன . எனினும் யாவரும் எளிதில் பயன்படுத்தும் வகையில், ஏற்றுக்கொள்ளும் வகையில் அமையப்பெறல் சிறந்தது . ஒலிபெயர்ப்பு, மொழிபெயர்ப்பு, கருத்துப்பெயர்ப்பு என்று மூன்றும் அடிப்படையானதாகும் . இவற்றில் முன்னொட்டு , பின்னொட்டு என்ற வகையால் மேலும் பல பகுப்புக்களில் உருவாக்கலாம் . எவ்வாறாயினும் கருத்துச்சிதைவு கூடாது என்பதில் தெளிவு வேண்டும் . (இராதா செல்லப்பன் :2011) சொல்லுருவாக்கம் பல நிலைகளில் நிகழும் ஆயினும் அறிவியல் மற்றும் தொழில்நுட்பத்துறைக்கு கலைச்சொற்கள் தவிர்க்கமுடியாத ஒன்றாகும் (சுந்தரம். இராம (1995:34) கலைச்சொல்லாக்கத் துறையில் பல்வேறு அறிஞர்களின் கருத்துக்களை ஒருங்கிணைத்து பார்க்க வேண்டும் .

தமிழாக்கம் செய்யவேண்டிய கலைச்சொற்கள்

செல்பேசியைத் திறந்தவுடன் நம் கண்ணில் படுவன கீழ்க்கண்ட சொற்களாகும் . கணினியில் ஏராளமான இடங்களில் நுழைந்த தமிழ் இந்த செல்பேசிக்குள்ளும் நுழைந்து தமிழ்மணம் வீசவேண்டும் என்ற நோக்கில் அவற்றுக்குரிய கலைச்சொற்கள் இங்கே உருவாக்கப்பெறுகின்றன . செல்பேசி இயங்கும் தன்மை, மற்றும் பொருண்மை, கருத்து, பயன்பாடு ஆகியவற்றின் அடிப்படையில் கலைச்சொல்லாக்கம் நிகழ்த்தப்பெறவேண்டும் . சூழல் பயன்பாடு இங்கு முக்கியமாகின்றது .

உருவாக்கிய கலைச்சொற்கள்

sweep screen to unlock-திரையைத் தடவி உள்நுழைக .	Service numbers- சேவை எண்
On-தொடங்கு	Settings-அமைப்புகள்
Off-நிறுத்து	Message-செய்தி
logs –அழைப்புகள்	Send- அனுப்பு
all calls-அனைத்து அழைப்புகள்	Add media-ஊடக இணைப்பு
missed calls-தவறிய அழைப்புகள்	Picture-படம்
received calls-ஏற்ற அழைப்புகள்	Video-காட்சிப்படம்
dialled calls-பேசிய அழைப்புகள்	Sound-ஒலி
copy to dialling screen-திரையிலிருந்து பேச ஒட்டுக .	Photo-நிழற்படம் எடு
add to reject list – தவிர்க்கும் அழைப்புப் பட்டியல்	Video record-காட்சிப்பதிவு செய்
Log manager- அழைப்பு நிர்வாகி	Sound record-ஒலிப்பதிவு
block by address-முகவரி தடு	Add text-பிற செய்தி
Message-செய்தி	Add subject-பல்லுடகச் செய்திவழி
Delete-அழி	Attach item-இணைப்பு வகை

More-பிற	File-கோப்பு
Contact- தொடர்புகள்	Name card-முகவரி அட்டை
Groups-குழுக்கள்	Calendar-நாட்காட்டி
Favourite-விருப்பமானவை	Task-திட்டமிடல்
Assign-தேர்ந்து	Memo-குறிப்பாணை
Create contact-தொடர்பு எண் உருவாக்கம்	Sending options-அனுப்பும் வசதி
Copy-ஒட்டு	Reply path-பதில்வழி
Send name card via- முகவரி பட்டை அனுப்பு வழி	Request delivery-சேர்ந்த நிலை
Keep a copy-படியெடு	
Create contact-தொடர்பு எண் உருவாக்கம்	Sending options-அனுப்பும் வசதி
Copy-ஒட்டு	Reply path-பதில்வழி
Send name card via- முகவரி பட்டை அனுப்பு வழி	Request delivery-சேர்ந்த நிலை
Keep a copy-படியெடு	
Music- இசை	Email- மின்னஞ்சல்
Push message-அனுப்பிய செய்திகள்	
All tracks-அனைத்துப் பாடல்	Voicemail number- குரல் அஞ்சல் எண்
Settings- அமைப்பு	Sending options- அனுப்புதல் வசதிகள்
Effects- கேட்பு இனிமை	Receiving options- ஏற்றல் வசதிகள்
Internet- இணையம்	Help- உதவி
Go to home page- முதல் பக்கம் செல்க.	Memory status- நினைவக நிலை
History- முகவரிகள்	Available- காலியிடம்
Messages-செய்திகள்	Busy count- பயன்படுத்திய பகுதி
Create message-செய்தி உருவாக்கு	Phone- செல்பேசியில்
Create email –மின்னஞ்சல் உருவாக்கு	Sim- சிம் அட்டை
Conversations-உரையாடல்	General settings- பொது அமைப்புகள்

Inbox-செய்திப்பெட்டி	Keep a copy- படியெடுக்க
Email inbox-மின்னஞ்சல் பெட்டி	Delete by folder- கோப்பினை அழிக்க
Sentbox-அனுப்பிய செய்திகள்	Block messages- செய்திகளைத் தடுக்க
Outbox-வெளிப்புறச் செய்திகள்	Alert repetitions- மறுமுறை நினைவூட்ட
Drafts-செய்திக்கோர்வை	Once- ஒருமுறை
Myfolder-செய்திப்பெட்டி	Every2minutues- ஒவ்வொரு இரண்டு நிமிடமும்
Voicemail-குரல்வழி அஞ்சல்	Myfiles- கோப்பி
Broadcast message-ஒளிபரப்பு வகை	Images- படங்கள்
Spam box-அழித்த செய்திகள்	Video-காட்சிப்படம்
Settings-அமைப்பு	Sound- ஒலிகள்
General settings- பொது அமைப்பு	Other- பிற
Textmessage- எழுத்துச்செய்தி	Create folder- கோப்பியை உருவாக்கு
Multimedia message- பல்லாடகச் செய்தி	Wallpaper- முகப்புத்தாள்
Details- விபரங்கள்	Theme- பின்னொளிர் படம்
Time- நேரம்	Widgets- அமைப்பு வசதி
Type-வகை	Shortcuts- எளிதில் செல்ல
Name-பெயர்	Font type- எழுத்துரு
Size-அளவு	Brightness- வண்ண அளவு
Calendar-நாள்காட்டி	Backlight- பின்னொளிச்செறிவு
Create- உருவாக்கு	Call- அழைப்பு
Appointment- நாள்குறிப்பு	Voice call- குரலழைப்பு
Anniverasary- ஆண்டுவிழா	Fake call- முகவரியற்ற அழைப்பு
Event-நிகழ்வு	Backup manager- மீளெடுப்பு பாதுகாப்பு
Listby- வரிசைமுறையாக்கு	Viewby- வகைகாணல்

Week- வாராந்திர	Backup- மீளெடுப்பு
Day-நாள்	Restore-மறு உள்வீடு
More-பிற	Help- உதவி
Go to day- இன்றைய நாளுக்குள் செல்	Security- பாதுகாக்க
Go to date- தேதிக்குச் செல்	Reset setting-மறுசீரமைப்பு
Delete- அழி	Master reset- மைய மறுசீர்
This month- இந்த மாதம்	Alarms-நினைவு ஒலி
Before date- இந்த தேதிக்கு முன்	Communities- சமுதாய ஒருங்கிணைவு
All events- அனைத்து நிகழ்வும்	Facebook- முகநூல்
Camera- படக்கருவி	Bluetooth- அருகமை பரிமாற்றம்
Sim card management- சிம்கார்டு நிர்வாகம்	Memo- குறிப்பாணை
Sound profile- ஒலிக்கோப்பு கையாளல்	Task- திட்டமிடு
Display-காண்திரை	FM Radio- பண்பலை
Time and date- நேரமும் நாளும்	Clock-கடிகாரம்
Network- இணைப்பு	Games- விளையாட்டு
Sim auto network- தானியங்கு வலையம்	Applications-பயன்பாடு
Registration- பதிவு	Voice recorder- குரல் பதிவன்
Sound-ஒலி	Calculator-கணக்கி
Normal- இயல்பு	Convertor-மதிப்பு மாற்றி
Silent-ஒலியிலா	Dictionary- அகராதி
Driving-வாகன நேரம்	Display- காண்திரை, காட்சித்திரை
Outdoor- வெளிவேலை	---
Offline-இயங்காநிலை	--

ஏறத்தாழ நூற்று எழுபது கலைச்சொற்கள் இங்கே உருவாக்கிக் கொடுக்கப்பட்டுள்ளன. இவையனைத்தையும் செல்பேசியில் கணினி நிரல் வல்லுனர்கள் ஏற்றினால் தமிழ் உலகமெலாம் பரவும் வகையில் எளிதில் சென்றடையும்.

நிறைவுரை

அறிவியலும் தொழில் நுட்பமும் வளர்ந்த இந்த சூழலில் செல்பேசிப் பயன்பாட்டுக்குத் தனித்தமிழ்ச்சொற்கள் உருவாக்கப்பெறல் வேண்டும் . அதற்குரிய முயற்சியில் தமிழார்வலர்களும் மொழிப்பற்றாளர்களும் செயல்படவேண்டும். முன்பே உருவாக்கப்பெற்ற சில கலைச்சொற்களைப் பயன்படுத்தியும் வழக்குச்சொற்களுக்கு முக்கியத்துவம் கொடுத்துக் கலைச்சொற்களை ஆக்கவேண்டும் . அதற்குரிய பணியினை தகவல் தொழில்நுட்பவியலாளர்களும் தமிழ் வல்லுனர்களும் ஏற்றால் தமிழ் தகவல் தொடர்புத்துறையில் மேலும் வளம்பெறும்.

ஆதாரநூல்கள்

1. இராதா செல்லப்பன், (1985) கலைச்சொல்லாக்கம், மங்கை, வாத்தியார் தோட்டம், சென்னை
2. இராதா செல்லப்பன் (1992), அறிவியல் தமிழ் இன்றையநிலை, உலகத்தமிழராய்ச்சி நிறுவனம், சென்னை.
3. இராதா செல்லப்பன் (2008), மொழியியல், கவிதை அமுதம், திருச்சி
4. இராதா செல்லப்பன் (2011), தமிழும் கணினியும், கவிதை அமுதம், திருச்சி
5. சண்முகம்.செ.வை. (1994), அறிவியல் தமிழாக்கம், என்.சி.பி.எச். சென்னை.
6. சுந்தரம்.இராம (1999), பொருள் புதிது, வளம் புதிது (1999), வசந்தம் பதிப்பகம், தஞ்சாவூர்
7. ஜானகிராமன் (2003), அறிவியல் கலைச்சொல்லாக்கம், இராகவன் பதிப்பகம். விருத்தாசலம்
8. ஜானகிராமன் (2004) அறிவியல் தமிழ் வரலாறு (2004), இராகவன் பதிப்பகம், விருத்தாசலம்.

களம்சார்ந்த கலைச்சொல்லாக்கம்

மு.சிவலிங்கம்

சென்னை, இந்தியா
musivalingam@gmail.com

தமிழக அரசு 2000-ஆம் ஆண்டில், பல்வேறு துறைகளுக்காகப் பதினான்கு கலைச்சொல் பணிக்குழுக்களை அமைத்தது. தகவல் தொழில்நுட்பத்துக்காக முனைவர் மு. ஆனந்தகிருஷ்ணன் அவர்கள் தலைமையில் அமைந்த பணிக்குழு, கலைச்சொல்லாக்க உத்திகளை வகுத்துக் கொடுத்தது. அதில் 8-வது உத்தியாக, ஒரே ஆங்கிலச் சொல் வேறு வேறு இடங்களில், வேறு வேறு பொருளில் பயன்படுத்தப்படும்போது, இடம் பொருள் எவல் அறிந்து, தமிழில் வேறு வேறு கலைச்சொற்களை உருவாக்க வேண்டும் என்று பரிந்துரைக்கப்பட்டிருந்தது.

ஓர் ஆங்கிலக் கலைச்சொல் வெவ்வேறு துறைகளில் வெவ்வேறு பொருளில் பயன்படுத்தப்படலாம். அதன் பொருளறிந்து, கவனமாகத் தமிழ்க் கலைச்சொல்லை உருவாக்க வேண்டும். எனவே, கலைச்சொல்லாக்கத்துக்குத் தமிழறிவு மட்டுமின்றித் துறைசார்ந்த அறிவு மிகவும் இன்றியமையாதது ஆகிறது. ஆங்கிலமும், தமிழும் தெரிந்துவிட்டால் ஆங்கிலச் சொல்லுக்கு ஈடான தமிழ்க் கலைச்சொல்லை உருவாக்கிவிட முடியாது. இயற்பியல், வேதியியல், கணிதம், மருத்துவம், பொறியியல், கணிப்பொறியியல், இன்னும் இதுபோன்று பல்வேறு துறைகளிலும் புழங்கும் ஆங்கிலக் கலைச்சொல்லுக்குப் பொருத்தமான தமிழ்க் கலைச்சொல்லை உருவாக்க அந்தந்தத் துறைசார்ந்த அறிவு அவசியமாகிறது.

Bus, Driver, Conductor ஆகிய மூன்று சொற்களை எடுத்துக் கொள்வோம். தமிழில் இச்சொற்கள் முறையே பேருந்து, ஓட்டுநர், நடத்துநர் என்று வழங்கப்படுவதை அறிவோம். சுவையான தகவல் என்னவெனில், இந்த மூன்று சொற்களுமே கணிப்பொறி இயலில் இடம்பெற்றுள்ளன. கணிப்பொறியியல் அறிவு இல்லாத ஒருவர், கணிப்பொறி அறிவியலுக்கான கலைச்சொல் அகராதி ஒன்றைத் தயாரிக்கிறார், அல்லது கணிப்பொறித் துறையில் ஒரு நூலை மொழிபெயர்க்கிறார் என வைத்துக் கொள்வோம். அவர் Bus என்ற சொல்லுக்குப் பேருந்து என்றும், Driver என்ற சொல்லுக்கு ஓட்டுநர் என்றும், Conductor என்ற சொல்லுக்கு நடத்துநர் என்றும் குறிப்பிட்டிருப்பார். அதைப் படிக்க நேருகின்ற கணிப்பொறி அறிவியலாளர்கள் குலுங்கிச் சிரிப்பர். இது ஏதோ நகைச்சுவைக்காகச் சொல்லப்பட்ட செய்தியன்று. இதுபோன்ற கலைச்சொற்கள் கொண்ட கணிப்பொறியியல் அகராதிகள் வெளிவந்துள்ளன என்பதை வருத்தத்தோடு இங்கே குறிப்பிட வேண்டியுள்ளது.

மின்னணுத் துறையில் இரு வயரில் அமைந்த மின் இணைப்பை Channel, Circuit, Line என்று குறிப்பிடுவர். இரு முனைகளுக்கு இடையே தகவல் பயணம் செய்யும் இணைப்பை Path, Link என்று கூடக் குறிப்பிடுவதுண்டு. இச்சொற்கள் தமிழில் பாதை, தடம், இணைப்பு, தொடுப்பு என்று அழைக்கப்படுகின்றன. இவற்றில் தகவல் பிட்டுகள் ஒன்றன்பின் ஒன்றாகப் பயணிக்கும். Bus என்பது 8, 16, 32 வயர்களின் இணைப்பில் அமைந்த தகவல் பரிமாற்றத் தடமாகும். ஒரே நேரத்தில் 8, 16, 32 பிட்டுகள் இணையாகப் பயணிக்கும். Bus என்பது அகலமான தகவல் பரிமாற்றப் பாதையைக் குறிக்கிறது. எனவே, Bus என்பதைப் பாட்டை என்கிறோம்.

கணிப்பொறியோடு இயைந்து செயல்படும், விசைப்பலகை (Keyboard), சுட்டி (Mouse), அச்சுப்பொறி (Printer), வருடி (Scanner), வரைவி (Plotter) போன்ற புறச்சாதனங்கள் பல உள்ளன. கணிப்பொறியானது, இந்தப் புறச்சாதனங்களோடு தகவல் பரிமாற்றம் செய்ய வேண்டுமெனில், அவற்றின் மொழியை அறிந்து வைத்திருக்க வேண்டும். இந்த மொழிபெயர்ப்புக்கு, அதாவது கணிப்பொறிக்கும் புறச்சாதனத்துக்கும் இடையேயான தகவல் பரிமாற்றத்துக்கு, Device Driver என்னும் நிரல் உதவுகிறது. சுருக்கமாக Driver என்பர். புறச்சாதனத்தை இயக்க உதவும் நிரல் என்பதால் Driver என்ற சொல்லுக்குக் கணிப்பொறியியலில் இயக்கி என்ற சொல் பயன்படுத்தப்படுகிறது.

பெரும்பாலான உலோகங்கள் மின்சாரத்தைக் கடத்துகின்றன. அவற்றுள் எளிதில் கடத்துபவையும் உண்டு. அரிதில் கடத்துபவையும் உண்டு. ஆங்கிலத்தில், அவற்றை முறையே, Good Conductor, Bad Conductor என்று குறிப்பிடுகின்றனர். எனவே, Conductor என்னும் சொல் கடத்தி என்னும் பொருளிலேயே பயன்படுத்தப்படுகிறது.

கணிப்பொறிக்கான மின்னணுச் சாதனங்கள் தயாரிக்க Semi-conductor பயன்படுகிறது. அதனை குறைக்கடத்தி அல்லது அரைக்கடத்தி என்று குறிப்பிடுகின்றனர்.

ஆக, கணிப்பொறித் தொழில்நுட்பத்தை அறிந்தவர்கள் Bus, Driver, Conductor ஆகிய ஆங்கிலச் சொற்களுக்கு முறையே, பாட்டை, இயக்கி, கடத்தி என்பது போன்ற தமிழ்க் கலைச்சொற்களையே பயன்படுத்துவர் . பேருந்து, ஓட்டுநர், நடத்துநர் என்றெல்லாம் கூறிக் கொண்டிருக்க மாட்டார்கள் . இதுபோன்று இன்னும் பல சொற்களைக் குறிப்பிடலாம். கணிப்பொறியியலில் விருப்பத் தேர்வுகளை உள்ளடக்கிய பட்டியல் அல்லது பட்டியைக் குறிக்கின்ற Menu என்ற சொல்லுக்கு உணவு விலைப் பட்டியல் என்று எழுதப்பட்ட கணிப்பொறிப் பாடப் புத்தகமும் வெளிவந்துள்ளது என்பதை இங்குக் குறிப்பிட வேண்டும் .

சாதாரண ஆங்கிலம்- தமிழ் அகராதியில் Line என்றால் கோடு என்றுதான் பொருளிருக்கும் . ஆனால், தகவல்தொடர்பியல் அகராதியில் Line என்றால் இணைப்பு என்றும், தடம் என்றும் பொருள் கூறப்பட்டிருக்கும். அதேபோல்தான், argument என்ற சொல்லின் நேரடிப் பொருள் வாதம் அல்லது விவாதம். ஆனால், கணிப்பொறி இயலில் arguments என்னும் சொல் , ஒரு செயல்கூறுவுக்கு (Function) அனுப்பப்படும் உருபுகளைக் குறிப்பதால் செயலுருபுகள் என்கின்றனர். இதுபோல ஏராளமான எடுத்த துக்காட்டுகளைக் கூற முடியும் . ஆக, கணிப்பொறித் துறைசார்ந்த கலைச்சொல்லாக்கம் , அத்துறைசார்ந்த புலமை இல்லையேல் நகைப்புக்கு இலக்காகும் என்பது கண்கூடு.

குறிப்பிடத்தக்க செய்தி என்னவென்றால் , ஒரே துறையிலேயே ஒரு குறிப்பிட்ட சொல் வெவ்வேறு துறைப்பிரிவுகளில் வெவ்வேறு பொருளில் பயன்படுத்தப்படுகிறது என்பதுதான் . இன்னும் ஆழமாகச் சென்று பார்த்தால், ஒரே துறைப்பிரிவில்கூட ஒரு சொல் வெவ்வேறு இடங்களில் வெவ்வேறு பொருளில் பயன்படுத்தப்படுவதைப் பார்க்க முடியும்.

Key என்னும் சொல்லுக்குச் சாவி , திறவுகோல் என்று பொருள் கூறுகிறோம் . ஆனால், Answer Key, Key Stone, Key Industry, Key Source, Key Map ஆகிய சொற்களுக்கு முறையே விடைக் குறிப்பு , நடுகல், அடிப்படைத் தொழில் , முதன்மை மூலம் , கைப்படம் எனப் பொருள் கூறப்படுகிறது . இசையியலில் Keyboard, Key Note என்று இரு சொற்கள் புழக்கத்தில் உள்ளன . இவற்றை அத்துறையினர் முறையே , இசைப்பலகை, தொடங்கு சுரம் என்று வெவ்வேறு சொற்களில் குறிப்பிடுகின்றனர் . ஒருசொல் வெவ்வேறு துறைகளில் வெவ்வேறு விதமாகவும் , ஒரே துறையில்கூட வெவ்வேறு களங்களில் வெவ்வேறு விதமாகவும் பயன்படுத்தப்படுகிறது என்பதற்கு Key என்னும் சொல் சரியான எடுத்துக்காட்டாகும் .

கணிப்பொறித் துறைசார்ந்த கலைச்சொல்லாக்கத்துக்கு , அத்துறைசார்ந்த புலமை போதும் என்பது பாதி உண்மைதான். சி, சி++, சி#, ஜாவா போன்ற கணிப்பொறி நிரலாக்க மொழிகள் (Programming Languages) பற்றிய பாடங்களையும், நூல்களையும் எழுத முனைந்தபோதுதான் , என்னால் இந்தப் பாதி உண்மையை முழுமையாகப் புரிந்து கொள்ள முடிந்தது . கணிப்பொறி நிரலாக்கத்தில் பயன்படும் ஆங்கிலக் கலைச்சொற்களுக்கு நிகரான தமிழ்க் கலைச்சொற்களை உருவாக்க மேம்போக்கான கணிப்பொறி அறிவு மட்டும் போதாது , கணிப்பொறி நிரலாக்கம் பற்றியும் தெரிந்திருக்க வேண்டு ம் என்பதை உணர முடிந்தது . பொறியியல் கல்லூரிகளில் கணிப்பொறித் தொழில்நுட்பம் பயிலும் மாணவர்களுக்கு ஒரு பாடமாக இருக்கின்ற , பிணையத் தொழில்நுட்பம் (Network Technology) பற்றிய ஒரு முழுமையான பாடநூலைத் தமிழில் எழுதியபோது , குறிப்பிட்ட புலம் அல்லது களம்சார்ந்த கலைச்சொ ல்லாக்கத்துக்கு, அந்தந்தப் புலம்சார்ந்த அறிவு (Field Knowledge) அல்லது களம்சார்ந்த அறிவு (Domain Knowledge) எவ்வளவு இன்றியமையாதது என்பதை அனுபவ பூர்வமாக உணர முடிந்தது .

கணிப்பொறி இயலில் , Bus, Drive, Driver, Argument, Operator, Compile, Platform, Exception, Switch, Key, Polling போன்ற பல்வேறு சொற்கள் வழக்கமான பொருளில் இல்லாமல் , வேறான பொருளைக் குறிக்குமாறு பயன்படுத்தப்படுகின்றன. Key, Socket, Gateway, Distribution, Interface போன்ற சொற்கள் , கணிப்பொறி இயலிலேயே வெவ்வேறு இடங்களில் வெவ்வேறான பொருளில் பயன்படுத்தப்படு கின்றன. இத்தகைய சொற்களுக்குத் தமிழில் கலைச்சொற்களை உருவாக்கும்போது துறைசார்ந்த அறிவோடு , குறிப்பிட்ட களம்சார்ந்த புலமையும் தேவைப்படுகிறது என்பதை நிறுவுவதே இவ்வாய்வுக் கட்டுரையின் நோக்கம் .

முதலில் Key என்ற சொல்லை எடுத்துக் கொள்வோம் . கணிப்பொறியியலில், Keyboard என்னும் சாதனத்தை விசைப்பலகை என்கிறோம். Tab Key, Shift Key, Ctrl Key, Alt Key, Insert Key ஆகியவற்றை முறையே தத்தல் விசை, நகர்வு விசை, கட்டுப்பாட்டு விசை, மாற்று விசை, செருகு விசை என்று சொல்கிறோம் . Key Press என்பதை விசையழுத்தம் என்கிறோம். அதாவது, Key என்ற சொல்லுக்கு ஈடாக விசை என்னும் சொல்லையே பயன்படுத்துகிறோம். கணிப்பொறிப் பிணையங்களைப் பற்றிப் படிக்கும்போது , Secret Key, Public Key, Private Key போன்ற சொற்களை எதிர் கொள்கிறோம் . தகவலை இரகசியமாக வைத்துப் பூட்டுவது , திறப்பது என்ற பொருளிலேயே இங்கு Key என்னும் சொல் பயன்படுத்தப்படுகிறது . எனவே, விசை என்ற சொல் அவ்வளவு பொருத்தமாகப் படவில்லை . சாவி எனப் பொருள்படும் திறவி என்னும் தூய தமிழ்ச்சொல் மிகவும் பொருத்தமாக இருக்கிறது. ஆகவே, Public Key, Private Key ஆகிய சொற்களைப் பொதுத்திறவி, தனித்திறவி என அழைக்கலாம் . Key Distribution, Key Management ஆகியவற்றைத் திறவி வினியோகம், திறவி மேலாண்மை எனலாம்.

Socket என்னும் சொல்லைப் பிணையத் தொழில்நுட்பத்தில் , பிணைய வன்பொருள்கள் , பிணைய மென்பொருள்கள் ஆகிய இரண்டைப் பற்றிப் பேசும்போதும் எதிர் கொள்கிறோம் . வன்பொருள்களில், இரண்டு மின்சார அல்லது மின்னணு சாதனங்களை இணைக்கும்போது , பெரும்பாலும் ஒன்றின் நுழைப்பகுதியை இன்னொன்றின் துளைப்பகுதியில் பொருத்துகிறோம் . இவற்றில் துளைப்பகுதியை Socket என்கிறோம். ஒரு சாதனம் அல்லது கம்பிவடத்தின் முனைப்பகுதி நுழைந்து பொருந்துகின்ற வாய்ப்பகுதி என்பதால் , Socket என்பதைத் தமிழில் பொருத்துவாய் என்கிறோம். ஆனால், பிணையத்தில் தகவல் பரிமாற்றம் தொடங்குவதற்கு முன், இரு முனைகளிலுமுள்ள சாதனங்கள், சரியாகச் சொல்வதெனில், இரு முனைகளிலுமுள்ள கணிப்பொறிகளில் செயல்படும் பயன்பாட்டு நிரல்கள் , தமக்கிடையே தகவல் பரிமாற்றம் மேற்கொள்ளும் முன்பு, ஒரு தருக்கநிலை இணைப்பை (logical connection) ஏற்படுத்திக் கொள்கின்றன . அந்தத் தருக்க இணைப்பின் இருபுறமுமுள்ள , கருத்தியலான (abstract) முனைகளே Socket எனப்படுகின்றன. பருநிலை இணைப்பில் (physical connection) பயன்படுத்தப்படும் பொருத்துவாய் என்னும் சொல்லை , தருக்கநிலை இணைப்பில் (logical connection) கருத்தியலான முனைகளுக்குப் பயன்படுத்துவது அவ்வளவு பொருத்தமாக இல்லை . என்னுடைய நூலில் நான் இணைப்புமுனை என்ற சொல்லைப் பயன்படுத்தியுள்ளேன் . ஆனாலும், இன்னும் சிறந்தவொரு பொருத்தமான சொல்லைக் கண்டறிய வேண்டும் என்று கருதுகிறேன்.

இதுபோலவே, Gateway என்னும் சொல்லைப் பிணையத் தொழில்நுட்பத்தில் , பிணைய வன்பொருள்கள், பிணைய மென்பொருள்கள் ஆகிய இரண்டைப் பற்றிப் பேசும்போதும் எதிர் கொள்கிறோம் . Gateway என்னும் சொல் நுழைவாயில் என்னும் பொருளைத் தருகிறது . இரண்டு பிணையங்களுக்கு இடையே தகவல் பரிமாற்றம் நடைபெறப் பல்வேறு சாதனங்கள் பயன்படுகின்றன . ஒத்தவகைப் பிணையங்களுக்கு இடையே , அனுப்பப்படும் தகவலைச் சரியான பாதையில் திசைவிக்க Router என்னும் சாதனம் பயன்படுகிறது . எனவே, அதனைத் திசைவி என்கிறோம். முற்றிலும் வேறுபட்ட பிணையங்களுக்கு இடையே தகவல் பரிமாற்றத்துக்கு Gateway பயன்படுகிறது. இது ஒரு சிறப்பு வகைத் திசைவி ஆகும். அதனை, நுழைவாயில் என நேரடிப் பொருளில் அழைக்காமல், ஒரு சாதனத்தைக் குறிப்பதற்கு ஏற்றவாறு , திசைவி என்ற சொல்லுக்கு இசைந்தவாறு , நுழைவி என்ற சொல்லால் குறிக்கிறோம் . ஆனால், அதேவேளையில், பிணைய மென்பொருளான நெறிமுறைகளைப் பற்றிப் படிக்கும்போது, இரண்டு தன்னாட்சிப் பிணைய முறைமைகளுக்கு இடையேயான தகவல் பரிமாற்றத்தில் , Border Gateway Protocol, Interior Gateway Protocol, Exterior Gateway Protocol என்றெல்லாம் படிக்கிறோம். இவற்றில் Gateway என்ற சொல் ஒரு சாதனத்தைச் சுட்டவில்லை . சாதனத்தின் கருத்தியலான தன்மையையே சுட்டுகிறது . எனவே, இங்கே நுழைவி என்ற சொல்லைவிட நுழைவாயில் என்ற சொல்லே பொருத்தமாக இருக்கிறது . Border Gateway Protocol என்பதை எல்லை நுழைவாயில் நெறிமுறை என்று குறிப்பிடுவதே அதன் பொருளைத் துல்லியமாக வெளிப்படுத்துவதாய் உள்ளது .

Distribution என்னும் சொல் கணிப்பொறியியலில் பல்வேறு இடங்களில் பயன்படுத்தப்படுகிறது . Distributed Systems, Distributed Database, Distributed Services என்றெல்லாம் பேசப்படுகிறது . இங்கெல்லாம் Distribution என்பதற்குப் பொதுவாகப் பகிர்மானம் என்றே பொருள் கொள்கிறோம் . பகிர்மான முறைமைகள் அல்லது பகிர்ந்தமை முறைமைகள் , பகிர்ந்தமை தரவுத்தளம் , பகிர்ந்தமை சேவைகள் என்று குறிப்பிடுகிறோம் . இங்கெல்லாம் பகிர்மானம் என்ற சொல்லை அடிப்படையாகக் கொண்ட பகிர்ந்தமை என்னும் சொல் பொருத்தமாகவே இருக்கிறது .

பிணையப் பாதுகாப்பு பற்றிப் படிக்கும்போது , ஒரு நிறுவனத்தின் எதிரிகள் அந்நிறுவனப் பிணையத்தின்மீது தொடுக்கும் பல்வேறு வகையான தாக்குதல்கள் பற்றிப் படிக்கிறோம் . அவற்றுள், சேவை-மறுப்புத் தாக்குதல் (Denial of Service Attack) என்பது ஒருவகையான தாக்குதல் ஆகும் . சுருக்கமாக, டிஓஎஸ் தாக்குதல் என்று கூறுவர் . பிணையங்களில் நுகர்வி (Client), வழங்கி (Server) கணிப்பொறிகளுக்கு இடையே , பொதுவாக, கோரிக்கை-பதிலுரை என்ற வடிவிலேயே தகவல் பரிமாற்றம் நடைபெறுகிறது . ஒரு நிறுவனத்தின் பிணையத் தைத் தாக்க நினைப்போர், அந்நிறுவன வழங்கிக்குப் போலியான கோரிக்கைகளைப் பல்லாயிரக் கணக்கில் அடுத்தடுத்துத் தொடர்ந்து அனுப்பிக் கொண்டே இருப்பர் . வழங்கியானது, அத்தனை கோரிக்கைகளுக்கும் பதிலுரை வழங்க முடியாமல் திக்குமுக்காடும் . அந்த நேரத்தில் வழக்கமான பயனர்களின் நியாயமான கோரிக்கைகளுக்குப் பதிலுரை அனுப்ப முடியாமல் வழங்கி திணறிப் போகும் . உண்மையான பயனர்களுக்கு நியாயமான சேவைகளை வழங்க முடியாத நிலை ஏற்படும் . அவர்களுக்கு வழக்கமான சேவைகள் மறுக்கப்படுவதால் இது சேவை-மறுப்புத் தாக்குதல் ஆயிற்று.

சில வேளைகளில் இந்தத் தாக்குதலை எதிரிகள் கடுமை ஆக்குவர் . ஒரு கணிப்பொறியிலிருந்து ஒருவர் தாக்குதல் நடத்துவதற்குப் பதிலாக , எதிரிக் குழுவைச் சேர்ந்த பலநூறு பேர் , உலகத்தின் பல்வேறு மூலைகளிலிருந்தும் , முன்பே திட்டமிட்டபடி , ஒரே நேரத்தில் சேவை மறுப்புத் தாக்குதலைத் தொடுப்பர் . இவ்வாறு பல முனைகளிலிருந்து தொடுக்கப்படும் சேவை மறுப்புத் தாக்குதல் ஆங்கிலத்தில் Distributed Denial of Service Attack (DDoS) எனப்படுகிறது. இதைத் தமிழில் எப்படிச் சொல்வது ? பகிர்ந்தமை சேவை மறுப்புத் தாக்குதல் எனலாமா? அவ்வளவு பொருத்தமாகப் படவில்லை அல்லவா ? பல்முனைச் சேவை-மறுப்புத் தாக்குதல் என்று கூறினால் மிகவும் பொருத்தமாக இருக்கும் . ஆங்கிலச் சொல்தொடரில் உள்ளடங்கியுள்ள பொருளும் தெளிவாகப் புலப்படும் .

எனவே, தகவல் தொழில்நுட்பப் பணிக்குழு வகுத்துத் தந்த கலைச்சொல்லாக்க உத்தியின்படி , ஆங்கிலக் கலைச்சொல்லின் நேரடியான மொழிபெயர் ப்பாக இல்லாமல் , அது உணர்த்துகின்ற உண்மையான பொருளைத் துல்லியமாக வெளிக்கொணருமாறு , இடம், பொருள், ஏவல் அறிந்து , தமிழ்க் கலைச்சொல்லை உருவாக்குவதே சிறந்த நடைமுறையாக இருக்கும் . அவ்வாறு, உண்மைமையான பொருளை வெளிக்கொணருமாறு கலைச்சொற்களை உருவாக்கக் களம்சார்ந்த புலமை (Domain Knowledge) கட்டாயத் தேவை என்பது , இதுகாறும் நாம் பார்த்த எடுத்துக்காட்டுகளின் மூலம் எளிதில் புலனாகும் .

மொழிபெயர்ப்புப் பணியில் தொழில்நுட்பத்தின் பயன்பாடு

சி.இளங்குமரன்

Sultan Idris University of Education & Malaysian Translators Association

வளர்ந்து வரும் தொழில்நுட்பமயமான உலகில் விவசாயம் தொடங்கி விண்வெளி ஆராய்ச்சி வரை தொழில்நுட்ப உதவிகளின்றி காரியம் ஆற்றுவதல் மிகச் சிரமமானதாக இருக்கின்றது . அந்த அளவிற்குத் தொழில்நுட்பம் அனைத்துத் துறைகளிலும் ஊடுருவி இருக்கின்றது . இத்தகைய சூழல் மொழிபெயர்ப்புத் துறை யிலும் ஊடுருவத் தவறவில்லை. உலகில் பல்வேறு நாடுகளைச் சேர்ந்த மக்கள் , பல்வேறு துறைகளில் கரை கண்டவர்களாக இருக்கின்றனர். சிந்தனை மற்றும் அறிவுப் பகிர்வினால்தான் நாடுகள் வளர்ச்சி அடையவியலும் . அதுதான் நடைமுறையும் கூட. இவ்வாறு உலக அளவில் வளர்ந்து வரும் பல்வேறு துறைகளின் வளர்ச்சியும், தொழில்நுட்ப மேம்பாடுகளும், குறிப்பிட்ட சில நாடுகளில் அபரிமித வளர்ச்சியைப் பதிவு செய்யும் வேளையில் , அதே வளர்ச்சி உலகின் பிற நாடுகளுக்கும் சென்று சேர மொழிபெயர்ப்புத் துறை இன்றியமையாததாகின்றது .

இரு துறைகளும் ஒன்றுக்கொன்று உதவியாய் இரு ந்து ஒன்றை மற்றொன்று வளர்த்து வருகின்றது என்றாலும் , இந்த ஆய்வில் மொழிபெயர்ப்புப் பணிக்குத் தொழில்நுட்பம் எந்த அளவு பயன்படுகின்றது என விவரிக்கப்படுகின்றது. தொழில்நுட்பப் பயன்பாட்டினால் தரமான மொழிபெயர்ப்புகள் உருவாவதற்கும் , மொழிபெயர்ப்புப் பணிகளை நிர்வகிக்கவு ம் மொழிபெயர்ப்பாளர்களுக்குத் தொழில்நுட்பம் எந்தெந்த வழிகளில் உதவிகள் புரிகின்றது; புரியலாம் என இவ்வாய்வு ஆய்கின்றது .

அடிப்படையில், மொழிபெயர்ப்புத் தொழில்நுட்பங்கள் மற்றும் அதன் கருவிகள் ஆகியன மொழிபெயர்ப்பாளர்களின் பணிகளை இலகுவாக்குவதற்காகத்தான் கண்டுபிடி க்கப்பட்டிருக்கின்றன. தனிநபர் மொழிபெயர்ப்பாளர்கள் மற்றும் மொழிபெயர்ப்பு நிறுவனங்களுக்கு இத்தகைய தொழில்நுட்பங்கள் பேருதவியாக அமைகின்றன.

மொழிபெயர்ப்புத் தொழில்நுட்பங்கள் எந்தெந்தப் பணிகளைச் செய்ய உதவுகின்றன?

1) தட்டச்சு – எந்தவொரு மொழி சம்பந்தப்பட்ட வேலைகளையும் போலவே மொழிபெயர்ப்புத் துறையிலும் தட்டச்சு செய்ய அடிப்படையான தொழில்நுட்பம் உதவுகின்றது (அடிப்படையான தொழில்நுட்பம் என்றாலும் கூட தட்டச்சு இயந்திரத்தில் இருந்து கணினிக்கு வளர்ந்திருக்கும் இன்றைய தொழில்நுட்பத்தில் பல விஷயங்களைக் கூடுதலாகச் செய்ய முடிகிறது)

2) சொற்களின் எண்ணிக்கையை அறிய- பொதுவான கண்ணோட்டத்தில் பார்க்கும் போது மிக அற்பமாகப் படும் இது, மொழிபெயர்ப்பாளர்களைப் பொருத்தமட்டில் அதி முக்கியமானதொன்றாக இருக்கின்றது . பெரும்பாலான மொழிபெயர்ப்புப் பணிகளில் சொற்களின் எண்ணிக்கையை அடிப்படையாகக் கொண்டே கட்டணத்தை நிர்ணயிப்பதே இதற்குக் காரணமாகும் . தட்டச்சு செய்தவுடனேயே ‘Microsoft word’ –இல் சொற்களின் எண்ணிக்கையைத் தெரிந்து கொள்ளலாமே என்பவர்களுக்கு , சில வேலைகளில் , PDF மற்றும் படங்களாகவும் மொழிபெயர்ப்புப் பணிகள் தரப்படுவதையும் அத்தகைய வேலைகளுக்காகப் பிரத்தியேகமாக எண்ணிக்கையை அறியும் மென்பொருட்கள் அவசியம் என்பதே பதிலாகும் .

3) சொற்பிழை திருத்தி – மொழிபெயர்கப்படும் ஆவணங்களைப் பிழையின்றி தர வேண்டியது ஒவ்வொரு மொழிபெயர்ப்பாளரின் கடமையாதலால் அதனை இலகுவாகச் செய்யவும் மென்பொருட்கள் உள்ளன . (சிறப்பாகத் தமிழ் மொழிக்கான சொற்பிழை திருத்திக்காகவே இங்கு இதனை இணைத்துள்ளேன்)

4) PDF – கோப்பாக மாற்றுதல் - மொழிபெயர்க்கப்பட்டு, தட்டச்சு செய்யப்பட்டாலும் குறிப்பிட்ட எழுத்துருக்கள் (குறிப்பாகத் தமிழ் மொழி) வாடிக்கையாளரிடம் இல்லாது போனால் அவ்வாவணத்தைப் PDF முறையில் அனுப்பவும் தொழில்நுட்பமே ஆங்கு உதவி புரிகின்றது .

5) மின்னியல் அகராதிகள் – மொழிபெயர்ப்புப் பணிகளுக்கு அகராதிகளின் பயன்பாடு அதி முக்கியமானதாகும் . அதிலும் தற்போதைய சூழலில் மொழிபெயர்ப்புகளைத் தாள்களில் எழுதி பின்னர் தட்டச்சு செய்யும் நிலை மாறி

நேரடியாகவே கணினிகளிலேயே உடன் மொழிபெயர்த்துத் தட்டச்சு செய்யப்படும் சூழலில் மின்னியல் வடிவில் இருக்கும் அகராதிகளே பெரிதும் உபயோகப்படுத்தப்பட்டு வருகின்றன .

6) முகப்புரை (Subtitle) – திரைப்படங்கள், தொலைகாட்சி நிகழ்ச்சிகள் ஆகியவற்றில் வழங்கப்படும் முகப்புரை மொழிபெயர்ப்புத் துறைக்கு தொழில்நுட்பத்தின் வாயிலாக விளைந்தேயாகும் .

7) உரைபெயர்ப்பு (Interpretation) – தொழில்நுட்பத்தின் உதவியின்றி உரைபெயர்ப்பு மிகவும் சிரமமானதொன்றாகிவிடும் . குறிப்பாக, மாநாடுகளில் ஒருவர் பேசும் உரையைப் பலரும் அவரவர் மொழிகளில் கேட்பதற்குத் தொழில்நுட்பத்தின் உதவி மிக அருமையானதாகும் .

ஆக, பயனீட்டாளர் அல்லது வாடிக்கையாளர்களிடமிருந்து மொழிபெயர்ப்புப் பணி ஒன்றினைப் பெருவதிலிருந்து அதனை முடித்து மீண்டும் அவரிடம் ஒப்படைப்பது வரை தொழில்நுட்பத்தின் பங்கு மிக அவசியமாக இருக்கின்றது.

இதனைத் தொடர்ந்து சில முறைமைகளின் (System) அறிமுகம் மொழிபெயர்ப்புப் பணிகள் செய்வதற்குப் பேருதவியாக இருக்கின்றன . இவற்றின் வழி விரைவானதாகவும் தரமானதாகவும் மொழிபெயர்ப்புப் பணிகள் அமைகின்றன என்றால் மிகையிலலை .

சொல்லியல் (Terminology) நிர்வாக முறைமை

இந்த முறைமையின் வழி மிக நீண்ட மொழிபெயர்ப்பு வேலைகளை நேர்த்தியாகச் செய்ய முடிகின்றது . எடுத்துக்காட்டாக, ஆண்டறிக்கைகள் போன்ற நீளமான மொழிபெயர்ப்புகளின் போது , மொழிபெயர்ப்பாளர் குறிப்பிட்ட ஒரு வார்த்தைக்குப் பல்வேறு சொற்களைத் தந்துவிட வாய்ப்புகள் உள்ளன . இதனால் பொருள் மாற்றம் ஏற்படாவிடினும் , வாசகர்களுக்கு அப்பகுதியைப் படிப்பதில் குழப்பம் ஏற்படலாம் . ஆகையால், மொழிபெயர்ப்பில் வார்த்தைகளின் நிலைப்பேற்றைத் (Consistency) தக்க வைக்க இது மிகவும் அவசியமானதாகின்றது . இவ்வொருங்கியத்தைத் தனிநபர் மொழிபெயர்ப்பாளர்கள் பின்பற்றலாம் . மேலும், பல மொழிபெயர்ப்பாளர்கள் இணையப் பரிமாறல்களின் வழி சொற்களஞ்சியங்களை இவ்வொருங்கியத்தின் துணைகொண்டு பகிர்ந்துகொள்ளலாம் . இதன் வழி, மொழிபெயர்ப்பின் தரம் உயர வாய்ப்பு ஏற்படுகின்றது .

மொழிபெயர்ப்பு நினைவக முறைமை

இந்த நினைவக (Memory) முறைமையின் வாயிலாகப் புத்தக மொழிபெயர்ப்பு அல்லது ஒரு குறிப்பிட்ட துறையில் தொடர்ந்து பல மொழிபெயர்ப்புகளைச் செய்துவரும் மொழிபெயர்ப்பாளர்களுக்கு இம்முறைமை வரப்பிரசாதமாகும் . இம்முறைமையின் வாயிலாகச் செய்யப்படும் மொழிபெயர்ப்புகள் தானியங்கியாக நினைவகத்தில் சேமிக்கப்படுகின்றன . பிரிதொரு தருணத்தில் ஏற் கனவே செய்த துறையில் மீண்டும் ஒரு சில வார்த்தைகள் வரும்போது இம்முறைமை முந்நாளில் செய்த மொழிபெயர்ப்பை நமக்குத் திரையில் காட்டும் . அப்படி அவை தேவைப்பட்டால் , அதனைத் தெரிவு செய்து தற்போது தாங்கள் செய்யும் மொழிபெயர்ப்புப் பணிக்குப் பயன்படுத்திக்கொள்ளலாம் . இல்லையேல் தொடர்ந்து புதிதாக மொழிபெயர்க்கலாம் . அனைத்து மொழிபெயர்ப்பாளர்களுக்கும் இஃது உதவும் என்றாலும் , சட்டத்துறையில் மொழிபெயர்ப்பு செய்பவர்களுக்கு இம்முறை பெரும் உதவியாக அமைகின்றது . இத்தகையப் பணிகளின் கோப்புகள் இணையம் வழி பகிர்ந்துகொள்ளப்படும்போது, அதிக எண்ணிக்கையிலான மொழிபெயர்ப்பாளர்கள் நன்மை அடைகின்றனர் .

ஒருங்கிணைக்கப்பட்ட மொழிபெயர்ப்பு கையாள்கை முறைமை

இந்த முறைமையின் வழி , கிட்டதட்ட அனைத்து மொழிபெயர்ப்பாளர்களும் பெரும் நன்மை அடைகின்றனர் . எப்படி? இம்முறைமையின் வாயிலாக, மொழிபெயர்ப்பு வேலைகளுக்கு பொதுவாகச் செய்யப்படும், விலைப்பட்டி தயாரித்தல், மொழிபெயர்க்கப்படவிருக்கும் சொற்களின் எண்ணிக்கையைக் கணக்கிடுதல் , மொழிபெயர்ப்பு விவரப்பட்டியல், மீள்-சோதனை/தேடல், தானியங்கியாகச் செய்யப்படும் விலைப்பட்டி (குறிப்பிட்ட பகுதியில் இருக்கும் சொற்களின் எண்ணிக்கையின் அடிப்படையி ல் விலையை , இம்முறையே நிர்ணயம் செய்துவிடும்) மொழிபெயர்ப்புப் பணிகள் வாடிக்கையாளர்களுக்குத் திரும்ப ஒப்படைக்க வேண்டிய நாள் , குறிப்பிட்டதொரு

பணிகளில் இடம்பெற்றிருக்கும் ஆவணங்கள் , இதுவே, பெரிய மொழிபெயர்ப்பு நிறுவனங்களாக இருப்பின் ; மொழிபெயர்ப்பாளர்களின் விபரம், இன்னும் மொழிபெயர்க்கப்படாத பக்கங்களின் எண்ணிக்கை , மொழிபெயர்ப்பு வேலைகள் முடிந்துவிட்டால் அதனைத் திருத்த அனுப்பியிருக்கும் பதிப்பாசிரியரின் விவரங்கள் போன்ற பணிகளை மிக இலகுவாக , மிக மிக குறைந்த கால அளவில் தயார் செய்துவிடலாம் . விலைப்பட்டி தயார் செய்து கொடுக்கப்பட்ட அனைவரும் அம்மொழிபெயர்ப்பாளரிடமே அப்பணியை ஒப்படைப்பர் என்ற கட்டாயம் இல்லை. பிற நிறுவனங்களுடன் ஒப்பீடு செய்வதற்காக மட்டுமே கோரப்படும் விலைப்பட்டிகள் ஏராளம் . அவ்வாறான தருணங்களில் , நிலையற்ற காரியங்களில் ஈடுபட்டு நேரத்தை விரையமாக்குவதை இதன் வழி தவிர்க்கலாம்.

இயந்திர மொழிபெயர்ப்பு முறைமை

இந்த இயந்திர மொழிபெயர்ப்பு குறித்து ஆராயும் வேளையில் , பல மொழிகளுக்கு இயந்திர மொழிபெயர்ப்பு முறைமை சற்று செம்மைப்படுத்தப்பட்டிருந்தபோதும் , தமிழ் மொழியைப் பொருத்த மட்டில் அஃது இன்னும் ஆரம்ப நிலையில்தான் உள்ளது . தமிழ் மொழியின் ஆழ்ந்த நுட்பம் இதற்குக் காரணமாக அமைகிறது . எவ்வாறாயினும், பல ஆண்டுகள் ஆராய்ச்சிகளுக்குப் பின்னரும் , இயந்திர மொழிபெயர்ப்பு முறைமையைப் பயன்படுத்தி செய்யப்படும் மொழிபெயர்ப்புகள் , எந்த மொழியிலும் 100% முழுமைத்தன்மையைத் தருவதில்லை , மனித மூளையின் பங்களிப்பு ஆங்கு அத்தியாவசியமானதாகின்றது . இருப்பினும் இந்த முறைமை மூல மொழியில் இருந்து மொழிபெயர்க்கப்படவிருக்கும் மொழிக்குக் குறிப்பிட்ட தரவுகளை மாற்றி , அதில் ஓர் ஒட்டுமொத்த விளக்கத்தைப் பெற ஏதுவாக இருக்கின்றது . அதன் பிறகு , மொழிபெயர்ப்பாளர்கள் தங்களின் உத்திகளைப் பயன்படுத்தி மொழிபெயர்ப்புப் பகுதிகளைச் செம்மைப்படுத்த வேண்டியிருக்கின்றது .

திறம்பட்ட இணைய இயங்குதளங்கள்

இத்தகைய இணைய இயங்குதளங்களின் (Platform) பணி என்னவென்றால், நாம் ஏற்கனவே பார்த்ததைப்போல் , தனிநபர் மொழிபெயர்ப்பாளர்கள் தாங்கள் செய்யும் மொழிபெயர்ப்புப் பணிகளை இத்தகு இயங்குதளங்களில் பதிவு செய்கின்றனர் . இதன் மூலம் அடுத்து வரக்கூடிய மொழிபெயர்ப்பாளர்கள் ஆரம்பக்கட்டத்தில் மொழிபெயர்த்தவர்களின் அளவிற்குக் கஷ்டப்படாமல் இயங்குதளங்களில் இருக்கும் மாதிரிகளைக் கொண்டு மொழிபெயர்ப்புப் பணிகளை மேற்கொள்கின்றனர் . இதனால் ஏற்கனவே இயங்குதளங்களில் தங்கள் பணிகளைச் சேமித்து வைத்தவர்களுக்குப் பலன் ஏதும் இல்லை என்றாலும் , தங்கள் மொழிக்கும் அதன் தர வளர்ச்சிக்கும் செய்யும் தொண்டாக இவர்கள் அதை எண்ணுகின்றனர் . இம்மாதிரியான நிலை தமிழ்மொழிக்கு இன்னும் ஏற்படாது இருப்பது வருத்தமளிக்கின்றது . இதனை தமிழ் மொழிக்கும் கொண்டு வர அனைவரும் ஒன்றினைந்து செயல்பட வேண்டும்.

ஆக, இங்கு கொடுக்கப்பட்டிருக்கும் யாவும் மொழிபெயர்ப்பாளர்களின் பணியை இலகுவாக்க கண்டுபிடிக்கப்பட்டு மேம்படுத்தப்பட்டவையாகும் . அவை ஒருபோதும் மொழிபெயர்ப்பாளர்களுக்கு மாற்றாக உபயோகப்படுத்தப்பட வேண்டும் என்ற எண்ணத்தில் செய்யப்பட்டவை அன்று என்பதை நன்கு அறிந்துகொள்ள வேண்டும். அது நடைமுறைக்குச் சாத்தியமும் அன்று . இத்தகைய தொழில்நுட்பங்கள் யாவையும் நன்முறையில் பயன்படுத்தி மொழிபெயர்ப்பாளர்கள் தரமான மொழிபெயர்ப்புகளைச் செய்து , பிற மொழிகளில் இருக்கும் சிறப்பான விஷயங்களை தமிழ் மொழிக்கும் , நம் மொழி , இலக்கண இலக்கியங்களின் தொண்மையையும் பெருமையையும் உலகறிந்திட செய்திடவும் வேண்டும்.

கையடக்கக் கணினிகளில் தமிழ்

செல்வ.முரளி

CEO, விசுவல்மீடியா டெக்னாலஜிஸ் விஎம் ரோபாடிக்ஸ்
Murali@visualmediatech.com

கட்டுரைச்சுருக்கம்

உலக அளவில் பல்வேறு வகையான இயங்குதளங்கள் கையடக்கக் கணினி பயன்பாட்டிற்கு வெளியிடப்பட்டு வருகின்றன. அவற்றுள் குறிப்பிடத்தக்கது ஆன்டிராய்டு இயங்குதளம். ஆன்டிராய்டு கட்டற்ற மென்பொருள் என்ற வகையின் கீழ் வெளியிடப்பட்டுள்ளது. ஆன்டிராய்டு ஐ பயன்படுத்தி பல்வேறு வகையான வன்பொருட்கள் வந்து கொண்டிருக்கின்றன. இருப்பினும், அதில் பிராந்திய மொழிகளுக்கான ஒத்திசைவு வசதி இல்லை. ஆன்டிராய்டு மட்டுமின்றி, அனைத்து இயங்குதளங்களிலும் தமிழைப் புகுத்துவதற்கு என்ன செய்யலாம்? என்பது பற்றிய யோசனைகளைச் சொல்வது கட்டுரையின் நோக்கம்.

கையடக்க கணினிகளில் தமிழ்

இன்றைய நவீன உலகில் ஆன்டிராய்டு, விண்டோஸ், ஐபோன் மற்றும் விரைவில் வரவிருக்கின்ற மீகோ (Meego) போன்ற இயங்குதளங்கள் பயன்பாட்டிற்கு வந்து கொண்டிருக்கின்றன. இப்படி உலகம் முழுக்கப் பலவிதமான இயங்குதளங்கள் வந்தாலும், அவற்றில் நம் தாய்மொழியான தமிழைப் பயன்படுத்துவதில் பெரும் சிரமங்களைச் சந்திக்க வேண்டியுள்ளது.

தற்போது ஆன்டிராய்டு இயங்குதளம் கைத்தொலைபேசிகளில் பயன்படுத்தப்படுகிறது. மிகக் குறைவான விலையில் ஆன்டிராய்டு வகை சாதனங்கள் கிடைப்பதால் எல்லாத் தரப்பினரும் அதைப் பயன்படுத்துவதில் ஆர்வம் காட்டுகின்றனர். ஆனால் தமிழ் வசதி இன்னும் முழுமையாகச் சேர்க்கப்படவில்லை.

மைக்ரோசாப்ட் விண்டோஸ் அலைபேசிகளிலும் இதே பிரச்சினை உண்டு. ஐபோனில் தற்போது தமிழில் படிக்கும் வசதி உள்ளது. எதிர்காலத்தில் வரவிருக்கும் மீகோ (MeeGO), Ubuntu போன்ற இயங்குதளத்தில் தமிழ் சாத்தியமா? என்பது பற்றிய தெளிவு தற்போது இல்லை.

சிக்கல்கள்

உலகில் பயன்பாட்டளவில் குறைவாக உபயோகப்படுத்தப்படும் மொழிகளுக்கு அம்மொழி சார்ந்த வசதியை இயங்குதளம் உருவாக்கும் நிறுவனங்கள் அளிப்பதில்லை. மிக அதிகமாக பயன்படுத்தப்படும் மொழிகளுக்கு மட்டுமே அந்நிறுவனங்கள் மொழி வசதியை அளிக்கின்றனர். ஆன்டிராய்டு பயன்படுத்தும் மின்னணுச் சாதனங்களும் அலைபேசிகளும் இதற்கு விதி விலக்கல்ல.

தற்போது உள்ள இயங்கு தளங்களில் ஆன்டிராய்டு Droidsan அடிப்படையிலான எழுத்துருக்கள் பயன்படுத்தப்படுகின்றன. அதில் ஆன்டிராய்டு அளிக்கும் எழுத்துருக்களில் யுனிகோடு அடிப்படையிலான தமிழ் எழுத்துகளின் இடங்களில் எந்த எழுத்துருவும் இல்லாத காரணத்தால் நம் தமிழ் எழுத்துக்களை முழுமையாகப் பார்க்க முடிவதில்லை. (படம் கட்டுரையின் கடைசியில்)

இதே நிலைதான் மற்ற விண்டோஸ் மற்றும் மீகோ இயங்கு தளங்களுக்கும். எழுத்துருக்களை நிறுவ ரூட்டிங் எனப்படும் மேலாண்மை அனுமதி தேவை. ஆனால் ஆன்டிராய்டு இயங்கு தளம் பயனாளர்களுக்கு அனுமதியை வழங்குவதில்லை.

எல்ஜி நிறுவனம் மற்றும் எச்சிசி மற்றும் சாம்சங் ஆகிய நிறுவனங்கள் தமிழில் ஒத்திசைவை வழங்கின. ஆனால் பிற நிறுவனங்கள் தமிழ் ஒத்திசைவை வழங்கவில்லை. நீண்ட தேடலுக்கு பின் நேரடியாக கூகிள் நிறுவனத்தின் மென்பொருள் பொறியாளர் எனக்கு அனுப்பிய பதில் உங்கள் பார்வைக்கு

I am not sure what version of Android you are using, but languages requiring complex character rendering (Indic, Arabic, Thai, and Hebrew) are supported in ICS (Ice Cream Sandwich, Android 4.0) in WebViews, but are not supported in TextViews and other parts of the System UI. The Android JB (Jelly Bean, Android 4.1) release added complex character support for TextViews, which will work with the Tamil and Devanagari fonts that are included in the release. Full support for adding additional complex character fonts may be supported in a future release.

Ying Wang (王映)

Software Engineer, Google Inc.

ஆன்டிர்யாடு முதல் பதிப்பு முதற்கொண்டு ஆன்டிர்யாடு பதிப்பு 4 வரை ஆன்டிர்யாடு இயங்குதளங்களும் தன்னியல்பாக தமிழுக்கு ஒத்திசைவு வழங்காது என்று அறியலாம். அப்படியே ஒத்திசைவு வழங்கினாலும் வெப் வியூவிற்கு மட்டுமே ஒத்திசைவு கிடைக்கும்.

இதன்அடிப்படையில் எங்கள் நிறுவனம் ஆன்டிர்யாடு இயங்குதளம் அடிப்படையிலான (www.CPADINDIA) டேப்ளேட் கணினிகளை 1.1.2012 முதல் சிபேடு CPAD என்ற பெயரில் வெளியிட்டு வருகிறது

அதில் எங்கள் நிறுவனம் முழுமையாக தமிழ் படிக்கும் வசதியினை அறிமுகப்படுத்தியுள்ளது. அதோடு தமிழ் தெரியாத மற்ற ஆன்டிர்யாடு அடிப்படையிலான அலைபேசிகளிலும், கையடக்க கணினிகளிலும் தமிழ் ஒத்திசைவை எப்படி கொண்டுவருது என்ற வழிகாட்டியையும் எங்கள் நிறுவனம் வெளியிட்டுள்ளது. இதனால் எல்லா தரப்பினரும் இந்த வசதியினை பயன்படுத்தி அவர்கள் கையடக்க கணினிகளிலும் தமிழ் கொண்டு வர முடிந்தது.

ஆன்டிர்யாடு ல் தமிழ் ஒத்திசைவினை எப்படி கொண்டு வருவது?

ஆன்டிர்யாடு, விண்டோஸ் அடிப்படையிலான கணிப்பலகைகள் (tablet) மற்றும் அலைபேசிகளுக்கு தமிழ் ஒத்திசைவை தரவில்லை. ஆகையால் ஒபேரா வை நிறுவி அதன் வழியாக மாற்றங்களை செய்து தமிழ் செய்திகளை படித்திருந்தோம்.

ஆன்டிர்யாடு 1.6 -ல் தமிழை பயன்படுத்தி தட்டச்சு செய்திட தமிழ்விசை தமிழா குழுவினரால் கொண்டுவரப்பட்டு பின் அனைவரும் பயன்படுத்த ஏதுவாக இருந்தது.

எனவே நிச்சயமாக இது எழுத்துரு சம்பந்தமான பிரச்சனையாகத்தான் இருக்கும் என்று என்று எண்ணி தேடியபோது நிச்சயமாக இது எழுத்துரு பிரச்சனையாகவே இருந்தது.

என்னவெனில் ஆன்டிர்யாடு தன்னுடன் வழங்கும் எழுத்துருக்களில் தமிழ் உட்பட பல மொழிகளுக்கான எழுத்துருக்கள் சேர்க்கப்படவில்லை . எனவே அதன் தன்னியல்பான எழுத்துருவான DroidSansFallback எழுத்துருவில் யுனிகோடு அடிப்படையிலான தமிழ் இடங்களில் தமிழ் எழுத்துக்கள் இல்லை . எனவே பாண்ட் கிரியேட்டர் வழியாக தமிழ் எழுத்துருவை சேர்க்கலாம் . ஆனால் வேறு ஒரு குழுவில் வேறு மொழி பிரச்சனைக்கு மைக்ரோசாப்ட்-ன் லதா எழுத்துருவை DroidSansFallback என பெயர் மாற்றம் செய்து system\font கோப்பில் இட்டால் எல்லா மொழியையும் காணலாம் என்று கூறியிருந்தார்கள் .

கூடவே தமிழ் விசை வழியாக தமிழும் சிறப்பாக தட்டச்சுவும் முடிந்தது . ஆனால் சில இடங்களில் முழுமையாக ரெண்டரிங் ஆகவில்லை என்றாலும் தட்டச்சு செய்து பதிந்த பின் மிகச்சரியாக தமிழ் வந்தது .

எனவே ஆன்டிர்யாடு எல்லா பதிப்புகளுக்கும் கீழ்க்க ண்ட எழுத்துருக்களை நிறுவி தமிழ் உட்பட எல்லா மொழியினையும் கொண்டு வரலாம் .

மேலும் திரு.தகடூர் கோபி அவர்கள் பல்வேறு விதமான எழுத்துருக்களை யுனிகோடு வடிவில் வெளியிட்டுள்ளார் . எனவே எழுத்துரு மாற்றம் வேண்டுபவர்கள் அந்த எழுத்துருக்களை தரவிறக்கம் செய்து பெயர் மாற்றி ஆன்டிர்யாடுல் பயன்படுத்தலாம் வித விதமான எழுத்துருக்களுடன் :)

எச்சரிக்கை சில மொபைல்களில் ரூட்டிங் வசதியை அனுமதிப்பதில்லை . இல்லையேல் உங்கள் மொபைலின் வாரண்டி போய்விடும். எனவே கவனமாக பயன்படுத்தவும்

ஆன்டிராய்டில் தமிழ் எழுத்துருக்கள் தரவிற்றக்க

<http://dl.dropbox.com/u/8548231/droidfonts.zip>

- ரூட்டிங்கிற்கு பயன்படுத்திய மென்பொருள்
- Super User/BusyBox - ரூட்டிங்கு தேவையான மென்பொருள்
- ரூட் பிரசவுசர் - பான்ட் ஐ நிறுவ தேவையான மென்பொருள்

https://play.google.com/store/apps/details?id=com.jrummy.root.browserfree&feature=search_result#?t=W251bGwsMSwxLDEslmNvbS5qc2VtbXkucm9vdC5icm93c2VyZnJlZSJd

நமக்கு பிடித்தமான எழுத்துருக்களை சேர்க்க

http://www.higopi.com/index.php?option=com_jdownloads&Itemid=4&view=viewcategory&catid=11

ஏற்கனே TAM அடிப்படையில் உள்ள எழுத்துருக்களை யுனிகோடுக்கு மாற்றி ஆன்டிராய்டு உட்பட எல்லா மொழிகளிலும் கொண்டு வர வழிமுறைகள் பற்றி திரு.தகடூர் கோபி (higopi.com) அவர்கள் எழுதியுள்ள கட்டுரை

<http://higopi.blogspot.in/2008/08/1.html>

அதன்பின் நாங்கள் தமிழ் யுனிகோடு எழுத்துருக்களை கொண்ட எழுத்துருக்களை ஆன்டிராய்டில் நிறுவியபின் தமிழில் படிப்பதற்கு வசதியாக இருந்தது.

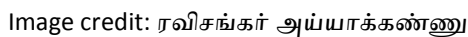
வரும்நாட்களில் வரக்கூடிய மீகோ, உபுண்டு, வெப்ஓஎஸ் என்று நிறைய இயங்குதளங்கள் வந்தாலும் அவை எல்லாமே லினக்ஸ் ன் அடிப்படையில் வரும் என்பதால் ஆன்டிராய்டுக்கு பயன்படுத்தக்கூடிய அதே முறைதான் எல்லாவற்றிலும் வரும் என்பதால் மேற்கண்ட இதே முறையினை பயன்படுத்தி எல்லாவற்றிலும் தமிழினைக்கொண்டுவரலாம்

தீர்வு

எந்த வகையான இயங்கு தளங்கள் வந்தாலும் அவற்றிற்கு என்று ஒரு பொதுவான எழுத்துருக்கள் உருவாக்கி அதை மேலாண்மை அனுமதி இல்லாமல் நிறுவிக் கொள்ள வசதி செய்யவேண்டும்.

1. எல்லா இயங்கு தளங்களுக்கும் பொதுவான மொழியான ஜாவா அல்லது QT யில் தமிழ் தட்டச்சு செய்யும் மென்பொருளை உருவாக்கி புதிய இயங்கு தளங்களுக்குத் தேவையான தட்டச்சு மென்பொருட்களையும் உருவாக்கிடவேண்டும். (உதாரணம்) தமிழா குழுவினரால் உருவாக்கப்பட்ட தமிழ் விசை
2. தமிழா குழுவினரால் ஏற்கனவே உருவாக்கி பயன்பாட்டில் உள்ள தமிழா தமிழ்விசை மென்பொருளை அரசாங்கமே எடுத்து அதை எல்லா அலைபேசிகளுக்கும் ஏற்றவாறு மாற்றித்தந்து யாவரும் பயன்படுத்திட கட்டற்ற மென்பொருளாக வெளியிட வேண்டும்.
3. ஆனால் டிஸ்கி (TSCII) அடிப்படையிலான எழுத்துருக்கள் ஆன்டிராய்டில் மிகச்சரியாக தோன்றுகின்றன . தமிழில் மென்பொருட்கள் உருவாக்கும்போது ஆன்டிராய்டு முதல் பதிப்பு முதல் 4ம் பதிப்பு வரை டிஸ்கி (TSCII) எழுத்துருக்களை பயன்படுத்தலாம்
4. எல்லா அலைபேசிகளிலும் தமிழ் தெரிய: ஓபேரா நிறுவியபின்னர் அதன் முகவரி தட்டச்சு செய்யவேண்டிய இடத்தில் about:config என்று கொடுத்து உள்ளே சென்றால் வரும் பட்டியலில் Use bitmap fonts

படங்கள்: ஆன்டிராய்டு தமிழ் எழுத்துரு பிரச்சனை பற்றிகள் பற்றிய விபரம்



- <http://thamizha.com/>
- www.android.com
- www.higopi.com
- Hard Core Hacker, Europe

தமிழில் எழுத்துருக்களின் உடற்கூறியலும் வனப்பெழுத்தும்

A.M.Ponraj

Associate Professor, Department of Computer Science, Madurai Kamaraj University
ponrajcomputer@gmail.com

Extended Abstract

வனப்பெழுத்து (Calligraphy) என்பது எழுத்துகள் அல்லது ஒவியங்களை மிக அழகாக வரையும் அல்லது எழுதும் ஒரு கலையாகும். இந்த கலையை உபயோகித்து ஒவ்வொரு நாளும் தங்களுக்கு என்று ஒரு தனித்த எழுத்துருக்களை உருவாக்கி கொண்டுள்ளனர். அவர் அவர் தாம் மொழிக்கு மட்டுமின்றி, இணைப்பு மொழியான ஆங்கில எழுத்துருக்களையும் தங்கள் மரபுக்கும் மொழிக்கும் எற்றவாறு வடிவமைத்துக் கொண்டுள்ளனர். அனால் இந்தியாவிற்கென்று இந்த கலையை கொண்டு ஒவியங்கள் மட்டுமே வரையப்பட்டுள்ளன.

இக்கலையை கொண்டு நமக்கென்று எழுத்துக்கள் உருவாக்கப்படவில்லை குறிப்பாக தமிழ் எழுத்துருக்களுக்கு இதுவரை இதுபோன்ற வரைகலை பயன்படுத்தப்படவில்லை.

இந்த கட்டுரையில், இதற்கான முயற்சி மேற்கொள்ளப்பட்டு தமிழ் எழுத்துருக்கள் வடிவமைக்கப்பட்டுள்ளது. இந்த எழுத்துருக்கள் ஆங்கில எழுத்துருவான (Old English) எழுத்துருவை அடிப்படையாக கொண்டு வடிவமைக்கப்பட்டுள்ளன.

எழுத்துருக்களின் உடற்கூறியல்

மட்டக்கோடு (arm): கிடைமட்ட அல்லது சாய்வு கோட்டில் ஒரு புறம் மேல் அல்லது கீழ் வளைந்தும் மறு புறம் ஒய்வாகவும் இருக்கும்.

எறுமுகஉயரம் (Ascender): ஒரு எழுத்தின் செங்குத்துக் கோட்டின் மேற்பகுதியின் உயரத்தைக் குறிக்கும்.

இறங்குமுகஉயரம் (Descender): ஒரு எழுத்தின் செங்குத்துக் கோட்டின் அல்லது சாய்வு கோட்டின் கீழ்பகுதியின் உயரத்தைக் குறிக்கும்.

இடைப்பட்டை (Bar): ஒரு எழுத்தின் நடுப்பகுதியின் கிடைமட்டக்கோட்டை குறிக்கும்.

சருவம் (Bowl): ஒரு எழுத்தில் உள்ள ஒரு வளைந்த கோடு உள்ளடக்கப்பட்ட இடைவெளியை உருவாக்குவது சருவம் எனப்படும்.

சருவமுகப்பு (Counter): ஒரு எழுத்தின் குறிப்பிட்ட எல்லைக்குள் பகுதியாக அல்லது முழுமையாக உள்ளடக்கப்பட்ட இடத்தை குறிக்கும்.

குத்துயரம் (Capheight): ஒரு எழுத்தின் செங்குத்துக் கோட்டின் உயரம் குத்துயரம் எனப்படும்.

உருச்செவி (Ear): ஒரு எழுத்தின் மேற்பகுதியில் உள்ள சிறிய கோடு அல்லது வளைந்த கோடு உருச்செவி எனப்படும்.

அகஇணைப்பு (Link): ஒரு எழுத்தின் மேற்பகுதியையும் கீழ்பகுதியையும் இணைக்கும் ஓர் கோடு அல்லது ஓர் வளைவு அகஇணைப்பு எனப்படும்.

உருவடி (Serif): உட்படுத்தப்படும் எழுத்துக்களின் மேற்பகுதியில் நீட்டிக்கப்பட்ட நேர் அல்லது வளைகோடு உருவடி எனப்படும்.

உருமுடி (Serif): உட்படுத்தப்படும் எழுத்துக்களின் கீழ்பகுதியில் நீட்டிக்கப்பட்ட நேர் அல்லது வளைகோடு உருவடி எனப்படும்.

உருதிமிழ் (Shoulder): ஒரு எழுத்தின் வளைகோடு உருதிமிழ் எனப்படும்.

வளைத்தண்டு (Spine): ஒரு எழுத்தின் முக்கிய வளைகோடு வளைத்தண்டு எனப்படும்.

சரிவடிப்பு (Stroke): நேர் அல்லது வளைகோடு சரிவடிப்பு எனப்படும்.

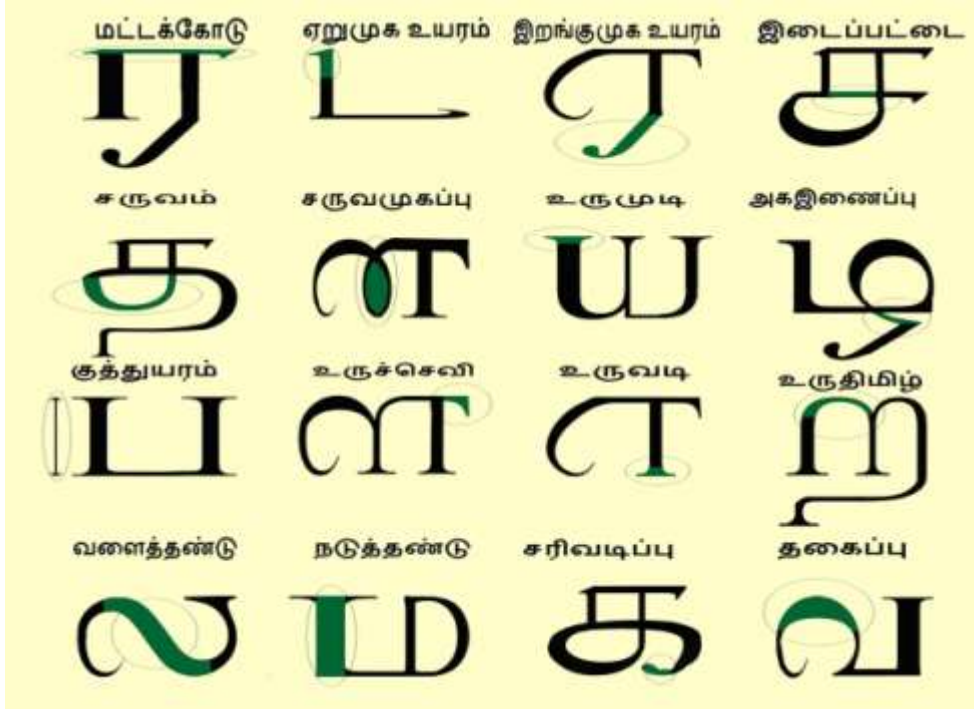
நடுத்தண்டு (Stem): ஒரு எழுத்தின் செங்குத்துக்கோட்டின் உயரம் நடுத்தண்டு எனப்படும்.

தகைப்பு (Stress): ஒரு எழுத்தின் விரிவடைந்த வளைகோடு தகைப்பு எனப்படும்.

Conclusion

The standard calligraphy fonts are used in some other countries. But, In India there is only calligraphy arts are available. Now we are designing the new calligraphy fonts in Tamil. In this project, we are designing calligraphy tamil fonts based on Old English font style.

Appendix A: anatomy of tamil fonts



Appendix B: Rules for calligraphy in tamil, Sample output



தேசிய மொழிகள் ஊடாக தகவல் தொழில்நுட்பத்தை வலுப்படுத்தல் தொடர்பான பங்களிப்பு

ஜெயராஜ்ஜன் யோகராஜ்

ICTA நிறுவனத்தின் மொழி அபிருத்திக் குழுவின் சிரேஷ்ட உறுப்பினர், ICTA நிறுவனத்தின் தமிழ் எழுத்துரு
மீளாய்வுக் குழுவின் சிரேஷ்ட உறுப்பினர், இலங்கை
yogaraj35@gmail.com

1. கணனி செயல்பாடுகளை உள்ளூர் மொழிகளில் மேற்கொள்வதற்கு பல நடவடிக்கைகள் எடுக்கப்பட்டுள்ளன . அந்த வகையில் , திறந்த மூலங்களான உபுண்டு (Ubuntu OS), பையர் பொக்ஸ் (Fire Fox), தண்டர் பேர்ட் (Thunderbird), ஓபன் ஒபிஸ் (Open Office) உட்பட பல முக்கிய மென்பொருட்களை தமிழில் பயன்படுத்த நடவடிக்கைகள் மேற்கொள்ளப்பட்டுள்ளன .
2. இந்த மென்பொருட்கள் தமிழில் மொழிபெயர்க்கப்பட்டது மட்டுமல்ல , அவற்றை மொழிபெயர்க்கும் செயல்பாடுகளில் கணனி துறை நிபுணர்கள் மற்றும் தமிழ் அறிஞர்களும் உள்வாங்கப்பட்டு உள்ளனர். அந்த வகையில், இந்த மொழிபெயர்ப்பாளர் சங்கம் ஒரு சமூகமாக விரிவடைவதற்கு ஆவண செய்யப்பட்டுள்ளது . இங்கு திறந்த மூல தயாரிப்புக்களுக்கான மொழிபெயர்ப்பு மிகவும் அவதானமாக மேற்கொள்ளப்பட்டது . அதாவது மொழிபெயர்ப்பு சொற்களுக்கான மீளாய்வு திட்டமொன்றும் அமுல்படுத்தப் பட்டுள்ளது. அதன் மூலம் குறிப்பிட்ட ஆங்கில சொல்லுக்கான திருத்தமான தமிழ் வடிவம் பெறப்பட்டது . இது தொடர்பான மேலதிக விபரங்கள் இணைப்பு ஆவணத்தில் உள்ளன .
3. மேலும் இந்த மென்பொருட்களை பயன்படுத்துவதற்கு தேவையான அனைத்து ஊக்குவிப்பு நடவடிக்கைகளும் மேற்கொள்ளப்பட்டுள்ளன. இலங்கையில் தேசிய இலத்திரனியல் ஊடகங்கள் , அச்சு ஊடகங்கள் , ஒலி-ஒளி ஊடகங்கள் உட்பட தனியார் ஊடகங்களும் இந்த பயன்பாட்டை ஊக்குவிக்க ஆக்கபூர்வமான நடவடிக்கைகளை மேற்கொண்டுள்ளது.
4. இதன் ஊடாக சிறு பிள்ளை பராயம் தொடக்கம் பெரியோர் வரை அனைத்துத் தரப்பினரும் திறந்த மூலங்களுக்கான தமிழ் வடிவத்தை நன்கு புரிந்துகொள்ள வாய்ப்பு ஏற்பட்டுள்ளது . இதன் மூலம் திறந்த மூல பயன்பாடும் கணிசமான அளவு அதிகரித்துள்ளது . திறந்த மூல தயாரிப்புக்கள் மூலம் கணனி பாவனையாளர்களுக்கு பல நன்மைகள் கிட்டுகின்றன . இந்த மென்பொருட்களை பயன்படுத்துவதற்கான அனுமதி பத்திரங்கள் இலவசம் . மேலும் மைக்ரோசொப்ட் தயாரிப்புக்களுக்கு நிகரான அனைத்து தயாரிப்புக்களையும் திறந்த மூலங்கள் கொண்டிருக்கின்றன .

Microsoft Products Open Source Products (equivalent)

Internet Explorer Fire Fox
Outlook Express Thunderbird
MS Office Open Office

மேற்குறிப்பிட்ட சில தயாரிப்புக்கள் இவற்றிற்கு சிறந்த உதாரணங்களாகும் . எனவே, திறந்த மூல தயாரிப்புக்களில் உள்ள மற்றுமொரு நன்மை என்னவெனில் , இவை வைரஸ் தாக்கங்களுக்கு உட்படாதவையாகும் . இவ்வாறு திறந்த மூல தயாரிப்பு பாவனையாளர்களுக்கு பல நன்மைகள் உண்டு . இந்த நன்மைகளை அறிவுறுத்தும் வகையிலும், பல செயலமர்வுகள் மேற்கொள்ளப்பட்டுள்ளன . ஊடகங்களும் இவற்றிற்கு கணிசமான பங்களிப்பு செய்துள்ளன .

5. இதன் அடிப்படையில் , திறந்த மூல தயாரிப்புக்களுக்கான மொழிபெயர்ப்பாளர்கள் மற்றும் பாவனையாளர் எண்ணிக்கையில், அண்மைக்காலத்தில் அதிகரிப்பு காணப்பட்டுள்ளது . இந்த செயல்பாடு ஒரு வரையறுக்கப்பட்ட செயல்பாடு அல்ல . மாறாக நீண்டகால செயல் திட்டமாக தொடர்ச்சியாக மேற்கொள்ளப்படுகிறது . எனவே, நீண்டகால நோக்கம் கருதி , மேற்கொள்ளப்பட்ட இந்த செயல்திட்டத்தின் மூலம் , எதிர்காலத்தில் இலங்கை வாழ்

தமிழ் பேசும் மக்கள் திறந்த மூல பயன்பாட்டை மேலும் விருத்தி செய்துகொள்வது என எதிர்பார்க்கப்படுகிறது . இது தேசிய ரீதியில் இலங்கைக்கு பல்வேறு அனுகூலங்களை பெற்றுத்தரும் . வருடாந்தம் வெளிநாடுகளில் இருந்து மென்பொருள் அனுமதிப் பத்திரங்களுக்காக கோடிக்கணக்கான வெளிநாட்டு செலாவணி பயன்படுத்தப்படுகிறது. திறந்த மூல பயன்பாட்டின் மூலம் இந்த வெளிநாட்டு நாணய செலவினம் கணிசமான அளவு குறைவடையும் என எதிர்பார்க்கப்படுகிறது .

*

மேலும் ICTAநிலையத்தினால் நியமிக்கப்பட்ட உள்ளூர் மொழி செயல்பாட்டுக் குழுவில் உள்ளூர் மொழி அபிவிருத்தி தொடர்பில் பல்வேறு பங்களிப்புக்களைச் திரு .ஜெயரஞ்ஜன் யோகராஜ் வழங்கியுள்ளார் . அந்த வகையில் **Ubuntu** இயங்குதளம், **firefox**, **Thunderbird**, **Open Office** உட்பட இதர திறந்த மூல மென்பொருட்களை மொழிப்பெயர்ப்பு செய்தமைக்கு திரு .ஜெயரஞ்ஜன் யோகராஜ் விசேட விருது வழங்கப்பட்டது . **Ubuntu** இயங்குதள மொழிப் பெயர்ப்புக்கு அப்பால் , “**Ubuntu Sri Lanka Tamil Translators**” (இணையத்தள முகவரி : <https://launchpad.net/~ubuntu-l10n-ta-lk>) எனும் சர்வதேச அங்கீகாரம் பெற்ற மொழிப்பெயர்ப்பாளர் சங்கத்தை உருவாக்கிய பெருமையும் இவரை சாரும் .

இதுவரை காலமும் கணினி துறைசார் மென்பொருட்களை தமிழில் மொழிப் பெயர்க்கும் பொழுது அதற்கான கூடுதல் செலவாக்கையும் ஆதிக்கத்தையும் இந்திய தமிழ் மொழிப்பெயர்ப்புச் சங்கங் களே கொண்டிருந்தன . அந்த வகையில் **Ubuntu** இயங்குதளத்திற்கான தமிழ் மொழிப் பெயர்ப்புச் செயற்பாடுகளையும் இந்திய கணினி தமிழ் குழுவொன்று மேற்கொண்டது . ஆனால், அந்த மொழிப் பெயர்ப்புக்கள் முழுமையாக பூர்த்தியடையவில்லை . மேலும் அவை இந்திய தமிழ் மொழி வழக்கில் காணப்பட்டன . இந்திய தமிழ் மொழி நடைக்கும் இலங்கை தமிழ் மொழி நடைக்கும் இடையில் சொற்கள் , வாக்கிய அமைப்புக்கள் , ஒத்தக கருத்துச் சொற்கள் உட்பட பல விடயங்களில் வேறுபாடுகள் காணப்படுகின்றன . எனவே, இவற்றை கருத்திற் கொண்டு இலங்கை தமிழ் எனும் இலங்கைக்கே உரித்தான தமிழ் மொழி வழக்கில் மொழிப் பெயர்ப்பதற்கு முயற்சிகளை திரு .ஜெயரஞ்ஜன் யோகராஜ் மேற்கொண்டார் .

இதற்கு ஏற்கனவே இந்தியாவில் உள்ள சர்வதேச அங்கீகாரம் பெற்ற **Ubuntu** மொழிப் பெயர்ப்பாளர்கள் எதிர்ப்பு தெரிவித்ததால் இதனை தீர்த்துக் கொள்ள **Ubuntu** இயங்குதளத்தின் மொழிப் பெயர்ப்பு தொடர்பான சர்வதேச அங்கீகாரம் மற்றும் அதிகாரங்களைப் பெற்ற திரு .டேவிட் பெனலாவை (**David Panela**) தொடர்பு கொள்ளவேண்டிய நிலை ஏற்பட்டது . தமிழ் மொழிப் பெயர்ப்புக்கு **TAMIL (IN)** எனும் இந்தியக் குழு இருப்பதனால் அந்தக் குழுவோடு இணைந்து செயல்படும்படி இலங்கைக் குழுவுக்கு வேண்டுகோள் விடுக்கப்பட்டது. ஆங்கிலம் ஒரு மொழியாக இருந்தாலும் **English (UK)**, **English (US)** என்று இரு வேறு நாடுகளுக்கு ஆங்கில மொழிகள் இருக்கும் பொழுது, தமிழுக்கு ஏன் இருக்கக்கூடாது என்ற கேள்வி இலங்கைக் குழுவிடமிருந்து எழுப்பப்பட்டது . மேலும் சில சொற்கள் மற்றும் வசனங்களும் இங்கே ஆதாரமாக காட்டப்பட்டன .

இதன் அடிப்படையில் திரு .ஜெயரஞ்ஜன் யோகராஜ் மேற்கொண்ட முயற்சியின் பலனாக **Ubuntu**, **firefox**, **Thunderbird**, **Open Office** உட்பட எதிர்காலத்தில் வெளிவரும் அனைத்து ஆங்கில மொழியில் அமைந்த மென்பொருட்களுக்கான தமிழ் மொழிப் பெயர்ப்புக்கள் கணினி இயந்திரங்களில் **TAMIL (IN)**, **TAMIL (SL)** என்ற பிரிவுகளில் வெளியாகும் . இதில் **TAMIL (SL)** என்பது இலங்கைக்கே உரித்தான தமிழ் மொழிப் பெயர்ப்பாகும் . இதற்கான பூரண புலமைச் சொத்துரிமை திரு .ஜெயரஞ்ஜன் யோகராஜ் உரித்தானது . இவரது தலைமையில் **Ubuntu Sri Lanka Tamil Translators** ' எனும் குழு உரிய மொழிப் பெயர்ப்புக்களை மேற்கொண்டு வருகிறது . இந்த பாரிய பங்களிப்புக்காகவும், தமிழ் விசைப்பலகை ; உருவாக்கம் மற்றும் எழுத்து ரு உருவாக்கம் தொடர்பாக ஆலோசனை தர பங்களிப்புக்களை ICTA நிலையத்தின் உள்ளூர் மொழி செயற்பாட்டுக் குழுவின் சிரேஷ்ட அங்கத்தவராக இருந்து, மேற்கொண்டமைக்காகவும் இந்த விருது, அவருக்கு வழங்கப்பட்டது .

இலங்கையில் தகவல்தொழில்நுட்பம், கணினி, மென்பொருள், கணினி-தகவல்தொழில்நுட்ப சட்டவாக்கம் உட்பட அனைத்து தகவல்தொழில்நுட்பம்சார் விடயங்களையும் இலங்கை அரசாங்கத்தின் சார்பில் உருவாக்கி , அமுல்படுத்தும் ஒரே ஒரு அமைப்பே இந்த தகவல்தொழில்நுட்ப முகவராண்மை நிலையம் (Information Communication Technology Agency of Sri Lanka – ICTA) என்பது குறிப்பிடத்தக்கது. இந்த நிலையம் ஜனாதிபதி செயலகத்தின் நேரடி மேற்பார்வையில் இயங்கும் நிறுவனமாகும் .

தேசிய மொழிகளை கணினியில் பயன்படுத்துவது குறித்து இலங்கையில் கடந்த ஏழு வருடங்களாக பல ஆக்கப்பூர்வ ஆய்வுகள் மேற்கொள்ளப்பட்டன . தேசிய மொழிகள் அல்லது உள்ளூர் மொழிகளாக சிங்களமும், தமிழும் காணப்படுகின்றன . இந்த மொழிகளை கணினி விசைப் பலகையின் ஊடாக பயன்படுத்தல் , அதற்கான விசைப் பலகை வரைபட உருவாக்கம் , இலங்கை கட்டளைகள் நிலையத்தின் விசைப் பலகை சட்ட விதிமுறைகள் , இலங்கைக்கே உரித்தான நியம எழுத்துருக்கள் , பிரபல மென்பொருட்கள் மற்றும் இயங்குதளங்களுக்கான சிங்கள மற்றும் தமிழ் மொழிப் பெயர்ப்புக்கள் உட்பட பல்வேறு கணினித்துறை சார்ந்த விடயங்கள் ஆராயப்பட்டன .

இந்த விடயங்களை ஆராய்ந்து அவற்றுக்கு உரிய தீர்வுகளை வழங்குவென , இலங்கை தகவல் தொடர்பாடல் முகவராண்மை நிலையம் உள்ளூர் மொழி செயல்பாட்டுக் குழுவொன்றை ஏழு வருடங்களுக்கு முன்னர் உருவாக்கியது . அந்தக் குழு கடந்த ஏழு வருடங்களில் பல்வேறு ஆய்வுகளை மேற்கொண்டு தீர்வுகளையும் முன்வைத்துள்ளது . இதன் அடிப்படையில் உள்ளூர் மொழிகளை கணினியில் சிறந்த முறையில் பயன்படுத்துவதற்கென சில தயாரிப்புக்களும் உருவாக்கப்பட்டுள்ளன . இந்த முயற்சிகளுக்கு சில அரசு தரப்பு அமைப்புக்களும் உதவி செய்தன .

தகவல் தொழிநுட்ப முகவராண்மை நிலைய தமிழ் நியம எழுத்துரு - ஸ்ரீ தமிழ் என்ற பெயரில் இந்த குழுவும் வெளியிடப்பட்டது . ஸ்ரீ தமிழ் எழுத்துரு உருவாக்கல் தொடர்பான மீளாய்வு குழு சிரேஸ்ட அங்கத்தவராகவும் திரு.ஜெயரஞ்சன் யோகராஜ் பணியாற்றியுள்ளார் .

Ubuntu தமிழாக்கத்தினால் இலங்கை வாழ் தமிழ் சமூகம் பெற்றுக்கொள்ளும் நன்மைகள் :

1. தமிழில் கணினி செயல்பாடுகளை சிறந்த முறையில் கற்றுக்கொள்வதற்கான வாய்ப்பு மேலும் அதிகரித்துள்ளது .
2. அதுவும் பிரபல திறந்த மூல தயாரிப்புக்களான Ubuntu OS, Firefox, Thunderbird, Open Office போன்ற மென்பொருட்களை இலகு தமிழில் புரிந்துக்கொள்ளக் கூடிய வாய்ப்பு ஏற்பட்டுள்ளது .
3. Microsoft தயாரிப்புக்களுக்கு இணையான பல தயாரிப்புக்கள் திறந்த மூலங்களில் வெளிவந்துள்ளன .செலவு குறைந்த அல்லது இலவச தயாரிப்புக் களாகவே திறந்த மூலங்கள் காணப்படுகின்றன .இவற்றை இலங்கைக்கே உரித்தான தமிழில் கற்றுக்கொள்ள கூடிய வாய்ப்பு இலங்கையர்களுக்கு கிடைத்துள்ளது .
4. இந்த தமிழ் மொழியாக்கம் அல்லது மொழிபெயர்ப்பை மேற்கொள்ள 'இலங்கைக்கே உரித்தான சர்வதேச அங்கீகாரம் பெற்ற மொழிபெயர்ப்பாளர் சங்கம்' உருவாக்கப்பட்டுள்ளது .
5. இந்த சங்கத்தின் மூலம் இந்த மொழிபெயர்ப்பு செயல்பாடுகள் தடையின்றி இடம்பெற்று வருகின்றன.தேவைப்படும்பட்சத்தில் புதிய சொற்கள் உரிய ஆதாரங்களுடன் உருவாக்கப்படுகின்றன .
6. துறைசார் விடயங்களுடன் தொடர்புடைய பலர் இந்த சங்கத்தில் இணைந்து கொள்ள சந்தர்ப்பம் உண்டு .
7. இதனுடாக சமூகத்தில் இந்த செயல்பாட்டிற்கான பங்காளர்கள் பலர் உருவாகி வருகின்றனர் .இதனால் சமூக விழிப்புணர்வு ஏற்படுகிறது .
8. இந்த மொழிபெயர்ப்பு செயல்பாடு முடிவில்லா செயல்பாடாக அமையவுள்ளது .எதிர்காலத்தில் இது போன்ற மேலும் பல திறந்த மூலங்களை தமிழாக்கம் செய்யக்கூடிய வாய்ப்பு ஏற்படும் என்று நாம் எதிர்பார்க்கிறேன் .
9. இலங்கை தமிழாக்க பாரம்பரியத்திற்கு அமைய , மொழிபெயர்ப்புக் கொள்கை விதிமுறை ஒன்றும் ஏற்கெனவே உருவாக்கப்பட்டுள்ளது .

https://docs.google.com/document/pub?id=16f-10g_Y5luzFA0YC7W-BPx4E_wulTEJImupDGeHKHI

10. மொழிபெயர்ப்பு விடயத்தில் கூடுதல் கவனம் செலுத்தப்பட்டுள்ளது .ஏற்கெனவே உருவாக்கப்பட்ட கொள்கை விதிமுறை கோவைக்கிணங்க, மொழிபெயர்ப்பு மீளாய்வு திட்டமும் அமுல்படுத்தப்பட்டுள்ளது .

11. இங்கு பல இந்திய சொற்கள் உள்வாங்கப்பட்டாலும் , அமுல்படுத்தல் வழிகாட்டல் ஆவணத்திற்கிணங்க இலங்கை தமிழ்சொற்களும் நடைமுறைகளும் தெளிவாக காட்டப்பட்டுள்ளன .
12. சில சொல் பரிமாற்றங்களையும் உதாரணங்களாக நான் இங்கே தந்துள்ளேன் .

ஆங்கில சொல்	இந்திய நடைமுறை	இலங்கை நடைமுறை (கொள்கை ஆவணத்திற்கிணங்க)
Yahoo!Mail	யாஹூ!மெயில்	Yahoo! மின்னஞ்சல்
Microsoft	மைக்ரோசொவ்ட்	Microsoft
Oracle	ஓராக்கிள்	Oracle
Yahoo	யாஹூ	Yahoo
Google	கூகிள்	Google
Gmail	ஜிமெயில்	Gmail
Firefox	பயர்பொக்ஸ்	Firefox
GIMP	கிம்ப்	GIMP
Debug		பிழை நீக்கு/ வழு நீக்கு
Dialogs		உரையாடல்கள்/ உரையாடல் திரைகள்;
Proxy	புரொக்ஸி	Proxy
Boot Option	துவக்க தெரிவு	இயக்க தெரிவு
Start	துவக்கு	ஆரம்பி
Main Server	மெயின் சர்வர்	முக்கிய சேவையகம்
Ubuntu	உபுண்டு	Ubuntu

எளிதாகத் தமிழ் மொழியில் மின்னூல்கள்

சுகந்தி வெங்கடேஷ்
vknsvn@gmail.com

மின் நூல்கள் என்பது இன்றைய பதிப்பக தொழில் நிறுவனங்களில் ஓர் அன்றாட விஷயமாகி விட்டது .பொழுது போக்குப் புத்தகங்கள் மட்டுமின்றிப் பாடப் புத்தகங்கள் அறிவியல் கட்டுரைகள் பத்திரிக்கைகள் எனப் பலதரப்பட்ட அச்சுப்பிரசுரங்கள் இப்போது மின்வடிவத்தில் உலகின் பெரும்பால மொழிகளில் வலம் வருகின்றன. அதுவும் ஆங்கிலத்தில் அவற்றின் எண்ணிக்கை நாளுக்கு நாள் பெருகி வருகிறது . உரைகளும் படங்களும் எண்ணியல் மயமாக்கப்பட்டு ஒரு புத்தக வடிவில் காட்டப்படும் போது அது மின்னூல் என்று அழைக்கப்படுகிறது . இன்றைய கணினி வளர்ச்சியினால் இத்தகைய நூல்களை மேஜைக் கணினி , மடிக்கணினி கைக் கணினி , சாமர்த்திய அலைபேசி ஆகிய பலவித மின்னணு சாதனங்களைக் கொண்டு படிக்க முடியும் . இதைத் தவிர மின்னூல்களைப் படிப்பதற்கேன்றே நடமாடும் புத்தக வடிவம் கொண்ட மின் வாசிப்பான்களும் (eReaders) இன்று விற்பனையிலும் புழக்கத்திலும் உள்ளன .

தமிழ் மின்னூல் களைப் படிக்க மின் வாசிப்பான்களைப் பயன்படுத்தும் போது மின் புத்தகங்களைப் படிக்க வெவ்வேறு மென்பொருட்களைத் தரவிறக்கம் செய்ய வேண்டியதில்லை. கைக்கணினியிலோ அல்லது சாமர்த்திய அலைபேசியிலோ மின்னூல்களைப் படிக்க அதற்கு உகந்த மென்பொருள் பயன்பாடுகளைப் (software apps) பதிவிறக்கம் செய்து கொள்ள வேண்டும். பொதுவாக் கிடைக்கும் மென்பொருள் பயன்பாடுகளைத் தவிர மின் வாசிப்பான்களுக்கே உரிய மென்பொருள் பயன்பாடுகளைக் கைக்கணினியிலோ, கைபேசியிலா பதிவிறக்கம் செய்தும் நாம் மின்னூல்களைப் படிக்கலாம். இன்று மக்கள் பயன்படுத்தும் கைகணினிகள் மற்றும் சாமர்த்திய அலை பேசிகளில் முக்கியப் பங்கு வகுப்பவை ஆப்பிள் நிறுவனத்தின் இயங்கு தளமும் (iOS6) ஆண்ட்ராய்டு இயங்கு தளமும் இரண்டும் தான் அடுத்தபடியாக மைக்ரோசப்ட்டின் விண்டாஸ் இயங்கு தளம் இருக்கிறது. மின் வாசிப்போன்களை எடுத்துக் கொண்டால் புத்தக விற்பனையாளர் அமேசானின் Kindleயும் ஆப்பிள் நிறுவனத்தின் Ipadயும் போட்டி போடுகின்றன. இது தவிர அமெரிக்காவின் பெரிய புத்தக நிறுவனத்தின் nook என்ற மின் வாசிப்பானும் பெரிய அளவில் உபயோகப் படுத்தப்படுகிறது. Infibeam என்ற நிறுவனம் இந்தியாவில் இந்திய மொழிகளுக்கென மின் வாசிப்பான்களை அறிமுகப்படுத்தியுள்ளது .

தற்போது ஆப்பிள் நிறுவனத்தின் கைகணினிகளிலும், ஆண்ட்ராய்டு கைக்கணினிகளிலும் தமிழில் எளிதாகத் தமிழில் படிக்க முடிக்கிறது. தமிழில் உள்ளிடவும் முடிகிறது என்பது அனைவரும் அறிந்ததே. nook என்ற மின் வாசிப்பானிலும் ஒருங்கீடு குறிமுறையில் எழுதப்பட்ட உரைகளை எளிதில் மின்னூல் வடிவம் கொடுக்கலாம்.அமேசானின் Kindleய் மட்டும் இன்னும் தமிழ் ஒருங்குறி குறியீட்டிற்கு ஆதரவு இல்லை. அதில் தமிழ் மின் னூல்களை ஒரு pdf ஆவணமாக மட்டுமே படிக்க முடிகிறது. ஆப்பிள் நிறுவனத்தின் iBook என்ற மென்பொருள் பயன்பாடும், ஆண்ட்ராய்டு கைக் கணினிகளில் Google Books என்ற மென்பொருள் பயன்பாடும் மின்னூல்களை வாசிக்கப் பயன் படுத்தப்படுகிறது.Aldiko Book என்ற மென்பொருள் பயன்பாடும் ஆண்ட்ராய்டு இயங்கு தளத்தில் நன்றாக வேலை செய்கிறது.விண்டாஸ் கைபேசியில் தமிழ் மின்னூல்களை Google Books மென்பொருளைப் பயன் படுத்திப் படிக்க முடியும்.

தமிழ் மொழியில் மின்னூல்களின் எண்ணிக்கை ஆங்கில மொழியோடு ஒப்பிடும் போது மிக மிகக் குறைவு என்றாலும் நம் மொழியில் மின்னூல்களை அறிமுகப்படுத்துவதன் மூலம் வளர்ச்சி தமிழ் மொழியில் ஒரு புரட்சியே உருவாக்கலாம். பழம் தம்மொழி நூல்களையும் பாடப் புத்தகங்களையும் தமிழ் மின்னூல்களாகக் கொண்டு வரலாம். தமிழில் ஒருங்குறி எழுத்தமைப்பு வருவதற்கு முன்னால் தமிழில் படிப்பதற்குப் பல வகை எழுத்துருக்களைக் கணினியில் பதிவிறக்கம் செய்ய வேண்டியிருந்தது. தமிழில் மின்னூல்கள் என்றால் பெரும்பாலும் pdf ஆவணங்களாக இருந்தது.சில இடங்களில் flash முறையில் படங்களைக் கொண்டு உருவாக்கப்பட்ட புரட்டும் புத்தகங்களாக மின்னூல்கள் இருந்தன. இந்த நூல்களில் வரும் எழுத்துருக்களின் அளவை மாற்றவோ, பக்க வடிவத்தைக் கைக் கணினிகள், சாமர்த்திய கைபேசிகளின் வடிவத்திற்கு ஏற்றவாறு மாற்றவோ முடியாது. ஆனால் ஒருங்குறியில் எழுதப்பட்ட தமிழ் எழுத்துரைகளை மின்னூல்களாக படிப்பதில் எந்த சிரமும் இல்லை.அது மட்டுமின்றிக் கேட்பொலி, காணொளி ஆகியவற்றையும் இந்த மின்னூல்களில் உள்ளிடலாம் .

மின்னூல்களின் வடிவூட்டங்கள் பல வகைப்படும் . மின்னூல்களைத் தயாரிப்பது எளிது என்றாலும் உருவாக்கப்படும் ஆவணங்களின் வடிவூட்டத்தை (format) கவனத்தில் கொள்வது மிகவும் அவசியம். ஒவ்வோரு மின்வாசிப்பானும் தனக்கென ஒரு வடிவூட்டத்தைக் கொண்டுள்ளது .ஆப்பிள் Ipadல் படிக்க உதவும் ஆவணத்தின் வடிவூட்டம் . IBA என்று முடிய வேண்டும். iBooks Authorல் மின்னூல்கள் உருவாக்கும் போது அவை .IBA என்ற வடிவூட்டத்தில் உருவாக்கப் படுகிறது. இந்த மின்னூல்களை ஆப்பிள் நிறுவனத்தின் மூலமே விநியோகிக்க முடியும்.Kindlலுக்கான ஆவண வடிவூட்டம் .AZW/.MOBI/.PRC ஆகியவை.Nook மின் வாசிப்பானில் பயன்படுத்தப்படுவது. Pdb வடிவூட்டம். பரவலான முறையில் எல்லோராலும் பயன்படுத்தப்படும் ஆவண வடிவமைப்பு .epub என்பதாகும்.இது பொது நியமக் கட்டுப்பாடு (open standard) விதிகளைக் கொண்டது இது xmlஐ அடிப்படையாகக் கொண்டிருப்பதால் இதை உருவாக்குவதும் , படிப்பதும் எளிது . எழுத்துருக்கள் ஒருங்குறியில் இருக்க வேண்டியது மிக மிக அவசியம்.இந்த வடிவூட்டம் எல்லா மின்வாசிப்பான்களிலும் வேலை செய்யும் .

மின்னூல்களின் உள்ளடக்கம் ஒருங்குறியில் எழுதப்பட்ட இணையப்பக்கங்கள் தான் . மின் வாசிப்பான்களும் கைக்கணினிகளும் அடிப்படையில் ஒரு கணினி மட்டு மே.ஒரு கணினியில் தற்போது நம்மால் தமிழில் ஒருங்குறியில் எழுதப்பட்ட இணையத் தளங்களை எளிதாகப் படிக்க முடியும் . இணையப் பக்கங்களின் தொகுப்பே மின்னூல்களாக வடிவம் மாறுகின்றன . (Loton, 2010) இணையப் பக்கங்களின் தொகுப்பை மின்னூல் வடிவில் மாற்றிக் காட்டுவது வது தான் மின் வாசிப்பான்களின் வேலை. எச்.டி.எம்.எல், எக்ஸ்.எம்.எல் சி , எஸ், எஸ் ஜாவாஸ்க்ரிப்ட் நான்கையும் பயன்படுத்தித் தமிழ் ஒருங்குறியில் தட்டெழுத்துச் செய்யப்படும் இணையப் பக்கங்கள் மின்னூல்களாக மாற்றப் படுகின்றன .

எக்ஸ். எம். எல் (XML) கணினி மொழி ஓர் இணையப் பக்கம் ஒரு சீராக இருக்க வேண்டும் என்பதை வலியுறுத்துகிறது. இதனால் இணையப்பக்கத்தில் கவனக் குறைவால் ஏற்படும் தவறுகள் குறைக்கப்படுகின்றன . எக்ஸ்.எம்.எல்.லின் விதிகள் ஓர் இணையப்பக்கம் எல்லாவித இணைய உலாவிகளிலும் மின் வாசிப்பான்களிலும் ஒழுங்காக வேலை செய்ய வழி செய்கின்றன. எச்.டி.எம்.எல் ஆவணத்தோடு எக்ஸ்.எம்.எல் கணினி மொழியை இணைக்கும் போது அந்த ஆவணம் எக்ஸ். எச்.டி.எம்.எல் ஆவணமாக மாறி விடுகிறது . எக்ஸ்.எச்.டி.எம்.எல் (xhtml) ஆவணங்கள் தான் மின்னூல்களின் அடிப்படை.<h1></h1> என்ற இழைகளைப் பயன்படுத்தி அத்தியாயங்களின் தலைப்புக்களை உருவாக்க வேண்டும் என்ற இழைகள் புத்தக அடையாளக் குறிகளை அமைக்க உதவுகின்றன.புத்தகத்தின் மற்ற பக்கங்களுக்குச் செல்லவும், நூலிலிருந்தே நேரடியாக இணையச் சுட்டிகளுக்குச் செல்லவும் இந்த இழைகளைப் பயன்படுத்தலாம். புத்தகத்திலிருந்தே நேரடியாக மின்னஞ்சல் செய்யவும் முடியும்.<p></p> என்ற இழைகள் புத்தகத்தின் பத்திகள் அமைக்கவும் உதவுகின்றன. இழைகள் பட்டியலிடப் பயன்படுத்தப் படுகின்றன.இவ்வாறாக அமைக்கப்பட்ட பக்கங்களை வடிவமைக்க உதவுவது சி.எஸ்.எஸ்.

சி.எஸ்.எஸ் (C.S.S) .புத்தகத்தின் பக்கங்களுடைய ஓர அளவுகளையும், தலைப்பு எந்த அளவு இருக்க வேண்டுமென்பதையும், எழுத்துக்களின் சீரமைப்பு எப்படி அமைய வேண்டுமென்பதையும் பத்திகளின் ஆரம்பத்தில் அமைய வேண்டிய ஓரத்தில் இடம் வெற்றிடம் எப்படி இருக்க வேண்டும் என்பதையும் சி.எஸ்.எஸ் குறிக்கிறது. @font-face முறையைப் பயன்படுத்தி தமிழ் ஒருங்குறி எழுத்துக்களை மின்னூல் ஆவணங்களோடு இணைக்கலாம். ஜாவாஸ்க்ரிப்ட் மூலம் கேட்பொலிகள் காணொளிகள் ஆகியவற்றை மின்னூல்களோடு இணைக்கலாம். அசைவூட்டப்படங்களைப் புத்தகத்தில் இணைக்கலாம் . . ஜாவாஸ்க்ரிப்ட் மின்னூலின் பக்க அளவுகளைக் கணினிக்குக் கணினி எளிதாக மாற்றி விடும். நூலை பக்க வாட்டில் திருப்பியோ நேராகவோ எளிதாகப் படிக்கலாம். அலைபேசிகளாகட்டும் மேஜைக்கணினிகளாகட்டும் அவற்றிற்கு ஏற்றது போல நூலின் அளவும் மாறிக் கொள்ளும்.

மின்னூல்களில் பயன்படுத்தப்படும் படங்கள் அட்டைப் படங்களாகவே பயன்படுத்தப்படுகின்றன . இப்படங்கள்72px பிரிதிநிலை (resolution) இருக்க வேண்டும்.மின்னூலின் அட்டைப்படங்களின் உயரம் 860 pxலாகவும் நீளம் 600 pxலாகவும் இருக்க வேண்டும் . (castro 2011) குழந்தை இலக்கியங்களுக்கான படங்கள் இந்த அளவை விடச் சிறியதாக இருக்க வேண்டும் .படங்களின் உயரம் நீளம் இவற்றின் விகிதாச்சாரம் மாறிவிடக் கூடாது. அட்டைப்படத்தில் வரும் தலைப்பு, ஆசிரியர் பெயர் ஆகியவை படங்களாகத் தான் இருக்க முடியும் ,

தமிழ் மின்னூல்களைத் தயாரிப்பது எளிது .epub என்பது .zip என்ற கோப்பி ன் மறு வடிவமே . ஒரு மின்னூல் கோப்பை எடுத்து அதன் நீட்சியை.epub லிருந்து.zip ஆக மாற்றி விட்டு கோப்பை விரிவு படுத்திப் பார்த்தோம்

என்றால் அதற்குள் mimetype, Content.opf, OEBPS folder என்று மூன்று வித கோப்புக்களைக் காண முடியும், (DeMont, 2011) . Mimetype ஒரு மின் வாசிப்பானுக்கு இந்தக் கோப்பின் உள்ளடக்கங்களை விவரிக்கிறது. இது வெறும் ASCII உரை மட்டுமே. META-INF என்ற உறையில் மின்னூலின் அடிப்படையான Content.opf ஆவணம் எங்கே இருக்கிறது என்று மின் வாசிப்பான்களுக்குச் சுட்டிக் காட்டும் எக்ஸ்.எம்.எல் ஆவணம் இருக்கிறது. இஃது எல்லா மின்வாசிப்பான்களுக்கூறிய மென்பொருட்களுக்கும் பொதுவானது. OEBPS உறையில் எக்ஸ்.எச்.டி.எம்.எல் ஆவணங்கள், நூலுக்குரிய படங்கள், சி.எஸ்.எஸ் ஆவணங்கள் ஆகியவற்றோடு Content.opf toc.ncx என்ற இரு வேறுபட்ட ஆவணங்கள் இருக்கும். இவை எக்ஸ்.எம்.எல் முறையில் எழுதப்பட்டவை. (DeMont, 2011)

Content.opf என்ற ஆவணத்திற்குள், நூலின் பெயர் ஆசிரியர், பதிகப்பத்தின் பெயர் ஆவணம் எழுதப்பட்ட நூல் அது யாருடைய உரிமம் எந்த மொழியில் எழுதப்பட்ட ஆவணம் ஒவ்வொரு அத்தியாங்களின் சுட்டிகள் ஆகிய குறிப்புக்கள் இருக்கும். toc.ncx ஆவணத்திற்குள் தலைப்புக்களின் அட்டவணை ஒன்று இருக்கும். இந்த எல்லா ஆவணங்களையும், கோப்புக்களையும் உரைப்பான் கொண்டு தயாரித்த பிறகு , ஒரு வெற்று .zip கோப்பை உருவாக்கி எல்லா ஆவணங்களையும் அதனுள் இட வேண்டும் . பின் .zip என்ற நீட்சியின் பெயரை .epub என்று மாற்றி விட வேண்டும். இந்த . epub ஆவணத்தை நாம் மின் வாசிப்பான்களில் படிக்க முடியும். (DeMont, 2011)

மற்ற முறைகளைப் பயன் படுத்தியும் மின் நூல்களை எளிதாகத் தயாரிக்கலாம் .மின்னூல்களின் அடிப்படை எச்.டி.எம்.எல் என்றாலும் மின்னூல்களைத் தயாரிக்க ஹெ.டி.எம். எல் தெரிந்திருக்க வேண்டியது இல்லை Microsoft word ஆவணத்திலிருந்தும் Open office ஆவணத்திலிருந்தும் ஒருங்குறி முறையில் தட்டச்சுச் செய்யப்பட்ட ஆவணங்களை மின்னூல்களாக மாற்றலாம், ஆப்பிள் நிறுவனத்தின் Pages, iBooks Author அடோபி நிறுவனத்தின் InDesign மென்பொருட்களின் உதவியுடன் மிக எளிதாக மின்னூல்களை உருவாக்கலாம். Sigil, Calibre போன்ற திறவூற்று மென்பொருட்களும் மின்னூல்கள் தயாரிக்க உதவுகின்றன. ஆப்பிள் நிறுவனத்தின் Pages, Sigil, Calibre ஆகிய மென்பொருட்களைக் கொண்டு மிக எளிதாக இந்த ஆவண வடிவமைப்பில் தமிழ் மின்னூல்களை உருவாக்கலாம். ஆப்பிள் நிறுவனத்தின் Pages ஒரு சாதாரணச் சொற்செயலி தான். அதில் ஆவணத்தைச் சேமிக்கும் போது .epub ஆகச் சேமிக்க முடியும். Sigil XHTML ஆவணங்களை வேகமாகத் தொகுத்து .epub ஆக எளிதாக மாற்றுகிறது. Calibre எந்த ஓர் ஆவணத்தையும் .epub என்று மாற்றித் தருகிறது. <http://www.epubbud.com/> என்ற இணையத் தளம் மூலமாகவும் இலவசமாகத் தமிழ் ஒருங்குறியில் எழுதப்பட்ட உரைகளை .epub ஆக மாற்றிக் கொள்ளலாம். மின்னூல்களைத் தயாரிக்கும் ஒவ்வொரு மென் பொருளும் எக்ஸ் .எம்.எல் (XML 1.1) தொழில் நுட்பத்தை இணையப் பக்க ஆவணத்தோடு இணைக்கிறது.

முடிவுரையாக மின்னூல்களைத் தமிழில் தயாரிப்பது இவ்வளவு எளிதாகவும் சிக்கனமாகவும் இருக்கும் போது தமிழில் மின்னூல்களைத் தயாரிக்கப் பதிப்பகத்தோடும் விற்பனையாளர்களும் முன் வர வேண்டும். தனிமனிதர்களுக்கு உதவும் வண்ணம் மின்னூல்களைத் தமிழில் தயாரிக்க உதவும் வண்ணம், தமிழ் சொல் திருத்திகள், இலக்கணத் திருத்திகள் வர வேண்டும். தமிழ் பேச மட்டுமே தெரிந்தவர்களுக்கு உதவும் வண்ணம் தமிழ் கேட்போலிகளை எழுத்துரையாக மாற்றும் வகைச் செய்ய வேண்டும் .படங்களாக வரும் தமிழ் உரைகளைத் தேர்வு செய்யக்கூடிய உரைகளாக மாற்றும் OCR தொழில் நுட்பம் தமிழ் மொழிக்கு ஓர் இன்றியமையாத தேவை . இணையத்தில் பயன்படுமாறு இணைய ஒருங்குறி எழுத்துருக்கள் தமிழ் மொழியில் தேவை . இந்தத் தொழில் நுட்பங்கள் தமிழ் மின்னூல் வளர்ச்சியுடன் பிணைந்து உள்ளன . தமிழ் மின்னூல்களில் வளர்ச்சியினால் தமிழ் கணினித் துறையும் மேலோங்கி வளரும் என்பதில் ஐயமில்லை.

Work Cited

1. Castro, E. (2012) . Elizabeth castro\'s epub, straight to the point: Creating ebooks for the apple ipad and other ereaders. Peachpit Pr.
2. DeMont, A. (2011) . *What is epub? why should i use it?*. Retrieved from <http://www.jedisaber.com/eBooks/Tutorial.shtml>
3. Loton, T. (2010) . e-book publishing diy. Lexington KY: LOTONtech Limited.

LANGUAGE STRUCTURE OF TAMIL WEB JOURNALISM

Dr.P.Rajendran

Department of Journalism & Mass Communication, DDE, Madurai Kamaraj University

Introduction

Language is a base of all type of media communication. Digital convergence of text, audio and image, on the internet make some impact on the content, structure and delivery on messages in the web media. The most of the Newspapers, Radio and television writers use different types of computers and software programmes for writing stories in the media.

Technology is the vital aspect of web journalism. This versatile medium has combined the features of traditional media, radio, television and newspapers. When one compares the media with other media the specific features of the web media are multimediality, Hypertextuality and interactivity.

Nature of the Tamil web journalism

The knowledge of online tools and technologies are very essential for making multimedia packages programmes for the online audiences. The nature of the beauty of online journalism are collecting, editing and informing the public about the events happened at the real time. The only difference between the E-Journalism and traditional media is using of computers, digital camera, internet and web creation software.

The Tamil web journalism general based on the real life experiences. The activities of web journalists are creating of text, image, sound and video. The Tamil web journalism is used for the audience as a source of reference, objects of research, mode of communication, and a way for delivering messages.

Characters of Tamil web writing

The web writings of the pages of Tamil are having the characters like 1. Attract the attention of the readers 2. Grab their interest 3.Pull them into content 4. Add real value to their work 5. Make them register or return 6.Increase the sense of trust in their community. This can be applicable for writing editorial, news item, announcement, feature or article in Tamil web pages. Further this Tamil web writing has the consistence on scannability or readability, brevity, information, and title. The good abstract and or synopsis of the Tamil writing encourage people to read the text.

Structure of Tamil web journalism

1. Body Text of the Tamil web page of the first paragraph contains key points. The second and third paragraph contains the supporting information. The using of the bullet points and lists help the readers to grasp the text quickly. At the end of the Tamil text there is a longer explanation or background which are very really need for the readers.**2. Embedding links** are used in the Tamil web writing to refer for outsource as additional information. Link for the text which is in the form of single word helps to click quickly to the targeted address. **3. Paragraphs:** The body of the text of the paragraph is quite short. 4. For the means of attractiveness, a **title** is a link and is best in keep them the same colour. 5. The **photograph** which ordered with the text is relevant to the story. There is a contrast between the back ground and fore ground of the text or graphics.

The level of interest of the Tamil web audience

In web writing the level of interest of the audience is based on title, one sentence summary, one paragraph summary for the major points, minor points, and link are detailed information writing that are interesting level. If the web page caters reader interest the result will get more satisfied readers. Further a web user is happy when he gets quickly and easily, the information he wants. For attracting the no interest Tamil readers, the writers should give clear and informative title to make links clear. He also provides accurate description and key words for the search engines.

Important constituents of Tamil web journalism

1. Title: The titles of the Tamil web pages generally contained the basic idea of the text. The title generally gives the answer to the questions one wants to ask. Through the link the readers to the web page, the title of the single piece of information is given in clear, meaningful to filter the users in the site.

2. One sentences summary: The sentence summary of the Tamil web page helps the readers to determine the level of interest in the text. The one sentence description of many hub are directory sites have separate links to that the pages. The one sentence summary of the text is very useful within the context of the page wire the searchers through the search engines to determine their interest in the subject quickly.

The very first sentences the summaries the complete page introduce the small interest level filter. It gives an insight into the major point of the text for those who want to skim the pages. The one sentence summary is the first sentence on the text of the web page and it is once sentence summary of the web text.

3. One paragraph summary: In Tamil web pages one paragraph summary is a higher quality version of the one sentence summary. It is used by high quality hubs. Readers who reached this level clearly have some interest in the page.

4. Major points headings: In Tamil web pages the heading are more useful to make major points. The reader scrolls down the page and read the points without even stopping. The headings of the text convey the main ideas presented the web page. Like the one sentence or the one paragraph summaries headings are more informative. In Tamil web page cute style headlines has won the users attention and trust and they impress the readers to make link with the other related pages. Each major heading of the Tamil web page has a sub-anchor. Sub-anchors allow links to be made straight to relevant section of a page. It makes the information in most specific and further the readers scan the information for downloading the page. This information of the each page provides useful information for the searchers.

4. Minor points of the writing emphasizing

- skim reader's pickup the information easily
- Highlighting points helps to read in details
- rereading page helps to brush quickly
- to find specific points is simple

The minor points of Tamil web page are behind the major point of the text and there are three principal techniques use to highlight minor points 1.Bold face font 2.topic sentences and 3. Bulleted lists. The using of combination of these techniques is more effective. Bold text is used for the easier scan and a topic sentence starts at each paragraph is useful for interpreting the matter.

The minor points of the web page highlights the sequence that in a logical and clear arguments. The text of the font of the page makes interest to the readers on own. The information of the Tamil text is understandable in writer's perspective and the presentations are unique. The most powerful features of the Tamil web writing is interaction and sharing. The headings and minor points interrupt a continuity of the page and these are used for the reinforcement of the text.

5. Hypertext of the web page helps the readers to know more information in detail. If one adds extra information with the text that may confuse the readers and loose their interest. Each web page has specific links with the relevant sections of the document which are having related and detailed sources of information. A web page is the first filter for readers that make interest with other web pages. The writings of the Tamil web page which are in logical manner levels are interesting for the readers. The web writing techniques which are used for the web pages are clear and effective. The language of Tamil web writing is scannable, concise and objectives and it may be interesting to the readers. The writings are honest, informal and personal and that are making the page more objective for the many users.

6. Inverted Pyramid style which used in the web pages are similar to the multi-level approach that helps for easy accessibility of information. This style of the pages is having more detail information that is creating more interest to

readers as long as they are interested. The inverted pyramid style is used for formatting the ordering of information within the story. These structures of the content with a multi-level approach can easily be produced.

This multi-level style captures the best web writing techniques in to a logical layout process. The text of the web pages has followed the writings of top-down methodology namely

- Presenting in major headings or sections
- Writing minor points for each sections and ordering them appropriately
- Putting relevant links next to the corresponding points
- Writing one sentence summary the text and then using that sentence to start a one paragraph summary
- Shortening the one sentence summary in to a short informative style

Texting and editing of any text in the manner of clarity and readability at each level of interest gives a broad overview on the topic

Classification of Tamil web journalism

The internet Tamil web classification system can be categorized into taxonomy and folksonomy. Taxonomy is the structure of information and the web site is designed by the web designers. Folksonomy is the big territory gained by users and social classifications.

Genre of Tamil web journalism

The genre of Tamil web journalism can be classified as 1. Informative genre 2. Interpretative genre 3. Dialogic genre 4. Opinion genre and 5.digital genre. The informative genre is the typical journalism text narration is first persons. The interpretative genres are the reports which are having reliability and closeness to objectivity. Dialogic genre is the interview format. Opinion genre is the views in the form of articles. Digital genre is the virtuality of the hypertext in multimedia potentiality and at the form of interactivity.

Conclusion

Web journalism is a versatile medium capability of combining the features of traditional media, radio, television and newspapers. This medium has limit less capacity of storage, real time updating, and interactive graphic elements, voice or video. This online journalism conserves the style, genres and technical of journalistic elements of the 'old' media. The online new stories generally contain the format of news structure, news values and pictures. The news topics of the online media are in general softer but stories are more dramatized and make to influence the readers' emotions. To conclude traditional practice of journalism contains such as reporting, editing, photo journalism and design but web journalism has apart from the features are having capacity, immediacy, flexibility, permanency, and interactivity. The web editing contains writing headlines side designing and managing the interactivity.

References

1. Diaz Noci, J. & Salaverria, R., "Manual de Redaccion ciber periodistica", Barcelona: Ariel comunicacion.
2. Joseba Ezeuz Ramos et al, "Towards a faceted Taxonomy to structure web genra corpora", Volumen 6 ano, 2011, pp.139-150.
3. D.W.Maguire, "Virtual organization to virtual product: structural challenges to online Newspapers, Publisher@Informingscience.org, June 2002, pp.987-994.

Topic B

Software

மென்பொருள்கள்

COMPUTER BASED TAMIL BRAILLE SYSTEM – A Review

Capt.Dr.S Santhosh Baboo (santhos2001@sify.com), Associate Professor

V.Ajantha Devi (ajay_press@rediffmail.com), Research Scholar

P.G and Research Department of Computer Science, D.G.Vaishnav College, Arumbakkam, Chennai

I. INTRODUCTION

Braille is used to represent text by means of tactile symbols. Braille is a system of writing that uses patterns of raised dots to inscribe characters on paper. It therefore allows visually-impaired people to read and write using touch instead of vision. It is a way for visually impaired people to participate in a literate culture. First developed in the nineteenth century, Braille has become the pre-eminent tactile alphabet. Its characters are six-dot cells, two wide by three tall. Any of the dots may be raised, giving 2^6 or 64 possible characters. Although Braille cells are used world-wide, the meaning of each of the 64 cells depends on the language that they are being used to depict. Different languages have their own Braille codes, mapping the alphabets, numbers and punctuation symbols to Braille cells according to need.

Braille provides a crucial means of literacy and independence, for people who are not able to read standard print. Through Braille many visually impaired people enjoy access to educational, professional and leisure materials. The Braille code has become the main system for the majority of those visually impaired people who read and write using tactile means, and can be found in many countries around the world. Thereby fully able to participate in and contribute to society on an equal footing with sighted colleagues. Also, the concept of Braille has been accepted as a universal approach that works across the boundaries of the world. Different countries of the world have adapted the system of Braille to suit their languages.

II. BRIEF INTRODUCTION TO BRAILLE

The Braille system is a method that is widely used by blind people to read and write, and was the first digital form of writing. Braille was devised in 1825 by Louis Braille, a blind Frenchman. Each Braille character, or cell, is made up of six dot positions, arranged in a rectangle containing two columns of three dots each. A dot may be raised at any of the six positions to form sixty-four (2^6), possible subsets, including the arrangement in which no dots are raised. It described by naming the positions where dots are raised, the positions being universally numbered 1 to 3, from top to bottom, on the left, and 4 to 6, from top to bottom, on the right as shown in Figure 1.

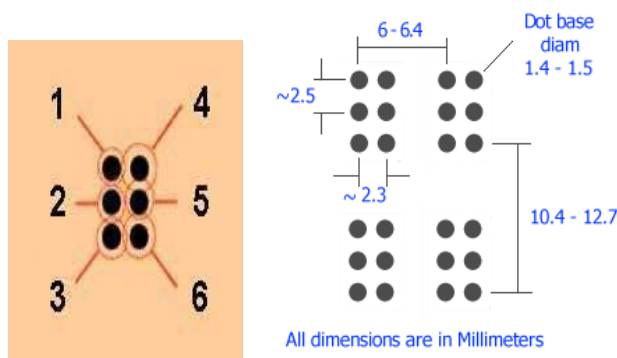


Figure 1: Braille Cell Figure 2: Braille Cell Dimensions

The lines of horizontal Braille text are separated by a space, much like visible printed text, so that the dots of one line can be differentiated from the Braille text above and below. Punctuation is represented by its own unique set of characters. In Braille, Figure 2, a cell dot pattern gives the letter to read. The dot height is about 0.5 mm; the space between dots is about 2.5 mm. A standard page in Braille has about 40 – 43 cells per line and about 25 lines. Larger cells are often used by those who have problems feeling the normal Braille cells.

III. BRAILLE SYSTEM

a. Braille Read

Standard Braille is an approach to creating documents which could be read through touch. This is accomplished through the concept of a Braille cell consisting of raised dots on thick sheet of paper. The protrusion of the dot is achieved through a process of embossing. A visually impaired person is taught Braille by training them in discerning the cells by touch, accomplished through their fingertips. The size of Braille cell is such that only one character at a time can be read by a single finger. The Figure 3 shows how this is done.

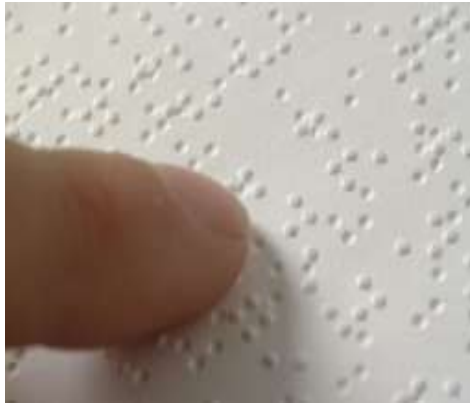


Figure 3: Braille sheet

b. Braille Write

Braille is almost always written with a slate and stylus as shown in Figure 4. Braille is written from right to left so that page can be read from left to right when it is removed from the slate and turned over. Learning to write Braille in this manner is really as the feedback is delayed until the paper is removed and then flipped over and read. Since the act of writing has no discernible and immediate effect, the reading and writing Braille conceptually challenging.



Figure 4: A Braille slate and stylus. Figure 5: A Perkins Brailler.

Braille is usually embossed with a six key typewriter known as a Brailier, which is fast and easy to use. The Perkins Brailler, shown in Figure 5, is a manual Braille typewriter. It has 8 buttons; one is used for sliding the paper, another for the movement of the 6-pin type head. The remaining 6 buttons are used for typing the Braille script.

IV. COMPUTERISED BRAILLE

Nowadays blind students are taught in mainstream classes, however most teachers of these students are not Braille literate. One method presently used to overcome this difficulty is that the students work is first sent to trained Braille transcribers, where the Braille is translated to literary text and then sent back to the teacher before it can be marked. This creates unnecessary delays and cost for the student, teacher and government. Problems also exist to be interpreted by other Braille illiterate persons, needs to be first translated by the Braille user themselves.

Today, Computers, Internet, and information explosion provide an informational structure which has changed the way people interact with the outside world. These rapid changes in this area have the potential to reduce the differences between disabled and normal individuals. In the early days of computers, visually impaired users had little or no access to the information in the computer screen. Since then, software developers of screen readers have been able to cope with the text-based environment of early operating systems. However, most modern operating system environments are now GUI. Hence, the information on the screen is no longer accessible to users who rely on screen readers or Braille displays.

There is a significant need for a system to computerize Braille documents in order to preserve them and make them available to more visually impaired people. In order to make the bi-directional communication between the sightless and sightless community feasible, it is required to transliterate the Braille documents to the corresponding text document in the corresponding language.

The Visually impaired person uses a variety of equipments and programs that enable him/her to enter data into computers or control them. Among these input devices are Braille keyboards, Optical Braille/Character scanners. Among the output devices used by the visually impaired are Braille displays, Screen readers, Braille Embossers and Screen magnifiers. There are also other assistive software packages and devices, designed exclusively for visually impaired people. Among these packages are Scientific Braille packages, Braille Note Taker.

The automatic translation process between Braille system and normal text is not a new research field. It started as early as mid 1960's (Kr, 1969) . To the best of our knowledge, the first Braille translation system was DOTSYS III (Sullivan, 1973) . One need for such translation systems is in mainstreaming, in which disabled and non-disabled students study together in the same classroom (Brule, 1985) .

V. TAMIL BRAILLE - A Review

The Ministry of Education established a Unit to deal with education of the visually impaired for developing a Uniform Braille Code and setting up Braille Printing Presses in the Country in 1947 (April) following the recommendations of the "Report of Blindness in India, ". When India gained independence (1947), 11 Braille codes were in use in various parts of the country, Bharati Braille was conceived of as early as 1951.

Bharati Braille: With the advent of computers preparation of Braille documents has been rendered easy and flexible. In the earlier days, Braille had to be printed using special Braille Printing units that worked more like typesetting printing presses. Computers have rendered the process simple where the required text can be typed normally on a computer terminal and automatically transcribed into Braille and printed. Transcription software will be language dependent but the rules of transcription can be programmed for each language. Bharati Braille may also be transcribed using computer programs by typing in the text in the vernacular.

The Bharati Braille software has taken a phonetic approach to representing Indian Language text and so it is quite easy to convert the text prepared using the multilingual editor into Braille codes. Just a simple table look up procedure is all that one would require and the program will convert text in the vernacular into appropriate Braille codes for use with an embosser connected to a computer. Bharati Braille assigns the cells to the basic sounds of the Indian languages (these are called aksharas) in a manner where vowels and consonants that find direct equivalents in English are given the same representation as in English. This way, with minimal effort one would be able to read both English text and Indian language text. This arrangement is essential if the visually handicapped are required to communicate with their counterparts in other countries.

The aksharas of Indian languages are divided into vowels and consonants. Across the many different languages of the country, one finds up to sixteen vowels and about forty consonants. The assignment of the cells is therefore applicable across all the languages though it must be stated that a few cells have to interpret based on the language.

அ	ஆ	இ	ஈ	உ	ஊ	ஏ	ஐ	ஓ	ஔ
க				ங	ச		ஜ		ஞ
ட				ண	த				ந
ப				ம	ய	ர	ல	வ	ள
ஸ	ஷ	ஸ	ஹ	க்ஷ			ற		
ன	ஃ		.	ஔ	எ	.	ழ		

Figure 6: Source: Bharati Braille – IIT Madras

This is a consequence of the fact that we have only sixty three cells available to us and reserving ten cells for punctuation leaves us in a tight situation.

VI. A Review - FINDING

In Bharati Braille, the basic vowels and consonants of the languages have been assigned individual cells. Across the languages of the country, between 13 and 18 vowels are in use and the consonants are between 33 and 37 in number. Thus more than 50 cells have been assigned for the basic vowels and consonants leaving the rest for special marks.

Tamil script consists of 12 vowels, 18 consonants and one special character, the āytam. The vowels and consonants combine to form 216 compound characters. Thus only 38 cells have been assigned.

There are many commercial text OCR products available and much research has been undertaken in text recognition, but little has been done to successfully produce a portable optical Braille recognition system. Some with greatest advantage of conversion from Braille to any of the natural languages depending on the conversion rule.

Although Braille forms exist for many languages including the students' primary languages of Kannada, Tamil etc, they are taught Braille in English first because it is the standard approach and relatively simple. Therefore most Software/ Computerization are limited to English Braille.

VII. CONCLUSION AND FUTURE SCOPE

Braille has been developed as the reading and writing system for the visually impaired. The attention was given on this is very difficult to teach a visually impaired people in the early stage and more training is needed for teaching them and converting documents to Braille, is costly and cumbersome work.

The country's first national Hindi periodical in the Braille script was launched from April 2012. there were 82, 222 newspapers and periodicals registered with Registrar of Newspapers, but Hindi Braille newspaper is the first national initiative for the visually-impaired. Hope to expect in other languages too.

VIII. REFERENCES

1. Bharati Braille URL: http://acharya.iitm.ac.in/disabilities/bh_brl.php
2. J.Bhattacharya, S.Majumder, G.Sanyal, "Automatic inspection of Braille Character: A vision based approach." International Journal of computer & Organization trends – Volume1 Issue 3 – 2011.
3. Manzeet Singh, "Automated conversion of English and Hindi text to Braille representatoin". Intrenational Journal of Computer Applications – Volume 4 – No.6, July 2010.
4. AbdulMalik S. Al-Salman: "A Bi-directional Bi-Lingual Translation Braille-Text System". J. King Saud University, Vol. 20, Comp. & Info. Sci., pp. 13-29, Riyadh (1428H./2008)
5. Al-Salman, AbdulMali; AlOhali, Yosef ; AlKanhall, Mohammed; and AlRajih, Abdullah, "An Arabic Optical Braille Recognition System", Proceedings of the First International Conference in Information and Communication Technology & Accessibility. Hammamet, Tunisia, April 12-14, 2007. pp. 81-86.
6. Zhang X., Ortega-Sanchez C., and Murray I., " Hardware-Based Text-to-Braille Translator ", the 8th International ACM SIGACCESS Conference on Computers & Accessibility, Portland, Oregon, USA, pp. 229-230, 2006.
7. Murray and T. Dias, "A Portable Device for Optically Recognizing Braille – Part I: Hardware Development " presented at ANZIIS 2001, Perth Western Australia, 2001.
8. Murray and T. Dias, "A Portable Device for Optically Recognizing Braille – Part II: Software Development " presented at ANZIIS 2001, Perth Western Australia, 2001.
9. Ebnnet, Matthew (2001), "Braille Challenge Gives Young Blind Students a Chance to Shine", Los Angeles Times, <http://articles.latimes.com/2001/jun/30/local/me-16960>.
10. Blenkhorn, Paul; System For Converting Print Into Braille, IEEE Transactions on Rehabilitation Engineering, Vol. 5, No. 2, 1997, pp. 121-129.
11. Sullivan, J. E.; Dotsys III: A Portable Braille Translator, ACM Annul Conf, 1973, pp. 398-403.
12. Braille Layout and Dimensions URL: <http://dots.physics.orst.edu/gslayout.html>

பயனர் நோக்கில் 'மென்தமிழ்' தமிழ்ச்சொல்லாளர்

இல. சுந்தரம்

M.A. (Tamil), M.A. (Ling.), M.Sc. (I.T), MCA., MBA., M.Phil., PGDCA., (Ph.D)

துணைப்பேராசிரியர், ஒருங்கிணைப்பாளர், கணினித்தமிழ்க் கல்வித்துறை, தமிழ்ப்பேராயம், SRM பல்கலைக்கழகம், ilasundaram@gmail.com

தமிழ்மொழியைக் கணினியில் பயன்படுத்துவதற்குத் தற்போது எழுத்துருக்கள் , விசைப்பலகைகள், எழுத்துரு மாற்றிகள் போன்ற தனித்தனி மென்பொருள்கள் கிடைக்கின்றன . ஆங்கில மொழியைக் கணினியில் பிழையறப் பயன்படுத்துவதற்கும் பதிப்பிப்பதற்கும் உதவும் மைக்ரோசாஃப்ட் வேர்டு (Microsoft Word) எனும் சொற்செயலி அல்லது சொல்லாளர் (Wordprocessor) என்றழைக்கப்படும் ஒரு முழுமையான தொகுப்பு மென்மத்தைப் போன்று தமிழுக்கும் சொற்பிழை திருத்தி , சந்திப்பிழை திருத்தி , இலக்கணப்பிழை திருத்தி போன்ற வசதிகள் கொண்ட முழுமையான மென்பொருளைத் தமிழுக்கு உருவாக்குவதற்குப் பத்தாண்டுகளுக்கும் மேலாகப் பலராலும் முயற்சிகள் மேற்கொள்ளப்பட்டுவருகின்றன.

தமிழ்மொழியமைப்பைக் கண்டறிவதிலும் கண்டறிந்த அமைப்பைக் கணினித் தொழில்நுட்பத்திற்கேற்ற வகையில் மாற்றிக் கொடுப்பதிலும் மொழியிலும் தொழில்நுட்பத்திலும் பொருளாதாரத்திலும் பல்வேறு நிலைகளில் சிக்கல்கள் எழு கின்றன. இத்தகைய நிலையில் தமிழ் இலக்கணம் (Tamil Grammar), கணினிமொழியியல் (Computational Linguistics), மொழித்தொழில்நுட்பம் (Language Technology) ஆகிய துறைகளின் அடிப்படையில் மென்தமிழ் (MenTamizh) எனும் தமிழ்ச் சொல்லாளர் (Tamil Wordprocessor) மென்பொருளை NDS Lingsoft Solutions Pvt. Ltd. எனும் நிறுவனத்தின் சார்பாகப் பேராசிரியர் முனைவர் ந . தெய்வசுந்தரம் அவர்கள் உருவாக்கியுள்ளார். இதன் மூன்றாம் பதிப்பு 2012 ஆம் ஆண்டு ஆகஸ்ட் மாதம் புதுப்பொலிவுடன் வெளிவந்துள்ளது. இதனை SRM பல்கலைக்கழகத்தின் தமிழ்ப்பேராயக் கணினித்தமிழ்க் கல்வித்துறை மக்கள் பயன்பாட்டிற்காக வழங்கிவருகிறது.

தமிழ்த் தட்டச்சு விசைப்பலகைகள் (Tamil Keyboards), ஒருங்குறி எழுத்துருக்கள் (Unicode Fonts), குறியேற்ற மாற்றி (Encoding Converter) சொற்பிழை திருத்தி (Spell Checker), சந்திப்பிழை திருத்தி (Sandhi Checker), தமிழ்ச்சொல் சுட்டி (Tamil Word Suggester), அகராதிகள் (Dictionaries), அகரவரிசைப்படுத்தம் (Sorting), சொல்லடைவு (Indexing), துணைநூற்பட்டியல் கருவி (Bibliography), எண்->எழுத்து மாற்றி (Number to Word Converter) போன்ற கணினியில் தமிழைத் தவறில்லாமல் பயன்படுத்துவதற்கு உதவும் பலவகையான தமிழ்மொழிப் பயன்பாட்டுக் கருவிகள் கொண்ட தொகுப்பு மென்பொருளாக மைக்ரோசாஃப்ட் வேர்டு 2010 எனும் சொல்லாளருடன் ஒப்புநோக்கிப் பார்க்கத் தகுந்த அளவுக்கு சிறப்புறத் திகழ்கிறது .

இந்த மென்தமிழ் மென்பொருளானது கணினியில் தமிழைப் பிழையறப் பயன்படுத்துவதற்கு எந்தெந்த வகைகளில் பயனர் நோக்கில் பயன்படுகின்றன என்பவற்றையும் இதன் சிறப்பியல்புகள் , செயல்திறன், திறன் எல்லை போன்றவற்றையும் இந்த மென்பொருள் இன்னும் எந்தெந்த வகைகளில் கூடுதல் வசதிகளுடன் சிறப்புடையதாக மேம்படுத்தப்படவேண்டும் என்பதையும் குறிப்பிடுவதாக இந்தக் கட்டுரை அமைகிறது .

விண்டோஸ் இயங்குதளங்களில் (எக்ஸ்பி - விண்டோஸ்8) 32 Bit, 64 Bit ஆகிய இரண்டிலும் செயல்படுகிறது . எம்.எஸ்.ஆபிஸ் வேர்டு 2010 எனும் மென்பொருளைப் போன்று காணப்படுவதால் இந்த மென்பொருளைப் பயன்படுத்துவதற்காக எந்தப் பயிற்சியும் தேவைப்படுவதில்லை . எளிமையாக இதனைக் கையாளமுடிகிறது .

விசைப்பலகைகள்

தமிழ்நெட்99 (TamilNet99), பழைய தட்டச்சு (Old Typewriter), தமிழ் ஒலிபெயர்ப்பு (Tamil Phonetics) பாமினி (Bamini), மாடுயுலர் (Modular), அஞ்சல் (Anjal), தமிழ்நெட்99 எம்.ஓ.இ (TamilNet99MOU), இன்ஸ்கிரிப்ட் (Inscript) மலேசியன் ஒலிபெயர்ப்பு (Malaysian Phonetics) சிங்கப்பூர் தமிழ் (Singapore Tamil), ரோமன் தமிழ் (Roman Tamil) விசைப்பலகை எனப் பல முறைகளில் தட்டச்சு செய்வதற்குரிய விசைப்பலகைகள் உள்ளன . இதனால் எந்த முறையிலும் தட்டச்சுப் பயிற்சி பெற்றவர்களும் இதன் வாயிலாக ஒருங்குறியில் (யுனிகோடில்) தட்டச்சு செய்ய இயலும்.

இந்நிலையில் தமிழ்நெட் 99 (TamilNet99) என்ற தட்டச்சு முறையே தமிழக அரசால் தரப்படுத்தப்பட்டது என்பதும் புதிதாகத் தமிழ்த் தட்டச்சு செய்யப் பழகுவோர் இதனைப் பின்பற்றவேண்டும் என்பதும் இங்கு நினைவுகூறத்தக்கது.

எழுத்துருக்கள்

தற்போது Adobe InnDesignமென்பொருளின் CS6 பதிப்பில் தமிழ் ஒருங்குறி செயல்படும் வசதி சேர்த்திருப்பதால் நூல் உருவாக்கம், வடிவமைப்புப் பணிகளுக்கு ஏற்ற சாய்வடிவு, தலைப்பு, குறுந்தலைப்பு, பாடல் எனத் தனித்தனி எழுத்துருக்களைப் பயன்படுத்தும் வகையில் அச்சு ஊடகத்திற்காக 20 தமிழ் எழுத்துருக்கள் இதில் இடம்பெற்றுள்ளன.

குறியேற்றமாற்றி

இது பல்வேறு தமிழ் மென்பொருள்களில் பயன்படுகின்ற அஸ்கி (ASCII) அடிப்படையிலான பல வகையான தமிழ் எழுத்துருக்களை ஒருங்குறி உள்ளீட்டு முறையில் மாற்றித்தரும் ; டெக்ஸ்ட் (.txt) கோப்புகளை மட்டுமல்லாமல், ஆர்டிஎஃப் (.rtf) கோப்புகளையும் மாற்றித்தருகிறது.

தமிழில் பரவலாகப் பயன்படுத்தப்படும் எந்த வகைத் தமிழ் எழுத்துக் குறியீட்டையும் ஒருங்குறிக்கு மாற்றலாம். பின்னர் சொற்பிழை, சந்திப்பிழை போன்றவற்றைச் சரிசெய்த பின்னர் பழைய குறியீட்டுக்கு மாற்றிக்கொள்ளலாம். இதற்காக டேம், டேப், வானவில், பாமினி, இன்டோவோர்டு, சாப்ட்வியூ, கபிலன், திபூமி, கணியன், ஸ்ரீடேம் (ஸ்ரீலிபி), இளங்கோ, மயிலை, டிஸ்கி, ரோமன், டயகிரிடிக்க, அனு, செந்தமிழ் போன்ற குறியேற்றங்களை மாற்றித் தரும் வசதி இடம்பெற்றுள்ளது. மேலும், ஒருங்குறிக்கு அடுத்த நிலையில் பதிப்புப் பணிகளுக்கும் இயல்மொழியாய்வுப் பணிகளுக்கும் பயன்படும் வகையில் தமிழக அரசின் தரமாக அறிவித்த டேஸ் (TACE) குறியேற்றத்திற்கும் மாற்றித்தருகிறது.

பெரும்பாலும் மேற்குறித்த குறியேற்றங்களில் பலவற்றை மாற்றித்தரும் பல குறியேற்ற மாற்றிகள் இருந்தபோதிலும் அவை .text கோப்புகளை மட்டுமே மாற்றித்தருவதாக அமைந்துள்ளது. கோப்பில் எழுத்துரு வடிவம், அளவு, நிறம், அட்டவணை போன்றவற்றை உள்ளடக்கிய .rtf கோப்புகளை வடிவமைப்பு மாறாமல் மாற்றித்தருவது கூடுதல் சிறப்பு.

சொற்பிழை திருத்தி

ஆங்கிலச் சொற்பிழை திருத்தியைப் போன்று தட்டச்சு செய்யும்போதே சொற்பிழையைச் சரிபார்க்கும் முறையில் இதை உருவாக்கவில்லை. தமிழ்மொழியின் சொற்பட்டியலை வரையறை செய்ய இயலாத நிலையிலும் ஆங்கிலத்தைப் போன்று அகராதியை மட்டும் வைத்து இத்தகைய சொற்பிழைத் திருத்தியை உருவாக்க முடியாத நிலையிலும் சொற்களையும் சொற்களோடு இணையத் தகுந்த இலக்கண விசுவகோணங்களையும் வைத்து விதிகளின் அடிப்படையில் சொற்பிழையைத் திருத்தும் வகையில் இது உருவாக்கப்பட்டுள்ளது என்பதை அறியமுடிகிறது. தட்டச்சு செய்து முடித்த பிறகு அல்லது வேறு ஏதேனும் மென்பொருளிலிருந்து படியெடுத்து (Copy) இதனுள் ஒட்டிய (Paste) பிறகு இரண்டு முறைகளில் இந்த மென்பொருளின் வாயிலாகச் சரிபார்க்கலாம்.

1. தட்டச்சு செய்தபிறகு, ஆவணத்தைப் பயனீட்டாளர்கள் ஆங்கிலச் சொற்பிழை திருத்தியைப் போன்று ஊடாட்டத்தின் (Interaction) வழியே திருத்திச் செம்மைப்படுத்தலாம். அதாவது 'சொற்பிழைதிருத்தி' எனும் வசதியின் வாயிலாகச் சென்றால் அந்தக் கோப்பில் உள்ள தவறான சொற்களை அடையாளம் கண்டு அதற்குரிய சரியான மாற்றுச் சொற்களைப் பரிந்துரைக்கும். இவ்வாறு காண்பிப்பவற்றுள் சரியானதைத் தேர்ந்தெடுத்து ஒவ்வொன்றாகச் சரிசெய்யலாம்.
2. அதிகப் பக்கங்கள் கொண்ட ஆவணமாக இருந்தால், 'சொற்பிழைகாட்டி' எனும் கருவி ஆவணத்தில் உள்ள தவறுகளைப் பயனீட்டாளர்களின் ஊடாட்டம் இன்றி, தானாகக் கண்டறிந்து, அவற்றைச் சிவப்புக் கோடிட்டுக் காட்டுகிறது. பின்னர் சிவப்புக் கோடிடப்பட்ட சொற்களை மட்டும் நாம் சொற்பிழை திருத்தியின் பரிந்துரை வாயிலாகவோ அல்லது தட்டச்சு வாயிலாகவோ திருத்திக்கொள்ள முடிகிறது.

சரவணன், ரவி, சுகுமார் போன்ற பெயர்ச்சொற்களில் சிலவற்றையும் மன்னார்குடி , திண்டிவனம், மயிலாப்பூர், அடையாறு போன்ற ஊர்ப்பெயர்களில் சிலவற்றையும் கண்டறிவதில்லை . பெயர்ச்சொற்கள் முழுமையாக இவ்வளவுதான் என்று தொகுத்து உள்ளிட முடியாது . காலந்தோறும் புதியபுதிய பெயர்ச்சொற்கள் உருவாகிக்கொண்டே வரும் என்பதையும் அறியமுடிகிறது . 'மலர்' 'கொடி' என்ற இரண்டு சொற்களைத் தனித்தனியாகக் கொடுத்தால் சரி என்று சொல்கிறது . 'மலர்க்கொடி' என்று இரண்டு பெயர்ச்சொற்களைச் சேர்த்து எழுதும்போது உருவாகும் தொகைச்சொற்களைத் தவறு என்றே காட்டுகிறது . ஏனெனில், இவையெல்லாம் இரண்டு சொற்கள் இணைந்து உருவாகும் தொகைச்சொற்கள் . இத்தகைய தொகைச்சொற்களைக் கையாளும் தொழில்நுட்பமும் இதில் சேர்க்கப்படவேண்டும் . எனினும் அடிப்படை நிலையில் பிழை திருத்துவதற்குப் பொருத்தமுடன் செயல்படுகிறது.

ஆங்கிலமொழிக்கான சொற்பிழைதிருத்தியும் இதில் தனியாக உள்ளது . ஆங்கிலச் சொற்பிழைகளைத் தனியாக ஒரு முறை கொடுத்துத் திருத்திக்கொள்ளும் வகையில் அமைந்துள்ளது .

சந்திப்பிழை திருத்தி

ஒற்று (சந்தி) இடப்படாத இடங்களில் இட்டும் , தேவையில்லாத இடங்களில் இட்ட ஒற்றினை நீக்கியும் தருகிறது . 'படித்து பார்' என்ற தொடரில் ஒற்று இடாமல் இருக்கிறபோது , அதைக் கண்டறிந்து 'படித்துப் பார்' என்று திருத்தித் தருகிறது . 'வந்துப் பார்' என்ற தொடரில் 'ப்' என்ற ஒற்று வரக்கூடாது . அதையும் கண்டறிந்து , 'வந்து பார்' என்று திருத்தித் தருகிறது . 'தேடிச் சென்றான்' என்று மாற்றி 'ச'க்குப் பதிலாக 'க்' போடப்பட்டிருந்தால் அதனைச் சரியாக மாற்றித் தருகிறது . 'வாழ்ந்தக் வீடு' ஒற்று வரக்கூடாத இடத்தில் ஒற்று போடப்பட்டிருந்தால் அவற்றை நீக்கியும் தருகிறது .இதே அடிப்படையில் 'லைப் ஜாக்கெட்', 'சி-டாக் நிறுவனம்' போன்ற சொற்களில் உள்ள 'ப்', 'க்' ஒற்றுகளை நீக்கிவிடுகிறது.

'தமிழ் கல்வி கழகம்' என்று இரண்டு அல்லது மூன்று பெயர்கள் ஒன்றுசேர்ந்து வரும்போது ஒற்று இடுவதில்லை . ஆனால், 'தமிழ்க் கல்விக்கழகம்' என்று ஏற்கனவே ஒற்று இடப்பட்டிருந்தால் சந்திப்பிழை திருத்தியில் அவற்றை நீக்குவதில்லை. அப்படியே விட்டுவிடுகிறது.

'கடலை சாப்பிட்டான்' என்ற தொடரைச் சந்திப்பிழை திருத்தம் செய்யும்போது கீழுள்ளவற்றில் ஒன்றைத் தேர்வுசெய்யவும் என்று 'கடலைச் சாப்பிட்டான் (கடல் + ஐ), கடலை சாப்பிட்டான் (கடலை)' ஆகிய இரண்டினைத் தருகிறது . மேற்குறித்த சொற்களின் வெளியமைப்பு ஒன்றாகத்தான் இருக்கின்றன . ஆனால், பொருள் வேறுபடுகிறது . கடலை என்பது முழுச்சொல்லா அல்லது ஐ வேற்றுமை விசுவாசத்தோடு சொல்லா என்பதைக் கண்டறியமுடியவில்லை . இத்தகைய முடிவைத் தொடரியல் ஆய்வின் அடிப்படையில் முடிவெடுக்க முடியும். இந்த ஆய்வின் அடிப்படையில் இந்த மென்பொருள் வளர்ந்தால் நம்மிடம் இந்த இரண்டு விருப்பத்தேர்வுகளைக் கேட்காமல் தானாகவே முடிவெடுப்பதற்கு வழிவகை செய்யும் என்பது இங்கு உணரப்படுகிறது. மேற்குறித்த அடிப்படையில் வேலை , முக்கிய, போன்ற சொற்களுக்கு இரண்டு பரிந்துரைகளைக் கேட்கிறது . இவ்வாறு ஏதேனும் ஒரு பரிந்துரையைக் கொடுத்த பிறகே அடுத்த சொல்லின் சந்திப்பிழை திருத்தத்திற்குச் செல்கிறது.

தமிழ்ச்சொல் சுட்டி

இன்றைய தமிழ் ஊடகங்களில் தமிழ்த் தொடர்களில் அயல்மொழிச் சொற்கள் மிகுதியாக இடம்பெறுகின்றன . ஆங்கிலச் சொற்களே அளவுக்கு அதிகமாக பயன்படுத்தப்படுகின்றன . அயல்மொழிச் சொற்களைக் கண்டறிந்து , அவற்றிற்கு இணையான தமிழ்ச்சொற்களை (விகுதிகளுடன்) பதிலீடுசெய்து தருகிறது . இத்தகைய வாய்ப்பு நல்ல தமிழ்ப் பயன்பாட்டுக்கும் , தமிழ் வளர்ச்சிக்கும் உதவும் . ஆங்கிலச் சொற்களை ஒலிபெயர்த்து எழுதினாலும் அவற்றுக்கு நிகரான தமிழ்ச் சொற்களாக மாற்றித்தருகிறது . 'நான் காலேஜில் டீச்சரைப் பார்த்தேன்' என்று தட்டச்சு செய்திருந்தால் , 'நான் கல்லூரியில் ஆசிரியரைப் பார்த்தேன்' என்று 'காலேஜ்' 'டீச்சர்' இரண்டையும் விகுதிகளோடு மாற்றித் தருகிறது . சொற்பிழைதிருத்தம் மேற்கொள்ளும்போது இந்தத் தமிழ்ச்சொல் சுட்டி கூடுதல்வாய்ப்பாகவே வழங்கப்பட்டுள்ளது . விரும்பினால் இதனைத்தேர்வு செய்துகொள்ளலாம் . அவ்வாறு தேர்வுசெய்கிறபோது பிறமொழிச் சொற்களைப் பச்சைநிற அடிக்கோடிட்டும் சொற்பிழைகளைச் சிவப்புநிறத்தில்

அடிக்கோடிட்டும் வேறுபடுத்திக்காட்டுகிறது. பிறமொழிச் சொற்களுள் சிலவற்றை மட்டுமே அடையாளம் காணுகிறது. இன்னும் பலசொற்கள் இவற்றோடு சேர்க்கப்படவேண்டும்.

இவ்வாறு சொற்பிழை திருத்தம், சந்திப்பிழை திருத்தம் போன்றவற்றைச் செய்வதற்கு வேறு ஒரு கோப்பிலிருந்து இதில் படியெடுத்தும் (copy) ஒட்டலாம் (paste) அல்லது .doc, .docx, .odt, .txt, .rtf, .html, .xml போன்ற கோப்புகளாகச் சேமிக்கப்பட்டவற்றைத் திறக்கவும் (Open) செய்யமுடிகிறது. மேலும், இதிலிருந்து மேற்குறித்த வேறு எந்த வகைக் கோப்புகளாகவும் சேமிக்கவும் (Save as) முடியும்.

அகராதிகள்

இருமொழி அகராதி (தமிழ்-ஆங்கிலம், ஆங்கிலம்-தமிழ்), தமிழ்நாடு அரசு ஆட்சிச்சொல் அகராதி (தமிழ்-ஆங்கிலம், ஆங்கிலம்-தமிழ்), மயங்கொலிச் சொல் அகராதி, அயற்சொல் அகராதி ஆகிய நான்கு வகை அகராதிகள் உள்ளன. 41 ஆயிரம் தற்காலத் தமிழ்ச் சொற்களைக் கொண்ட தமிழ்-ஆங்கிலம், ஆங்கிலம்-தமிழ் அகராதி இதில் இடம்பெற்றுள்ளது. கோப்பில் காணப்படும் சொற்களுள் ஏதேனும் ஒரு சொல்லுக்குப் பொருள் அறிய விரும்பினால் அந்தச் சொல்லின்மேல் இடப்புறம் சொடுக்கினால் அதற்குரிய பொருள், இணைச்சொல், எதிர்ச்சொல், அதன் இலக்கண வகைப்பாடு போன்றவற்றைத் தருகிறது.

இணைச்சொல் அகராதி: 'வெட்கம்' என்ற சொல்லுக்குரிய இணையான சொல்லைத் தேடும்போது 'கூச்சம், நாணம், தயக்கம்' போன்ற இணைச்சொற்களை வழங்குகிறது.

எதிர்ச்சொல் அகராதி: 'வெறுப்பு' என்ற சொல்லுக்கு எதிரான சொல்லைத் தேடும்போது 'விருப்பு, அவா, அன்பு, பாசம், நாட்டம், பிரேமை, பிரேமம், விருப்பம்' போன்ற எதிர்ச்சொற்களைத் தருகிறது.

மயங்கொலிச்சொல் அகராதி : தமிழில் ஏற்படும் 'ல, ள, ழ' - 'ந, ண, ன' - 'ர, ற' எழுத்துகளில் ஏற்படும் மயக்கங்களைத் தவிர்க்கப் பயன்படும் பொருள் வேறுபாட்டுக் குறிப்புகளை வழங்குகிறது. 'Pain' என்ற சொல்லுக்கு இணையான 'வலி' என்ற தமிழ்ச்சொல்லிற்கு எந்த 'லி'- 'ளி'- 'ழி' வரும் என்கிற நிலையில் ஏதாவது ஒரு 'லி'யைக் கொடுத்துத் தேடும்போது இவை மூன்றிற்கும் என்னென்ன பொருள் வேறுபாடு, பெயர் (Noun), பெயரடை (Adjective), வினை (Verb), வினையடை (Adverb) போன்ற நிலைகளில் இலக்கணக் கூறுகளின் அடிப்படையில் என்னென்ன பொருள்களில் பயன்படுத்தப்படுகின்றன என்னும் பொருள் வேறுபாட்டை வழங்குகிறது.

தமிழ்நாடு அரசு ஆட்சிச்சொல் அகராதி: தமிழ்நாடு அரசின் தமிழ் வளர்ச்சித்துறை வெளியிட்டுள்ள ஆட்சிச்சொல் அகராதியும் இடம்பெற்றுள்ளது. இதன் வாயிலாக ஆட்சிச் சொல்லுக்குரிய பொருளைப் புரிந்துகொள்ளமுடிகிறது.

அகரவரிசைப்படுத்தி: ஆவணத்தில் இடம்பெற்றுள்ள சொற்களைத் தமிழ் நெடுங்கணக்கின் அடிப்படையில் அகரவரிசைப்படுத்தித் (Sorting) தருகிறது. இது கோப்பில் உள்ளவற்றை ஒவ்வொரு தனித்தனிச் சொல்லாகக் கணக்கில் எடுத்துக்கொள்கிறது. பல வரிகளையோ பத்தியையோ கொடுத்து வரிசைப்படுத்தச் சொன்னால் அவற்றைத் தனித்தனி சொல்லாகப் பிரித்து அவற்றை அகரவரிசைப்படுத்தித் தருகிறது. இது ஒரு முறை. மற்றொரு முறையில் வரியையோ, பத்தியையோ அப்படியே அகரவரிசைப்படுத்தும் முறை வழக்கில் உள்ளது.

சொல்லடைவி: சொற்பட்டியலுடன் சொல் வரும் பக்கங்களை வழங்குகிறது. தனிச்சொல் மட்டுமல்லாமல், இந்திய இராணுவம், இந்திய விமானம் போன்ற இருசொல் இணைகளையும் அடைவு செய்துதருகிறது.

எண்->எழுத்து மாற்றி : பயன்பாட்டில் இருக்கும் எத்தகைய எண்களையும் தமிழ்ப் புணர்ச்சி விதிப்படி தமிழ்ச் சொற்களாக மாற்றித்தருகிறது. ஆவணங்களின் இடையில் '56894' போன்ற ரோமன் எண்கள் இடம்பெற்றால், அவற்றை 'ஐம்பத்தாறாயிரத்து எண்ணூற்றுத் தொண்ணூற்றுநான்கு' என மாற்றித் தருகிறது. இதே எண்களை 'fifty-six thousand eight hundred ninety-four' என்று ஆங்கிலச்சொற்களாகவும் மாற்றித் தருகிறது. இதே பணியைப் பின்னோக்கி எழுத்திலிருந்து எண்ணாகவும் மாற்றித் தருகிறது.

Pdfகோப்பு மாற்றி : ஆவணக் கோப்புகளை இடம்பெற்றுள்ளது.

Pdfகோப்புகளாக மாற்றும் வசதியும்

இம்மென்மத்தில்

துணைநூற்பட்டியல்: ஆய்வாளர்களுக்குப் பயன்படும் வகையில் துணைநூற்பட்டியல் தயாரிப்புக் கருவி ஒன்றும் இடம்பெற்றுள்ளது. MLA, API, Chicago போன்ற பலவகை ஆய்வுநடைகளில் துணைநூற்பட்டியலை எளிதாகத் தயாரிக்க உதவுகிறது.

மேற்கட்டியவாறு பல்வேறு சிறப்பியல்புகள் கொண்டதாக மென்தமிழ் திகழ்கிறது. எனினும் தமிழ்மொழியின் இலக்கணத்தைக் கணினிக்கு விதிகளாகக் கொடுக்கிற நிலையிலும் இலக்கணத்தில் காணப்படும் விதிவிலக்குகள், சூழல் தன்மையிலான பொருள் குழப்பங்கள் போன்றவற்றாலும் தொழில்நுட்பங்களைப் பயன்படுத்துவதற்குரிய பொருளாதார நெருக்கடிகளாலும் முழுமையான தன்னிறைவு பெறும் நிலையை நோக்கியே பயணிக்கிறது. உலகளாவிய பன்னாட்டு மென்பொருள் உருவாக்கம் மற்றும் பயனர் எளிமையாக்க (User Friendly) முறைகளில் இந்த மென்பொருள் இன்னும் மேம்படுத்தப்பட வேண்டும்.

தமிழ்ச்சொற்பிழை திருத்தியும், சந்திப்பிழைதிருத்தியும் ஏறத்தாழ 90 விழுக்காட்டுக்குமேல் சரியான முடிவுகளைத் தருகின்றன. 100 விழுக்காடு சரியான முடிவுகளைத் தருவதற்குத் தமிழ்த் தொடரியல், பொருண்மையியல் ஆய்வுகளும், தமிழ் அகராதியும் வளரவேண்டும் என்பது உணரப்படுகிறது. தொகைச்சொற்களை கையாளுவதில்லை. தமிழ்மொழியமைப்பைக் கண்டறிவது ஒருபுறம்; கண்டறிந்த அமைப்பைக் கணினித் தொழில்நுட்பத்திற்கேற்ற வகையில் மாற்றிக் கொடுப்பது என்பது மறுபுறம். இவை இரண்டும் தொடர்ந்து வளரக்கூடியவையாகும்.

‘யானைகள் வந்தது’, ‘போன்றவை பயன்படுத்தப்பட்டு வருகின்றது’ என்று தவறாக ஒருமை-பன்மை மயக்கம், உயர்திணை-அஃறிணை மயக்கங்களைக் கண்டறிந்து திருத்தத்தரும் இலக்கணப் பிழை திருத்தியும் இதில் இருந்தால் கூடுதல் பயன் தரும் என்று எதிர்பார்க்கப்படுகிறது. உருபனியல் ஆய்வின் அடிப்படையில் இந்த மென்பொருள் உருவாக்குவதற்குப் பல ஆண்டுகள் ஆயின் என்பதும் மேற்குறித்த இலக்கணப் பிழைத் திருத்தியை உருவாக்கத் தொடரியல் (Syntax) ஆய்வு மேற்கொள்ளப்படவேண்டும் என்பதும் உணரப்படுகின்றன.

தமிழில் சொற்களைச் சேர்த்து எழுதுகிற, பிரித்து எழுதுகிற கட்டுப்பாடுகள் விதிக்கப்படவில்லை. அவரவர் விருப்பத்திற்கேற்ப ‘வந்துகொண்டிருந்தான்’ என்பதை ‘வந்து கொண்டு இருந்தான்’ என்று இரண்டு மூன்று சொற்களாகத் தனித்தனியாகப் பிரித்து எழுதுகிற நிலையும் பிரித்துத் தனித்தனியாக எழுதவேண்டிய சொற்களை இடைவெளி விடாமல் சேர்த்து எழுதுவதாலும் மொழியில் பல்வேறு குழப்பங்கள் ஏற்பட்டு, பொது இலக்கண விதிகளை உருவாக்க முடியாமல் போகிறது. தற்போதைய சொற்பிழைதிருத்தி, சந்திப்பிழைதிருத்தி ஆகியவற்றின் திறன் எல்லை தொடர்ந்து விரிவடையவேண்டும். இதற்கான முயற்சியில் என்.டி.எஸ் லிங்க்சாஃப்ட் சொல்யூஷன்ஸ் நிறுவனம் தன்னைத் தொடர்ந்து ஈடுபடுத்திக்கொண்டுவருகிறது.

இதுபோன்ற மென்பொருள் தொடர்ந்து உருவாகாமல் போனதற்குத் தமிழ் மென்பொருள் உருவாக்கத்திற்குப் போதிய பொருளாதார ஆதரவும் ஒரு மென்பொருளை உருவாக்கினால் அதனை விலை கொடுத்து வாங்குவதற்கு அரசு, தனியார் நிறுவனங்கள், தனியாள் எனப் பலரும் முன்வரா ததால் இத்தகைய முயற்சிகள் பல தொடக்கநிலையிலேயே முடங்கிக்கிடக்கின்றன. இத்தகைய பணிகளுக்கு ஆதரவும் விற்பனைக்குரிய சந்தையும் இருந்தால் தமிழுக்கு விரைவில் எழுத்து பேச்சு மாற்றி, பேச்சு எழுத்து மாற்றி, எந்திர மொழிபெயர்ப்பு போன்ற பல்வேறு மென்பொருள்கள் உருவாவதற்கு வழிவகை ஏற்படும்.

On implementing a single tap keypad software for Tamil

Dr V Krishnamoorthy

Learn Fun Systems. Chennai, India
prof.vkrish@gmail.com

Abstract

This paper describes the differences found in implementing a single tap method for the 12 key phone keypad for Tamil, with that of English. The major difference being the number of extensions possible for a noun or verb. There are only about half a dozen extensions for a word in English, it is in thousands in Tamil. Also in Tamil we have concepts like, a noun being formed from a verb (vinayaal anaiyum peyar) . Many times we find that two words are combined and written as one word. A good software has to take care of all these. Along with these, the software should be able to run in real time in phones which have less powerful processors and smaller size memory. These lead to a complicated algorithmic solution.

Also we discuss the small changes we have made to the layout to make things easier to operate.

Introduction

There are several methods in English, to type in quickly using the 12 keys available in the mobile phones. T9 is a well known example. In these methods, for each letter a key is pressed only once. These software look at the different possibilities for a word that can be formed using one letter from each key, and selects the ones which are meaningful words. In general they are called "Dictionary Methods". This is due to fact that the words are searched from a built in dictionary.

Normally, all the words need not be in the dictionary. For example, storing the word "play" is enough. The words "playing", and "played" can be easily got from the root "play". Storing only the root words, and forming the derived words from the roots, will reduce the size of the dictionary.

In English, the number of such extensions are very much limited. They can be enumerated as - s, es, ed, ing, ting, ging, ness etc.. Processing a small number of different extensions is easier.

The differences

In Tamil, the number of different extensions for a noun or verb root runs into thousands. If we take a noun, there are the case markers which are added. Other extensions include the plural marker, emphasis marker "thaan", the endings "aa", "ae" and "oo", extensions like "vida", "kooda", "illai", "vum". These ending can come in different ordering.

The number of such extensions for a root can even go for more than ten. The ordering in these have to be taken care of. This is a major difference between English and Tamil.

The second is that verb can get transformed into a noun. For example, "vandhavan". Once it becomes a noun, it can take many other extensions available for a noun. For example, "vandhavanidaththilumkooda".

A root verb changes depending on tense, person and place. There are about thirty combinations for this. Each one can take some extensions. For example "vandhaayaa".

A root verb has many variations called "vaaipaadu". Some of them are "seiyya vaaipaadu", "seithu vaaipaadu". Each one of them also take more extensions. For example "paarththathanaal".

The above shows that implementation of a single tap method for Tamil is not straight forward, and it is a tricky one.

It is obvious that all the variations of a word cannot be kept in the dictionary. So the formation of words from the roots have to be done algorithmically, and, this algorithm is a complicated one. We have implemented such an algorithm.

While implementing this process, another issue crops up. Since this method is supposed to work in real time, the response of the software has to be really very quick. It has to respond within a few mille seconds, after each tap. Since this software has to work in the mobile phone environment, the restrictions in the speed of the processor becomes an important bottleneck. Also the size of the software should be such that it is possible to install this in many phones.

Next problem encountered is that a word can end with any tap. So when the last tap is done for a word, the system should predict a meaningful word. Not only that, it should be the most appropriate word. In other words, most of the time the user should get what he intended to type. For such prioritized full word output, further intelligence has to be built in. After each tap, a meaningful and most appropriate word should be predicted. The software we have developed has gone through al these stages, successfully.

In English, auto completion is usually implemented. That is, when a few letters have been typed/tapped, the software suggests the possible words with that starting. This is possible since in a dictionary such words are very much limited in number. But in Tamil, due to the very many possible extensions for a word, this seems to be practically very difficult. Hence we do not attempt to implement it.

The changes to the layout

The Government On Tamil Nadu, instituted a committee, to study the Tamil layout for mobile phones. It received suggestions from different parties, implemented about half a dozen of them, and got the user feedback from college students. Based on the feedbacks, and considering the merits in different layouts, the committee has recommended a layout. This is shown in Figure 1.



Figure 1 Figure 2

The software we have implemented is based on this recommendation. The delinking key is in the star key and the symbol table key is in the "9" key. The delinking key is to be pressed once, to avoid a vowel combining with the previous consonant. Since the symbol table showing function is not a character producing function, it is not good to have this in a number key. In English, this and other functions like going to a menu containing various options for system configuration, layout change etc., are provided in * or # key.

So, in order to be consistent, the symbol table key is transferred to the * key and the delinking key is shifted to the "9" key, as this is connected to formation of letters.

A software which uses this modified configuration has been tested with some school children, studying fifth standard and above. They had learnt it easily, without any problem. They were using a version of this software modified to work with the number pad in the normal PC keyboard.

When we were trying out the software in-house, we thought of a further modification. This layout is shown in Figure 2.

The "pulli" is used many times (about 8%) in Tamil. In the recommended layout it is in the # key, in the first position. This was due to the following reasons. 1. The first position will reduce the number of taps for pulli in the multi tap. Keeping it along with all the sparingly used grandha letters may make things easier for a single tap software.

We felt that in single tap, the finger travels lot to the # key which is in one extreme corner. It would have been easier if it is in the center. Also the keys for 7, 8 and 9 had only one/no vowel/aaitham, resulting in only 3/2 letters per key.

To change the above, we shifted the pulli to the centre key 5. The vowels from ai are shifted to the next positions. The nedil ai is kept in key 6. o, oo and ow went to 7. Aaitham went to 6. The six grantha letters were shifted to 8 and 9.

This made the layout compact, within 9 keys. # is not used. The new configuration is almost the same as the recommended configuration. The changes are minimal. It does not change the basic theme of placing the vowels and consonants in the same lexical order.

We have not noticed any perceptible change in the efficiency of the software in predicting the words. We found that the new layout is easier than the previous one, since it reduces the movement of the finger considerably.

Online handwritten Tamil word recognition using segmentation, bigram models and verification

A G Ramakrishnan and Bhargava Urala

MILE Laboratory, Dept. of Electrical Engineering, Indian Institute of Science, Bangalore

The term 'online' refers to the fact that the handwritten data is a series of (x, y) co-ordinates captured by a digitizer (Tablet PC) using a pen-like interface. There are 'pen-down' and 'pen-up' signals, which can be used to separate the captured online data into strokes. The co-ordinates are stored in chronological order as opposed to 'offline' handwritten data, which has the nature of an image with no chronological order of points available.

Preprocessing

This consists of 3 steps: (i) Smoothing – reduces the amount of high frequency noise in the input resulting from the capturing device or jitters in writing; (ii) Normalising - eliminates variability due to size differences; (iii) Resampling - obtains a constant number of data points, and makes the data independent of local and global variations in writing speed. Figure 1 (a) shows a raw character and (b), after it is preprocessed.

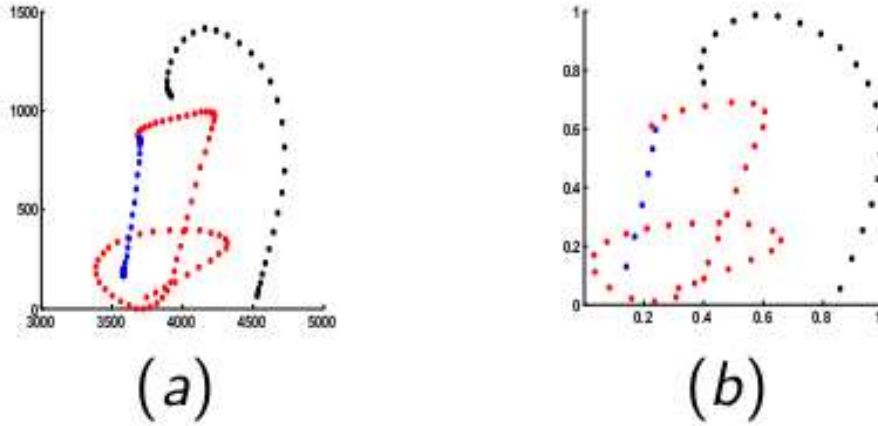


Fig. 1. (a) Original acquired sample of /ki/. (b) its preprocessed version.

Segmentation

The segmentation of individual Tamil symbols from the words (a sequence of strokes) is accomplished by two successive steps – dominant overlap criterion segmentation (DOCS) and attention-feedback segmentation (AFS) .

The DOCS module segments using the degree of horizontal overlap between bounding boxes of consecutive strokes. If the measured overlap is greater than an experimentally determined appropriate threshold, the strokes are merged to be part of the same stroke group; otherwise, they are split into different stroke groups.

The horizontal overlap between successive strokes is determined as the maximum of the two ratios obtained by dividing the x-overlap by the x-ranges of each stroke separately. The overlap is obtained as $x_{max}^{S_k} - x_{min}^c$, whereas the x-range of S_k and c are obtained as $x_{max}^{S_k} - x_{min}^{S_k}$ and $x_{max}^c - x_{min}^c$, respectively, where $x_{max}^{S_k}$, $x_{min}^{S_k}$, x_{max}^c , x_{min}^c denote the maximum and minimum x-values of the previous stroke group and the current stroke, respectively. In this method, depending on the way the characters are written in a word, cases of oversegmentation (i.e. a stroke group being a part of a valid symbol) and undersegmentation (i.e a stroke group being a merger of two or more valid symbols) can arise occasionally. Figure 2 (a), (b) and (c) show one example each of correct segmentation, oversegmentation and undersegmentation, respectively performed by dominant overlap criterion segmentation module.

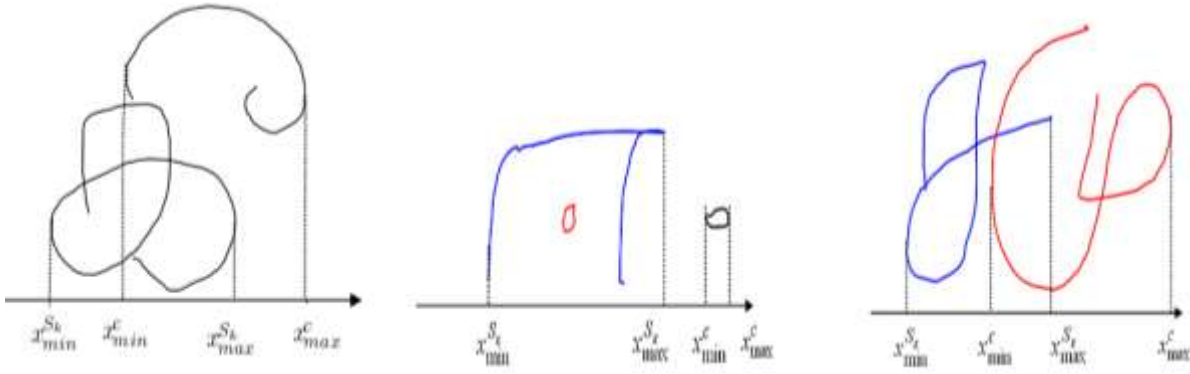


Fig. 2. Correct segmentation, over-segmentation and under-segmentation by DOCS.

Possible segmentation errors are detected by paying attention to specific features such as pen displacement, bounding box to stroke displacement and the number of dominant points (the minimum number of points to be retained to represent the character's shape). After detecting these errors, recognition likelihood from the main SVM (support vector machine) classifier and certain statistical features measured over large data such as inter-stroke displacement are used as feedback to correct the segmentation. This is known as attention-feedback segmentation (AFS). The overall AFS scheme is shown in Fig. 3.

Main Classifier

Support Vector Machine with a radial basis function kernel is used as main classifier. It is trained on preprocessed (x, y) coordinates of the data from IWFHR database.

Bigram Models

Bigram statistics are generated from a large Tamil text corpus (Emille-CIIL corpus and MILE OCR corpus), which consists of about 14 million words. The Unicode sequence of each word is converted to class label sequence and the following statistics are generated:

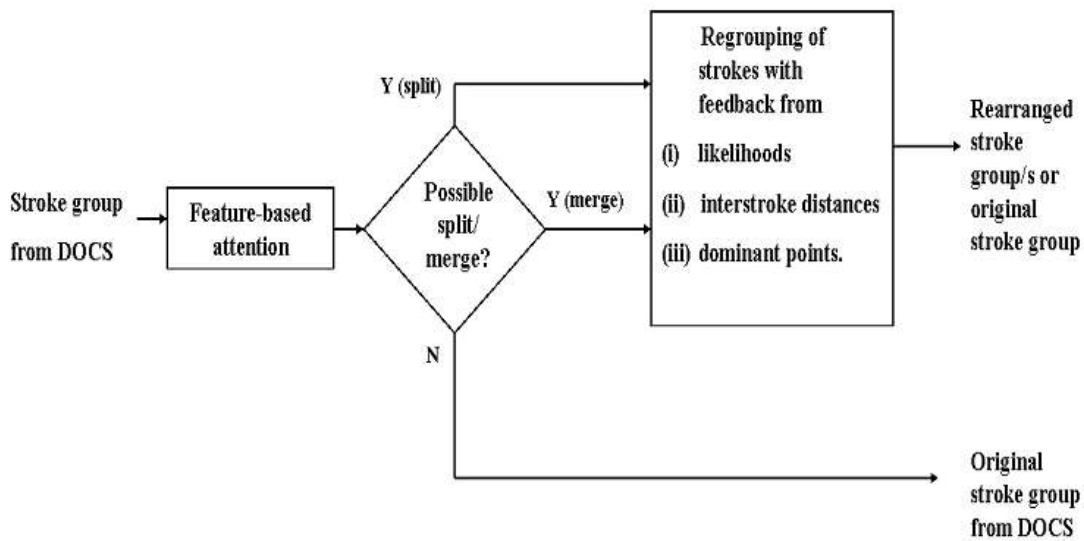


Fig. 3. Principle behind attention-feedback segmentation scheme.

N_T : Total count of all the words in the entire text corpus.

$N_1(s_i)$: Count of number of occurrences of the single symbol s_i in the corpus.

$N_2(s_1, s_2)$: Count of joint occurrences of the two symbols s_1 and s_2 contiguously.

Any handwritten word W can be considered to be a first-order Markov process, where any symbol depends only on the previous symbol. Joint symbol (bigram) probabilities can be obtained as,

$$P(s_i|s_{i-1}) = N_2(s_{i-1}, s_i) / N_1(s_{i-1})$$

Probabilities of a symbol being at the start or end of a word are obtained as,

$$P_b(s_i) = N_2(b, s_i) / N_T$$

$$P_e(s_i) = N_2(s_i, b) / N_T$$

where b denotes 'space'. In order to effectively use bigram models, we consider the 3 class labels with the topmost SVM recognition scores for every symbol in the given word X . Let W represent the set of all possible words. The most likely symbol sequence W^* to represent the unknown word is obtained as that which maximizes $P(W|X)$.

Using Bayes' rule we get, $W^* = \arg \max_W \{P(X|W)P(W) | P(X)\}$

where, $P(X|W) = \prod_{j=1}^k P(x^{s_j} | w_j)$ represents the likelihood of the handwritten word given by the SVM and $P(W)$ is obtained from bigram statistics previously computed. Neglecting $P(X)$, the equation can be rewritten in terms of logarithms as follows.

$$W^* = \arg \max_W \{ \log P(X|W) + \log P(W) \}$$

The most likely sequence W^* is obtained by backtracing the lowest cost path using the Viterbi algorithm.

Lexicon based correction

A lexicon is constructed by extracting all the unique words from the corpus of Tamil text. Total size of the lexicon is 2.4 lakh words, which is divided into several smaller lexicons, depending upon the number of symbols contained in each word.

In this postprocessing, the number of symbols k in the input word is noted down, as reported by the attention-feedback module. Then the class label sequence of the input word is obtained using the main classifier. The lexicon L_k , containing all words with k number of symbols, is loaded into the memory. Using Levenshtein distance metric, the distance d_i between the recognized sequence of symbols and every word W_{pi} in the lexicon is computed. If the distance to any entry in the lexicon is zero, it represents an exact match and the lexicon search is terminated. Otherwise, the minimum distance $d_m = \min(d_i)$ over all the entries of the lexicon L_k is found out, and the corresponding lexicon word is taken as the recognized word.

Verification

In Tamil script, there are many pairs or sets of symbols that are visually similar, which are often confused by the main classifier. Some of these symbol sets are /mu/ and /zhu/; /na/ and /La/ and /ai/; /ni/ and /Li/; /ki/ and /chi/; /la/ and /va/. In order to disambiguate between these frequently confused sets, an expert classifier is trained on the discriminating features between the set of confused symbols we use a technique called reevaluation where. Every time the AFS module assigns a class label corresponding to any one of the confused pair symbols we use the expert classifier to decide its final class label. We have examined and chosen 6 possible confusion pairs in Tamil.

Results

To see the real potential of the bigram language models and lexicon on the recognition accuracy, 2000 words with quality labels B and C are taken. These are relatively badly written words, with unusual number of strokes, unexpected overlaps, etc. The raw symbol recognition accuracy is 75.4% and the word recognition accuracy is 34.9%. With bigram models, these numbers improve to 78.8% and 42.5%. The accuracies using the lexicon are 74.1% and 46.6%, respectively. It was analyzed that the limited improvement obtained by the use of lexicon is due to the fact that nearly 25% of the test words were not found in the 2.4 lakh vocabulary, due to the morphological richness of Tamil language. On the other hand, only by using verification of 5 confusion sets, the symbol and word recognition accuracies improved to 76.9% and 45.1%, respectively. A combination of bigram models, verification and lexicon based postprocessing could lead to a more significant improvement.

Neural Networks based Recognition of Unicode Characters for identification of handwritten scripts in applications

Magesh Kasthuri (magesh.kasthuri@wipro.com)

Research Scholar, SCSVMV University, Enathur & Tech Analyst, Wipro Technologies, Bangalore,

Vigneshwar Manokar (vignesh.manohar@gmail.com)

Assistant Professor, Department of Computer Science and Engineering, Karpagam University, Coimbatore

Nandhini Vigneshwar Manokar

Technical Consultant, Bharath Corporate

Abstract - This paper discusses the concept of Neural Networks, character recognition and various technologies used for character recognition. The aim of this research on 'Neural networks and Character recognition' is to study different techniques used for character recognition and to list the differences between them. This enables us to learn about the novel algorithms and also the behaviour of the algorithms under various methods and circumstances. This highlights the inter-relation between them and also bringout the advantage of each technique with their area of specialization (how easy to use the specified technique for a specific requirement) . This proposed paper will help in easily identifying one of the best suited Handwritten Character Recognition technique for any requirement (digitizing cheques, palm Scripts/manuscripts, extract data from old documents to name a few) .

I. INTRODUCTION

One of the most important effects the field of Cognitive Science can have in the area of Computer Science is the development of technologies making tools more humanistic in nature. A relevant traditional and current existing field of natural interface research's handwriting recognition technology. Evidenced by the fact of not currently using Tablet computers, accurate hand writing recognition is clearly a difficult problem to solve. For a system which needs to be considered as good and acceptable it must be extraordinarily sorting out the issues and providing solutions, atleast to an approximated extent if not to an highly accurate and apt level. An analysis of user acceptance of handwriting technology performed by IBM showed that an accuracy rating under 97% was considered unacceptable by most users. In computer science world, AI a.k.a artificial intelligence attempts to give computers human abilities. Humans train computer to

Think artificially. One of the primary means by which computers are endowed with humanlike abilities is through the use of soft Computing Technologies, Artificial Intelligence, Mining based technologies, etc., one such broadly used is the neural network -human brain is the ultimate example for this. The human brain

Consists of a network of over a billion interconnected element called neurons. These are combinations of so many individual cell that processes small amounts of information and then activate other neurons to continue the process and simulate a network called neural network. **Neurons** - Biological neurons use pulses or spikes to encode information. Neurological research also shows that the biological neurons store information in the timing of spikes. Spiking neural networks belong to the third generation of neural networks and like their biological counterparts use spikes to represent information flow. The applications are used as spatiotemporal information in communication and computation similar to biological neurons. They use pulse coding for information processing. They are much faster than rate coding which implies some averaging mechanism, and is typical to represent information flow. This doesn't mean though that the rate coding scheme is never used but it means that pulse coding is used whenever faster speed is desired. The neuron (the granular unit of the brain) is the basibuilding block of the neural network. A neuron is communication conduit that both accepts input and produces output. The neuron receives its input either from other neurons or the user program. Similarly the neuron sends its output to other neurons or the user program.

II NEURAL NETWORK RECOGNITION

This need for accuracy is so demanding. ANN is so strong that it has even caused millions of people to learn an entirely new way to write. It has also made way that was easier for computers to detect. Unistroke recognition algorithms, like the popular Graphiti used on Palm devices require the user to adapt instead of the device, essentially the antithesis of

natural interface design. The technology has proven to be very accurate: "Each character is written with a single stroke. This solves the character level segmentation problem that previously plagued handwriting recognition. The curve drawn between pen down and pen up events can be recognized in isolation. Unistroke recognition algorithms can be relatively simple because there is no need to decide which parts of the curve belong to which character or to wait for more strokes that belong to the same character as is often the case when we try to recognize conventional handwriting. There have been many studies in the past using spiking neuron models to solve different problems. They provided a biologically plausible learning algorithm for realizing RBFs (Radial Basis Functions), which themselves are quite powerful in function approximation, pattern classification etc. In this study, spiking neurons were used to compute RBFs by storing information in their delays. The time difference between the pre and the post synaptic spikes was used to learn these delays. Neural networks use a set of processing elements (ornodes) analogous to neurons in the brain. These processing elements are interconnected in a network that can then identify patterns in data once it is exposed to the data, i.e the network learns from experience just as people do. This distinguishes neural networks from traditional computing programs, which simply follow instructions in a fixed sequential order. There are many types of neural networks. One famous concept isotones neural network, a two-level network. In this concept, all images are down sampled before being used, which prevent the neural network from being confused by size and position. To develop a handwriting recognition system that is both as reliable as Unistroke, and natural enough to be comfortable, the system must be highly adaptable. Creating software that is as adaptable as its users are unique is a very challenging problem for conventional computer algorithms. This is why many people in the field of handwriting recognition are turning to neural networks to perform the recognition processing. Adaptable by their very nature, neural networks can bring to the computing world software that molds and conforms in ways algorithms like Unistroke never could. To line up the neurons with their recognized letters, each letter image in the network will be trained and fed into the network and the winning neuron is determined. The output neuron with the largest output value is considered the winner. For example, if you were to feed the training image for "J" into the neural network and the winning neuron were neuron #4, we will know that it is the one that had learned to recognize J's pattern. This is done by getting an array of characters. The index of each array element corresponds to the neuron number that recognizes that character.

III. HANDWRITTEN CHARACTER RECOGNITION

Handwritten character recognition is one of the popular fields of research in pattern recognition. The need of handwritten character recognition is to digitize the manual written documents from hard copies of old documents like manuscripts, palm scripts and important age-old documents (bonds and registers maintained for long time) .Some of the classical examples of handwritten character recognition usage are: 1)

Recognize signatures from cheques and match them with specimen copy to identify validity 2) Recognize data from cheques and interpret them for transactions 3) Digitize old manuscripts and palm scripts (earlier palm scripts were stored in digital films and rolls from which manually we interpret the content of the scripts) . With this technology, we can digitize the content at the same time interpret the content during extraction and recognition. The process of handwriting recognition involves extraction of some defined characteristics called features to classify an unknown handwritten character into one of the known classes. A typical handwriting recognition system consists of several steps, namely: pre-processing, segmentation, feature extraction, and classification. Several types of decision methods, including statistical methods, neural networks, structural matching (on trees, chains, etc.) and stochastic processing (Markov chains, etc.) have been used along with different types of features. Many recent approaches mix several of these techniques together in order to obtain improved reliability, despite wide variation in handwriting.

A The Hopfield neural network

The Hopfield neural network is perhaps the simplest of neural networks. The Hopfield neural network is a fully connected single layer auto-associative network. This means it has one single layer, with each neuron connected to every other neuron. Consider an example program that creates the Hopfield network with four neurons. This is a network that is small enough that it can be easily understood, yet can recognize a few patterns.

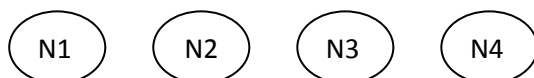


Fig. 1: Neuron example for Hopfield neural Network

	Neuron 1 (N1)	Neuron 2 (N2)	Neuron 3 (N3)	Neuron 4 (N4)
Neuron 1 (N1)	(N/A)	N2- >N1	N3- >N1	N4- >N1
Neuron 2 (N2)	N1- >N2	(N/A)	N3- >N2	N4- >N2
Neuron 3 (N3)	N1- >N3	N2- >N3	(N/A)	N4- >N3
Neuron 4 (N4)	N1- >N4	N2- >N4	N3- >N4	(N/A)

Table – 1 Connection on a Hopfield Neural Network

A Hopfield neural network has every Neuron connected to every other neuron. This means that in a four Neuron network there are a total of four squared or 16 connections. However, 16 connections assume that every neuron is connected to itself as well. This is not the case in a Hopfield neural network, so the actual number of connections is 12.

B) How the Neural Network Learns

Learning is the processes of selecting a neuron weight matrix that will correctly recognize input patterns. A Kohonen neural network learns by constantly evaluating and optimizing a weight matrix. To do this, a starting weight matrix must be determined. This matrix is chosen by selecting random numbers. Of course, this is a terrible choice for a weight matrix, but it gives a strain point to optimize from. Once the initial random weight matrix is created, the training can begin. First the weight matrix is evaluated to determine what its current error level is. This error is determined by how well the training input (the letters that is created by user in training the network) maps to the output neurons. The error is calculated during this process. If the error level is low, say below 10%, then the process is complete. Once a neural network has been trained it must be evaluated to see if it is ready for actual use. This final step is important so that it can be determined if additional training is required. To correctly validate a neural network validation data must be set aside that is completely separate from the training data. Neural networks can be applied to different application like Pattern Recognition and Auto-association. These two areas are explained below. Though this article concentrates more on Neural network application in Java, a brief introduction on Pattern Recognition will be helpful in understanding Neural Network application and Hopfield networks. Pattern Recognition Pattern recognition is one of the most common uses for neural networks. Pattern recognition is simply the ability to recognize pattern. The pattern must be recognized even when that pattern is distorted in a way. Pattern recognition in human life is used in many ways practically. A driver driving a vehicle should be able to accurately identify a traffic light. This is an extremely critical pattern recognition procedure carried out by countless drivers every day. But not every traffic light looks the same. Even the same traffic light can be altered depending on the time of day or the season. In addition, many variations of the traffic light exist. This is not a hard task

for a human driver. Hard task would be to write a computer program that accepts an image and tells you if it is a traffic light? This would be a very complex task.

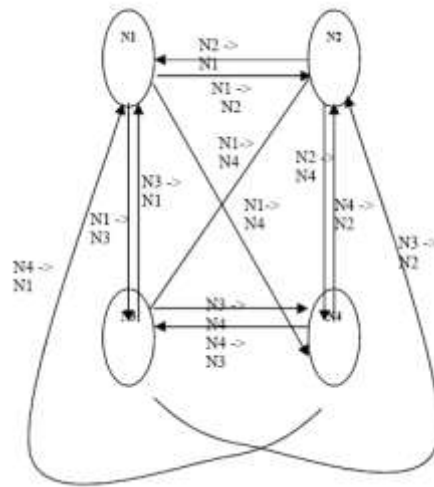


Fig. 2: A Hopfield neural network with

IV PRE – PROCESSING AND SEGMENTATION

The implementation process involved the use of a plain A4 sized paper to collect sample handwriting. A single document typically includes 10 lines with 3-4 words in each line. The documents are scanned at a resolution of 100 dpi and binarized using an adaptive thresholding technique. The image was then segmented into constituent text lines using the horizontal projection profile. For each line of text, four reference lines are extracted namely: upper line, upper baseline, lower baseline and lower line. These four lines determine three zones namely: the upper, core and lower zones. Zero values in the projection profile correspond to horizontal gaps between lines. The maximum and minimum zero value positions adjoining a text line are taken as the line boundaries corresponding to the lower line and upper line respectively. The upper baseline and lower baseline are identified using the first derivative of the horizontal projection profile. The local extreme of the first derivative in the two halves of the text line image are assumed to correspond to the two baselines. In some text lines, the upper or lower zones do not contain any character, in which case, the direct application of the above method may fail. Therefore, some heuristic rules were included in the segmentation procedure to handle such situations. To simplify the process of reference line extraction, a pre-formatted paper is used to collect handwriting, which contained all four reference lines on it. However, these lines are completely eliminated during the binarization of the image and have no effect on the segmentation. After the reference lines have been found, the individual words and characters are extracted using the vertical projection profile of each text line. Once the characters are segmented, the minimum bounding box of each character was identified eliminating the white space around it. Upper and lower boundary values of the minimum bounding box relative to the character line are sending to the next stage for preliminary classification.

VI INTERMEDIATE CONVERSION

There are some intermediate conversion types processed in –order to obtain a processed Unicode format image.

VI CLASSIFICATION RESULTS

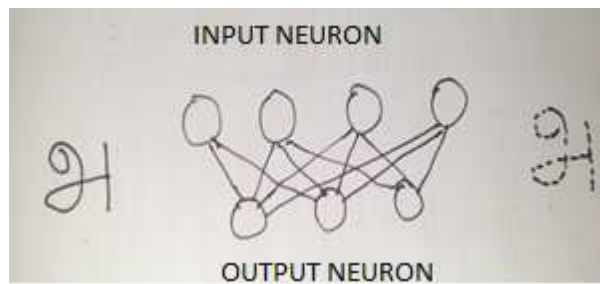


Fig. 3: A successful recognition

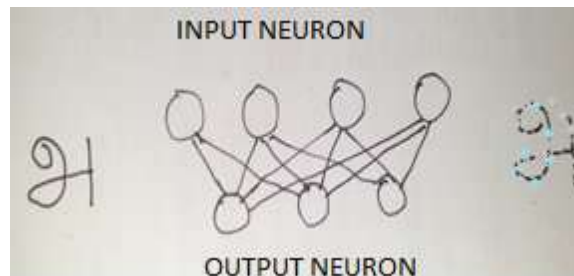


Fig. 4: A failed recognition (partially recognition pattern – pattern with distortion)

Neural Networks are one of the most commonly used systems in Artificial Intelligence. Neural Networks are particularly adept at recognizing patterns. This allows them to recognize something even when distorted. A Neural Network may have input, output and hidden layers. The input and output layers are the only required layers. The input and output layer may be the same neurons. Neural networks are typically presented input patterns that will produce some output pattern. If a Neural Network mimics the input pattern it was presented with, then that network is said to be auto associative. For example, if a neural network were presented with the pattern "0110" and the output were also "0110", then that network would be said to be auto associative. A neural network calculates its output based in the input pattern and the neural network's internal connection weight matrix. The values for these connection weights will determine the output from the neural network, based upon input pattern. The outputs in the above Fig. 1 and Fig. 2 gives brief details about the experimentation results concluded so far.

VII. REFERENCES

1. Hilla Ben-Yaacov, David Malah, Life Fellow, IEEE, and Meir Barzohar Recognition of 3D "Objects Basedon Implicit Polynomials" IEEE Transactions on Pattern Analysis And Machine Intelligence, Vol. 32, No. 5, May 2010.
2. Latent Fingerprint Matching Hilla Ben-Yaacov, Anil.K.Jain, Life Fellow, and Jianjiang Feng, IEEE Transactions on Pattern Analysis and Machine Intelligence January 2011 (vol. 33 no. 1) pp. 88-100
3. Artificial Neural Networks in Real-life Application JuanR.Rabunal, Pg. 11-30
4. Neural Networks Algorithms, Applications, and Programming Techniques - James A. Freeman and David M. Skapura - Addison- Wesley Publishing Company, Pg. 44-73
5. Neural Networks Theory - Alexander I.Galushkin - Springer Press, Pg.31-40, 110-113
6. S. M. Metev and V. P. Veiko, *Laser Assisted Microtechnology*, 2nd ed. R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.
7. Neural Networks Algorithms, Applications, and Programming Techniques - James A. Freeman and David M. Skapura - Addison- Wesley Publishing Company.
8. Fundamentals of Neural Networks - Architectures, Algorithms and Applications - Laurene Fausett.
9. Neural Networks for Pattern Recognition - Christopher M.Bishop - Charendon Press-Oxford-1995.
10. Neural Networks Theory - Alexander I.Galushkin - Springer Press.

TAMIL PUBLISHING FOR MOBILE PLATFORMS - TOOLS & UTILITIES

Hari Prasath (hari@mobileveda.com)
Ganesh Ram (ganesh@mobileveda.com)

Context

Given the rapid pace of Mobile Platform proliferation and they being part of people's life for vast majority of People non-English, but conversant in the regional language - like Tamil. For effective reach of content and to unleash the potential of technology for masses: say Basic Mobile Phone to Smart Phones to Tablets – Android, , iOS for iPad, ..., it is of utmost importance that content rendered on the device is in regional language - like Tamil.

The Need

Let me start with the below Note:

India Country Growth Manager of Facebook, **Mr. Kevin D'Souza** has been reported to have said it as:

"With over 50 million people in India on Facebook, we want to make sure that everyone has a great Facebook mobile experience regardless of the device that they choose to use, The launch of 'Facebook for every phone' mobile application in Tami, Hindi and the other regional languages enables more people to connect and share anytime and anywhere, "

- Classic testimony for the need mobile content in Tamil!



This clever strategy intends to attract more consumers in the Indian Market, providing a better growth rate of the Mobile Internet, as well as targeting the potential users of regional Languages including Tamil. And this is only scratching the surface; Tamil based mobile content right from phone menu as soon as it boots up to reading and communicating in Tamil to Power off message in Tamil – will be order of the day.

Challenge

Tamil Language character-set has universally adopted Unicode representation in Computers - yet how far!

For a moment let's look back at Tamil Language support in Computers in widely used Operating systems - the first Windows OS dates back to 1985 and only by the version Windows 2000 was one of the earliest OS to support Indic Languages including Tamil – still never supported Complex Script rendering using Unicode, until the very recent Windows 8 . While Linux as well had its lag in Tamil Language Support.

Critical factor of Tamil language content access is the Native operating system support – and above data shows the lengthy time period taken for Tamil content access in Computers, which has a much mature eco-system from developer point of view and realization by the operating system vendors.

Whereas, Mobile Phone platforms has a very diverse operating system by plethora of vendors ranging from basic Java Mobile OS [J2ME], Feature phones with proprietary OS, Symbian OS, BlackBerry OS, Bada OS, ...list goes on up until reasonably mature Mobile OS for Phones/Tablets such as Android and iOS – most of them evolving with new releases quite rapidly. Imagine they prioritizing Tamil Language Native Mobile OS support in the scheme of things - a true Challenge and a significant Gap indeed !

And then the eco system of Mobile App developers, Services targeting regional content, Government communication, Leisure content like magazines, Books and ocean of content in Tamil to be made mobile-ready and being made available to readers Globally - is a Long Haul indeed.

Scenario

In the context of regional language support Mobile Web Initiative in India conference was held in March, 2012 - organized by W3C India & IAMAI as well articulated by Mr.Venkataraman, Former Chairman of INFITT:

Most of the time the Device OEM's Engineering & Head Office (say in US, Europe, Japan or Taiwan) is ready to do Indian Languages (when they do tens of languages worldwide this is routine to them), it is the Indian Marketing & Operations office that throws the spanner. They sit in their glass offices in Gurgaon & Bangalore and think everyone in India follow English including Drivers, Maids, Cooks, Factory workers, ... By doing this they are not only killing our languages (but most of Indian Languages are classical languages surviving over thousands of years of external invasion), but also depriving the productivity & economic advantage that better communication through Mobile enables for common man.

This is the scheme of things. Yet, before few years, from a quiet corner of Vellore, bunch of Tamil Folks armed with Technology & Expertise and Passion to enable regional language Tamil content ubiquitously onto the mobile platforms, found a way out !

Pioneering Effort of Tamil Rendering in Mobile Platforms

Why wait for a Global device vendor to enable Tamil on mobiles in the scheme of their priorities - realizing this ; **As early as year 2007 Team at MobileVeda took up this mission and very first project was to render Thirukurral in Mobile Phones** based on J2ME OS which was the key mobile platform those days. It was not attempted before and no precedent to follow in the Mobile spectrum - No forums or Developer support from the device vendors or mobile OS J2ME community - as App Dev community thriving today, yet MobileVeda managed to fill the Pieces of this puzzle block - by - block for of Native Tamil Language rendering on mobile platforms, such as:

- **Design of each Tamil Character rendering:** Programmers finding all chars and combinations for each language
- **Font Designing** by graphic designers
- **Validation with Linguistic Experts**
- **Seamless UI design** for display resolution vs. Mobile Phone Screen Size
- **Mobile API Development** suitable to variety of Mobile OS
- **Readability controls** of rendered content
- **Developing Library as Font Tool so that not just us, Tamil Content and Mobile App eco-system can take advantage of.**
- **Due Academic review** [at VIT University, Tamil Nadu



Content rendered using MV Font for
Tamil Version 1.0 on 25-Jun-2007
[\[Source\]](#)

This penchant of Tamil Publishing was a key driver of the project roadmap in MobileVeda – that has led to seamless content rendering across Mobile Platforms – Gadgets like: Phones, Tablets, Portable Devices powered by Android and -

Mobile OS: from basic J2ME Java phones, Symbian, BlackBerry ...all the way upto iOS and Android through an ambitious initiative targeted at Tamil Content - Books, Magazines and other format of publishers – saying

Why just PUBLISH is pass get ready to **FUBLISH** - A Futuristic Publishing Platform

and

MV Tamil SMS Utility – An easy to use mobile messaging framework.

Tamil Publishing - Tools and Utilities

By virtue of continued innovation and sustained efforts, Tamil Language content rendering across mobile platforms is a reality today, through our Tools & utilities such as:

[1]. **Fublish – Tamil Content Mobile Publishing Platform:** Very simple mobile publishing platform: all it needs is content wide variety of supported file formats like ePub, PDF, ... – Which will be readily available to be uploaded via simple web interface and FUBLISH seamlessly delivers to cross-platform reader mobile Apps and includes sales & payment collection. Content Publishers in the following sectors can exploit its potential:

1. Build magazines in your native language.
2. Book Reading application.
3. Build Blog applications in your native language.
4. Create an application for News Reader.
5. Product catalog for Mobiles.
6. Build SMS based applications to receive SMS in your preferred Indian languages

Fublish Platform - to instantly publish eBooks, eMagazines, eJournals, eReports, eCatalog, eBrochures... Commercial Business can reach customers via product/service in Tamil.

[2]. **MV Font Library for Tamil:** Using the Application Programming Interface [API], Tamil Language content in mobile application with zero coding and content developers can package contents seamlessly without worrying about the native language support in mobile. Based on Unicode Tamil Character set rendering, This API seamlessly enables rendering of wealth of content in Tamil effortlessly: adapting content to various screen sizes of Mobile Phones. API is available for J2ME, BlackBerry, Nokia - Samsung Phones and works in most Java phones irrespective of the Vendors. [Ref: [Click on Supported Devices List](#)]

[3]. **MV Tamil SMS Utility:** Messaging is a basic use-case of a Mobile Phone and when done in Tamil to the near & dears to Government 2 Citizen communication to Marketing & Service messages by Business houses, ... will be that much more effective. This utility powers such possibility.

Conclusion

Essentially, Tools & Utilities presented in this Paper enables all possible avenues of Tamil Publishing avenues across Mobile Platforms seamlessly & enables Content and App Developers ecosystem. MobileVeda invites Tamil Fraternity across the Globe to take advantage of the tools and disseminate / consume in TAMIL way.

Snapshot of Tamil Publishing Tools & Utilities

Tamil Books / Magazine Publishing

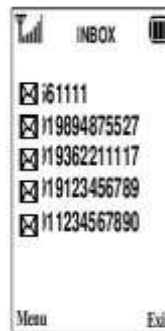


Popular Tamil Publications such as AnandaVikatan, Nakeeran, Rajesh Kumar Novels, ... are powered by FUBLISH

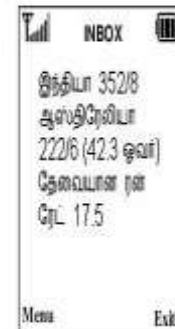


MV Tamil SMS Utility

- SMS Framework for eGovernance, Business, Individuals.



Message with the content of chosen service is received in the inbox and shown in chosen language



தமிழ் மென்பொருள்களும் மக்கள் பாவனையும்

அனுராஜ் சிவராஜா

இலங்கை

tamilambu@yahoo.com

கணினி மற்றும் இணையப் பாவனையில் தமிழ் மென்பொருள்களின் இருப்பும் பாவனையும் என்பது தொடர்பாக இந்த கட்டுரை ஆராயவிருக்கிறது. கடந்த 10 வருடங்களாக கணினி, இணையம் போன்றவற்றின் தமிழ் பாவனையாளர்களுடன் எனக்கு இருந்துவந்த நெருக்கமான தொடர்பு இந்தக் கட்டுரையை வரைவதற்கு காண்டிகலாக அமைந்ததுடன் இங்கே ஆராயும் விடயங்களின் உண்மைக்கன்மையையும் நியாயப்படுத்தும். மேலும், கடந்த 2009ஆம் ஆண்டு ஜேர்மனியில் நடைபெற்ற தமிழ் இணைய மாநாட்டில் இதே தலைப்பில் ஒரு கட்டுரை படித்திருந்தேன். அந்த கட்டுரையில் அலசிய விடயங்களின் இன்றைய நிலை பற்றி எனது இந்த கட்டுரை அமையவுள்ளது.

இந்தக் கட்டுரையில் ஆராயப்படும் விடயங்களை அடிப்படையாக வைத்து , எதிர்காலத்தில் தமிழ் மென்பொருள்களை சரியான முறையில் உருவாக்கவும் அவற்றினை உரிய முறையில் மக்களிடம் கொண்டுசேர்ப்பதற்கும் அதன்மூலம் மக்கள் உயரிய பயனை அடைவதற்குமான ஒரு பொறிமுறை உருவாக்கப்படவேண்டியதன் அவசியத்தினை உணரலாம்.

முதலாவதாக, கணினி மற்றும் இணையப் பாவனையில் தமிழ் மென்பொருள்களின் இருப்பு தொடர்பாக ஆராயும்போது பிரதானமாக விடயங்களாக பின்வருவன எடுத்துக்கொள்ளப்படுகின்றன .

தற்பொழுது பாவனையில் உள்ள தமிழ் மென்பொருள்கள்

தற்பொழுது பாவனையில் பல தமிழ் மென்பொருள்கள் உள்ளன. இவற்றில் தமிழினை பாவிப்பதற்கான ஒருதொகுதி மென்பொருள்களும் தமிழில் பாவிப்பதற்கான ஒரு தொகுதி மென்பொருள்களும் அடங்கும் . தமிழினை பாவிக்கும் மென்பொருள்களில் தமிழில் தட்டச்சு செய்வது , மின்னஞ்சல் அனுப்புவது , இணைய அரட்டையின்போது எனவும் தமிழ் சொல்திருத்தி , கையெழுத்து உணரி , தமிழ் அச்செழுத்து உணரி எனவும் பலவகையானவை உள்ளன . அடுத்து, ஏனைய அனைத்து மென்பொருள்களும் தமிழ் இடைமுகப்புடன் வரும்பொழுது அவை தமிழில் பாவிக்கும் மென்பொருள்கள் என கருதப்படும் .

அந்த மென்பொருள்கள் தற்போதைய தேவையை பூர்த்திசெய்கிறதா?

தற்போது பாவனையில் உள்ள இவ்வாறான தமிழ் மென்பொருள்கள் மக்களின் தற்போதைய தேவையை பூர்த்திசெய்கிறதா என்றால் , இல்லை என்பதே அதற்கான பதிலாக கிடைக்கும் . இப்பொழுது உள்ள மென்பொருள்களில் பல பரிசோதனை நிலையிலும் , மக்களின் உண்மையான தேவையை பூர்த்திசெய்வதாக இல்லாமலுமே இருக்கின்றன .

மேலதிகமாக எவ்வாறான தேவைகள் உள்ளன.

தமிழ் மென்பொருள்களை பொறுத்தவரையில் மக்களின் உண்மையான தேவைகள் என்று பார்க்கும்பொழுது பிரதானமாக அவர்கள் சார்ந்துள்ள பிரதேசத்தைப்பொறுத்து இரண்டு வகைப்படும் . புலத்தில் உள்ள மக்கள் , அதாவது இலங்கை மற்றும் இந்தியாவில் உள்ள மக்களைப் பொறுத்தளவில் தமிழில் பாவிக்கும் மென்பொருள்களின் தேவையே அதிகமாகும் . ஏனெனில் ஆங்கிலத்தில் உள்ள ஏனைய மென்பொருள்களை பாவிப்பதில் அவர்களின் மொழி அறிவு தற்பொழுது பெரும் தடையாக இருக்கிறது . ஆனால் புலம்பெயர் மக்களை பொறுத்தளவில் தமிழினைப்பாவிக்கும் மென்பொருள்களே பிரதான தேவையாக இருக்கின்றது . இந்த இரு வகைகளிலும் இப்பொழுது பாவனையில் உள்ளவை ஒருசிலவே .

அவற்றினை எவ்வாறு நிவர்த்திசெய்யலாம்

தற்பொழுது உள்ள மென்பொருள்களை சரியான முறையில் வரிசைப்படுத்துவதன் மூலம் இன்னமும் மக்களுக்கு தேவையாக உள்ள மென்பொருள்களை இலகுவாக அடையாளப்படுத்த முடியும். அவ்வாறு அடையாளப்படுத்தப்பட்ட மென்பொருள்களை சரியான முறையில் ஆய்வுசெய்து மென்பொருள் வடிவமைப்பாளர்களிற்கு கொடுப்பதன்மூலம் சிறந்த மென்பொருள்களினை மக்களிற்கு தரமுடியும். அடுத்ததாக, கணினி மற்றும் இணையப் பாவனையில் தமிழ் மென்பொருள்களின் பாவனை என்று பார்த் தால் பாவனையாளர்களே கருத்தில் கொள்ளப்படவேண்டியவர்கள். அந்தவகையில்,

பாவனையில் உள்ள தமிழ் மென்பொருள்கள் சரியான முறையில் பாவனையாளர்களிடம் சென்றடைந்திருக்கிறதா

தற்பொழுது பல விதமான உயர் பாவனைத்திறன் கொண்ட தமிழ் மென்பொருள்கள் உள்ளபோதிலும் அவை பெரும்பாலான மக்களிடம் சென்றடையவில்லை என்பதே உண்மை. இதற்கு உதாரணமாக, சாதாரணமான தமிழ் தட்டச்சுக்கு உதவும் மென்பொருள்கூட பெரும்பாலான தமிழ் கணினி மற்றும் இணையப் பாவனையாளர்களுக்கு தெரிந்திருக்கவில்லை என்பதையே குறிப்பிடலாம்.

இவ்வாறான மென்பொருள்கள் சரியான முறையில் பாவனையாளர்களிடம் சென்றடையாமைக்கான காரணங்கள்

எமது தமிழ் மொழியானது பல பெருமைகளுக்கு உரிய மொழி. தமிழினை ஒரு மொழியாக மட்டும் கருதுவதோடு நிற்காமல் எமது கவுரவமாகவும் அதை பார்க்கிறோம். எனவே தமிழ் மொழியில் நாம் உருவாக்கும் மென்பொருள்களினை வர்த்தகரீதியாக பார்க்காமல் மொழிக்கு ஆற்று ம் ஒரு சேவையாகவே கருதப்படுகிறது. இதனால் அந்த மென்பொருள்களினை உருவாக்குவதுடன் தமது கடமை முடிந்துவிடுவதாக பலர் கருதுகின்றனர். மக்களுள் இவ்வாறான மென்பொருள்களினை காட்சிப்பொருள்களாக பார்க்கிறார்களேயன்றி பாவனைக்கானதாக உணரவில்லை.

எவ்வாறு சகல தமிழ் மக்களிடமும் இந்த தமிழ் மென்பொருள்களை கொண்டுசேர்க்கலாம்

இந்த தமிழ் மென்பொருள்களை மக்களிடம் கொண்டுசேர்க்க வேண்டுமானால் முதலாவதாக அவற்றினை சரியானமுறையில் வரிசைப்படுத்தி அனைவரும் தெரிந்துகொள்ளும்வகையில் வைக்கவேண்டும். மேலும் மக்களின் தேவைகள் அறிந்து அவற்றினை பூர்த்திசெய்யக்கூடியான மென்பொருள்களினை உருவாக்கவேண்டும். அதுமட்டுமல்லாமல், கணினிப்பாவனை மற்றும் தமிழ் மென்பொருள்கள் தொடர்பாக மக்கள் விழிப்புணர்வினை ஏற்படுத்தவேண்டும்.

பாவனையாளர்களிடம் இருந்தான சரியான பின்னூட்டம்

இவை எல்லாவற்றிற்கும் மேலாக, தற்பொழுதுள்ள மென்பொருள்களின் பாவனையாளர்களிடம் அவைதொடர்பான சரியான பின்னூட்டங்களை பெற்று பாவனையிலுள்ள மென்பொருள்களினை உரிய முறையில் மேம்படுத்துவதும் ஒரு பயனுள்ள செயற்பாடு.

மேலும், புலத்தில் உள்ள தமிழர் (இலங்கை, இந்தியா), புலம்பெயர் தமிழர் என இரண்டு பிரிவுகளாக தமிழ் மென்பொருள் பாவனையாளர் களை பார்க்கலாம். தமிழ் மென்பொருள் தொடர்பில் இந்த இரண்டு பிரிவினருக்குமான தேவைகள், பாவனைமுறை என்பன மிகவும் மாறுபட்டவை. அதனை கருத்தில்கொண்டு எவ்வாறான புதிய மென்பொருள்களின் உருவாக்கம் தொடர்பான கருத்துக்கள்.

மொத்தத்தில், கணினி மற்றும் இணையப் பாவனையில் தமிழ் மென்பொருள்களின் இருப்பும் பாவனையும்
என்பது தொடர்பாக ஆராய்வதன்மூலம் தற்போது உள்ள தமிழ் மென்பொருள்களை சரியான மக்கள் பாவனைக்கு
கொண்டுசெல்வதுடன் மக்களுக்கு தேவையான சரியான புதிய தமிழ் மென்பொருள்களை உருவாக்குவதிலும்
கவனத்தை செலுத்தமுடியும்.

தமிழ் தொடர்புசால் தரவுதளம் மற்றும் நிரலாக்கம்

முனைவர் முத்துக்குமார் ஆறுமுகம் Ph.D (amk_in_2000@yahoo.com)

பல்கலைக்கழக இயக்குனர், சிக்கிம் மனிப்பால் பல்கலைக்கழகம், ஆக்ரா, கானா, மேற்கு ஆஃப்ரிக்கா

நிஸங்கா முத்துக்குமார் பர்வத் பி.இ.

காக்கிணிஷனன்ட் கணினி நிறுவனம், கீரனத்தம், கோவை.

கிறிஸ்டினா லவ்லீனா, பி.டெக்

கோவை.

சுருக்கக் குறிப்பு

ஒரு கணினியில் தரவுதளம் எனபது அதிமுக்கியம் . ஏனென்றால் இன்றைய காலகட்டத்தில் , ஒரு நிறுவனம்பற்றியோ, ஒரு அமைப்புபற்றியோ தகவல்களைத் திரட்டி வழிவகை செய்ய நமக்கு ஒரு கணினியும் அதனைச் சார்ந்த ஒரு தரவுதளமும் அவசியம். மேலும், அந்த தரவுதளத்திலிருந்து , அறிக்கையிடு பெற நாம் சில கட்டளைகள் பிறப்பிக்க வேண்டும் . அதற்கு நமக்கு கணினி நிரலாக்க மொழி தெரிந்திருத்தல் அவசியம் . மேலும் கணினி நிரலாக்க மொழி , நமது தமிழ் அல்லது ஆங்கிலம் போல அல்லாமல் ஒரு சில விசுவகருக்கும் இலக்கணங்களுக்கும் கட்டுப்பட்டு இருக்கும் . எனவே அனைத்து தரப்பினருக்கும் , குறிப்பாக பள்ளிப் படிப்பு படிக்காதவர்கள், ஆங்கில அறிவற்றவர்கள் அதை இயக்குவது கடினம் . எனவே, இக்கட்டுரை, தமிழ்மொழி வழியாக, தொடர்புசால் தரவுதளத்தை உருவாக்கவும் , அதனை தமிழ்மொழி வழியாக கட்டளைகள் பி றப்பித்து தகவல்களை சேமிக்கவும் , பின் திரும்பப் பெறவும் இயலும் . மேலும், Quasi-Natural Language எனப்படும், அரைஅளவான ஒரு இயற்கை மொழி அதாவது , இங்கு தமிழ்மொழி வழி நிரலாக்க உதவி கொடுக்கப்படும் . இது, Java, PHP, MySQL, XML, WebServices போன்ற தொழில்நுட்ப அறிவினால் அமையப்படுகிறது. மேலும் இது Cloud Computing எனப்படும் சேவை வழியாக பின்னர் தமிழ்கூறும் நல்லுலகத்திற்குத் தரப்படும் . இதன் கால அவகாசம் ஒரு வருடம். இந்த செயல்முறைத் திட்டம் இன்னும் குறைந்தது ஓராண்டு கால அவகாசத்தில் நிறைவு பெறும் .

முன்னுரை

மேற்கண்ட கட்டுரைச் சுருக்கத்தில் கூறியவாறு , தமிழ் மட்டுமே தெரிந்த ஒருவர் கணினியில் தரவுதளத்தை உருவாக்க வேண்டும் என்றால் அது கடினம் . எப்படி ஒரு MS-Word, MS-Powerpoint, MS-Access, MS-Excel போன்ற மென்பொருள்கள் நிரலாக்க உதவியின்றி ஒருங்குறியில் தமிழில் அடித்து அதனது வடிவத்தை அமைக்க முடியுமோ அவ்வாறே, ஒரு தரவுதளத்தை எளிதாக்கி அதன் நிரலாக்கத்தைக்கூட தமிழிலேயே ஏற்பாடு செய்து அதன் வெளி வடிவமைப்பைப் பெற முடியும். அதுவே இந்த கட்டுரையின் நோக்கமாகும் .

தரவுதள மொழி

SQL எனப்படும் ஒரு தரவுதள மொழி தெரிந்திருந்தால் மட்டுமே ஒருவரால் , தரவுதளத்தில் சேமித்து வைத்துள்ள தகவல்களைப் பெற முடியும் . எனவே மேற்கூறிய மென்பொருள்களிலிருந்து தரவுதள ஏற்பாடு இவ்வாறு வேறுபடுகிறது. இந்தப் படிப்பு கணினியில் இளநிலை , முதுநிலை மாணவர்களால் மட்டுமே திறமையாக கையாள இயலும். ஆனால் அவர்கள் ஒரு சிறு வியாபாரியிடமோ , ஒரு சராசரி தமிழ் தெரிந்த வியாபாரியிடமோ வேலை பார்ப்பதில்லை. அவர்கள் ஒரு தேசிய அல்லது பலநாட்டு வியாபார நிறுவனத்தில் ஆங்கில நாடுகளுக்காக மென்பொருள் தயாரிக்கும் துறையில் இறங்குகிறார்கள் . எனவே மேற்கூறிய தமிழக வியாபார நிறுவனத்தினரை ஆங்கிலம் வழியாகவே தகவல்களை தற்போது பெற இயலுகிறது . இதனைப் போக்க தொடர்புசால் தரவுதள நிரலாக்க மொழியை அமைக்க வேண்டியிருக்கிறது .

தரவுதள நிரலாக்க மொழியாக்கம்

கீழே, தமிழ் மற்றும் அதனது சரிசமமான ஆங்கில தரவுதள மொழியாக்கம் கொடுக்கப்பட்டிருக்கிறது . இது mysql எனப்படும் ஒரு மொழியை ஒட்டி உள்ளது . ஒரு உதாரணம் மூலம் இந்த மொழியாக்க எளிமையை நாங்கள் இங்கு விளக்கியுள்ளோம் .

பட்டியல் = Table in Database

காந்தி-நேரு பட்டியலை பின்வருமாறு உருவாக்கு {

(பெயர் எழுத்து 50),

(கதவு-எண் எழுத்து 20),

(கிராமம் எழுத்து 60),

(ஊர் எழுத்து 60),

(பின்-கோடு எண் 6)

}; // Create Database Gandhi-nehru

காந்தி-நேரு பட்டியலிருந்து பெயர், கதவு-எண், தெரு-பெயர், கிராமம், ஊர், பின்-கோடு போன்றவற்றை எடு;
(Select Statement)

காந்தி-நேரு பட்டியலில் பெயர்='முத்துக்குமார்' என்பதை பெயர்='ராம்குமார்' என மாற்று; (update statement)

காந்தி-நேரு பட்டியலில் ('ராஜாஜி', '50', 'ராமநகர் 4 வது வீதி ', 'ராக்கியாபாளையம்', 'திருப்பூர்', 638602)
என்பதைச் சேர்; (insert statement)

காந்தி-நேரு பட்டியலில் பெயர்='வீரப்பன்' என்பதை நீக்கு; (delete statement)

காந்தி-நேரு பட்டியலை பெயர் படி ஏறுமுகமாக வரிசைப்படுத்து; (Sorting Ascending)

காந்தி-நேரு பட்டியலை பெயர் படி இறங்குமுகமாக வரிசைப்படுத்து; (Sorting Descending)

//create procedure –people name listing

மக்கள்-பெயர்பட்டியல் என்ற ஒரு செயல்திட்டத்தை பின்வருமாறு உருவாக்கு {

மனிதவள தரவுதளத்தை திறக்க; // Open HR db

ஒருவேளை காந்தி-நேரு பட்டியல் (பூஜ்ஜியமாக இல்லாதிருந்தால்)

{ // when not empty

காந்தி-நேரு பட்டியலிருந்து ஒன்றன்பின் ஒன்றாக

{ // check the table one record at a time

(உபெயர், உகதவு-எண், உதெரு-பெயர், உகிராமம், உஊர், உபின்-கோடு)

என்பதில் இதைக் கொட்டு (பெயர், கதவு-எண், தெரு-பெயர், கிராமம்,

ஊர், பின்-கோடு) ; // move to a local variable

அறிக்கையீடு (உபெயர், உகதவு-எண், உதெரு-பெயர், உகிராமம்,

உஊர், உபின்-கோடு) ; // report using local variable

அடுத்ததை எடு; // get another record

} // until the table is over.

} // when finish doing non-empty process

இல்லை (பூஜ்ஜியமென்றால்)

{ // if it is empty, give error reporting.

பிழைச்செய்தி-அறிக்கியீடு (“காந்தி-நேரு பட்டியல் பூஜ்ஜியமாக உள்ளது”);

}

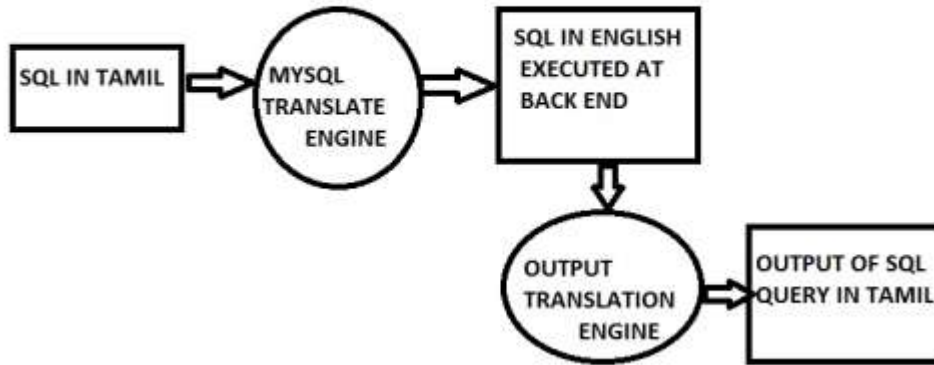
மனிதவள தரவுதளத்தை மூடு; // close HR db

}

எவ்வாறு கணினியில் அமைக்கப்படுகிறது?:

கணினியில் ஆங்கில கட்டளைகளே பெறப்படுகிறது . எனவே, நாங்கள் முதலில் பயனாளியிடமிருந்து தமிழ் மொழி கட்டளைக் களஞ்சியத்தைப் பெற்று அதனை கீழ் கொடுக்கும் முறையில் , ஆங்கில கட்டளைக் களஞ்சியங்களாக மாற்றி, அதனை ஆங்கிலத்தில் , செயல்பட வைத்து , அதனது தகவல்களை ஆங்கிலத்திலோ அல்லது தமிழில் சேமிக்கப்பட்டிருக்குமாயில் அந்த தமிழ்த் தரவுகளைப் பெற்று , தமிழிலேயே வெளியீடு செய்கிறோம் .

GENERAL ARCHITECTURE OF TAMIL SQL THROUGH MYSQL WITH TAMIL OUTPUT



முடிவுரை

தமிழ்கூறும் நல்லுலகில் தமிழில் ஒரு கணினியை ஒருங்குறியீட்டுக்கு மட்டும் பயன்படுத்தாமல் , ஒரு மொழி நிரலாக்கத்திற்கும் பயன்படுத்தி மக்களுக்குப் பயன்படச் செய்வதே இக்கட்டுரையின் நோக்கமாகும் . எதிர்காலத்தில், இதே கட்டமைப்பு மட்டும் செயல்முறையை ஆராய்ந்து வெற்றி கொண்டேயேமானால் , தமிழ் ஆப்ரேட்டிங்க் சிஸ்டம் என்பதை லினக்ஸ் மூலம் இதே முறையில் அமைக்கப்பாடுபடுவோம் . இதற்கு கூட்டுமுயற்சி மட்டும் ஆலோசனைகள் தேவைப்படுமானால், அதைத் தர நாங்கள் தயாராக உள்ளோம் .

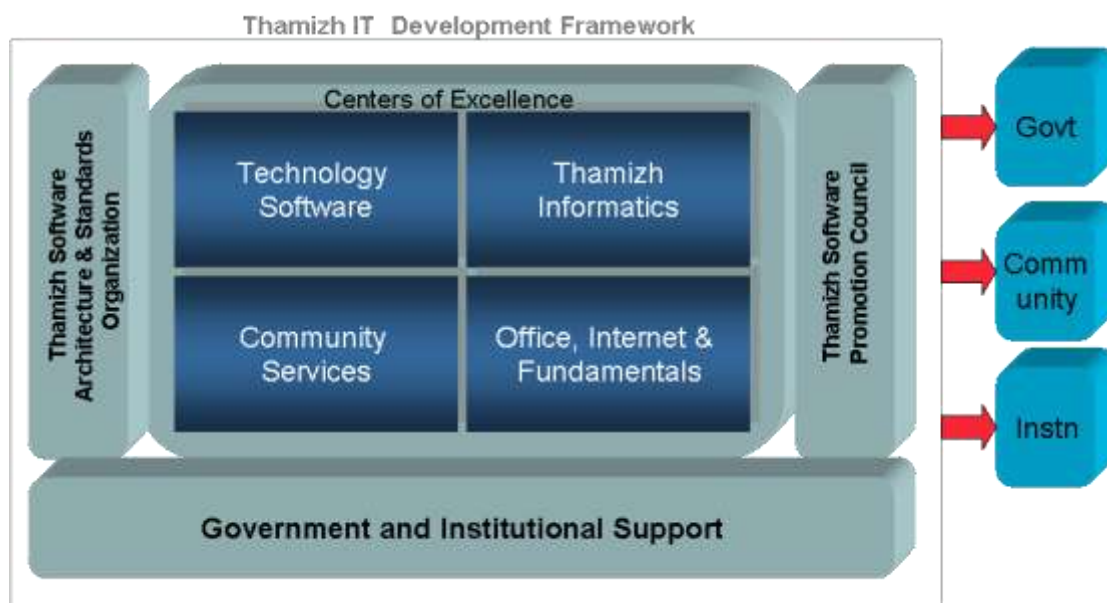
Topic C
Corpora
தரவகங்கள்

An Introduction to Thamizh Informatics and Thamizh Literature Intelligence

Naga Elangovan

This paper introduces Thamizh Informatics as a branch of computing for the benefits of Information Management of Thamizh Language Records. It emphasizes the need for maintaining clean, integrated and preserved Thamizh data to obtain valuable information through systematic development of Thamizh Informatics. In order to accelerate the growth of Thamizh-IT as a whole, this paper proposes a reference framework for development of Thamizh-IT, and places Thamizh Informatics within it and describes the same. Overall, this paper discusses a framework for the development of Thamizh-IT, Thamizh Informatics and System of Records, and Thamizh Literature Intelligence application.

A Reference framework for Thamizh-IT



Picture-1: Thamizh-IT Development Framework

Global growth of “using” Information Technology (IT) has been exponential year-on-year. If growth of Thamizh-Information Technology also has to occur, it largely depends upon adhering to the global industry model of working. The current practiced model of “Sporadic Thamizh Computing Initiatives” does not help in rapid and systematic growth. The present approach lacks constructive framework, functional model and engineering. In fact, there is no market, systematic application of software development, and long term identification of needs. Therefore, it mandates a definite reengineering, if Thamizh Computing (TC) is intended to create an industry and market for the same. Hence this reference framework is now proposed and described.

Centers of Excellence (COE): IT/Computing has scores of branches globally. Some of them are applicable and important as well for the foundation of the Thamizh-IT Industry. Currently, the term Thamizh Computing (TC) is very loosely used for all Thamizh-IT initiatives, be it Unicode character standards, font creation, arbitrary data record creation, websites creation, video/audio data creation etc. Anything Tamil to do with computers is called TC. Such an approach is too general and lacks precision. TC initiatives need to be more specific and branch wise so that a systematic focus is applied for the

growth. The foundation branches of Thamizh-IT can broadly be classified into four viz. Technology, Informatics, Community services, Office ware & Fundamentals.

Technology Software refers to the research & development of “solutions” like OCR, Speech Recognition, grammar technology etc. *Thamizh Informatics* deals with turning a vast amount of Thamizh data into useful information. Integration of literature, archaeology, manuscript and research data, Image and document data management are some of the areas within Informatics. *Community Services* refers to all the online interactions with people like eGovernance, Thamizh call centres etc. *Office ware and Fundamentals* is another vast area within the Thamizh-IT development framework dealing with Office and Publishing software, Fonts etc.

COE will focus on developing various solutions to all the above needs, assembling appropriate vendors and resources. 4 COE can be owned by 4 different sets of Government or Academic Institutions that participate in Funding and Governance support. COE for software development on these branches individually will enable a rapid development of Thamizh-IT Industry and Market. These branches will grow each other sharing the benefits of each other. These branches being core of TC, it becomes essential to pay focused attention in terms of governance, policies and support in order to serve the needs of TN over a very long time.

Thamizh Software Architecture & Standards Organization (TS-ASO): When focusing on creating an industry and market for Thamizh-IT, efficient planning is imperative. It requires an office to establish decision making principles and standards. It will integrate technology practices and standards across various branches of TC. Also it will identify issues and conflicts, and provide possible resolutions. It will interact and coordinate with world institutions and bring the best practices, guidelines and standards to Thamizh-IT industry.

Thamizh Software Promotion Council (TS-PC): While the TS-ASO supports TC with technology and engineering, Promotion Council supports TC with Market for Thamizh Software. TS-PC will work on identifying the areas requiring Computerization and Thamizhization solutions. It creates the market for Thamizh-IT. It understands the market issues and conflicts and strategizes the solutions. The advantage of positioning TS-PC within the framework is to earn revenue for the Government and other funding organizations. The funding support of the Government and other organizations can not go always as expenses. There should be returns for Thamizh-IT initiatives wherever possible. It will help the industry sustain and grow in a healthy way.

Government and Institutional Support: The whole Thamizh-IT initiative can only be successful if the government of Thamizh Nadu, in collaboration with various Academic and Research Institutions, fully supports Thamizh-IT by enabling the other components of the framework. GoTN may consider it as a potential revenue earning opportunity. Revenue opportunity can be shared by self and the other participating funding Institutions. GoTN support shall be in terms of Funding and Governance. Revenue driven Thamizh-IT development will offer more benefits than a non-revenue approach.

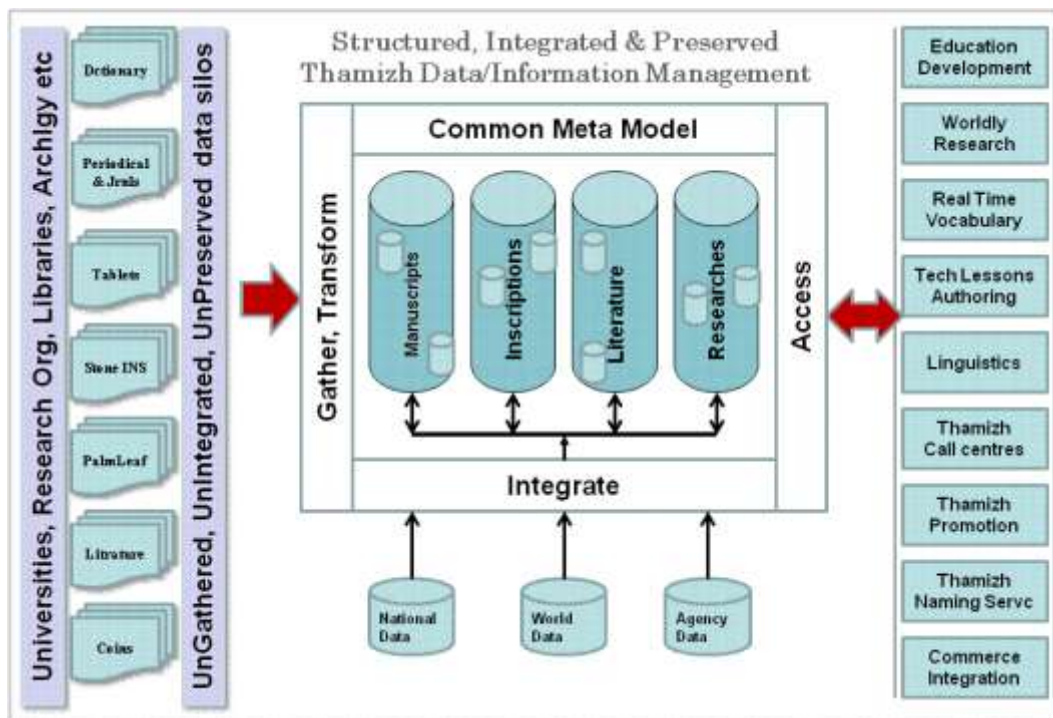
Thamizh Informatics

Informatics is a broad field within IT that deals with large amount of data, integration and processing to obtain useful information. The business industry world-wide has used it and benefited beyond imagination. Tens of social and technology fields also have used Informatics in a big way. In the context of Thamizh-IT, this branch of IT can be adapted to Thamizh Language computing where huge amount of data is available; and hence it will turn out to be a vast field of Thamizh Informatics.

Lots of Thamizh data are available in Literature, Archaeology, Manuscripts, Research Periodicals and Journals, and others. These data exist in different media/formats like stone, palm-leaves, paper, Picture, Digital, Multi-Media etc. Oldest data is reported to be dated 500 BC. Importantly these data exist in silos across multiple organizations like Universities, Libraries, Research Institutions, Private Organizations and individuals of TN and the world.

Be it business or language domain, every civilized country trusts that their data are assets. They use Informatics as a technological solution to Gather, Cleanse, Arrange, Integrate, Process, Analyse, Research and Use the data for meaningful information. More importantly preservation of data for a very long period of time is another attractive part of

Informatics. Thamizh Language domain as of today has every reason to migrate into an information system in an integrated way, and preserve the ancient evidences too. The status of Thamizh data can be summarized as under:



Picture-2: Thamizh System of Records

- Not captured on a common model
- Digital versions of literature are unreliable due to flood of mistakes
- No record of Inscriptions and Coins
- Research papers are not accessible
- Palm-leaves have lost their life and are decayed worst in various libraries. There is every chance of loosing these evidences of ancient literature
- Any existing small digital records of Thamizh data are not accessible to all. They are mostly in silos and unused.
- Cross reference between any two Thamizh data is impossible
- Very few Dictionaries are seen on the net but ownership by Institutions of TN is very rare; an Integrated dictionary system is absent
- Tens of Dictionaries are yet to be digitized and no Word/Alphabet/Sign concordance system is available
- Thamizh Data Corpora is a continuing dream; and no systematic base for qualitative and quantitative analysis and research
- Thamizh texts and books are available a lot in the internet but no way an integration with the rest of Thamizh data is possible

It can be observed that there are unavailability and inconsistency issues in all the Thamizh related data and related initiatives. This means that “Thamizh System of Records” (SOR) is absent in Thamizh-IT world. Another important concern is that there is a general belief that publishing in the internet is the way of Thamizh growth. It is true to some extent but with the inconsistent and flawed versions of any Thamizh data is not of any help. There is a pressing need for SOR for Thamizh data. There is also a belief that Thamizh has grown a lot and is the most popular language in the internet after English. This is a very misplaced myth that can endanger the prospects of true Thamizh growth. The fact is that Thamizh is nowhere in the top ten languages of the Internet.

Picture-2 as above depicts the sources of Thamizh data, Thamizh System of Records, and possible uses of clean Thamizh Information System. It is very natural that any information system evolves from disparate sources of data and heterogeneous forms of existence. Hitherto, such data in the Thamizh context have been handled in individualistic way towards computerization/digitization. The results are proving that there is a need to adhere to proven industry standards and model of implementation. The following are some of the important tasks to be done to create a Thamizh Informatics System with proper SOR:

- Plan and Architect the Infrastructure (HW/SW). While cost effective database and software solutions are available, it should be taken into consideration that going for cheap solutions and free solutions will pose serious constraint over long time.
- Establish a common Meta Model for all Thamizh data
- Strategise Data gathering (or extraction or integration from source)
- Clean and Transform the source data
- Data Model the logical structures of the TSOR and design Physical Structure and Store the data
- Establish and Maintain Reference data and Meta data
- Integrate the various types of data like Literature, Inscriptions, Images etc
- Establish policies and processes to bring in external data from the National databases, World Language Informatics and Archaeological data sources.
- Build User Systems through internet that uses the data to build it's information; and input the data it has gained from the world

It is important that the Technologists and Thamizh Subject matter experts work together on these tasks in an "Incremental and Iterative" model of development for the best results.

Potential benefits of Thamizh Informatics: Thamizh Informatics when implemented using Information Architecture, Standards, methodology and models, it will set up a golden platform for Thamizh development. Some of the pointers towards its potential uses are:

- Common, integrated, collaborative platform of Thamizh Information
- Collaboration enables the students, teachers, researches and community working together on a world class language platform
- Standard of research and research education will grow exponentially
- Community services like Call centres for every Thamizh need of people
- Create a world class real-time science and technology translation and vocabulary creation and help Thamizh based teaching in science and engineering studies
- Integrate with commerce, science, law fields to serve their Thamizh needs
- Perform world class quantitative and qualitative linguistics, literary and archaeological researches
- Develop large corpora, build Alphabet/Signs/Words Concordance system.
- Integrate tens of dictionaries of different fields
- Integrate with eGovernance for all Thamizhization requirements

Thamizh Literature Intelligence

Thamizh Literature Intelligence is the software application part of Thamizh Informatics. The software application enables Analysis and Research of Thamizh literature. It works on top of the Thamizh System of Records created as described earlier in this article. It enables deep insight into the Literature from quantitative and qualitative perspectives. Further, it allows working with other form of Thamizh data like Inscriptions, Coins, Palm-Leaves etc. for comparative studies. The software allows multi dimensional analysis and data mining for deep researches.

The author has developed a Literature Intelligence application with some features to demonstrate. It grows gradually. The software works on the SOR that contains 27 literature titles and few inscriptions as of now. ThirumurugAtruppadaai, NedunalvAdai, Natrinai, Thirukural, Thirumurai 1 to 9, Prabandam 1 to 3 and Kalladam are added to the SOR as of now. Gathering of more literature and Inscriptions continues. This application enables graphical presentation of Frequency,

Distribution of letters, syllables, yAppu etc. for a given literature and across all literature. Some of the features that can be demonstrated are

- Frequency analysis of all Thamizh letters
- Frequency of vowels, consonants and composite letters
- Distribution of consonant categories
- Letter usage in literature – Highly used (top 10), Less used (low 10), unused letter analysis
- Author wise yAppu usage
- Data Mining – Inherent patterns of syllables across literature
- Concordance across literature and Inscriptions for given string of letters, syllables, words
- Pattern Analysis - It allows building various algorithms and investigate if there are any inherent patterns and similarity among poems

Using this platform the author is researching to find if a signature can be obtained for every literature and author. A few algorithms have been tried and the effort is gradually progressing. One of the advantage of this could be to find the differentiating factors between multi-author literature like Natrinai and single-author literature like Thirukural. It can also help identifying the possible insertions.

Due to article limitations, detailed discussion on various features can be a separate part. However, this application gives two messages. The first is that a disciplined and engineering way of setting up Thamizh System of Records can allow endless modern analysis and research on Thamizh data. It is required for Thamizh. The second message is that any Thamizh researcher and student can learn building Thamizh Informatics and Literature Intelligence if they follow the simple methodology of the author. They can participate in the initiatives of creating larger version of Thamizh Informatics as a more technical person.

Conclusion

All Thamizh data must be gathered under a defined Meta Model which does not exist today. Capturing them in bits and pieces all around will be waste of effort and money. Thamizh Informatics will place Thamizh on par with global language analysis status which is important for Thamizh to be competitive. Past evidences of all Thamizh data should be preserved on priority basis. Thamizh-IT development initiatives should classify Thamizh Computing and pay focused attention on the 4 foundation areas with the support of Technology and Market Promotion. Thamizh-IT development framework is essential for accelerated growth of Thamizh-IT industry and market. Spending on Thamizh-IT will yield if and only if it follows the standards and practices of the world that is very successful on its scientific approach.

தமிழ்மொழியில் எழுத்துகளின் பயன்பாட்டு-நிலை: நிரல்களும் வரைபடங்களும்

நூ த லோகசுந்தர முதலி

சென்னை, இந்தியா

உலகில் இன்று பெருமளவிலான மக்களால் பேசப்படும் மொழிகளின் நிரலில் முதல் இருபது என்பதிலே காணும் மொழி தமிழ்மொழி . அம்மொழிகளின் தொன்மை , இடையின்றி மக்கள் நாளும் பேசி எழுது மொழி எனும் இருநிலை, இயல்பான பரிணாம வளர்ச்சி, என்பனவுடன் முன்னிற்பவையில் தமிழுக்கு இணை சீனம் மட்டும் என அறிஞர் பெருமக்கள் அறிவர்.

இயற்கைவழி மாறிவரும் உலகில் கணினி கடிதே ஊடுருவி அளவில் மிகப் பெரிதாகி தொடர்ந்து மிக வேக வளர்ச்சியுற்று வரும் இக்காலகட்டத்தில் தமிழ்மொழி தன் பங்கினை நன்காற்றி வந்தாலும் தொடர்ந்து நிலைத்து நிற்க பல்வேறு துறைகளின் வல்லுநர் நன்றுணர்ந்து மடியின்றி உழைப்பதும் அறிவீர் .

தமிழ் மொழியில் உள்ள நூல்களில் பயன்கொண்ட சொற்களின் எழுத்துக்களின் சதவிகித பயன்பாட்டு நிலைகள் முதலில் ஓரளவிற்கு ஆயப்பட்டு அதற்கேற்பவே கணினி உள்ளீடு விசைப்பலகைகள் அப்பொது வடிவமைக்கப்பட்டன என்பதை நன்கறிவோம்

முதல்முதல் கணினியின் தொடக்க நிலையில் இருந்த அதன் பயன்பாட்டுப் போக்கு அதாவது சொற்களை /எழுத்துக்களை எழுதி படிக்க மட்டுமிருந்தது . இந்நாள் ஓர் நூலினுள் சொற்களை தேடுதல், ஒன்றிற்கு ஒன்று மாற்றி அமைத்தல், அகரமுதலிபோல் அவற்றை நிரல்படுத்தல் , மொழியின் இயல்வழி (இலக்கணம்) பல்வகை ஆய்வுகள் செய்ய ஒலிவழியே உள்ளீடு செய்ய , கேட்கும் ஒலியிலிருந்து பார்க்கும் எழுத்துக்காக எழுத , ஓர் மொழியிலிருந்து மற்றோர் மொழிக்கே முழு மொழிமாற்றம் செய்ய , கையால் கீறும் எழுத்துக்களை உணர, தொடுதிரை வழி உள்ளீடு என பலவகையிலும் தொழில் நுட்பம் வளர்ந்து வருவதையும் நாம் அறிவோம் .

நாம் நன்கே பழகிய கணினியின் QWERTY விசைப்பலகை பலகாலம் தொடரும் என்றாலும் மிகமிக வேகமாக வளர்ந்துவிட்ட கைபேசியின் விசைப்பலகைகள் பற்பல விதங்களாக நுழைந்துள்ளன . இவற்றிற்கெல்லாம் ஈடுகொடுத்து தமிழ்மொழி வளர அதற்காக உழைக்க தனி நபர் நிலையில் முயற்சிகள் அங்கொன்றும் இங்கொன்றுமாக தோன்றுகின்றன .

இவ்வகை இயல்பான தொடர் வளர்ச்சியில் ஒலிவழி உள்ளீடு போன்றவற்றின் நுண்ணிய பார்வையில் தமிழ்மொழியில் உள்ள எழுத்துக்களின் சதவிகித பயன்பாட்டில் ஓர் மீள்பார்வை பயன்படும் எனும் கருத்தில் அவற்றை ஆய்வு செய்ய எண்ணம் பிறந்தது . அதான்று, ஒலிவழி உள்ளீட்டில் தந்தோதுள்ள தனிஎழுத்துப் பார்வை மட்டுமல்லால் எழுத்துக்களில் தன் இனமே இணைந்து வரும் நிலைகளையும் (தன்னிலை மெய்மயக்கம் க்க , ன்ன, ய்ய என வருவன) தன் துணையாகவே பலபோதும் வரும் துணைநிலை (ங்க, ஞ்ச, ண்ட, ந்த, ம்ப, ன்ற) எழுத்துக்களின் சதவிகித பயன்பாட்டின் ஆய்வு ஒலிவழி உள்ளீட்டிற்கு கைகொடுக்க மென்பொருள் வடிவமைப்பில் உதவ அதன் செயல்திறனில் துணையாகும் எனக் கருதப்பட்டது .

முதற்கண் சங்கநூல்களின் (8+10+18) எழுத்தெண்ணல் துவங்கப்பட்டு அவைகளின் உயிர், மெய், உயிர்மெய் அட்டவணை நிரல் செய்யப்பட்டமைக் காண்க . ஒரே பார்வையில் காண விகிதாசார தூண்- வரைபடமாகவவும் வைக்கப்பட்டுள்ளது

மெய்மயக்க எழுத்துக்களும் துணைநிலை எழுத்துக்களின் நிரல்கள் அன்றி சில சொற்களில் வரும் மெய்- மெய்- உயிர்மெய் (C, C, C/V) என சே>>ர்ந்த<<, பா>>ய்ந்தே<<, வா>>ழ்க்கை<< போன்று மூன்று எழுத்துக்களில் தொடரும் இடையின எழுத்துக்களின் பயன்பாடுகளையும் நிரல்படுத்த முயற்சிக்கள் மேற்கொள்ளப்படும் .

இவ்வாறு ஒன்றுக்கு மேற்பட்ட எழுத்துக்கள் இணைத்த கூட்டெழுத்துக்களுக்கு என புதிய CODE POINTS அமைத்தலும் உதவலாம் கூட்டெழுத்துக்கள் என்றால் தமிழில் இதுவரை இல்லாத மற்ற இந்திய மொழிகளில்

காணப்படும் பிணைந்து கீறும் காண்வழிக் கூட்டெழுத்து தோற்றம் அல்லவே அல்ல . கணினிக்குள் பயன்படும் நிலைகளில் மட்டும் . உத்தமம் வடிவமைத்து தற்போது யூனிகோடு நிரலில் தேர்வுநிலைகளுக்காகவே உள்ள PUA தளத்தில் தேவையான வைப்பிடமும் உள்ளன

எழுத்தெண்ணும் இம்முயற்சியில் பல காலகட்டத்தில் உருவாகிய தமிழ்நூல்களை நான்கைந்து காலநிலைப் படிவங்களாக கொள்ளும் நினைப்பில் தொடங்கிய போதிலும் சங்க நூல்களை மட்டும் இந்நாள் வரை எடுக்க முடந்தது. இம்முயற்சி தொடரும்.

உயிர்

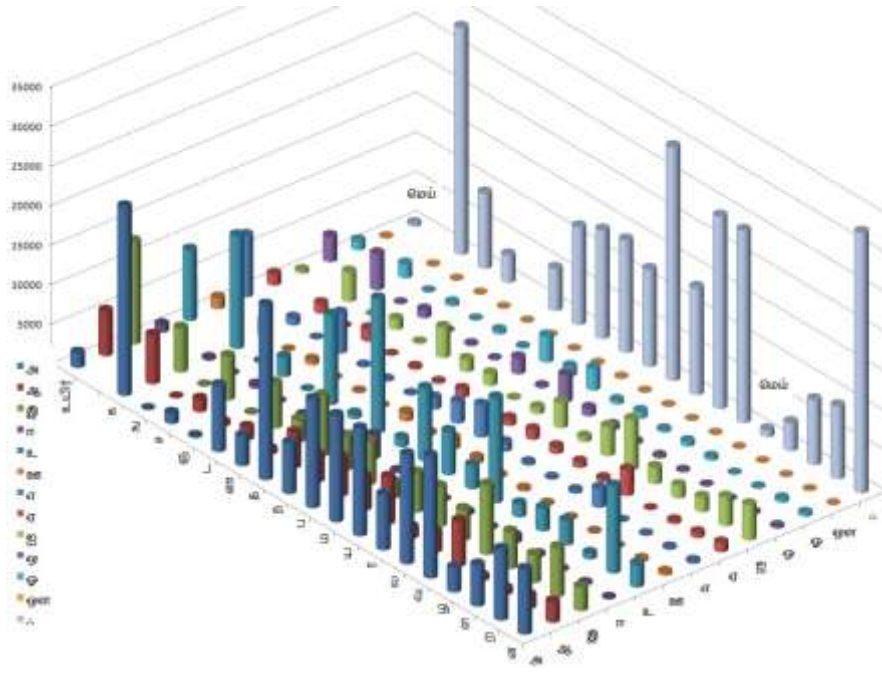
அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ	ஔ	ஃ
1823	5631	12943	948	8910	1343	7768	1381	359	3195	1148	49	348

மெய்

க	23882	6137	5551	226	14253	2149	863	1209	3727	4546	1750	34	28715
ங	10	0	0	0	0	0	0	0	0	0	0	0	9537
ச	1300	1603	5522	400	2420	590	5048	1578	1315	986	393	0	3472
ஞ	159	413	32	0	1	0	122	0	86	1	0	0	3454
ட	8356	1593	5711	103	11318	112	129	247	3821	134	428	0	5331
ண	3728	560	2797	107	567	21	96	116	1526	30	67	0	12355
த	22065	4468	7661	1017	16932	1026	1434	1013	1477	2032	3282	1	13721
ந	6385	3433	4317	2920	1055	216	2692	211	63	148	1174	0	14240
ப	13725	4979	5403	141	9169	1359	4090	870	761	3415	2760	23	12280
ம	13475	4272	2069	363	5484	634	1093	1228	3159	945	445	10	29571
ய	13556	5916	4991	47	2972	150	316	855	602	559	842	0	13556
ர	6982	1329	5029	257	13267	260	155	680	3677	215	309	0	24344
ல	13826	2551	3909	47	1630	116	347	379	6351	198	561	0	24327
வ	15403	5704	8789	810	2858	93	2403	3255	2083	113	387	0	869
ழ	3063	487	4639	263	3073	89	4	16	1548	19	10	0	3565
ள	5205	683	3505	72	705	39	96	356	1872	48	77	0	8219
ற	8982	1631	6189	172	11007	237	199	640	3634	150	444	0	9222
ன	8197	2588	2983	67	2885	369	279	1112	4452	207	396	0	32921

கீழ்க்காணும் உயிர் அளபெடைகள் உயிர் நிரலில் இணைக்கப்பட்டுள்ளன

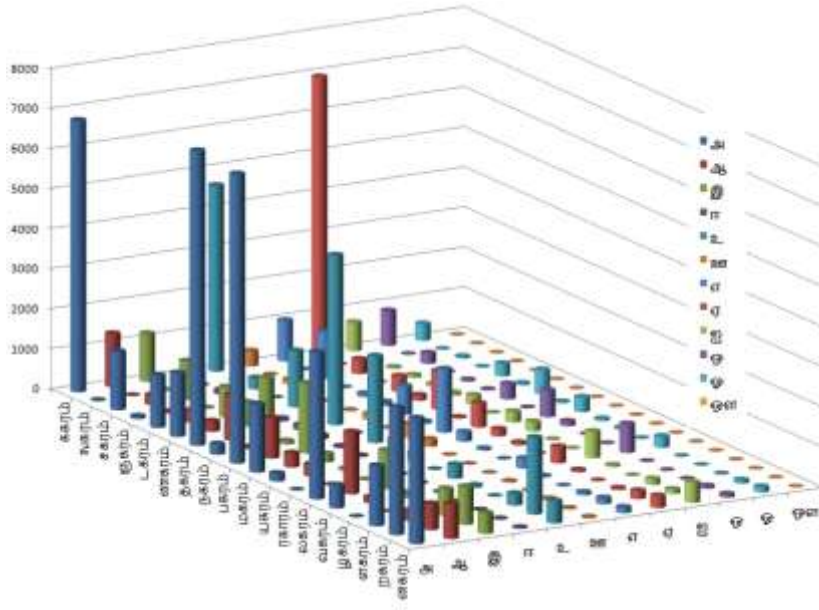
(1) அ/ஆஅ	(2) இ/ஈஇ	(3) உ/ஊஉ	(4) எ/ஏஎ	(5) ?-ஐஇ	(6) ?-ஓஔ	(7) ஔஉ
174/33	710/1	598/0	136/29	0	3	0



தன்னிலை மெய்மயக்கங்கள்

(எடுத்துக்காட்டு க்க, க்கா, க்கி >>>>>> க்கொ, க்கோ, க்கௌ)

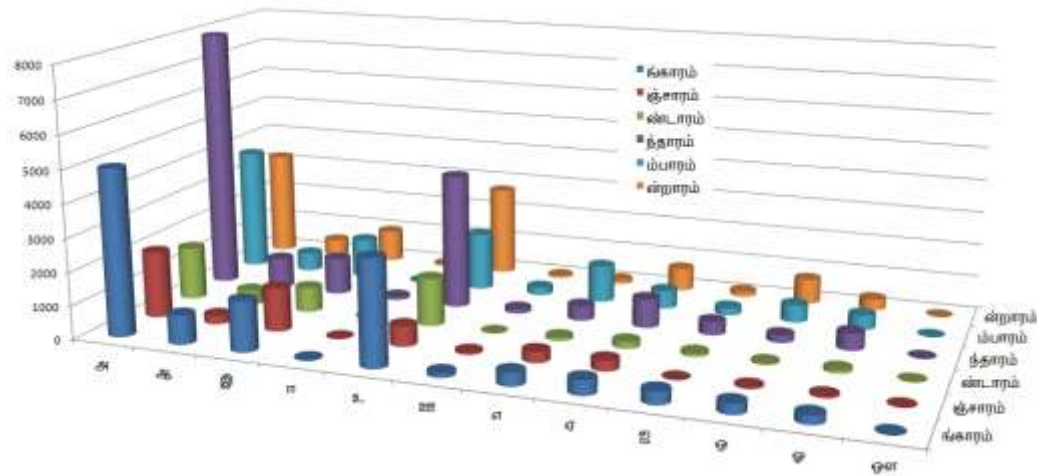
	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ	ஔ
ககரம்	6777	1325	1197	74	4639	379	1013	6967	706	871	414	0
ஙகரம்	10	0	0	0	0	0	0	0	0	0	0	0
சகரம்	1453	239	959	104	279	110	1171	360	33	258	46	0
ஞகரம்	60	56	4	0	0	0	0	0	59	0	0	0
டகரம்	1312	298	758	16	1400	13	78	385	54	35	309	0
ணகரம்	1587	256	597	64	359	9	98	134	68	15	44	0
தகரம்	7351	1126	1430	129	4233	141	707	953	225	379	602	0
நகரம்	259	65	77	90	18	11	50	27	0	3	27	0
பகரம்	7216	1008	1738	40	2160	342	1581	599	263	666	364	0
மகரம்	1702	351	200	28	320	178	230	179	232	115	49	0
யகரம்	196	290	55	1	193	2	42	63	9	9	47	0
ரகாரம்	0	0	0	0	0	0	0	0	0	0	0	0
லகரம்	3661	1546	953	13	345	1	240	409	659	719	277	0
வகரம்	576	140	127	3	55	25	24	48	40	1	1	0
ழகரம்	0	0	0	0	0	0	0	0	0	0	0	0
ளகரம்	1502	253	669	9	291	35	68	43	143	8	10	0
றகரம்	3186	652	945	11	1916	19	149	202	106	62	80	0
னகரம்	3134	869	512	22	555	21	141	297	507	81	137	0



துணைநிலை மெய்யக்கங்கள்

(ங்க, >>>>>>>> ங்கோ,)

	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ	ஔ
ங்காரம்	4995	873	1515	22	3168	130	399	438	377	313	252	0
ஞ்சாரம்	2004	261	1303	16	611	67	325	316	2	48	53	0
ண்டாரம்	1604	452	741	26	1447	9	104	204	99	58	105	0
ந்தாரம்	7917	920	1119	99	4079	138	448	869	401	224	514	0
ம்பாரம்	3752	567	1205	22	1756	238	1123	544	225	521	428	0
ன்றாரம்	3222	494	984	46	2718	65	128	691	156	751	331	0



தமிழ் மென்பொருள் உருவாக்கத்திற்கு அடிப்படை தேவையான தமிழ்த்தரவுகள்

முனைவர் அ.ஜேம்ஸ், அறிவுரைஞர்
முனைவர் ப.அர.நக்கீரன், இயக்குநர்
தமிழ் இணையக் கல்விக்கழகம், சென்னை

முன்னுரை

தமிழில் மென்பொருள் (software) நல்ல முறையில் உருவாக்கப்பட வேண்டும் என்ற எண்ணத்துடன் தமிழ் இணையக் கல்விக்கழகம் அடிப்படை தேவையான தமிழில் தரவுகள் இல்லாததை உணர்ந்து, ஒரு பிரத்தி யேக விரிதரவினைத் (corpus) தயாரித்து வருகின்றது என்பதையும் அதனை எவ்வாறு செயல்படுத்தி வருகின்றது என்பதையும் தெளிவாக விளக்குவது இக்கட்டுரையின் நோக்கம்.

நோக்கம்

இன்றைக்குத் தமிழில் எவ்வளவோ மென்பொருள்கள் பழக்கத்தில் இருந்து வருகின்றன. ஆனால், சில காரணங்களால் அவை முழுமையாகப் பயனாளர்களுக்கு உதவவில்லை என்பது திண்ணம். ஏனெனில் ஒரு மென்பொருள் தயாரிப்பவர் தமிழை உணர்ந்தவராக மட்டும் இருந்தால் அவர்களால், நல்ல தமிழ் மென்பொருளை உருவாக்க முடியாது. தகுந்த தமிழ்ப்புலமை இருந்தால் மட்டுமே தமிழ்மென்பொருளை உருவாக்க முடியும்.

ஒரு மென்பொருளை உருவாக்க வேண்டுமென்றால் அது யாருக்குப் பயன்படும் என்பதையும், அது எந்தவிதமானப் பயனைத் தரும் என்பதையும் கருத்தில் கொண்டு பின்பு உருவாக்குவது அவசியமானது. இன்றையளவில் இயல்மொழி செயல்பாடு NLP (Natural Language Processing) ஒன்று அவசியமாகிறது. அந்த இயல்மொழி செயல்பாட்டுக்கு எவ்வகையான தரவு அமைந்திருக்க வேண்டும் என்பது பற்றி ஆராய்ந்து பின் தரவினை தயாரித்து இயல்மொழி செயல்பாட்டுக்கு தேவையான கூறுகளை (module) உருவாக்க வேண்டும். இந்தத் தரவினை உருவாக்குவது என்பது அவ்வளவு எளிமையான ஒன்று அல்ல. ஏனெனில் இயல்மொழி செயல்பாடு என்பது மிகப் பெரிய கூறு. அதற்கு ஏற்றாற்போல் தரவினை அமைத்தல் அவசியம். இத்தரவினை அமைக்க பல்வேறுபட்ட கலைகள் அவசியமாகின்றன. அவற்றுள் தமிழ்ப்புலமை (scholarship on Tamil), மொழியியல் பேரறிவு (linguistics), அகராதிக்கலை நுட்பம் (lexicography), ஆங்கில இலக்கணம் (English grammar pattern) கணினி அறிவு (proficiency on computer) மிக இன்றியமையாததாகும். மேலேக் கூறப்பட்ட கலைகளில் பயிற்சிப் பெற்றவர்கள் ஒருங்கே த.இ.க.வில் அமைந்திருப்பதால் த.இ.க. இத்தகையத் தரவினைத் தயாரித்து வருகின்றது.

இத்தரவினைத் தயாரிக்க எண்ணி த.இ.க., கீழ்க்கண்டத் திட்டங்களை வகுத்து தமிழக அரசிடமிருந்து நிதிப்பெற்று செயல்பட்டு வருகிறது. அவையாவன

1. மொழியியல் அடிப்படையிலான இலக்கணக் குறிப்புடன் கூடிய விரிதரவு. (Annotated Corpus for Tamil literature)
2. தொடரணியியல் மற்றும் பொருண்மையியல் அடிப்படையிலான தமிழ் விரிதரவு. (Syntactically and Semantically Annotated Corpus for Tamil)
3. உச்சரிப்புடன் கூடிய தமிழ் மின்அகராதி (e-Dictionary for Tamil with pronunciation)
4. வாய்மொழித் தரவு (Speech Database)

திட்டம் - ஒன்று: மொழியியல் அடிப்படையிலான இலக்கணக் குறிப்புடன் கூடிய விரிதரவு

த.இ.க.வில் ஒரு சிறப்பான மின்நூலகம் (Digital Library) அமைந்திருக்கிறது. இம்மின்நூலகத்தில் சுமார் 400க்கும் மேற்பட்ட நூல்கள் மின்மயமாக்கப்பட்டு இணையதளப்படுத்தப்பட்டுள்ளன. இம்மின்நூலகத்தில் சங்ககால இலக்கியங்கள், சங்கம் மருவிய கால இலக்கியங்கள், காப்பியங்கள், பக்தி இலக்கியங்கள், நீதி நூல்கள், நெறி

நூல்கள், திரட்டு நூல்கள் , பழங்கால, இடைக்கால இலக்கண நூல்கள் , தற்கால இலக்கியங்கள் உள்ளன . இந்நூல்களுக்கு தேடுதல் வசதியும் (search facility) செய்து கொடுக்கப்பட்டுள்ளன .

இந்த மின்நூலகத்தில் அனைத்து நூல்களிலுள்ள சொற்களுக்கு மொழியியல் அடிப்படையிலான இலக்கணக் குறிப்புகள் கொடுக்கவேண்டும் என்று எண்ணி , த.இ.க. செயல்பட்டு வருகிறது . அதே சமயத்தில் மரபிலக்கணத்திலிருந்து (traditional grammar) வழுவாமலும் இருக்க உரிச்சொல் , உவமை உருபு போன்றவைகளையும் இட்டு விரிதரவு தயாரிக்கும் பணி நடைபெற்று வருகிறது . இவ்வாறு அறிவியல் முறைப்படி சரியான விரிதரவு இல்லாதக் காரணத்தால் நல்ல மென்பொருள் தமிழில் உருவாகாமல் இருக்கின்றது .

இன்றையளவில் தமிழில் மென்பொருள் தயாரிப்பவர்கள் தற்காலத் தமிழை மட்டும் எடுத்துக்கொண்டு அதற்கு பலவகையான வழிவகை வடிவமைத்து , கிடைக்கின்ற மென்பொருள்களை வைத்து தமிழை , ஒரு குடையின் கீழ் கொண்டு வந்துவிடலாம் என எண்ணி கற்பனையோடு மென்பொருள் தயாரிக்கிறார்கள் . இவ்வாறு தயாரிப்பதால் தான் சரியான மென்பொருள் நமக்கு இன்றைக்கு கிடைக்காமல் இருக்கின்றது . இந்நிலையில் தான் த .இ.க., பழங்கால இலக்கியங்கள் முதல் தற்கால இலக்கியங்கள் வரை யான நூல்களை எடுத்துக்கொண்டு அறிவியல் அடிப்படையில் கணினி படிக்கும் படியான (Machine readable text) தரவு தயாரிக்கப்பட்டு வருகிறது .

இந்த விரிதரவினைத் தயாரிக்கும் வேலையில் ஒரு தேடு பொறியும் (Search engine) தயாரித்து வருகிறது . இந்த தேடுபொறி நான்கு வகையான தேடுதல் வசதியையும் ஒரு வரைபடம் (Graph) கிடைக்கும்படியும் அமைக்கப்பட்டுள்ளது .

1. சொல் தேடல் (word search)

இங்கு 'மலர்' என்ற ஒரு சொல்லைத் தேடினால் 'மலர்' என்ற சொல்லுடன் அதனுடைய அடுக்கான (paradigm) மலர்கள், மலரின், மலரோடு, மலருக்காக, மலரால் போன்ற அத்துனைச் சொற்களையும் காட்டும் .

2. இலக்கணக் குறிப்புப் பெற (to get a grammatical category)

ஒரு சொல்லைக் கொடுத்து அதனின் இலக்கணக் குறிப்பை பெறலாம் . இல்லையெனில், ஒரு இலக்கணக் குறிப்பைத் தேர்ந்தெடுத்து அவ்விலக்கணக் குறிப்பின் தொடர்புடைய அத்துனைச் சொற்களையும் பெறலாம். எ.டு: 'வணங்கி' என்ற சொல்லுக்கு இலக்கணக் குறிப்புத் தேடினால் 'வினையெச்சம்' என்று காட்டும் . அல்லது இலக்கணக் குறியீட்டில் வினையெச்சச் சொற்களைத் தேடினால் தேவையான வினையெச்சச் சொற்களைத் தேடி கொணரும்.

3. சொல் சூழமைவு (Keyword in context)

இங்கு ஒரு சொல் எந்த சூழலில் அமைந்து வருகின்றது என்பதைச் சுட்டிக் காட்டுகிறது . எ.டு: 'அன்ன' என்ற சொல்லைத் தேடினால் அச்சொல் வரும் இடங்களைப் பட்டியலிட்டுக் காட்டும் .

ஓவத்து அன்ன வினை பனை நல் இல் – பதி. 61-03

பாவை அன்ன நல்லோல் கணவன் - பதி. 61-04

பொன்னின் அன்ன பூவின் சிறப்பிலை – பதி. 61-05

4. ஒப்பீடு (compare)

இங்கு ஒரே பொருளுடைய 'அடி, தாள்' என்ற இரு சொற்களைத் தேடினால் , அடி எந்த சூழலில் எந்த சொல்லோடு சேர்ந்து வருகிறது, தாள் எந்த சொற்களோடு வருகின்றது என்பதைத் தேடிப் பெறமுடியும் .

5. வரைபடம் (graph)

ஒரு சொல்லின் ஆட்சி ஒரு இலக்கியத்தில் எப்படி அமைந்துள்ளது என்பதை வரைபடம் மூலம் பெறலாம் .

திட்டம் - இரண்டு: தொடரணியியல் மற்றும் பொருண்மையியல் அடிப்படையிலான தமிழ் விரிதரவு

த.இ.க. தயாரிக்கும் மொழியியல் அடிப்படையிலான இலக்கணக் குறிப்புகள் மட்டும் இருந்தால் போதாது என எண்ணி, தொடரணியியல் மற்றும் பொருண்மையியல் அடிப்படையில் ஒரு விரிதரவு முக்கால் இலக்கியங்களுக்கும் தேடுபொறி வசதியுடன் தயாரிக்கப்பட்டு வருகிறது . இத்திட்டத்தை த .இ.க. இரண்டாகப் பிரித்து தொடரணியியல் மற்றும் பொருண்மையியல் எனப் பிரித்து விரிதரவு தயாரித்து வருகிறது .

இத்தரவினில் தொடரணியியல் படி , எழுவாய் (subject), பயனிலைகளைக் (predicate) கண்டறிய முடியும் . அதுமட்டும் அல்லாமல் பெயர்த் தொடர் (noun phrase), வினைத் தொடர் (verb phrase), பெயரடைத் தொடர் (adjective phrase), வினையடைத் தொடர் (adverb phrase), சொல்லுருபுத் தொடர் (postposition phrase), அளவடைத் தொடர்கள் (quantifier phrase), வினையெச்சத் தொடர் (verbal participle clause), நிபந்தனை எச்சத் தொடர் (conditional participle), குறையெச்சத் தொடர் (infinitive clause) இவைகளோடு ஒரு சொற்றொடர் (sentence) எங்கு ஆரம்பித்து எங்கு முடிகிறது என்பதையும் கிளைவடிவம் (tree diagram) மூலம் அறிந்து கொள்ளவும் முடியும் .

அவ்வாறே பொருண்மையியலிலும் உணர்பொருள் (connotation) குறிப்பொருள் (denotation), உவமை (simile), உருவகம் (metaphor) போன்றவைகளையும் கண்டறிய முடியும் .

திட்டம் - மூன்று: உச்சரிப்புடன் கூடிய தமிழ் மின்அகராதி

விரிதரவுகளைத் தயாரிக்கும் பணியோடு தமிழ் இலக்கியங்களில் உள்ள சொற்களுக்கு என்னென்னப் பொருள் இருக்கிறதோ அவற்றையும் உடனுக்குடன் அளித்து ஒரு தரவினையும் தயாரித்து வருகிறது . இத்திட்டத்திற்காக ஒரு தேடுபொறியையும் அமைத்து வருகிறது .

இத்தேடுபொறியின் மூலம் ஒரு சொல்லுக்கான பொருளை மிக எளிதில் கண்டறிய முடியும் . இதுநாள் வரையில் தமிழ் மொழிக்கு உருவாக்கப்பட்ட அகராதிகளில் ஒலியன் வடிவிலுள்ள வரிவடிவம் (phonemic transcription) மட்டும் தான் இருந்துவந்தது . ஆனால், த.இ.க. முதன்முதலில் ஒலியியல் அடிப்படையில் அதாவது phonetic transcription - narrow transcription முறையில் ஒரு அகராதியைத் தயாரித்து வருகிறது . இவ்வகராதியில் தலைச்சொல்லுடன் (headword) ஒரு சொல் எவ்வாறு உச்சரிக்கவேண்டும் (pronunciation) என்பதற்காக phonetic script, phonemic script தமிழில் பொருள் (Tamil meaning), ஆங்கிலப் பொருள் (English meaning) எந்த இலக்கியத்தில் அச்சொல் எவ்வாறு பயன்படுத்தப்பட்டு வருகின்றது (citation / usage), போன்றவற்றுடன் வட்டார வழக்கு (dialect variation) மற்றும் தமிழ் பண்பாட்டு சொற்களையும் (cultural terms) படங்களுடன் தயாரித்து வருகின்றது .

திட்டம் - நான்கு: வாய்மொழித் தரவு

த.இ.க. மின்நூலகத்தில் சுமார் ஒரு லட்சம் பக்கங்கள் கொண்டத் தரவுகள் இருக்கின்றன . இவற்றுள் B.A. தமிழியியலுக்கான 24 தாள்களுக்கான 576 பாடங்களும் அடங்கும் . த.இ.க. மின்நூலகத்திலுள்ள சுமார் 400 இலக்கண இலக்கியங்கள் மற்றும் 576 பாடங்களுக்கு மொழியியல் அடிப்படையிலான இலக்கணக் குறிப்புடன் கூடிய விரிதரவு தயாரிப்பது போன்று , 576 பாடங்களுக்கும் மின் விரிவுரை (e-lecture) தயாராகி வருகின்றது . இந்த மின் விரிவுரையை ஒலிவடிவத்திலிருந்து வரிவடிவமாக மாற்ற வேண்டும் என்று எண்ணி , Speech to text க்கிற்காக ஒரு மென்பொருளை உருவாக்கி வருகின்றது . இத்திட்டத்தை த .இ.க. இரு வழிகளில் உருவாக்கி வருகின்றது . முதலாவது DRAGON என்ற மென்பொருள் மூலம் மின்விரிவுரைகளை வரிவடிவமாக மாற்றுவது , இரண்டாவது யார் பேசினாலும் வரிவடிவமாக மாற்றுவது போன்ற பணிகளில் ஈடுபட்டுவருகிறது .

முடிவுரை

இத்தகைய தரவுகள் ஒருங்கிணைந்து ஒரு விரிதரவாக த .இ.க.வால் தயாரிக்கப்பட்டப் பின் அது பலவிதங்களில் பயன்பாட்டினைத் தரும். சான்றாக

- மரபிலக்கணக்காரர்களுக்கும், மொழியியலாரருக்கும் உள்ள நீண்டநாள் பிரிவினை சரிசெய்யப்படும் .
- வரலாறு அடிப்படையிலான ஒரு தமிழ் இலக்கணத்தை எழுத முடியும்.
- பல ஆராய்ச்சிகள் மரபிலக்கண வழியாகவும், மொழியியல் அடிப்படையிலும் ஆராய்ந்திட முடியும்.
- கணக்கற்ற அகராதிக ளை ஒரே மொழியிலும் (mono-lingual), இரு மொழியிலும் (bi-lingual) தயாரிக்க முடியும்.
- எல்லாவற்றுக்கும் மேலாக தமிழில் NLPக்கு வேண்டுமளவிற்கு பல கூறுகளை நல்ல முறையில் உருவாக்க முடியும்.

தரவகம் (Corpus)

முனைவர் இரா. பன்னிருகைவடிவேலன்

சொல்தொகுப்பாளர், தமிழ்ப் பேரகராதித் திருத்தப்பணித்திட்டம், சென்னைப் பல்கலைக்கழகம்
pann1973@gmail.com

கணினித் தொழில்நுட்பமும் வலைத்தளத் தொழில்நுட்பமும் அகர வேகத்தில் வளர்ந்துவருகின்றன . கணினி குக்கிராமங்களிலிருந்து பெருநகரம் வரை பரவிக்கிடக்கிறது . இதன் விளைவாகக் கணினி அனைத்துநிலைகளிலும் அனைத்துத்துறைகளிலும் பயன்படுத்தப்பட்டு வருகிறது . பல்கலைக்கழங்கள் தொடங்கிச் சிறு பள்ளிகள் கணினியின் பயன்பாடு மிகுந்துள்ளது . இந்தச் சூழலில் கல்வி கற்கவும் கற்பிக்கவும் கணினியையும் கணினி சார்ந்த தரவகங்களையும் பயன்படுத்திக்கொள்வது தேவையாகிறது . தமிழில் தரவக மொழியியல் என்பது வளர்ந்துவரும் ஓர் ஆய்வுப் பிரிவாக உள்ளது . இந்நிலையில் இக்கட்டுரை தரவகம் குறித்து விளக்குகிறது .

தரவகம்

கணினியில் பயன்படுத்தும் வகையில் முறையாகச் சேகரித்து வைக்கப்பட்டுள்ள ஒரு மொழியில் அல்லது பல மொழிகளில் அமைந்த பல்வேறு பனுவல்களின் (texts) தொகுப்பு தரவகம் எனப்படும் . இது ஒரு மொழியின் பல்வேறு பரிணாமங்களை எதிரொலிப்பதாகவும் பலதரப்பட்ட புலங்களுக்கு (Fields) முதன்மை அளிப்பதாகவும் அறிவியல் முறைப்படியும் இருக்க வேண்டும் .

தரவகத்தில் இடம்பெறும் பனுவல்கள் முழுமையாகவோ , பனுவல்களின் சிற்சில பக்கங்களாகவோ , தொடர்ச்சியின்றி ஆங்காங்கு எடுக்கப்பட்ட வாக்கியங்களாகவோ இருக்கலாம் . தரவகம் எழுத்துமொழிக்கென்றும் பேச்சுமொழிக்கென்றும் தனித்தனியாகவோ இரண்டும் சேர்ந்ததாகவோ உருவாக்கப்படலாம் . முதலாவதாக, தரவகம் மொழிப் பயன்பாட்டின் அதிகாரபூர்வமான தரவுகளை அளிக்கிறது . மொழியியல் மற்றும் இயற்கை மொழி ஆய்வு ஆராய்ச்சியாளர்களுக்குத் தேவையான குறிப்பிட்ட வகைத் தரவுகளைச் சில வரிசை ஒழுங்குடன் கணினி அறியும் வகையில் சேகரித்து வைத்திருப்பதால் எந்தக் குறிப்பிட்ட தகவல்களையும் எளிதில் தேர்வுசெய்து பெறமுடியும் . அதாவது, கணினி நிரல்கள் மூலம் தரவுகளைத் தேவையான தகவல்களாக மாற்றிப் பயன்படுத்தவும் முடிகிறது . இச்செயல் வல்லமைகளால் தரவகம் மொழியியல் ஆய் விற்கும் அவ்வாய்வுகளைக் கொண்டு மென்பொருள் தயாரிப்பதற்கும் அகராதி ஆக்கத்திற்கும் , மொழிகற்கவும் கற்பிக்கவும் ஒரு சக்திவாய்ந்த மூல ஆதாரமாக அமைகின்றது .

தொகுக்கப்படும் பனுவல்கள் பின்வருமாறு தெரிவுசெய்யப்படுகின்றன .

1. கலை, மொழி, இலக்கியம், அறிவியல், தொழில்நுட்பம், மருத்துவம் போன்ற பலதரப்பட்ட பொருண்மைகளில் அமைந்த பனுவல்கள் .
2. பல்வேறு ஆசிரியர்/படைப்பாளிகளது படைப்புக்கள் .
3. பல்வேறு வயதினருக்காக உருவாக்கப்பட்ட பனுவல்கள் .
4. பல்வேறு படிப்புக்கள், பயிற்சிகளுக்கென உருவாக்கப்பட்ட பாடங்கள் .
5. முறை சார்ந்த மற்றும் முறைசாரா உரைகள் .
6. செய்தித்தாள்கள், இதழ்கள் போன்று பொதுமக்களுக்காக உருவாக்கப்பட்ட உரைகள் .

இவை போன்று பல்வகை உரைகளைத் தேர்வுசெய்து முறைப்படி தொகுக்க வேண்டும் . அகராதியைப் போன்றே தரவகம் ஒருமொழித் தரவகம் , பன்மொழித் தரவகம் என வகைப்படுத்தப்படுகின்றன . சில நேரங்களில் ஒரு மொழிக்கெனத் தரவகம் உருவாக்கும்பொழுது அப்பனுவல்களைப் பிறமொழிகளில் மொழிமாற்றம் செய்து ஒருங்கிணைந்த பன்மொழித் தரவகம் உருப்படுத்தப்படுகிறது .

தரவுக்கள உருவாக்க முயற்சிகள்

தமிழ் இலக்கியங்களைப் பலரும் அறியச்செய்யும் விதமாகப் பல இணையதளங்கள் உருவாக்கப்பட்டுள்ளன . சான்றாக, சுவிட்சர்லாந்தைச் சேர்ந்த கல்யாணசுந்தரம் என்பவர் மதுரைத்திட்டம் (Project Madurai) என்ற ஒன்றை உருவாக்கியுள்ளார். மேலும் இவரது ஆதரவுடன் தமிழ் மரபு அறக்கட்டளை (Tamil Heritage Foundation) என்ற அமைப்பும் இயங்கிவருகிறது . இதன் உருவாக்கத்திற்குக் காரணமானவர்கள் ஜெர்மனியைச் சேர்ந்த கணினித்தமிழ்ப் பொறிஞர்களான கண்ணன் மற்றும் சுபாஷினி ஆவர் . தமிழ் ஓலைச்சுவடிகள் இந்த இணையத்தில் (Tamil Heritage Foundation) பாதுகாக்கப்படுகின்றன. ஜெர்மனியிலுள்ள பெர்ன் நூலகத்தில் மட்டும் 1000 தமிழ் ஓலைச்சுவடிகள் உள்ளன . இதைப்போன்று பிரிட்டிஷ் மற்றும் ஆம்ஸ்டர்டாம் நூலகங்களிலுள்ள தமிழ் ஓலைச் சுவடிகள் மற்றும் தமிழகத்திலுள்ள ஓலைச்சுவடிகளின் பிரதிகளை இணையத்தில் இணைக்கும் முயற்சியும் நடைபெற்று வருகிறது . மேலும் இதனுள் வில்லுப்பாட்டு , கரகாட்டம் போன்ற தமிழர்களின் மரபார்ந்த கலாசார அடையாளங்களைப் பல்லாடக முறையில் பதிவு செய்து இணையத்தில் உலா விடுவது ஆகிய பணிகளை இந்த அறக்கட்டளை செய்து வருகிறது. (<http://ccat.sas.upenn.edu/plc/tamilweb/tagger.html>).

தொடக்க காலத்தில் இருந்துவந்த மொழியியலும் அடிப்படையில் தரவக மொழியியலே (Corpus Linguistics) ஆகும். இருப்பினும், இன்றைய கணினித் தொழில்நுட்பத்தின் வல்லமையால் பெரிதும் விரிவுபட்ட தரவுமூலத்தொகுப்பு மொழியியல் பல்வேறு வகையில் வேறுபட்டுப் பல புதிய சிந்தனைகளையும் முடிவுகளையும் கொடுக்கவல்லதாக வளர்ந்துள்ளது. பொதுவாக, இப்போது பேசப்படும் தரவக மொழியியல் , விளக்க மொழியியலிலிருந்து (Descriptive Linguistics) மிகவும் வேறுபட்ட குறிக்கோளைக் கொண்டது .

தரவக மொழியியல்: 1980 - களுக்கு முன்

இயற்கை மொழி என்பது ஒரு வரையறைக்கு அப்பாற்பட்டது . தரவக மொழியியல் களஆய்வில் சேகரித்த சொற்களையும், தொடர்களையும், சொற்றொடர்களையும் அடிப்படையாகக் கொண்டு மொழியை விளக்க முற்பட்டது. இதில், குறிப்பாக ஒரு மொழியில் உள்ள தொடர்கள் ஒரு குறிப்பிட்ட எண்ணிக்கைக்குள் அடங்கக் கூடியன. அவற்றைச் சேகரித்து வரிசைப்படுத்த இயலும் என்ற நம்பிக்கையைக் கொண்டிருந்தது . மொழியியல் வளர்ச்சியில் சாம்ஸ்கியின் அணுகுமுறை புதிய திருப்பத்தை ஏற்படுத்தியது. அவரது அறிதிறன் மற்றும் செய்திறன் (Competence Performance) எனும் கொள்கை மேலோங்கியது . அறிதிறன் என்பது ஒருவரது மொழிஅறிவை வகைப்படுத்தி விளக்கவல்ல நிலையைக் குறிக்கும் . அவரது மனத்தளவில் உள்ளடக்கிய மொழியறிவைக் குறிக்கும். செய்திறன் என்பது ஒருவரது மொழியறிவை அவர் உருவாக்கிய தொடர்களைக் கொண்டு அளவிடுவது . அவர் உருவாக்கிய கோவை அவரது இலக்கண அறிவை மட்டும் பிரதிபலிப்பதல்ல ; அது அச்சுமூலக்கு ஏற்புடையது என்பதையும் விளக்கவல்லது . இருப்பினும் மொழியின் செய்திறன் என்பது உள்ளார்ந்த மொழியின் அறிதிறன் அறிவின் முழுமையையும் வெளிப்படுத்துவது அல்ல.

தரவுமூலத்தொகுப்பு மொழியியலாளர்கள் வெளிப்பாட்டு நிலைத் தரவுகளையே ஆய்வுக்கு எடுத்துக் கொண்டனர் . மொழியின் அறிதிறனுக்கான மாதிரியை வரையறுப்பதே ஒரு மொழியியலாளரின் முதற்பணி ஆகும் என்றார் . அப்படியென்றால் உள்ளார்ந்த மொழியின் அறிதிறனுக்கு மொழியின் செய்திறனுக்கும் உள்ள முக்கியமான வேறுபாடு, மொழிக்கு அப்பாற்பட்ட கருத்துக்களை உள்ளடக்கிய அனுமானமே ஆகும் . அளவுசார் தரவு (quantitative data) என்பது அளவு, பருமன், பயிலெண் மற்றும் மொழி அம்சம் ஆகிய தகவல்களைக் கொண்டதாகும். தரவகம் இத்தகைய அளவுசார் தரவுகளைக் கொடுக்கவல்ல சிறந்த மூலமாகும் . ஆனால் மொழி பேசும் மனிதர்கள் இவற்றைக் கொடுக்க இயலாது . ஸெனான் மற்றும் வீவர் (Shannon and Weaver: 1949) அளவுசார் தரவுகளை மூலமாகக் கொண்டு பல்வேறு அடிப்படை அளவுசார் ஆய்வுக் கோட்பாடுகளை அமைத்துப் புள்ளியியல் வாயிலாக மொழியை விவரிக்க முயன்றுள்ளனர் . இருப்பினும் சாம்ஸ்கி (1965) போன்ற அறிஞர்கள் அளவுசார் தரவுகள் தற்செயலாக, எதேச்சையாக அமையக் கூடியன; அவற்றைக் கொண்டு மொழிப் பயன்பாட்டின் உண்மை நிலையை அறிய இயலாது என நேரடியாகவே மறுத்துள்ளார்கள் .

அறிதிறன் நிலையில் பெரிதும் ஏற்றுக்கொள்ளப்படுகிற வாய்க்கிய வகைகள் செய்திறன் நிலையில் தோன்றாமலிருந்தால் அதன் பயிலெண் பூஜ்யமாகவே அமையும் என்பதாலும் அளவுசார் தரவு ஆய்வு முழுமையானது அல்ல என்பதாலும் 1980-களில் இவ்வாய்வுமுறை கைவிடப்பட்டது .

நவீன தரவகத்தின் பிறப்பு

ஆங்கில மொழியின் தரவக மொழியியல் 1960 - களின் தொடக்கத்திலேயே தொடங்கப்பட்டது. குயிர்க் (Quirk:1960) தனது ஆர்வம் மிகுந்த ஆங்கில மொழிப் பயன்பாட்டுக் கணிப்பீட்டை (Survey of English Usage) 1961 - இல் தொடங்கினார். அதே காலகட்டத்தில் ப்ரான்சிஸ் (Francis) மற்றும் குசேரா (Kucera) ஆங்கில மொழிக்கான தரவகம் அமைப்பதற்கான பணியில் இறங்கினர். தற்பொழுது பெரிதும் புகழடைந்துள்ள 'பிரவுன் கார்பஸ்' (Brown Corpus) இவருடைய முயற்சியால் தொடங்கப்பட்டதே ஆகும். இப்பணி ஏறத்தாழ 20 ஆண்டுகள் நடைபெற்றது. ஜான் ஸ்வர்த்விக் (Jan Svartik) என்பவர் ஆங்கிலப் பயன்பாட்டுக் கணிப்பீட்டை (SEV) பிரவுன் தரவகத்தில் (Brow Corpus) தொடர்ந்து பணிபுரிந்து இலண்டன் - இலண்ட் (London -Lund Corpus) தரவகத் தொகுப்பை உருவாக்கினார்.

ஆங்கிலப் பயன்பாட்டுக் கணிப்பீட்டில் தொடக்க காலத்தில் பணி புரிந்த ஜீயோஃப்ரே லீச் (Jeoffrey Leech) என்பவர் இலங்காஸ்டரில் ஒரு தரவக மையத்தை நிறுவினார். இம்மையமே பிற்காலத்தில் தோன்றிய பல்வேறு தரவக உருவாக்கத் திட்டங்களுக்கு மூலமாக அமைந்தது. தற்போது மிகவும் பிரசித்தி பெற்ற பிரிட்டிஷ் தேசியத் தரவகத்தின் (British National Corpus) மூலமான இலங்காஸ்டர் - ஓஸ்லோ - பெர்கன் (Lancaster - Oslo - Pergen, LOB) அத்தகைய பெருந்தரவகங்களுள் ஒன்றாகும். பிரிட்டிஷ் தேசியத் தரவகம் இப்போது 200 மில்லியன் சொற்களுக்கு மேல் கொண்டுள்ளது. இத்தகைய வளர்ச்சியின் மூலகாரணம் கணினியின் வல்லமையே ஆகும். தற்சமயம் உலக மொழிகளில் பெரும்பாலானவை எழுத்து மொழிக்கென்றும் பேச்சு மொழிக்கென்றும் தரவகங்களை உருவாக்கி மொழி ஆய்வில், குறிப்பாக அளவுசார் ஆய்வில் ஒரு புதிய திருப்பத்தையே ஏற்படுத்தியுள்ளன.

இன்றைய தரவக மொழியியல், செய்திறன் மொழி நிலையின் (Linguistic Performance) மீது கருத்தைச் செலுத்துகிறது. அறிதிறன் மொழிநிலை (Linguistic Competence) யின் மேல் ஆய்வு மேற்கொள்வது நடைமுறையில் இயலாத ஒன்று. மேலும் மொழியின் செய்திறன் மற்றவர்களின் கவனத்திற்கு உட்பட்டது. தரவக மொழியியல் ஓர் அளவுசார் ஆய்வு முறையாகும். தரவகத்தைக் கொண்டு பல்வேறு வகையான பயில்வெண்களை உருவாக்கி ஆய்வு செய்யப்படுகிறது. அளவுசார் ஆய்வுமுறைக்கு அடிப்படைத் தேவை தரவு. அதை ஒரு முறையான தரவகம் கொடுக்கின்றது. இத்தரவக மொழியியல் ஆய்வுமுறைவழி வெளிப்படும் முடிவுகள் அனைத்தும் விளக்க மொழியியல் போன்ற மற்ற மொழி ஆய்வுகளிலிருந்து வேறுபட்டவை ஆகும். இது சமுதாய மொழியியல், உளவியியல் மொழியியல், ஒலியியல், தொடரியல் போன்ற பிரிவுகளுடன் ஒப்பானது அன்று. இது ஒரு ஆய்வு முறை; மொழியியல் கூறுகள் சிலவற்றை ஆய்வதற்கான அணுகுமுறையை விளக்குவதே ஆகும். இது கணினி மொழியியலுடன் நெருங்கிய தொடர்பு கொண்டதாக அமைகிறது. தரவகம் என்றாலே அது ஒரு கணினி அறியும் (Machine Readable) வகையிலோ அல்லது பயன்படுத்தும் வகையிலோ அமைந்ததாகும்.

தரவகத்தின் பயன்கள்

தரவகங்கள் மொழிக் கொள்கைகளைச் சோதனை செய்வதற்கும், மொழியில் புதிய விதிகளை உருவாக்குவதற்கும், பழைய விதிகளை மதிப்பீடு செய்வதற்கும், புள்ளிவிவரப் பகுப்பாய்வு செய்வதற்கும், மொழியைக் கற்கவும் கற்பிக்கவும் அகராதி போன்ற மொழிக்கருவிகளை உருவாக்குவதற்கும் உறுதுணையாக அமையும்.

சொல்லின் இலக்கணக் கூறுகளை அடையாளப்படுத்தும் Parse Tagging, அடிச்சொல்லை (Base) வெளிப்படுத்தும் Tags, சொற்றொடர் மற்றும் வாக்கியத்தின் கட்டமைப்பை வெளிப்படுத்தும் Structural Level Tags ஆகியன தரவகத்தின் தரத்தை மேம்படுத்திக் காட்டுகின்றன.

மொழிவல்லுநர்கள் மட்டுமன்றிக் கணினிநிரலாளர்கள் மொழிப்பயன் பாடுகளுக்கென உருவாக்கும் மென்பொருளின் வடிவமைப்பிற்கும் உதவுகிறது. தகவல் தொழில் நுட்பத்தில் உலகமொழிகளைக் கையாள்வதற்கு வடிவமைக்கப்பட்டுள்ள மென்பொருள்கள் போன்று அந்தந்த மொழிகளுக்கான மென்பொருள்களை வடிவமைக்க இக்கருவிகள் செயற்படுத்தப்படுகின்றன.

ஒரு மொழியில் உருவாக்கப்படும் மென்பொருள் எந்த அளவுக்குப் பயன்படும் என்பதைச் சோதித்தறிய இயலும். மொழிக்கான மென்பொருளின் செயல், பயன்பாட்டுத் திறன் ஆகியவற்றின் துல்லியத்தைச் சோதனையிடத் தரவகம் ஒரு ஆய்வுத்தரவகமாகவும் சோதனைக்களமாகவும் விளங்கும்.

தமிழில் தரவகம்

தமிழில் தரவக உருவாக்கப் பணி 1987 - இல் தொடங்கியது . மைசூரிலுள்ள இந்திய மொழிகளின் நடுவண் நிறுவனம் ஜப்பானிலுள்ள அயல்நாட்டு ஆய்வுகளுக்குரிய டோக்கி யோ பல்கலைக்கழகத்துடன் (Tokyo University Of Foreign Studies) இணைந்து பள்ளிப் பாட நூல்கள் அனைத்தையும் கணினியில் ஏற்றி முதன் முறையாகத் தரவகத்தைத் தமிழில் உருவாக்கியது . இதனைத் தொடர்ந்து 1991 - இல் இந்திய அரசின் மின்னணுத் துறையில் தொடங்கப்பட்ட இந்திய மொழிகளின் தொழில்நுட்ப வளர்ச்சி (Technology Development in Indian Languages, TDIL) எனும் திட்டத்தின்கீழ் இந்திய மொழிகளின் தரவக உருவாக்கம் என்னும் ஆய்வுத்திட்டம் 18 அட்டவணை மொழிகளுக்கும் தொடங்கப்பட்டது . தமிழ் மொழிக்கான தரவக உருவாக்கம் இந்திய மொழிகளின் நடுவண் நிறுவனத்தால் மேற்கொள்ளப்பட்டது . சொல் நிலையில் இலக்கணக் குறியீடுகளைத் (Grammatical Tagging) தொகுப்பில் இணைக்கத் தேவையான மென்பொருள் உருவாக்கம் , பயில்வண் கண்டறிதல் , சூழலுடன் தொடரடைவு (Key Word In Context (KWIC) Concordance) உருவாக்கத் தேவையான மென்பொருள் சாதனங்கள் தயாரிப்பு, தரவக மேலாண்மை மென்பொருள் (Corpus Manager) தயாரிப்பு ஆகியவை இவ்வாய்வுத்திட்டத்தின் முக்கிய குறிக்கோள்களாக இருந்தன . இவ்வாய்வுத் திட்டம் 1995 -இல் முடிவடைந்தது . தமிழில் சுமார் 36 இலட்ச அளவில் சொற்களைக் கொண்ட தரவுகள் இத்திட்டத்தில் தொகுத்து வைக்கப்பட்டு உள்ளன .

1993 - இல் எரிக் பெடர்சன் (Eric Pederson) என்பவரின் பேச்சுத்தமிழ் ஆய்வுக்கான தரவுகளைப் படியெடுத்துக் கணினியில் ஏற்றிப் பேச்சுத் தமிழுக்கான தரவகம் ஒன்றையும் இந்திய மொழிகளின் நடுவண் நிறுவனம் உருவாக்கியது .

சென்னையில் 'க்ரியா' என்ற அமைப்பும் 'மொழி' என்ற அமைப்பும் எழுத்துத் தமிழுக்கான ஒரு தொகுப்பையும் உருவாக்கி அதனை அகராதித் தயாரிப்பில் பயன்படுத்தியுள்ளன . தொடரடைவிகளிலிருந்து பெறப்படும் தரவுகள் குறிப்பிட்ட சொற்கள் இடம்பெறும் தொடர்களைக் கொண்டிருக்கும் . இந்தப் பட்டியல் தரவுத்தளத் தொடரடைவிகள் உருவாக்கியதாகவோ அல்லது வேறு வழியில் பெறப்பட்டதாகவோ இருக்கலாம் . இவ்வாறு கணினியால் கையாளப்படும் வகையில் சொற்களின் தரவுத்தளங்களைக் கொண்டிருப்பவைதாம் தரவகங்கள் (Corpus) எனப்படுகின்றன . இவ்வகைத் தரவகங்கள் சிறப்பாக உள்ள ஒரு மொழியில் சிறந்த கணினிசார் மொழிக் கருவிகளை உருவாக்கமுடியும் . இந்தத் தரவகங்கள் ஒரு மொழியில் ஏற்படும் மாற்றங்களுக்கு ஏற்ப அடிக்கடி மேம்படுத்தப்படவும் வேண்டும் . இவ்வகையான தரவகங்கள் கணினிசார் மொழிப் பயன்பாட்டிற்கு மிகவும் அடிப்படையாகும் .

“க்ரியாவின் சொல்வங்கியிலிருந்து தற்காலத்தமிழில் வழங்கும் சொற்கள் , அவை பயன்படுத்தப்படும் வாக்கியங்களுடன் பெறமுடியும் . நாளேடுகள், புனைகதை, அறிவியல், குழந்தை இலக்கியம் என்று பல துறைகளிலிருந்து பிரதிகள் சொல்வங்கியில் இடம்பெற்றிருக்கின்றன . பல வகையான பிரதிகள் தொடர்ந்து சொல்வங்கியில் சேர்க்கப்படுகின்றன.” (www.crea.in, corpus, html)

செம்மொழித் தமிழாய்வு மத்திய நிறுவனத்தின் முதன்மைத் திட்டப்பணிகளுள் ஒன்றாக , “பழந்தமிழ் நூல்களுக்கான தரவகம் ” என்னும் ஒரு திட்டம் செயல்பட்டுவருகிறது . இவற்றிற்கான பணிகள் நடைபெற்றுவருகின்றன . இப்பணி வெகுவிரைவில் சிறப்பாக உருவாகித் தமிழுக்கும் தமிழ் ஆய்வுலகி ற்கும் பெரிதும் பயன்படும் என எதிர்பார்க்கலாம் . (www.cict.in) .

முடிவுரை

தரவகம், தரவகத்தின் பயன்கள் , தரவுக்களம் உருவாக்குதல் , தரவக மொழியியல் , நவீன தரவகத்தின் பிறப்பு , தமிழில் தரவகம் ஆகியன விளக்கப்பட்டுள்ளன . தமிழில் உருவாக்கப்பட்டுள்ள தரவகங்கள் முழுமையான தரவுகளைக் கொண்டதாக அமையவில்லை . எனவே தமிழில் சிறந்த தரவக உருவாக்க முயற்சியில் பல்கலைக்கழகங்கள் ஈடுபட முன்வர வேண்டும் .

References

- <http://icame.uib.no/brown/bcm.html>
- http://www.chs.nihon-u.ac.jp/eng_dpt/tukamoto/kwic_e.html
- <http://logos.uoregon.edu/faculty/pederson/index.html>
- <http://ildc.in/Tamil/Tindex.aspx>
- <http://benjamins.com/#catalog/books/tufs.3/main>
- <http://www.aber.ac.uk/media/Documents/short/trans.html>
- http://www.euralex.org/elx_proceedings/Euralex1992_1/007_Jan%20Svartvik%20Lexis%20in%20English%20language%20corpora.pdf
- http://en.wikipedia.org/wiki/Corpus_linguistics

தமிழில் ஓலைச்சுவடி தரவுத் தளம்

கோவி. சக்கரபாணி

தொழில்நுட்ப இயக்குநர், தேசிய தகவலியல் மையம், தலைமைச்செயலகம், சென்னை

கல்தோன்றி மண்தோன்றா காலத்தே முன் தோன்றிய மூத்த குடி எனும் பெருமை வாய்ந்த தமிழினமாயினும் சரி அல்லது உலகில் தோன்றிய வேறு எந்த மனித இனமாயினும் சரி தங்களின் உணர்வை அல்லது தேவையை சைகை கொண்டு தெரிவித்த காலம் தொட்டு படிப்படியாக வளர்ந்து மொழிகள் தோன்றின . அம்மொழியின் மூலம் ஒருவருக்கு இருக்கும் பரந்த அறிவை பிறருக்கு தெரிவித்து பயன்படுத்துவதற்காக பதிவு செய்ய வேண்டியிருந்தது .அதற்காக ஆரம்பத்தில் கோடுகள், வரைவுகள், குறியீடுகள், ஆகியவை தோன்றி அவை எழுத்துக்களாக உருமாறின. பனையோலையில் எழுதுவதற்கு எளிமையாக இருப்பதாலும் அவற்றை சரியான முறையில் பராமரித்து வந்தால் நீண்ட நாட்கள் அழியாமல் இருக்கும் தன்மை கொண்டதாலும் பனையோலைகள் பயன்படுத்தப்பட்டு வந்தன .

பழங்கால ஓலைச்சுவடிகளில் பொதிந்துள்ள வரலாற்று உண்மைகளான தமிழர்களின் வீர வரலாறு , கலை, கலாச்சாரம், பண்பாடு போன்ற கூறுகளை இன்றைய சந்ததியினர் உணர்ந்து கொள்வது அவசியம் . அக்கால ஓலைச் சுவடிகளை ஆன்றோர் சிலர் பாதுகாத்து வந்ததினால் அவைகள் பிற்காலத்தில் சுவடி படிப்போரால் படிக்கக் கேட்டு அச்சு பதிப்பைப் பெற்றன . ஓலைச் சுவடிகளை பாதுகாத்தலின் மூலம் எதிர்கால சந்ததியினருக்கு மொழி, கலை, பண்பாடு, அக்காலகட்டத்தின் சமூக நிலைமை ஆகியன குறித்த அரிய உண்மைகள் கிடைக்கச் செய்வதற்கு வாய்ப்பாக அமைகிறது . எனவே ஓலைச் சுவடிகளை பேணி பராமரித்தல் மிகவும் அவசியம் .

நம்முடைய கலாச்சார தொன்மையின் மதிப்புமிக்க சொத்தாக சுவடிகள் கருதப்படுவது மாத்திரமல்ல சரித்திர தொன்மையையும், அறிவையும் இந்த சமுதாயத்திற்கு வழங்குகிறது. மனித இனத்தின் கலாச்சார தொன்மையை அறிந்து கொள்ள மிக முக்கிய பொருளாக சுவடிகள் கருதப்படுகிறது. சுவடிகளில் மறைந்து கிடக்கும் அறிவுச் செல்வத்தின் துணையினால் மட்டுமே சமூக கலாச்சார புரிந்துணர்வின் இடைவெளியை நிரப்ப முடியும் என்கிற உண்மையை சமீப காலங்களில் தான் உணரமுடிந்தது. எனவே சமுதாயத்தின் இன்றைய மற்றும் எதிர்கால தேவைக்கு சுவடிகளைப் பாதுகாப்பது மிக மிக அவசியம். பழங்காலத்தில் மிகப்பெரிய அறிஞர்கள் தங்களுடைய வாழ்நாட்களில் பெரும் பகுதியை செலவழித்து அனுபவ அறிவை எழுதி வைத்து பாதுகாத்தார்கள். பல்வேறு வடிவங்களில் எழுதி பயன்படுத்தினார்கள். தற்சமயம் இந்த விலை மதிப்பற்ற சொத்துக்கள் பாதுகாப்பின்மையால் காணாமல் அழிந்து போகும் நிலையிலுள்ளது. பல்வேறு பகுதிகளில் ஓழுங்கற்றும் பராமரிப்பின்றியும் சிதறிக்கிடப்பது மாத்திரமன்றி தனியார் பாதுகாப்பில் வெறுமனே கிடக்கிறது. இந்த விலை மதிப்பற்ற செல்வத்தை பாதுகாப்பது தேசிய பிரச்சினையாகும். சுவடி ஆவணங்கள் ஒரு நாட்டின் சரித்திர மற்றும் கலாச்சாரத்தை பாதுகாக்கும் தன்மையைக் கொண்டது. சுவடிகள், கலாச்சாரம் மற்றும் அறிவை சமுதாயத்துற்கு வெளிப்படுத்த உதவும் மிக முக்கிய இணைப்பாகும். நிர்வாகம், கலை மற்றும் கலாச்சாரம், புவிசாத்திரம், கவிதைகள், மதம், தத்துவம், வானசாத்திரம், வேத மந்திரங்கள், தந்திர கதைகள் ஆகிய பாடங்களை உள்ளடக்கியது

வெப்பமும் ஈரப்பதமும் கொண்ட தென்னாட்டு தட்ப வெப்ப நிலையில் ஓலைச்சுவடிகள் பல நூற்றாண்டுகள் அழியாமல் இருப்பதில்லை. ஓலைச்சுவடிகளின் வாழ்நாள் சுமார் முந்நூறு ஆண்டுகளாகும் . எவ்வளவு தான் பராமரித்து வந்தாலும் அவைகள் நாளடைவில் உருமாற்றமடைந்து , மக்கி, எழுத்துக்கள் மங்கி ஒடிந்து தூள் தூளாக போய் விடும். அவ்வாறு பனையோலைகள் சேதமுறுவதற்கு முன்பதாக அவைகளை அச்சுப் பதிப்பாகவோ அல்லது இலக்கமாக்கியோ பாதுகாத்துக் கொள்ளுதல் மிக மிக அவசியம் .

இந்நாட்களில், அரசு மற்றும் தனியார் பல்கலைக் கழகங்களிலும் , அரசு நூலகங்களிலும் , அரசு மற்றும் தனியார் கல்லூரித் தமிழ்த் துறைகளிலும் , தனியார் நிறுவனங்களிலும் , தனியார் நூலகங்களிலும் , அரசின் ஆராய்ச்சி மையங்களில் ஓலைச் சுவடிகளும் காகிதச் சுவடிகளும் பாதுகாப்பில் வைக்கப்பட்டிருக்கின்றன . இவைகள்

சுவடிகளை பராமரித்து வருகின்றன . பராமரிப்பு மட்டும் போதாது . அவைகளில் பொதிந்து கிடக்கும் தகவல்களை மீட்டெடுக்க வேண்டும் . அதற்கு தேவையானவற்றை உடனடியாகச் செய்தாக வேண்டும் . அது யாதெனில் , சுவடிகளில் இருக்கும் தகவல்களையெல்லாம் இலக்கமாக்கி முறையான தரவுத் தளத்தில் பாதுகாத்து இணையத்தின் வாயிலாக ஒருங்கிணைத்து பொது மக்கள் பயன்பாட்டிற்கு கொண்டு வர வேண்டும் .

உலக முழுவதிலுமுள்ள பெரும்பாலான தமிழ் அறிஞர்கள் இணையத்தைப் பற்றிய அறிவும் , கணினி பற்றிய அடிப்படை மற்றும் பயன்பாடு செயல்பாடு பற்றிய அறிவும் கணினி வழி சுவடிகளில் உள்ள தகவல்களை எவ்வாறு இலக்கமாக்குவது என்கிற தெளிவான அறிவும் உடையவர்களாகவே இருக்கிறார்கள் . இருந்தும் அவர்களின் சிந்தனைகள் வேறு விதமாக இருக்கிறது . சிலருடைய எண்ணமெல்லாம் தங்கள் வசம் இருக்கும் சுவடிகள் தங்களுடைய மூதாதையர் சொத்து . இதை ஏன் மற்றவர்களுடன் பகிர்ந்து கொள்ள வேண்டும் என்கிற எண்ணத்தின் வெளிப்பாட்டால் சுவடிகள் அழிந்தாலும் பரவாயில்லை என்று எண்ணுகிறார்கள் . இந்த மனோபாவம் மாற வேண்டும் . அவர்கள் சேகரித்த சேகரங்களை பொதுவான இடத்தில் வைக்கும் அளவுக்கு அவர்கள் கீழிறங்கி வரவேண்டும் . அது பொதுமக்களுக்கு பயன்படவேண்டும் என்கிற எண்ணம் வேண்டும் . 'யாம் பெற்ற இன்பம் பெறுக இவ் வையகம்ய எனும் பரந்த மனப்பாண்யுடன் தங்களிடமிருக்கும் சுவடிச் செல்வங்களை அனைவரும் பயன் பெறும் வண்ணம் பார்த்துக் கொள்ள வேண்டும் . அதற்கு தேவையான முன்னெடுப்புகளில் ஊக்கமுடன் பங்கேற்க வேண்டும் . திக்கு எங்கும் சென்று சுவடிகளை திரட்ட வேண்டும் . திரட்டிய சுவடிகளை காலம் தாழ்த்தாது இலக்கமாக்கி பாதுகாத்துக் கொள்ள வேண்டும் .

மறைந்த மறைக்கப்பட்ட தமிழர் வரலாற்றுத் தேடல்கள், தூய தமிழ் வழக்கு, தமிழ் மொழிக்கல்வி தமிழைப் பற்றிக் கவலைப்படும் தனித்த இலக்கமாக்கல் ஆவணக் காப்பகங்கள் ஏதும் இல்லை. இந்திய கூட்டரசின் அமைப்புகள் இந்தி மொழியில் உள்ள ஆவணங்களை காப்பது பற்றி அக்கறை கொள்கிறது. அரசு ஆவணங்கள் சட்டப்படி பாதுகாக்கப் படவேண்டும். அவை நீதி மன்றங்களில் முறையாக அரசின் உறுதி கொண்ட ஆவணங்களாக நிறுவப்படவேண்டும். அதற்கான நெறிமுறைகள் வகுக்கப்பட வேண்டும்.

தமிழகத்திலும் அயல்நாடுகளிலும் பல்வேறு நூலகங்களிலும் உள்ள ஓலைச்சுவடிகளை இலக்கமாக்கம் செய்து வாசிப்புக்கு உட்படுத்தவதற்கு ஏதுவாக தகவல் தொழில்நுட்ப கருவிகளைப் பயன்படுத்தி மின்பதிப்பு பக்கங்களையும் அதற்கேற்ற அட்டவணையையும் தயார் செய்து இணையத்தில் பதிப்பித்து , உலக முழுவதிலுமுள்ள தமிழர்கள் பயன் பெற தமிழில் தேடு பொறி ஒன்று உருவாக்கி வழி வகை செய்ய வேண்டும் . இலக்கமாக்கம் செய்யும்பொழுது ஏற்படக்கூடிய தொழில் நுட்ப பிரச்சினைகளை கையாள கையேடுகள் உருவாக்கப் படவேண்டும் . சரியாக திட்டமிட்ட தரவுத் தளம் உருவாக்கப்பட வேண்டும் தகவல் தொழில் நுட்பம் அனைத்து தமிழ் மக்களுக்கும் கிடைக்க தரப்படுத்தப்பட்ட விசைபலகையுடன் அகராதிகளையும் தரவுத் தளங்களையும் கணிப்பொறியில் நிறுவி இணைய வழி பயனடைவோம் .

மின்னாக்கம் அல்லது இலக்கமாக்கல் சிறப்பாக செயல்பட தமிழக அரசு ஒரு சிறப்பு அரசு ஆணை மூலம் பொது மக்களுக்கு ஒரு வேண்டுகோள் வைக்கலாம் . அதாவது தங்களிடமிருக்கும் தொல் ஆவணங்களை ஓலைச்சுவடியாயினும் அல்லது காகிதச் சுவடியாயினும் அரசு குறிப்பிடும் நிறுவனத்தில் ஒப்படைத்து உதவிட வேண்டும் . அரசு இதற்கென ஒரு துறையை ஏற்படுத்தி அவர்கள் தாங்கள் சேகரித்த சுவடிகளை இலக்கமாக்கி இலக்க நூலகத்தில் சேர்க்க வேண்டும் . தமிழ் இணையப் பல்கலைக் கழகம் அல்லது உலகத் தமிழ் செம்மொழி ஆராய்ச்சி மையம் அல்லது தஞ்சாவூர் தமிழ்ப் பல்கலைக் கழகம் போன்ற ஏதேனும் ஒரு நிறுவனம் இந்த பணியை ஏற்றுக் கொள்ள முன்வரவேண்டும் . அதற்கான திட்ட வரைவை அரசுக்கு அனுப்பி ஒப்புதல் பெற வேண்டும் . அரசும் இத் திட்டத்தக்கான ஒரு பொது நிதி உருவாக்க வேண்டும் . அதற்கு ஒரு கால திட்டத்தை வகுத்துக் கொண்டு செயல் பட வேண்டும் . இலக்க நூலகத்தை இணையத்தின் வழி செயல்பட வைக்க வேண்டும் . இணையத்தின் வழியாக உலக முழுவதிலுமுள்ள மக்கள் பயன்படுத்திக்கொள்ள ஏதுவாக இது வடிவமைக்கப்பட்டு சிறப்பு அம்சங்களுடன் செயல்படுத்தப்பட வேண்டும் . ஓலைச்சுவடிகளின் தகவல்களை தட்டச்சு மூலம் கணிப்பொறியில் சேமிப்பதினால் பல்லாயிரக் கணக்கான ஓலைச் சுவடிகளிலுள்ள செய்திகள் இலக்கமாக்கப்படலாம் . தற்காலக் கணிப்பொறிகளில் படப்பகுப்பாய்வு மூலம் படங்களாகப் பதிவு செய்யவும் ஆய்வு செய்யவும் இயலும் . இவ்வாறு கணிப்பொறியில் பதிவு செய்யப்பட்டுள்ள செய்திகளை படமாகவோ அல்லது அச்சாக்கியோ பெற்றுக் கொள்ளும் வசதி இருப்பதால் ஓலைச் சுவடிகளை பாதுகாப்பது அவசியமாகிறது .

கணிப்பொறி மூலம் ஓலைச் சுவடிகளின் தகவல்களை பாதுகாத்து எளிதில் தொகுக்க இயலும்.. உலகம் முழுவதிலுமிருந்து கிடைக்கக் கூடிய ஓலைச் சுவடி உரைகளையும் அவை சார்ந்த நூல் அடைவுகளையும் பிற விவரங்களையும் கணிப் பொறி வழி தொகுக்கலாம். பின்னர் பொருள் மற்றும் பாடத் தலைப்புகளில் அவற்றை வகைப் படுத்த இயலும். ஆசிரியர், தலைப்பு, பொருள், உரையாசிரியர், முதலியனவற்றைக் கொண்டு செய்திகளைப் பகுத்தும் தேவைக் கேற்பத் தெரிவு செய்யவும் உரிய அட்டவணைகளை உருவாக்கலாம். உலகெங்கும் பரவிக் கிடக்கும் தமிழினம் தகவல் தொழில் நுட்ப மாற்றங்களின் வாயிலாக தங்கள் மொழி இன கலாச்சார செல்வத்தை பயன்படுத்திக் கொள்ளுதல் சாலச்சிறந்தது . பொறியியல், மருத்துவம் ஆகியன நம்மிடையே பெரும் தாக்கத்தை ஏற்படுத்தியிருந்தாலும் தகவல் தொழில் நுட்பம் வாயிலாக அவற்றை வசப்படுத்தி வாழ்வாங்கு வாழ்வோம் . தொன்மையான மொழியாம் நம் தாய் மொழி தமிழின் அனைத்து கலைச் சொற்களையும் புழக்கம் மற்றும் பயன்பாட்டில் கொண்டு வருவற் கு தகவல் தொழில் நுட்பம் வழி வகுக்கும் . உலக அளவில் தமிழின் பெருமை உயர்வது மட்டுமல்ல பயன்பாடு அதிகரிக்கும். எனவே தகவல் தொழில் நுட்ப கலைச் சொற்களை வளர்த்திட்டால் உலக அளவில் தமிழன் மத்தியில் ஒருங்கிணைப்பும் பயன்பாடும் உயரும் .

தமிழனின் தொன்மையான கலாச்சாரத்தை தாங்கி நிற்பவை பழம் பெரும் சுவடிகளேயாகும் . இத்தகைய சுவடிகளில் பொதிந்து கிடக்கும் உண்மைகளையும் அனுபவங்களையும் பரந்து பட்ட உலகில் பரவிக் கிடக்கும் தமிழர்கள் பயன்படுத்திக் கொள்ளுவதற்கு தகவல் தொழில் நுட்பம் வழிவகை செய்யும் . புழக்கத்திலிருந்து படிப்படியாக மறந்து மறைந்து போகும் தமிழ் சொற்களை பாதுகாத்து பகிர்ந்து கொள்ளுதல் சாத்தியப்படும் . இணையத்தின் துணை கொண்டு உலக முழுவதிலுமுள்ள தமிழர்கள் இலகுவாக பயன்படுத்த ஏதுவாக இருக்கும் .

Issues while developing annotated corpora for Modern Tamil

Dr.K.Umaraj

Assistant Professor, Dept. of Linguistics, Madurai Kamaraj University
umarajk@gmail.com

Aim

The goal of the Natural Language Processing (NLP) is to build Computer software that will analyze, understand and generate natural language like a human. There are two ways in which we can build software 1) Rule based modeling of the natural language 2) Corpus based modeling of the natural language. In rule based modeling of the natural language, the natural language information is to be given in the form of rules (Phonotactic, Morphotactic and Syntactic rules) . In corpus based modeling of the natural language, the algorithm itself derives the rules from the annotated corpora (electronically stored tagged linguistic data) .The present paper discusses the issues while developing annotated corpora at word level for the Modern Tamil.

Annotated Corpora

Annotated corpora means the electronically stored text or speech data can be marked or assigned with some notations or tags in such a way that a machine can understand the text or speech data. The marking may be at different levels of a language i.e Phonological, Morphological, Syntactic, semantic and Pragmatic levels.

Annotated corpora at word level

Each word has their own class and parts of speech (POS) tagging are simply the problem of placing words into equivalence classes. Garside (1995) says POS tags to each word is passed through the stages of morphological analysis and grammatical interpretation.. Generally, specially designed codes carrying grammatical information are assigned to the words to indicate their part-of-speech with regard to their use in the text (Leech and Garside 1982) . In most cases, a well-defined set of linguistic rules are used to identify and assign POS tags to words to determine their lexico-semantic entities and syntactico-grammatical functions in the text.

Uses of POS tagging:

POS tagged corpus is a valuable resource for the works of Natural Language Processing, Language Technology and Machine Learning

It is useful for describing a language.

It is very useful for testing a new linguistic hypothesis.

At lexical level it allows to analyse the morphological structure of words represented in their surface forms.

At orthographic level it draws distinction among the homographic forms used in the same text or similar other texts to make distinctions in their semantic roles, and

At syntactic level it allows to identify syntactico-grammatical functions of words.

Procedure for doing POS tagging

In general, the process of POS tagging may be carried out on a piece of text at three separate stages as the followings

(a) Stage 1: Manual or automatic pre-editing of text,

(b) Stage 2: Manual or automatic tag assignment to words, and

(c) Stage 3: Manual post-editing of tagged text.

Stage1: Corpus usually contains some extra symbols, spelling mistakes, some missing words and sentences etc this should be cleared. Tokenize the corpus properly by separating punctuations from preceding tokens and by splitting sentences or phrases into their constituent tokens.

Stage2: In this stage, only one tags should be assigned to each word based on their syntactic and grammatical function of the word in the sentence.

Stage3: At this stage the entire tagged text database is manually post edited to verify if words are properly tagged.

Tagset

Tagset are names given to a set of tags from which tags are to be given to the input words in a text. The tagset consists of grammatical tags, which may include the morphological, morpho-syntactic, semantic, and discourse level of tags.

There are different types of tagsets are available 1) Flat tagset 2) Hierarchical tagset 3) BIS tagset. Following is the list of tagsets used for the Tamil language

Ganesan's POS tagset, Annamalai University,
Amrita POS Tagset. Coimbatore,
AUKBC tagset, Anna University,
Kongu Engineering College POS Tagset, Erode
Rajendran's ILMT-Tagset, Tamil University,
Vasu Ranganathan's Tamil Tagset
kathambam of RCILTS-Tamil tagset
LDC-IL-Tamil hierarchical tagset CIIL, Mysore,
IIIT-Hyderabad tagset
Microsoft research labs-bagalore tagset

The present paper identified the problem in POS tagging based on the following 14 tagsets developed by CIIL Mysore

1. Pronoun (P)
2. Demonstrative (D)
3. Noun (N)
4. Nominal Modifier (J)
5. Verb (V)
6. Adverb (A)
7. Participle (L)
8. Postposition (PP)
9. Particle (C)
10. Numeral (NUM)
11. Reduplication (RDP)
12. Residual (RD)
13. Unknown (UNK)
14. Punctuation (PU)

Issues in POS tagging

The issue of assigning part-of-speech to words, although appears to be simple, straightforward and one dimensional, is in fact, embedded with several theoretical and technical complexities with regard to identification of actual lexico-semantic entities as well as syntactico-grammatical functions of words used within a piece of text.

1) In Modern Tamil, the word 'enRu' occurs in the following way. So based on the syntactic of the sentence we have to assign POS tags for the words.

1.wh-adjective

Example: enta naaLil nii vantaai " in which day you came "

2.Adverb

Example: tollai enRu tiirum? 'when will this problem solve'

3.Conjunction

a.Quatative

Example: avan naaLai varuvan enRu connaan 'He said that he will come tomorrow'.

b.Called

Raman enRu oru aaL irukiRaara? 'a person called raman is here?

c.Purposive

The subordinate verb end with a future tense marker

Example:

naan paNattai ungaLukku koTukkaalam enRu ninaitteen 'I thought I can give the money to you '

d.Onomotpoeic

Example: avan kutu kutu ena ooTinaan 'He rans quickly quickly'

e. Giving focus

naan unakku enRu vaangiya peena? 'I brought the pen for you only'

2) Tamil has adverbial suffix which is responsible to make any word into adverb example cuttamaaka (cleanly), cukamaaka (happily), nicaimaaka (surely) amaitiyaaka (silently) but there are other cases in Tamil where this adverbial marker 'aga' can come with the Proper noun avan raamanaka vantaan (he came as Lord Rama) .If we tag it as adverb the important information like proper noun will be missed out in POS tagging.

3) Compound

whether all the compound words should be written as one word or two separate is a problem. V.I.Subramaniam quote as follows for compounds in a texts "In some places segmentation of the head word was problematic". The principles of segmentation adopted are: If the units A and B mean X and Y, A is listed with meaning x and B with y:

If A and B mean only Z then A and B are listed as one word. In some cases A and B may give a different sense also. For example: the word 'Vilaimagal' 'prostitute', it is possible to assign the meaning 'price' to 'vilai' and the daughter to 'makal', the prostitute is a woman who sells her body. Here, he says the word 'vilaimagal' should entered as a single entry.

If we follow the theory, then we have to split the following words also: *maṅkoṭi* 'mango trunk' 1) *mā* 'mango tree' and *koṭi* 'trunk' and 2) the word *māṅkaṇi* 'mango fruit' *mā* 'mango' and *kaṇi* 'fruit'. But it is not necessary to split these compounds. Because these compounds are occurring in few places and it was lexicalized in Tamil.

Conclusion:

The present paper discusses only a few problems in POS tagging for the Tamil, still a lot more problems exist in Tamil language.

Bibliography:

1. Andronow, M. 1969. 'A standard grammar of modern and classical Tamil' Madras: *New Century Book House*.
2. Arulmozhi. P, Sobha L, A Hybrid POS Tagger for a Relatively Free Word Order Language. In proceedings of MSPIL-2006, Indian Institute of Technology, Bombay. pp 79-85, 2006
3. Dash, Niladri Sekhar (2005) *Corpus Linguistics and Language Technology: with Reference to Indian Languages*, New Delhi: Mittal Publications.
4. Lehmann, Thomas (1989) *A grammar of Modern Tamil*. Pondicherry: Pondicherry Institute of Linguistics and Culture.

இணைய வழி தமிழ் ஆய்வு வங்கி

கா. ஆனந்த குமார்

உதவிப் பேராசிரியர், தமிழ்த்துறை, நந்தா கலை மற்றும் அறிவியல் கல்லூரி, ஈரோடு
tamizhananth@gmail.com

சுருக்கவுரை

உலகெங்கும் வாழும் தமிழர்களும் பிற நாட்டறிஞர்களும் தமிழ் மொழி பண்பாடு கலாச்சாரம் வரலாறு போன்ற பல தளங்களில் ஆய்வுகளை மேற்கொண்டு முடிவுகளைக் கண்டறிக்கின்றனர் . இத்தகைய பணியில் தன்னார்வ அமைப்புகளும் கல்வி நிறுவனங்களும் ஈடுபடுகின்றன . ஆய்வுலகில் கண்டறியப்படும் முடிவுகள் தனித்தனியே முறைப்படுத்தப்படாமல் சிதறுண்டு கிடக்கின்றன . கணினித் தொழில் நுட்பம் வளர்ச்சி அடைந்துள்ள தற்காலத்திய சூழலில் இதுவரை நிகழ்த்தப்பட்டுள்ள ஆய்வுகளின் தொகுக்கப்பட்ட தரவுப்பட்டியல் (Research Bank-ஆய்வு வங்கி) இன்றியமையாததாகிறது .

முகவுரை

கணினித் தொழில் நுட்பம் இன்று மக்களின் அன்றாட வாழ்வோடு பிணைந்துள்ளது . வெகுசன மக்களுக்கான ஊடகமாக கணினி மாற்றம் பெற்றுள்ளது . அச்ச வடிவில் காகிதப் பிரதிகளாக அறியப்பட்ட தமிழ் குறித்த விடயங்கள் இன்று கணினி வழியே பரவலாக்கம் பெற்றுள்ளது . அறிவியல், சமூகவியல் போன்ற துறைகளோடு தமிழ் மொழியும் கணினியுடன் இணைந்து இன்று உலகளாவிய அளவில் பல்வேறு விடயங்கள் விவாதிக்க மிகப்பெரும் இடம் (Space) இணையத் தொழில் நுட்பம் வழியாக உருவாகியுள்ளது . எனினும் தமிழில் ஆய்வினைச் செய்யப் புகும் ஆய்வாளர் ஒருவருக்கு அதற்கு முன்பாக நிகழ்த்தப்பட்ட ஆய்வுகள் குறித்த தகவல்களை அறிந்து கொள்ள இன்றைய நிலையில் வசதிகள் (அச்சப் பிரதி/இணைய வழி) எதுவும் உருவாக்கப்படவில்லை . அத்தகைய வழிமுறைகளை அளிக்கக் கூடிய ஆய்வு வங்கியின் (Research Bank) உருவாக்கம் பற்றியும், ஏற்படுத்துவதில் உள்ள சிரமங்கள் பற்றியும் இங்கு காணலாம் .

1. இணைய வழி தமிழ் ஆய்வு வங்கி

இணைய வழி பயன்பாட்டின் காரணமாகத் தமிழுலகில் நிகழ்ந்து வரும் ஆய்வுச் செயல்பாடுகள் விரிவாக்கம் பெற்றுள்ளன . இதன் ஒரு பகுதியாக அச்சில் கிடைக்கப்பெற்ற பழந்தமிழ் நூல்கள் இன்று இணைய வழியாக (மதுரை தமிழ் இலக்கிய மின்பதிப்புத் திட்டம்) அனைவரும் அறிந்து கொள்ளும் வகையில் பதிப்பிக்கப்பட்டுள்ளது . அரிய ஒலைச் சுவடிகள் மின்- நூலாக்கம் (தமிழ்மரபு அறக்கட்டளை) செய்யப்பட்டுள்ளன . கற்றல்-கற்பித்தல் ஆகியவற்றைச் செயல்படுத்தும் வகையில் இணைய வகுப்பறைகள் (தமிழ் இணையக் கல்விக்கழகம்) தோற்றுவிக்கப்பட்டுள்ளன . உயர்கல்வியிறுவனங்கள் (பல்கலைக்கழகங்கள் . . .) ஆராய்ச்சி நிறுவனங்கள் போன்றனவும் தனி நபரின் வலைப் பூக்கள் தன்னார்வ அமைப்புகளின் இணையதளங்கள் ஆகியனவும் பல்வேறு ஆராய்ச்சித் தகவல்களை அளிக்கின்றன . ஆனால் ஆராய்ச்சி குறித்த விடயங்களைப் பற்றி அறிந்து கொள்ள ஆய்வுகளின் தொகுக்கப்பட்ட பட்டியல் இன்னும் உருவாக்கப்படவில்லை . இணையம் வழியாக இத்தகைய ஆய்வுகளுக்கான பட்டியலை பதிவேற்றம் செய்யும்போது முந்தைய ஆய்வுகளின் நிலையும் சமகாலத்திய நிகழ்வுகளும் நமக்கு அறிய கிடைக்கின்றன . இதனைப் பல்வேறு தளங்களில் விவாதிக்கப்படுகின்ற புதிய ஆய்வு விடயங்களைச் சேர்ப்பித்தும் ஒரு வங்கிபோன்று பயன்படுத்திக்கொள்ளலாம் .

2. இணைய வழி தமிழ் ஆய்வு வங்கியின் தேவை

இன்று தகவல் தொழில் நுட்பம் எல்லா நிலைகளிலும் பரவலான வளர்ச்சியினைக் கண்டுள்ளது . இத்தகைய வளர்ச்சியின் காரணமாக இணையம் வழியே . தமிழ் இலக்கிய ஆராய்ச்சிக் கட்டுரைகளை (www. nilacharal. com, www. keetru. com . . . போன்றன) வெளியிட்டு வருகின்றன . வார்ப்பு (www. vaarppu. com) இணையதளம் கவிதை இலக்கியப் படைப்புகளை வெளியிடும் அதன் மீதான ஆராய்ச்சிக் கட்டுரைகளை வெளியிடும் வருகின்றது . புதிய பனுவல் என்னும் உலகளாவிய ஆய்விதழ் தமிழ்மொழி பண்பாடு போன்ற பல்வேறு தளங்களில் சிறந்த

அறிஞர்களின் கட்டுரைகளை வெளியிட்டு வருகின்றது . இத்தகைய செயல்பாடுகள் நடைபெற்று வந்தாலும் இவையாவும் தனித்தனியே காணக்கிடைக்கின்றன . எனவே தொகுக்கப்பட்ட ஆய்வுகளின் தரவுப் பட்டியல் அவசியமாகின்றது .

உயர்கல்வி நிறுவனங்களான பல்கலைக்கழகங்கள் , தமிழ் உயராய்வு மையங்கள் ஆகியன முனைவர்பட்ட (Ph. D) ஆராய்ச்சிப் படிப்புக்களை நடத்துகின்றன . இவற்றின் மூலம் ஆண்டிற்கு நூற்றுக்கணக்கான மாணவர்கள் தங்களின் ஆய்வேடுகளைச் சமர்ப்பிக்கின்றனர் . இத்தகைய ஆய்வுகளுக்கான ஆய்வுப் பட்டியல் முறையாகப் பல்கலைக்கழகங்களால் உருவாக்கம் பெறவில்லை . கோவை பாரதியார் பல்கலைக்கழகம் சிறு அளவில் திரட்டப்பட்ட செய்திகளைக் கொண்டு ஆய்வுப் பட்டியல் நூல் ஒன்றினை அச்சுப் பிரதியாக வெளியிட்டுள்ளது . மதுரை காமராசர் பல்கலைக்கழகம் 'தமிழியல் ஆய்வு ' என்னும் பொருளில் 1986ஆம் ஆண்டு வரையில் நிகழ்த்தப்பட்ட ஆய்வுகளைப் பட்டியலிட்டுள்ளது . கேரள பல்கலைக்கழகம் "சங்க இலக்கிய ஆய்வுகள் செய்தனவும் செய்ய வேண்டுவனவும் " என்னும் பெயரில் சங்க இலக்கியம் குறித்த ஆய்வுகளைமுன்னிறுத்தி நூல் வெளியிட்டுள்ளது . சென்னைப் பல்கலைக்கழகத்தில் ந . சஞ்சீவி-யால் சங்க இலக்கிய ஆய்வு அட்டவணை என்னும் பெயரில் நூல் வெளியிடப்பட்டது . பாரதிதாசன், அழகப்பா, பொடியார் பல்கலைக்கழகங்கள் எவ்விதமான ஆய்வுப் பட்டியலையும் நூலாகவோ இணையம் வழியாகவோ வெளியிடவில்லை . மனோன்மணியம் சுந்தரனார் பல்கலைக்கழகம் மட்டும் விதிவிலக்காக 2002-2008 ஆம் ஆண்டு வரையில் நிகழ்த்தப்பட்ட ஆய்வுகளின் ஆய்வுப் பட்டியலை தமது இணையதளத்தில் வெளியிட்டுள்ளது . தில்லி பல்கலை, ஜவர்கர்லால் நேரு பல்கலை உட்பட பிற பல்கலைக்கழகங்கள் எவையும் தங்கள் நிறுவனத்தில் நிகழ்த்தப்பட்ட ஆய்வுகளின் பட்டியலை வெளியிடவில்லை .

தமிழில் உயர்ந்த ஆராய்ச்சி நிறுவனங்களாகக் கருதப்பெறும் உலகத் தமிழாராய்ச்சி நிறுவனம் , செம்மொழித் தமிழாய்வு மையம் ஆகியனவும் இன்று வரையில் நிகழ்ந்து வரும் ஆய்வுப் பட்டியல் எதுவும் வெளியிடவில்லை . தமிழ் மொழிக்கென தனியே துவக்கப்பட்ட தஞ்சை தமிழ்ப் பல்கலைக்கழகம் ஆராய்ச்சி நிகழ்வுகள் குறித்து பதிப்புத்துறையின் மூலமோ , இணையம் வழியாகவோ எத்தகைய குறிப்புக்களையு ம் வெளியிடவில்லை . ஓர் ஆய்வாளருக்கு ஆய்வினைத் துவக்கும் முன்பாக ஆய்வுப் பொருள் குறித்த முந்தைய ஆய்வுகளைப் பற்றிய புரிதல் அவசியமாகத் தேவைப்படுகின்றது . இதன் மூலமாக சா ட்யான நெறிமுறைகளோடு சிறந்த ஆய்வுகளை மேற்கொள்ள முடியும் . எனவே ஆய்வு வங்கி ஒன்றினை உருவாக்குதல் சிறந்த பயனை அளிக்கும் எனலாம் .

பிற துறைகளுக்கு உள்ளதைப் போன்று தமிழ் ஆய்வுக்கென உலகளாவிய அளவில் ஆய்விதழ்கள் (Journals) எதுவும் இணைய வழி உருவாக்கப்படவில்லை . புதிய பனுவல் என்னும் ஆய்விதழ் தற்போது பன்னாட்டு ஆய்விதழாக (International Journal) வெளிவருகிறது . முத்துக்கமலம் (www. muthukamalam. org) என்னும் ஆய்விதழ் இணைய வழி (e-journal) இலக்கிய ஆய்விதழாக வெளியிடப்பட்டுள்ளது . இவை மட்டுமே தற்போது இணையம் வழி கிடைக்கப் பெறுவனவாக உள்ளன . கல்வி மற்றும் ஆய்வு நிறுவனங்களான தஞ்சை தமிழ் பல்கலைக் கழகம் , சென்னை பல்கலைக்கழகம், உலகத் தமிழாராய்ச்சி நிறுவனம் ஆகியன முறையே தமிழ்க் கலை (Tamil Culture), அரிமா நோக்கு (Arima nokku), தமிழியல் (Journal of Tamil Studies) ஆகிய ஆய்விதழ்களை அச்சுப் பிரதிகளாக வெளியிடுகின்றன . தன்னார்வ அமைப்புகளும் ஆய்விதழ்களை (செந்தமிழ்-மதுரை தமிழ்ச்சங்கம், தமிழ்ப்பொழில்-கரந்தைத் தமிழ்ச்சங்கம்) வெளியிடுகின்றன . இலக்கியச் சிற்றிதழ்களான தமிழ் இலெமுரியா , குமுதம்-தீரா நதி , காலச்சுவடு , உயிர்மை, உயிர் எழுத்து போன்றவையும் சிறந்த ஆய்வுக் கட்டுரைகளை வெளியிடுகின்றன . இவையன்றி நாளிதழ்களான தினமணி (தமிழ்மணி), தி ஹிந்து ஆகியனவும் தமிழ் ஆய்வுக் கட்டுரைகளை வெளியிட்டு வருகின்றன . தமிழறிஞர்களால் பதிப்பகங்கள் மூலமும் ஆய்வு நூல்கள் வெளியிடப்படுகின்றன . ஆனால் மேற்குறித்த ஆய்வுகளையெல்லாம் அறிந்து கொள்வதற்கு அதிக முயற்சியும் காலமும் தேவைப்படுகின்றது . இவை அனைத்தையும் ஓரிடத் தில் முறையாக தொகுத்து சேமிக்கும் போது எளிதாகப் பயன்படுத்த முடிகின்றது . எனவே தற்கால சூழலில் இத்தகைய ஆய்வுகளுக்கான வங்கி (Research Bank) என்பது மிகத் தேவையான ஒன்றாக அமைகின்றது . இதனைப் பட்டியல் என்ற முறையில் மட்டுமே அமைக்காமல் தரவுகளோடு சேமிக்கின்ற பகுதியாக அமைத்தல் சிறந்த பயனை அளிக்கும் .

3. இணைய வழி தமிழ் ஆய்வு வங்கி உருவாக்க நெறிமுறை

இணைய வழியில் ஆய்வு வங்கியினை உருவாக்குதல் எளிதானதல்ல . இதற்கு அதிகப் பொருட் செலவும் , காலமும் தேவைப்படும் . ஆய்வு வங்கியினை உருவாக்கும்போது பலவித காரணிகளைக் கருத்தில் கொள்ள வேண்டும் . இதற்கென தனியாக இணைய தளம் உருவாக்கப்பட வேண்டும் . இவ்விணைய தளம் பொது நிலையில்

அனைவராலும் பயன்படுத்திக் கொள்ளத்தக்க வகையில் (OPAC-online open access catalogue) அமைய வேண்டும் . இணையதள ஆய்வு வங்கியின் மொத்தச் செயல்பாடுகளையும் உருவாக்கும் பொறுப்பினை தமிழ் இணையக் கல்விக் கழத்திடம் (Tamil Virtual University) அளித்துவிடலாம். உலகத் தமிழாராய்ச்சி நிறுவனம் , தஞ்சை தமிழ்ப் பல்கலைக்கழகம், சென்னை செம்மொழித் தமிழாய்வு மத்தி ய நிறுவனம் ஆகியன கூட்டாக தமிழ் இணையக் கல்விக்கழத்தோடு இணைந்து நிர்வாகம் , பராமாஃபு போன்ற பணிகளை மேற்கொள்ளலாம் . இதர பல்கலைக்கழகங்கள், ஆய்வு நிறுவனங்கள், கல்லூரி தமிழ் உயராய்வு மையங்கள், தன்னார்வ அமைப்புகள் (தமிழ்ச் சங்கங்கள் போன்றன) ஆகியன வழி நடத்தும் ஆலோசனைக் குழுவாக இயங்கலாம் .

3.1 முதற்கட்ட பணி

தமிழ் இணையவழி ஆய்வு வங்கியின் முதற்கட்ட பணி தேசிய அளவில் பல்கலைக்கழகங்கள் , கல்லூரிகள், ஆய்வு மையங்கள் அகியவற்றில் நிகழ்த்தப்பட்டுள்ள தமிழ் முனைவர்பட்ட (Ph. D Dissertations) ஆய்வுகளைத் திரட்டி அதனைப் பதிப்பிக்க (உள்ளீடு செய்தல்) வேண்டும். பதிப்பிக்கும் நிலையில் துறை (specific filed) அடிப்படையில் பதிப்பிக்கலாம். ஆய்வு நிறுவனங்கள் (Institutions) அடிப்படையிலும் பதிப்பிக்கலாம். உதாரணமாக,

ஆய்வாளர் பெயர்: க. தங்கமுத்து
ஆய்வுத்தலைப்பு: பத்துப் பாட்டில் புறக் காட்சிகள்
நெறியாளர்: ரா. சீனிவாசன்
ஆண்டு:1986
ஆய்வுத்துறை: சங்க இலக்கியம்
ஆய்வு நிறுவனம்: காஞ்சி மாமுனிவர் பட்ட மேற்படிப்பு மையம், புதுச்சேரி

(*மேற்குறிப்பிடப்பட்டுள்ளது எடுத்துக்காட்டிற்காக மட்டுமே) .

என்ற படி நிலைகளில் ஆய்வுப் பட்டியலைத் தொகுக்கலாம் . முதற்கட்டப் பணியில் இதுவரை நிகழ்த்தப்பட்ட ஆய்வுகளின் பட்டியலை மட்டும் பதிப்பிக்கலாம் . அதற்கு அடுத்த நிலையில் ஒவ்வொரு ஆய்வின் தலைப்பினைச் சுட்டும் போது அதன் உட்பொருள் , கண்டடைந்த முடிவுகள் ஆகிய விபரங்களை சுருக்கமாக அறிந்து கொள்ளும் வகையில் உருவாக்கலாம் . இறுதிக் கட்டமாக ஆய்வேட்டினை முழுமையான அளவில் பிடிஎஃப் (PDF) கோப்புகளாக பதிப்பிக்கலாம் . அவ்வாறு செய்யும் போது ஆய்வேட்டினைப் பார்வையிட (view) மட்டுமே அனுமதியளித்தும் அச்சிடுவதற்குத் (Print) தடை விதித்தும் பதிப்பிக்கலாம் . இதனால் ஆய்வுக் கருத்துக்கள் தவறான மு றையில் பயன்படுத்தப்படுவது தவிர்க்கப்படும் . இந்திய பல்கலைக்கழகங்கள் , கல்லூரிகளில் தமிழ் குறித்து முனைவர்பட்ட ஆராய்ச்சி செய்ய விழையும் எவரும் தங்களின் ஆய்வுத் தலைப்பை (Research Topic) இணைய வழி ஆய்வு வங்கியில் பதிவு செய்த பின்னரே ஆய்வினைத் தொடரும்படி செய்ய வேண்டும் இதற்குப் பல்கலைக்கழக நிதி நல்கைக்குழு (University Grant Commission) மூலம் இணைந்து விதிமுறைகள் உருவாக்கிக் கொள்ளலாம். இதன் மூலம் ஒரே வகையான ஆய்வுகள் மீண்டும் வெளிவரும் நிலை தவிர்க்கப்பட்டு புதிய புதிய ஆய்வுச் சிந்தனைகள் தோன்றுவதற்கு வழிவகை ஏற்படு ம். முந்தைய ஆய்வுகளிலிருந்தும் ஆய்வுகளை மீளாய்வு செய்வதற்கும் இடமளிக்கும் . ஆய்வேட்டினைச் சமர்ப்பிக்கும் போது அதன் பிரதி (Copy) ஒன்றினையும் ஆய்வு வங்கியிடம் சமர்ப்பிக்கும் நடைமுறையினை ஏற்படுத்த வேண்டும் . ஆண்டிற்கு ஒரு முறை அவ்வாண்டில் நிகழ்த்தப்பட்ட முனைவர்பட்ட ஆய்வுகளின் பட்டியலைத் தொகுப்பாக நூலாக்கம் செய்து வெளியிடலாம் . இதன் மூலம் தேசிய அளவில் நிகழ்த்தப்பட்ட ஆய்வுகளின் விபரங்களை அறிந்து கொள்ள இயலும் . ஆய்வு வங்கியில் பதிவு செய்யப்பட்ட ஆய்வுத் தலைப்புகளையும் , ஆய்வுகள் குறித்த பட்டியலையும் தொடர்புடைய பல்கலைக்கழகங்களின் இணையத்தோடு இணைக்க (Link) வேண்டும். மட்டுமன்றி மேலை நாடுகளில் உள்ளதைப் போன்று ஆய்வேடுகளைப் பரிசோதனை (scanning) செய்யும் முறையினையும் ஏற்படுத்த வேண்டும் . இதன் மூலம் ஆய்வேடுகளின் உண்மைத் தன்மை காக்கப்படும் .

3.2 இரண்டாம் நிலைப் பணி

தமிழாய்வு தொடர்பாக வெளிவரக் கூடிய ஆய்விதழ்கள் இலக்கிய சிற்றிதழ்கள் போன்றவற்றில் வெளிவரும் கட்டுரைகளையும் ஆய்வு நூல்களாக வெளியிடப்படும் ஆய்வுச் செய்திகளையும் இணைய வழி தமிழ் ஆய்வு வங்கியில் வெளியிடலாம். இதற்கென தனியே இணைய வழி தமிழ் ஆய்விதழை (Online Tamil Journal) உருவாக்கம்

செய்யலாம். புதிய ஆய்வுக் கட்டுரைகளைப் படைப்பதற்கும் வாய்ப்பினை ஏற்படுத்தலாம் . இதன் மூலம் தமிழ் ஆய்வு குறித்த முழுமையான செய்திகளை ஓரிடத்திலிருந்து பெறக் கூடிய வாய்ப்பு ஏற்படும் .

3.3 எதிர் கொள்ள வேண்டியவை:

இணைய வழி தமிழ் ஆய்வு வங்கியினை உருவாக்கும் வேண்டியுள்ளது.

போது பல்வேறு சவால்களை எதிர்கொள்ள

- இதுவரையில் நிகழ்த்தப்பட்ட ஆய்வுகளின் பட்டியலையும் விபரங்களையும் திரட்டுதல் கடினமான காரியமாகும். மேலும் இணைய வழியில் அவற்றை தொகுத்து வெளியிடும் போது பராமரித்தல் கடினமாகிறது.
- உருவாக்கும் காலமும், பொருட்செலவும் அதிகமாகிறது.
- இணைய வழி தொழில் நுட்பக் கோளாறுகள் ஏற்படும்போது பயன்படுத்துவது கடினம் .
- நாடெங்கும் நிகழும் ஆய்வுத் தரவுகளைப் பதிவு செய்தல் சிரமமான பணியாகும் .
- கணிப்பொறி அறிவுடையோரால் மட்டுமே இதன் பயனை அடைய இயலும் .
- ஒவ்வொரு ஆய்வு நிறுவனங்க ளும் சரியான முறையில் இதனைச் செயல்படுத்தாத நிலையில் தடைபட நேரிடும்.
- ஆண்டிற்கு ஒரு முறை இதன் பணிகள் ஆய்வுக்கு உட்படுத்தப்பட்டதாக அமைய வேண்டும் .
- காலத்திற்கு ஏற்றவாறு மாற்றங்களை (update) செய்ய வேண்டும்.

3.4 இணைய வழி தமிழ் ஆய்வு வங்கியின் பயன்கள்

- ஒட்டுமொத்த ஆய்வுகளின் தகவல்களை அளிக்கக் கூடிய ஆய்வுத் தகவற் களஞ்சியமாக இதனைப் பயன்படுத்தலாம்.
- தமிழாய்வு குறித்த ஆய்வுப் பரிமாணங்களை விளங்கிக் கொள்ளத் துணை புரியும்
- இதன் மூலம் ஆய்வுச் செய்திகளைத் தேடிச்செல்லும் காலமும் , பொருளும் மீதமாகும் . பயன்படுத்த எளிதானது.
- அடிப்படை கணினி அறிவு உடையோரால் இதனைப் பயன்படுத்த இயலும் .
- ஆய்வேடுகளை கணினியில் கோப்புகளாக வைக்கும்போது அதிக அளவில் தேவைப்படும் இடப் பரப்பின் அளவு குறைந்து ஓரிடத்தில் காணுகின்ற வாய்ப்பு ஏற்படும் .
- பிற நாட்டினரின் ஆய்வுகள், புதிய ஆய்வுக் கண்டுபிடிப்புகள் போன்றவற்றை எளிதில் பிறர் துணையின்றி தாமே அறிந்து கொள்ளும் வசதி இதன் மூலம் கிடைக்கின்றது .
- இணையம் இன்று மடிக்கணினி , செல்பேசி, கையடக்கக் கணினி போன்ற பல்வேறு ஊடகங்களுக்கான ஒன்றாக மாறி வருகிறது . எனவே இணைய வழி ஆய்வு வங்கி உருவா கும் போது பரவலான பயன் கிடைக்க ஏதுவாகிறது .

முடிவுரை

தமிழ் ஆய்வு உலகில் புதிய புதிய ஆய்வுச் சிந்தனைகள் தினந்தோறும் வெளிவருகின்றன . ஆய்வுகள் பெருகிவரும் சூழலில் அதனப் பயன்படுத்த தொழில் நுட்பம் சார்ந்து அடுத்த நிலையை அடைய வேண்டித் தேவை ஏற்பட்டுள்ளது. பழைய முறைகளைத் தவிர்த்து காலத்திற்கேற்ற வகையில் மாற்றம் பெற வேண்டிய சூழலும் தற்போது ஏற்பட்டுள்ளது . எனவே புதிய ஆய்வுச்சிந்தனைகளைப் பெறவும் , ஒரே நிலையில் நடைபெற்று வரும் ஆய்வுகளைத் தவிர்க்கவும் , பழைய ஆய்வுகளை மீளாய்வு செய்யவும் புதிய மாற்றங்கள் தேவப்படுகின்றது . இதனை அடைய இணைய வழியில் ஆய்வுகளைப் பதிவு செய்யும் இணைய வழி தமிழ் ஆய்வு வங்கி துணை புரியும். ஆய்வுகள் தினந்தோறும் பெருகுவது நன்மை அளிக்கக் கூடியதல்ல . அவை சிறந்த முறையில் புதிய சிந்தனைகளை விளைவிக்கக் கூடியதாக அமைய வேண்டும் . அதனை இணைய வழி தமிழ் ஆய்வு வங்கி சிறப் பாக நிறைவேற்றும் எனலாம் .

Topic D
Computational Linguistics
கணினி மொழியியல்

Building an Intelligent Rule Based System for Tamil Computing

Dr. M. Ganesan

CAS in Linguistics, Annamalai University, Annamalai nagar
ganesan_au@yahoo.com

Introduction

Computational Linguistics (CL) or Natural Language Processing¹ (NLP) has been developed as a separate discipline involving the subjects, linguistics, computer science, statistics, and logic. This branch of study mainly concentrates on the development of suitable models to represent the linguistic knowledge along with other knowledge to the system for the purpose of analyzing, understanding and generating a text in a natural language. Worldwide many of the tools, developed for these purposes impress our imagination. Programs for orthography and grammar correction, information retrieval from document databases, and translation from one natural language into another, among others, are nowadays available for many languages. However, we have to admit that such programs still lack real intelligence. The ambitious goal of creating software for deep language understanding and production, which would provide tools powerful enough for fully adequate automatic translation and man-machine communication in unrestricted natural language, has not yet been achieved, though attempts to solve this problem already have a history of nearly 50 years.

This suggests that in order to solve the problem, technologists will need to use the methods and results of a fundamental science, in this case linguistics, rather than the tactics of *ad hoc* solutions. 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. In this paper I discuss a computer tractable model, which I have designed and adopted in my software viz. Morphological Analyzer, Paradigm Generator, Morph and POS Tagger, and Spellchecker in Tamil, for representing Tamil morphology. I explained, in brief, the representation of morphological information in terms of Machine Readable Dictionary (MRD) through the unification based grammar formalisms, called feature structures (attribute – value systems) and the morphotactics through the mathematical formalism, called Finite State Automata (FSA) and Augmented Transition Network (ATN) .

Tamil Morphology

Tamil is a verb final language with SOV pattern. It shows the Subject-Predicate agreement. It is a left branching language. The head of a phrase finds its place at the right most position in that phrase and all the attributes are placed to the left of the head. The words are classified into 4 groups viz. *peyarccol* (Noun), *vinaiccol* (Verb), *iTaiccol* (Particle) and *uriccol* in traditional grammar and are now grouped into 8, viz. Nouns, Pronouns, Verbs, Adjectives, Adverbs, Postpositions, Conjunctions and Interjections in modern linguistics. Tamil is an agglutinative language; A stem can take many suffixes one after another to have different grammatical forms and semantic notions. A stem with one or two suffixes acts as a stem and further takes more suffixes. And Tamil is a morphologically rich and complex language. Particularly Tamil verbs have a long string of morphemes that express a range of meanings like tense, mood and aspects as well as sense of assertion, negation, interrogation, reflection, emphasis, etc.

Morphology is the branch of grammar that deals with the internal structure of words (Mathews, 1974) . i.e. It is the study of meaningful parts of words. The structures of Tamil words are

- i) Root (payyan ‘boy’)
- ii) Stem + suffix (payyan+kaL ‘boys’)
- iii) Stem + suffix +suffix +... (paTi+tt+a:n+a: ‘did he read?’)
- iv) Prefix + stem + suffix + ... (‘am+maNavan+ai ‘that student (acc.) ’)
- v) Stem + stem (kai+viral ‘hand’s finger’)

In Tamil verbs and nouns are more productive in the sense that they can produce more word forms. The noun takes plural marker to make plural forms and it declines to different forms by taking case markers. Particles and postpositions can be

suffixed to a noun. Similarly, a verb can be conjugated to different forms by suffixing tense markers, person-number-gender markers, infinitive, conditional, relative participle, verbal participle, etc. A verb in Tamil can have more than 1600 forms including forms with clitics, particles, postpositions, etc.

Tamil words are formed by the three morphological processes; inflection, derivation and compounding. The formation of word forms are normally discussed with the following three components, viz. morphological lexicon², morphotactic rules and rewrite rules. The morphological lexicon of Tamil need not have entries for different word forms like *paTitta:n* 'read – he', *paTitta:L* 'read –she', *paTikka* 'to read', *paTitta* 'on who read', *paTittu* 'having read', *paTitta:l* 'if one reads', etc. It only needs to list the unique morphs {paTi}, {tt}, {a:n}, {a:L}, {a}, {u}, and {a:l}. The suffixes apply to {paTi}'read' and other verbs like {koTu}'give', {izhu}'pull', {muRi}'break', etc. as well.

How does one know that */ttpaTia:n/, as oppose to /paTitta:n/ 'read-past-he', from the morphemes {paTi}, {tt} and {a:n} is invalid? The legitimate order of morphemes are expressed by morphotactic rules such as

finite verb → verb stem + tense + PNG marker
relative participle → verb stem + tense + rp marker

The order of morphemes in the words is explained by the morphotactic rules . The syntagmatic pattern adds suffixes to the right in the fixed order in an isomorphic way for each added grammatical meaning.

The legitimate addition of suffix to a stem or suffix to a suffix needs more clarification. The suffixation is not always very simple. For example, the morphemes {kal}+{ai} = /kallai/ 'stone (acc.) ' and not */kalai/. There are some changes in the morphemes and they are conditional. When {ai} is added to {kal} 'stone' the final [l] in {kal} is doubled. i.e. the final [l] of a monosyllabic word is doubled when it is followed by a segment starting with a vowel. Such changes are called morphophonemic changes and are explained by rewrite rules.

$l \rightarrow ll / (C) v \rightarrow V$ [where v refers to short vowel and V to any vowel]

The morphophonemic processes that are involved in Tamil morphology are assimilation, insertion, deletion and gemination.

Computational Linguistics

Computational linguistics is an interdisciplinary field dealing with the statistical and/or rule-based modeling of natural language from a computational perspective. The grammar of a language is mainly written for human use and they as such cannot be used for computers to process the language. Keeping the linguistic rules as base computer tractable rules or grammatical model has to be evolved. Linguistics aims at to describe the structure of a language in terms of rules. The précised set of rules which generate infinite sentences are considered to be a good grammar. Whereas in computational linguistics how precise or elaborate the grammar given to the system is not very important, but the grammar must be computer tractable and should analyze / generate a text correctly in a short span of time. Accuracy and speed are the main criteria for assessing the efficiency of the model represented to the system. Rest of this paper will discuss the model that represents the morphological information of Tamil to a system.

Computational Morphology

Computational morphology is a subfield of computational linguistics. Computational morphology concerns itself with computer applications that analyze words in a given text, such as determining whether a given word is a verb or noun. As shown above words in a language are not simple lexeme; may be inflected or derived or compounded. It is practically not possible to list the entire word forms in a language, especially for language like Tamil. Then the dictionary will have few lakhs of entries, which will be difficult to the computers to use. The efficient way to represent the word forms would be listing all the unique stems and suffixes in the dictionary and having a program for deriving words from the morphological components. Almost all practical applications that deal with natural language must have morphological component. After all, an application must first recognize the word in question before analyzing it syntactically, or semantically or whatever the case may be.

The typical morphological analyzer for any language tackles mainly three issues: the morphological lexicon, rewrite rules, and morphotactic rules. Apart from these, there are other issues which should be addressed when we think of any automatic processor for Tamil. First, the definition of a word. It is very difficult to define word meaningfully. For example, the utterance /paTittukkontirukkiRa:n/ 'is reading – he' can also be written as /paTittuk kontu irukkiRa:n/ or /paTittuk konTirukkiRa:n/, or /paTittukkonTu irukkiRa:n/. All the four ways of writing are available and acceptable. Second, writing a word with and without sandhi operation (morphophonemic changes) . The word /nu:RkaL/ 'books' can also be written as /nu:lkaL/ 'books'. Third, choice from many alternant forms. There are different past tense markers in Tamil, {t}, {tt}, {nt}, {in}, {R}, {T}, {n} and {i}³. The selection of a correct one for a given verb is not always a rule governed, mostly semantic based. Forth, more of exceptions in paradigm generation. Particularly, a considerable number of verbs in Tamil do not have many word forms. For example, /vil/ 'sell', /paya/ 'be afraid', /ini/ 'sweet', etc. do not have imperative forms; the second one does not have conjugation other than past tense. Any model developed for Tamil must tackle all these issues.

Approaching a Word

For computational purpose it is considered that a word is a chunk of characters between two consecutive spaces or punctuation marks. A word can be a lexeme or inflected or derived or compounded. When it is not a lexeme, it has to be segmented for meaningful components. A word is normally approached from the beginning, i.e. Left to Right (L to R) . But it can also be approached from the end, i.e. Right to Left (R to L) . In my analysis the latter approach is followed (Ganesan, 1994, 2003) . There are three reasons for opting R to L approach: one, the word final suffixes are limited when comparing to the stems, which are unlimited. Second, the suffixes are almost unique that they can indicate what could be the stem, whether a noun or verb, etc. Whereas many stems belong to more than one category. Third, a part of a stem can be a stem. For example, in the verb stems such as /aTikol/, /aTipaTu/, aTipaNi/, atipo:Tu/, aTipiTi/, aTiyeTu/, aTivai/ etc the first part /aTi/ forms a stem. These lead to backtracking⁴ and slow down the system processing.

Morphological Lexicon

A list of unique morphs constitutes the morphological lexicon of a language. In case of Tamil the list of morphs are not adequate for computational purposes. The morphological lexicon must have more information about the morphological components. They can be provided in the form of *feature structure* or *category – value* matrix. For example, take a noun /maram/ 'tree'. The feature structure of the stem is

/maram/	<div> cat = noun num = sing per = 3 gen = neut head = maram </div>
---------	--

These features are necessary to analyze the inflected word forms like /marama:/ 'is (it) tree?' /marankaL/ 'trees' /marattai/ 'tree- (acc) ' etc. and also to verify the agreement with the predicate in a sentence. These features are represented in the form a database, which I call a Machine Readable Dictionary (MRD) in my analysis. There are two MRDs: one for stems and the other for suffixes. The structure of the stem MRD is
Stem/type/grl.cat/status

The stem alternant gets a separate entry in the MRD. For example, the word /maram/ 'tree' has two entries, as given below.

maram/01/ian/id/
mara/02/ian/id/

The MRD has four fields. The first field gives the value / head of the stem. Second one gives the noun type that it belongs to. I have classified the noun into 19 types and the verb into 48 types based on the first suffix that the stem takes. Many scholars classified the verbs based on the past tense markers that the verb takes. But that classification is not handy to work computationally. These stem type numbers account for many features that are necessary to parse the word meaningfully and to have *feature unification* (explained in the next section) which is exhibited by the components. The

third field gives the grammatical category of the stem, in this case it is an inanimate noun (ian) and the forth field shows whether the stem can be an independent (id) or dependent (de) component. Similarly, the suffix MRD has the following structure.

Suffix/type/grl.cat/word tag /

Here too there are four fields. The first one gives the value of the suffix. The second field, “type” gives the direction to the machine, where the possible morphological components, that can go with this suffix in the next (towards left) level, are stored. While the third one gives the grammatical value of the suffix, the forth one marks the word level tag information, if the suffix is the determining component. Otherwise it is given as /0/.

a:L/36/3sf/FV/
um/10/part/0/

The selection of the first morpheme after the stem is only phonologically conditioned, but also, in some cases, grammatically and semantically conditioned. As all the words are enumerated in the MRD, the type number tackles the selection of the its adjacent component. The type number unifies the stem or a morphological component and its next morphological component. For example,

/maram/	$\left[\begin{array}{l} \text{cat} = \text{noun} \\ \text{per} = 3 \\ \text{no} = \text{sing} \\ \text{type} = 1 \\ \text{head} = \text{maram} \end{array} \right]$	/mara/	$\left[\begin{array}{l} \text{cat} = \text{noun} \\ \text{per} = 3 \\ \text{no} = \text{sing} \\ \text{type} = 2 \\ \text{head} = \text{mara} \end{array} \right]$
{nkaL}	$\left[\begin{array}{l} \text{cat} = \text{noun} \\ \text{type} = \text{S1} \\ \text{head} = \text{nkaL} \end{array} \right]$	{a:}	$\left[\begin{array}{l} \text{cat} = \text{noun} \\ \text{type} = \text{S2} \\ \text{head} = \text{a:} \end{array} \right]$

Here the feature matrix for /mara/ ‘tree’ and {nkaL} ‘plural marker’ unify on the basis of the category “type”. The stem /mara/ belongs to noun type 2 and {nkaL} belongs to S1, where S1 is the stem group which includes the noun classified as type 2. Similarly the feature matrix for /maram/ ‘tree’ and the morpheme {a:} ‘interrogative marker’ unify, as the stem group S2 includes the noun type 1. But, /maram/ + {nkaL} or /mara/ + {a:} does not unify. Therefore the formation of ungrammatical forms are avoided through feature based grammar.

Rewrite rules

The rewrite rules explain the phonological changes that take place when two morphological components are combined. This issue of rewrite rules is tackled by taking the alternant forms of morphemes into account in the morphological lexicon and by assigning proper value for the category called “type”, both in the stem MRD and the respective suffix MRD. The Finite State Automaton which models the morphological processes of Tamil is explained below.

An FSA is usually modeled by a program. The program receives a string from an input tape. It reads one character at a time from left to right. After reading the last character, it either accepts or rejects the string. An FSA consists of a finite number of *states*. First the machine will be in a state called *initial state*. As it scans the input string character by character it takes different states. One or more states will be marked as *final states*. The states are represented by small circles and are labeled q0, q1, q2, etc. as given in the following example. Here q0 is the initial state. q4 is the final state and is indicated by double circle. The transitions from one state to another is indicated by a *labeled arc*. The machine makes discrete moves from one state to another, the possible moves being just those given in the transition arc After the last character of the input is scanned if the automaton is in the final state, the input is *accepted*: otherwise, it is *rejected*. (For more details see Jurafsky, 2000) .

The information on the output tape i.e. PNG, past, and verb are reversed (as the input word was analyzed from the end) and are arranged as verb, tense and PNG. Now the ATN given below for the FV takes these as input and accept the given word is a finite verb.

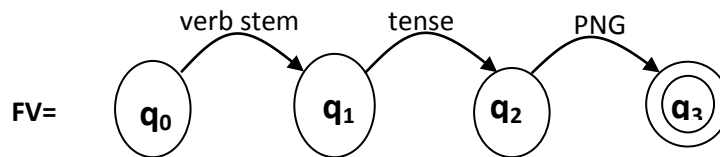
Morphotactic Rules

The morphotactic rules explain the arrangement order of morphemes in the word forms. The morphotactics is modeled through an Augmented Transition Network (ATN), an abstract machine like Finite State Automata (FSA) .

Similarly in ATN the transitions are labeled with strings (terminal or non terminal) . For example the morphotactic rule for a Finite verb is

FV = verb stem + tense marker + PNG marker.

This can be mapped to an ATN as follows:



Consider the following Finite verbs in Tamil.

- 1) /{pa:T}+ {in} + {a:n}'sing-past-he'
- 2) /{pa:T}+ {i} + {yatu}'sing-past-it'
- 3) /{pa:Tu}+ {kiR} + {a:n}'sing-pre-he'

The mapping for the three FVs are one and the same, but the morphemes combined are not the same. In (1) and (2) the past tense markers are {in} and {i} respectively, which are determined by the PNG markers, {a:n} and {yatu}⁶. Similarly in (1) and (3) the tense markers are {in} and {kiR}, which determine the choice of the stems {pa:T} and {pa:Tu}. Here the unification of features are on the category "type". In the former {a:n} belongs to type 36 and the morpheme {in} is listed in the next level under 36, but not the morpheme {i}. So, the system makes correct analysis and never allows the combination of morphemes {i} and {a:n} or {in} and {yatu} .

Conclusions

The twenty-first century will be the century of the total information revolution. The development of the tools for the automatic processing of the natural language spoken in a country or a whole group of countries is extremely important for the country to be competitive both in science and technology. To develop such applications, specialists in computer science need to have adequate tools to investigate language with a view to its automatic processing. One of such tools is a deep knowledge of both computational linguistics and general linguistic science. The models designed for Tamil morphology is purely based on linguistic description and those rules are converted into the abstract models.

Notes

1. The terms computational linguistics and Natural Language Processing are used synonymously, though there are differences between them.
2. Morphological lexicon is the inventory of all morphs in a language.
3. {t}, {tt}, {nt} and {in} are the four past tense markers. The others can be derived from these using some phonological rewrite rules. But, here all are taken as markers.

4. Backtracking a procedure where the system comes back to the same position a number of times when it fails in its previous steps.
5. In S file the verb types / noun types which can go with a particular type of suffix are listed as S1, S2, For example, if a suffix is marked with S1 in the type comes to the stem MRD, and if there is a match for the verb stem, then it compares whether the type number of the verb is included in the S1 list. If so, the word is accepted.
6. In this system {yatu} is taken as an alternant form of the PNG marker {atu}. It first segments {yatu} as a morpheme and later realizes to {atu}.

References

1. Ganesan, M. 1994 *"A Scheme for Grammatical Tagging of Corpora"*. In B. B. Rajapurohit (Ed.) **Technology and Languages**. Mysore: Central Institute of Indian Languages.
2. Ganesan, M. 2003 . *"Computational Grammar of Tamil"*. In R. Hari (Ed.) **Word Structure in Dravidian**. Kuppam: Dravidian University.
3. James, Allan. 1995. **Natural Language Understanding**. United States of America: The Benjamin/ Cummings Publishing Company.
4. Jurafsky, Daniel. and H. James Martin. 2000. **Speech and Language Processing**. Singapore: Pearson Education.
5. Kiraz, George Anton, 2001 *Computational Nonlinear Morphology*, London:CUP
6. Martin Kay, 1992 *"Unification"* in Michael Rosener and Roderick Johnson (Ed) **Computational Linguistics and Formal Semantics**, London:CUP
7. Methews, P 1974. *Morphology*, Cambridge

A Rule Based Converter of Formal Tamil to Colloquial Tamil (FT2CT) - An Application of Turing Machine

K. Arul Deepa (karuldeepa@gmail.com)

BNEC, Theni, Tamil Nadu, India

Dr. C. Deisy (cdcse@tce.edu)

TCE, Madurai, Tamil Nadu, India

ABSTRACT

Recent Machine Translation (MT) has focused on translation of non-colloquial language. Teaching the diglossic situation of a language like Tamil is always a challenging work, and requires e-learning tools to express the difference between the multiple faces of the language. Here the idea is to make use of the existing rules that defines the convergence of colloquial language from the formal one. This paper is an attempt to develop a fully automatic NLP system to translate the Formal Tamil (FT) to Colloquial Tamil (CT). Automation of the translation is achieved by expressing the conversion rules in terms of Turing Machine of Finite Automata paradigm. Though the Turing Machine handles word by word, the overall system is capable of converting a FT document into a CT document. Here the system performance is around 82%.

1. Introduction

In many Languages the colloquial version as found in typical conversations is different than text found in technical manuals, newspapers and books. The chance for a language which has two forms (colloquial and formal) is termed as diglossic situation. Tamil is one such language which has diglossic situation namely Formal Tamil (FT), mostly for writing and Colloquial Tamil (CT), for speaking. The language of writing differs considerably from the language of everyday conversation – so much so that there is no universally accepted way of writing the colloquial variety in Tamil script. Handling this diglossic situation within Tamil Nadu, would not be a major problem. But for the people who knows only writing Tamil (like Tamilians who have born and brought up in foreign countries) faces some difficulties to learn the colloquial Tamil. As most of the Tamil course materials available are for writing Tamil, those who do not have the practice of speaking in Tamil will find it difficult to identify the equivalence between these two forms. And another problem is that most of the times it is not possible to avoid the influence of formal Tamil in their speech. More than 40% of Tamil NRI kids are coming from Non-Tamil speaking backgrounds to primary schools [4]. This implied a lack of authentic context of the usage of the Mother Tongue Language at home.

A great deal of Machine Translation (MT) has focused on translation of non-colloquial language. Thus it would be a noteworthy work for Tamil Language and Tamil Computing researchers to focus on the development of colloquial Tamil teaching aid which gives the equivalent formal Tamil. The part of Tamil Language researchers have started considering this issue during 1980s itself [2]. A well defined collection of rules have been established for the conversion of colloquial Tamil to formal Tamil. So far 29 rules have been derived based on phonology and morphology of Tamil. Now it is our turn to automate these rules.

Thus a relationship between complexities in learning the oral language in parallel with formal version is an emerging field in the current research of natural language processing. Aim of this research paper is to present the automation of the derived rules of Formal Tamil (FT) to Colloquial Tamil (CT) converter using a virtual Turing Machine Automata as an implementation aid. This work is expected to bridge the gap between the formal and colloquial Tamil of NRI Tamil learners.

2. System Design – FT2CT

Any formal Tamil document will be the input for the system and the expected outcome is a corresponding colloquial Tamil document without any change in the meaning. Figure 1 shows the overall architecture for the converter. The word splitter splits the document into individual words. Before splitting, all the punctuation marks have to be hidden. Turing machine is constructed as a rule based system of FT2CT. Each word of the document will be the input for the Turing machine. TM construction is discussed in the fore coming chapters.

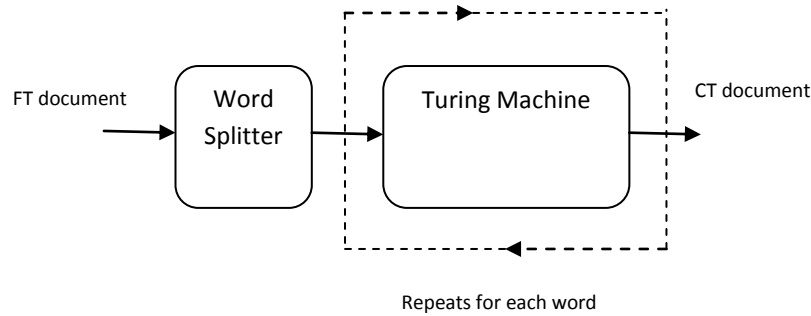


Figure 1 – FT2CT Architecture

3. Why TM?

Turing Machine is essentially a finite automaton that has a single tape of infinite length of which it may read and write data. One advantage of Turing machine over programs as representation of what can be computed is that the Turing machine is sufficiently simple that we can represent its configuration precisely. And comparing to programming languages the state of the variables can be followed in a more precise and understandable manner. [6]

Turing Machine has great computational capabilities that can be used as a general mathematical model for modern computers. It can model even recursive enumerable languages. Thus a Turing machine can model all the computable functions as well as the languages for which the algorithm is possible. Turing Machines can be programmed. Hence TM accepts a very large class of languages. So it is considered as a most powerful computational model.

Conceptually a Turing machine, like finite automata, consists of a finite control and a tape. At any time it is in one of the finite number of states. The tape has the left end but it extends infinitely to the right. It is also divided into squares and a symbol can be written in each square. However, unlike finite automata, its head is a read-write head and it can move left, right or stay at the same square after a read or write. Given a string of symbols on the tape, a Turing machine starts at the initial state. At any state it reads the symbol under the head, either erases it or replaces it with a symbol (possibly the same symbol). It then moves the head to left or right or does not move it and goes to the next state which may be the same as the current state. One of its states is the halt state and when the Turing machine goes into the halt state, it stops its operation.

3.1. Formal Definition for Turing Machine

$$M = (Q, \Sigma, \tau, \delta, q_s, B, F)$$

Where,

M: The Turing Machine

Q: The finite set of states of the finite control.

Σ : The finite set of input symbols.

τ : The complete set of tape symbols; Σ is always a subset of τ .

δ : The transition function $\delta(q, X) = (p, Y, D)$

Where,

P & q are states in M, q is the source and p is the destination

X is an input symbol

Y is a symbol in τ , for replacement

D is the direction for next move (left or right)

q_s : The start state, a member of Q, in which the finite control is found initially

B: The blank symbol. This symbol is in τ but not in Σ
F: The set of final or accepting states, a subset Q

4. System Construction

Main focus of this paper is the construction of Turing Machine for predefined rule.

4.1. Rule 1

இகரத்தில் தொடங்கும் ஒரு சொல்லில் இகரத்தை அடுத்து அகரம் அல்லது ஐகாரம் ஏறிய மெய்யொலிகள் வந்தால் இகரம் எகரமாக மாறுகிறது [1] & [2]. (எ.கா: திகைப்பு – தேகைப்பு, இதழ் – எதழ்)

TM1= ($\{q_s, q_1, q_2, q_f\}$, {TL}, {TL, B}, δ , q_s , B, $\{q_f\}$)

Where δ is,

$\delta(q_s, \text{Eagaram}) = (q_1, \text{Eagaram}, R)$
 $\delta(q_1, \text{Agaram}) = (q_2, \text{Agaram}, L)$
 $\delta(q_1, \text{Aigaram}) = (q_2, \text{Aigaram}, L)$
 $\delta(q_2, \text{Eagaram}) = (q_f, \text{Yegaram}, R)$

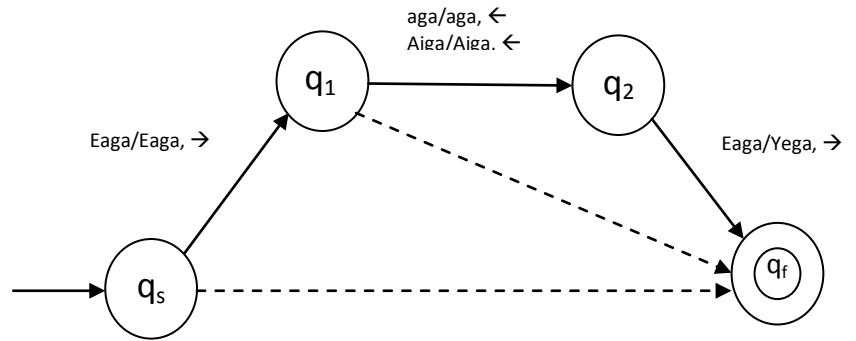


Figure 2 – TM for Rule 1

4.2. Rule 2

உகரத்தில் தொடங்கும் சொல்லுக்கு அடுத்து அகரம் அல்லது ஐகாரம் ஏறிய உயிர்மெய் எழுத்துகள் வந்தால் உகரம் ஒகரமாக மாறும். [1] & [2] (எ.கா: குடல் – கொடல்)

TM2= ($\{q_s, q_1, q_2, q_f\}$, {Tamil Letters}, {Tamil Letters, B}, δ , q_s , B, $\{q_f\}$)

Where δ is,

$\delta(q_s, \text{Vugaram}) = (q_1, \text{Vugaram}, R)$
 $\delta(q_1, \text{Agaram}) = (q_2, \text{Agaram}, L)$
 $\delta(q_1, \text{Aigaram}) = (q_2, \text{Aigaram}, L)$
 $\delta(q_2, \text{Vugaram}) = (q_f, \text{Vogaram}, R)$

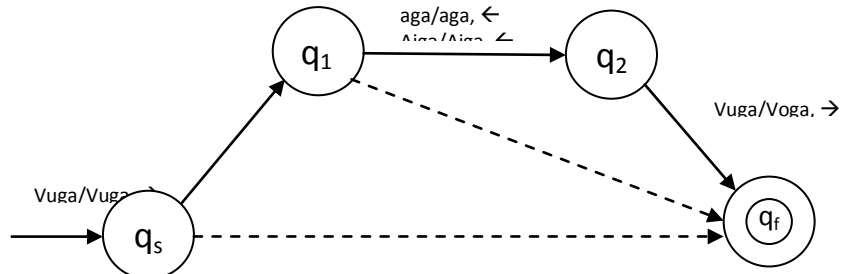


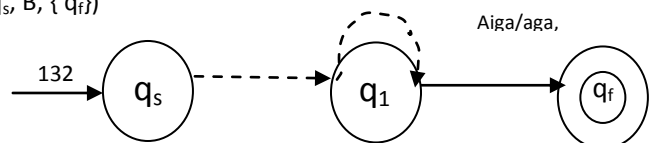
Figure 3 – TM for Rule 2

This machine is only for Rule1, the solid lines denote the actual transition and the dashed lines indicate the transitions for the fore coming rules.

4.3. Rule 3

ஒரு சொல்லின் இறுதியாக அல்லது இடையில் வரும் ஐகாரம் பேச்சுவழக்கில் சிதைந்து அகரத்திற்கும் எகரத்திற்கும் இடைபட்டு ஒலிக்கின்றது. (எ.கா: இசை – எச, பாதை – பாத)

TM3= ($\{q_s, q_1, q_f\}$, {Tamil Letters}, {Tamil Letters, B}, δ , q_s , B, $\{q_f\}$)



Where δ is,

$$\delta(q_1, \text{Aigaram}) = (q_f, \text{Agaram}, R)$$

Figure 4 – TM for Rule 3

4.4. Rule 4

இறந்தகால இடைநிலையாகிய “த்த” என்னும் இடைநிலை இகரத்தை அடுத்து வரும்பொழுது “ச்ச” என்றும், “ந்த” என்பது “ஞ்ச” என்றும் மாறுபடுகின்றன. (எ.கா: படித்தான் – படிச்சான், ஒழிந்தான் – ஒழிஞ்சான்)

$$TM4 = (\{q_s, q_1, q_2, q_3, q_4, q_f\}, \{\text{Tamil Letters}\}, \{\text{Tamil Letters}, B\}, \delta, q_s, B, \{q_f\})$$

Where δ is,

$$\begin{aligned} \delta(q_1, \text{Eagaram}) &= (q_2, \text{Eagaram}, R) \\ \delta(q_2, \text{th-th}) &= (q_3, \text{ch-ch}, R) \\ \delta(q_2, \text{ndh-th}) &= (q_4, \text{nj-ch}, R) \end{aligned}$$

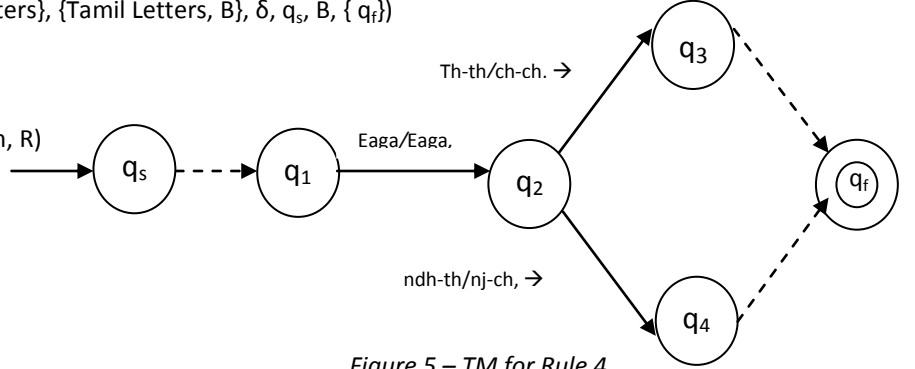


Figure 5 – TM for Rule 4

4.5. Rule 5

தனிக்குறிலை அடுத்துவரும் ணகர , னகர, ளகர, லகர மெய்களோடு முடிவடையும் சொற்களில் மெய்யெழுத்து இரட்டித்து உகரம் சேர்ந்து முடிவடைகின்றது . நெடிலைத் தொடர்ந்து வரும் லகர, ளகர மெய்யோடு முடிவடையும் சொற்களில் லகரம் உகரத்தோடு சேர்ந்து ஒலிக்கப்படுகிறது . (எ.கா: பல் – பல்லு, பால் – பாலு)

$$TM5 = (\{q_s, q_1, q_2, q_3, q_4, q_5, q_6, q_f\}, \{\text{Tamil Letters}\}, \{\text{Tamil Letters}, B\}, \delta, q_s, B, \{q_f\})$$

Where δ is,

$$\begin{aligned} \delta(q_s, \text{SVS}) &= (q_1, \text{SVS}, R), \delta(q_s, \text{LV}) = (q_4, \text{LV}, R) \\ \delta(q_1, n) &= (q_2, n, R), \delta(q_1, N) = (q_2, N, R) \\ \delta(q_1, l) &= (q_2, l, R), \delta(q_1, L) = (q_2, L, R) \\ \delta(q_2, B) &= (q_3, B, L), \delta(q_5, B) = (q_6, B, L) \\ \delta(q_4, l) &= (q_5, l, R), \delta(q_4, L) = (q_5, L, R), \\ \delta(q_3, n) &= (q_f, n-nu, R), \delta(q_3, N) = (q_f, N-Nu, R), \\ \delta(q_3, l) &= (q_f, l-lu, R), \delta(q_3, L) = (q_f, L-Lu, R), \\ \delta(q_6, l) &= (q_f, lu, R), \delta(q_6, L) = (q_f, L-Lu, R), \end{aligned}$$

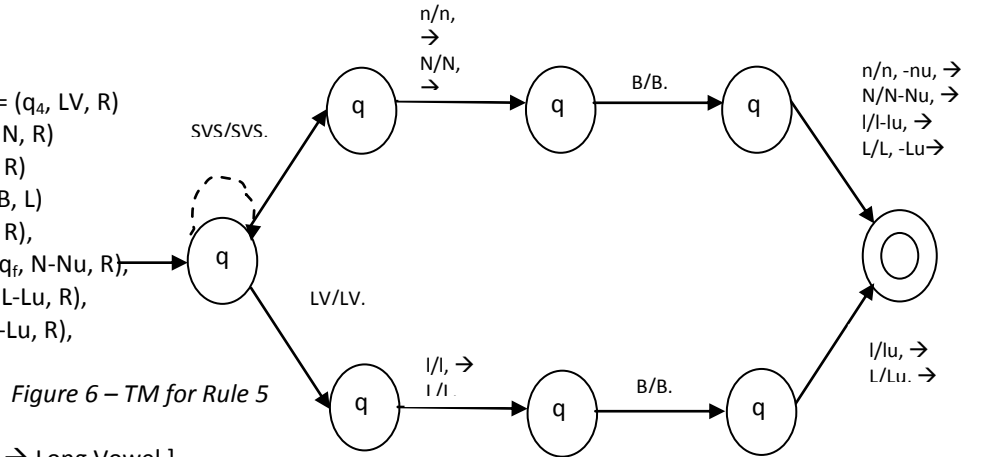


Figure 6 – TM for Rule 5

[SVS → Short Vowel in Solo and LV → Long Vowel]

4.6. Rule 6

தனிக்குறிலைத் தவிர்த்து னகர மெய்யோடும் மகர மெய்யோடும் முடிவடையும் சொற்களில் னகர, மகர மெல்லின ஒலிகள் கெட்டு , அதன் மூக்கொலித்தன்மை முந்தைய உயிரொலியில் ஏறுகின்றன . (எ.கா: வந்தான் – வந்தா~, போனோம் – போனோ~)

$$TM6 = (\{q_s, q_1, q_2, q_3, q_f\}, \{\text{Tamil Letters}\}, \{\text{Tamil Letters}, B\}, \delta, q_s, B, \{q_f\})$$

Where δ is,

$$\begin{aligned}\delta(q_s, !SVS) &= (q_1, !SVS, R), \\ \delta(q_1, n) &= (q_2, n, R), \delta(q_1, m) = (q_2, m, R), \\ \delta(q_2, B) &= (q_3, B, L), \\ \delta(q_3, n) &= (q_f, B, R), \delta(q_3, m) = (q_f, B, R)\end{aligned}$$

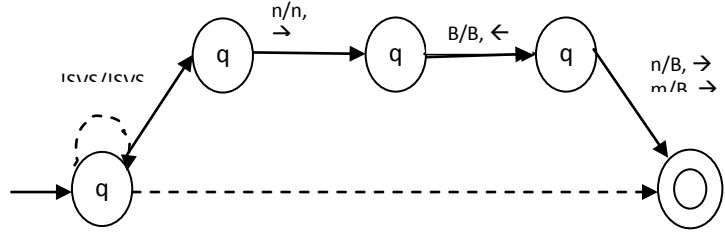


Figure 7 – TM for Rule 6

[!SVS \rightarrow Except SVS]

4.7. Rule 7

நெடிலைத் தொடர்ந்து யகர மெய்யில் முடிவடையும் சொற்கள் இகரத்துடன் சேர்ந்து முடிவடைகின்றன . (எ.கா: காய் - காயி) . குறிலைத் தொடர்ந்து யகர மெய்யில் முடிவடையும் சொற்கள் இரட்டித்து இகரத்துடன் சேர்ந்து முடிவடைகின்றன. (எ.கா: பொய் - பொய்யி) . ஐகாரம் ஏறிய ஓரெழுத்து ஒரு மொழிகளும் இகரம் ஏறிய யகரத்தோடு முடிவடைகின்றன. (எ.கா: கை - கையி) .

TM7= ({ $q_s, q_1, q_2, q_3, q_4, q_5, q_6, q_7, q_f$ }, {Tamil Letters}, {Tamil Letters, B}, $\delta, q_s, B, \{q_f\}$)

Where δ is,

$$\begin{aligned}\delta(q_s, SV) &= (q_1, SV, R), \delta(q_s, LV) = (q_4, LV, R), \\ \delta(q_s, lgaram) &= (q_7, lgaram, R), \delta(q_7, B) = (q_f, Y_i, R), \\ \delta(q_1, Y_i) &= (q_1, Y_i, R), \delta(q_2, B) = (q_3, B, L), \\ \delta(q_4, Y_i) &= (q_5, Y_i, R), \delta(q_5, B) = (q_6, B, L), \\ \delta(q_3, Y_i) &= (q_f, Y_i-Y_iE, R), \delta(q_6, Y_i) = (q_f, Y_iE, R)\end{aligned}$$

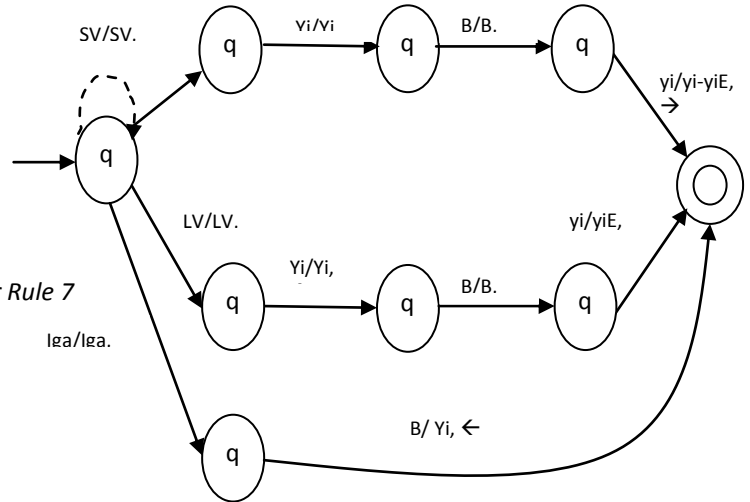


Figure 8 – TM for Rule 7

5. Performance Analysis

The rule based system of colloquial to Formal Tamil conversion is tested with 1282 words totally. Table 1 gives the performance of each rule.

Rule	# of Test words in FT	# of Correct CT words
1	243	203
2	231	190
3	152	134
4	271	220
5	289	233

6	172	141
7	167	129

TABLE 1 – Rule wise test results

Overall system performance is 81.97%. Reason for degradation is that for some words the converted version changes the intended meaning of the input word. For example as per rule1 the term “Eadai” will get converted into “Yadai”, which has completely different meaning (“eadai” means “in between” and “yadai” means “weight”) . Such special cases need to be handled separately.

6. Conclusion and Future Work

In this work, Turing Machine is constructed for 7 rules of FT2CT conversion. This system can act as a convenient tool to develop e-tutor kit to teach colloquial Tamil to NRI Tamil students. Likewise there are totally 29 rules exists. By implementing all the 29 rules a complete version of formal Tamil to colloquial Tamil converter can be published. It is highly modular allowing rapid extension to new target languages which has diglossic situation, provided the relevant rules need to be identified in advance.

Here the work is limited only for words, and is extendable to sentences and documents which are in colloquial form. Further work is planned to focus on the reverse process (i.e. Colloquial Tamil to Formal Tamil), which can create great impact on formalizing the output of speech to text converter and many like documentation works. This can also be added as a add-on feature in a Tamil Text editor for special purposes.

References

1. “பேச்சு தமிழை கற்பிப்பதில் கணினியின் பயன்பாடு”, Dr A Ra Sivakumaran, INFIT Tamil Language Conference.
2. N. Deiva Sundaram (Diglossic Situation in Tamil – A Sociolinguistic approach, Ph.D. Thesis submitted to the University of Madras, 1980, Chennai)
3. “E=mt2 E-Learning = M-Learning makes Tamil learning Squared” S.SWARNALATHA, INFITT Tamil Internet Conference, 2009.
4. Enhancing Learning of Tamil Language in a One-to-One Computing Environment Sivagouri Kaliamoorthy, INFITT Tamil Language Conference.
5. A Lexicalist Approach to the Translation of Colloquial Text, Fred Popowich et al.,
6. “Introduction to Automata Theory, Languages, and Computation”, John E. Hopcroft, Rajeev Motwani, Jeffery D. Ullman, Third Edition, Pearson Education.
7. “Colloquial Tamil – The Complete Course for Beginners”, R.E, Asher, E. Annamalai, Routeledge, Newyork

A Parsing Engine for the Study of Tamil (Finite –Tensed) Verb Structure

M.Somathasan

R.Saranya

CAS in Linguistics, Annamalai University

Introduction

Tamil is one of the South Dravidian Languages in India. It is agglutinative in nature and as a result, highly rich in morphological processing like inflection and derivation which cause many orthographic changes. Further, Tamil language has a rich literary and grammatical tradition. The earliest extent of Tamil work *tolkaapiyam* is a grammatical treatise which is considered to have been written before the beginning of the Christian era. It is very gratifying to note that the author of this great grammatical work was well known for many problems of Tamil morphology and morpho-phonemics that deal with internal organization of words including sandi rules. It is not claimed that it is a complete grammar. However, the attempts in morphological works like *morphological parsing of verbs* are still being done by many scholars in order to achieve a complete grammar in Modern Tamil in the aspect of natural language processing (NLP) .

Relating to parsing, exhaustive and detailed study of parsing Tamil verbs under various head words like infinitive, time expression, relative participle, verbal participle and person, number gender etc. is being worked out by various scholars at the Center of Advanced Study in Linguistics (Annamalai University), Tamil University, Madras University, and so on.

In general, to parse a word is to provide an explanation of its grammar. It includes the morphotactic structure of word and its rules. In parsing, the word is grammatically named; and the components or morphs of it are explained as well.

Once a formal parsing was experienced in the study of grammar to great extent; but now it has a less important place in the academic arena. For instance, the term *parsing* is now-a-days rarely used in the school level examinations, as well. In a number of examinations, however, the students are still asked to explain the *grammatical function* of certain words in a given context, and on the other hand it really amounts to *parsing* process. As a result, the students are subjected to follow a scheme of parsing properly. For instance, if we are asked to explain the grammatical function of particular words, we will generally find it necessary to give the scheme and we are free to add other notes to make it clear that we fully understand the grammatical points involved. Though the vital grammatical points are given below, for this paper the grammatical category *verb* is considered for the study of its internal structure.

- *Nouns*: Number, case, reason for case
- *Pronouns*: Kind, person, number, case, reason for case
- *Verbs*: Kind - transitive or intransitive or even ditransitive, conjugation - strong or weak, voice, mood, tense, person, number, agreement with subject.
- *Auxiliaries* should be taken with the verb to which they belong and the compound verb parsed as a whole.
- *Participles*: Tense, voice, word qualified
- *Gerunds*: Tense, voice, case (as for nouns)
- *Infinitives*: Tense, voice, if verb noun – case (as for nouns), if qualifying infinitive- word qualified
- *Adjectives*: Kind, degree (if any), word qualified
- *Adverbs*: Kind; degree (if any), word qualified (In the case of relative adverbs, name the antecedent)
- *Preposition*: Word governed

Objectives of the Paper

The objectives of this paper are:

- To parse the finite verbs (tensed - present, past and future) found in Modern Tamil with the labeling of its component tags through a computer program written in VB6.
- To create a *Tree Diagram* or *Parse-Tree* of the given verb, according to morphotactic rules. The visual parse tree will make us understand the components of the parsed verb easily.

- To compile a lexicon consisting of nearly 800 root verbs for this article. The lexicon is a must for returning/ extracting information, wherever as well as whenever necessary, while parsing the given verb.

Reviews

The various works on morphological analysis like morphological parsers have been developed by many scholars in India. But, for all we know, the works have not been seen on market, yet. Thus, as we, a student of linguistics at CAS in linguistics, Annamalai University and teacher of English at Trincomalee, Sri Lanka and the Associate Professor of linguistics at CAS in Linguistics, Annamalai University are very much interested in the field of NLP, we would like to engage ourselves into developing this user-friendly and detailed Tamil Verb Parsing Engine tool for modern Tamil finite verbs (tensed) as our complimentary contribution to our mother tongue, i.e. Tamil language spoken in India, Sri Lanka, and other countries in which the Tamil diasporas live.

Methodology of Morphological Parsing Engine

Generally morphological parsing engine is a computer program that performs the parsing process of the words based on morphotactic rules or analysis. In our program, the engine (program) obtains an input as a string (verb tokens) from the user for parsing process and it almost verifies whether the given string is a valid construction of the source language (Tamil) .

As far as the syntactic analysis is concerned, parsing is explained in two terms which are top - down parsing and bottom - up parsing. Similarly, left to right parsing and right to left parsing, are advocated for morphological analysis. To this seminar paper, the right to left parsing and the top - down parsing are applied for recognizing the verb root, verb stem and also the suffixes that are affixed to the stem, and for creating a parse-tree, respectively. Further, right to left parsing is an attempt to find the suffixes affixed with an input string and by starting from the right to left order and since it is an attempt to find the left most inflecting base for an input string, a right to left parser may require to do back tracking, i.e. the repeated scan of the given input until the base of the input string is obtained. Top - down parsing starts from root node of the given string and proceeds towards leaves.

Considering this article, the following three steps are involved in the methodology of the Parsing Engine for Modern Tamil Verbs (Tensed) .

a. Data Collection: For the present study, most of the finite verbs in modern Tamil have been collected from the *KRIYA* (Tamil Dictionary), text books and dailies as the primary source of data.

b. Coverage: In finite verbs, only the tensed (present, past, and future) verbs have been taken into consideration. The rest of the verbs like imperative, hortative and optative that belong to tense less finite verbs, are not taken for account for this paper.

c. Data Analysis: The above mentioned data are analyzed by applying computational linguistic approach for identifying the components or morphemes of the finite verbs like present, past and future tense markers; Person- Number - Gender suffixes; and the stem or its alternants that have to be parsed through the computer program.

Hypothesis

After implementing our concept in this program, the completion of our task (the product, i.e. *Tamil Verb Parsing Engine*) will be fully successful without any complication.

Description

1. Parse Tree (Tree Diagram or Verb Inflection Tree) of a String

Parse tree refers to an instance of a non-terminal in a given string by the right hand side of the production rule, which left hand side is the non-terminal to be replaced. It produces a new string from a given string, therefore the process

(inflection) can be used repeatedly to obtain a new string from a given string, and if a string obtained as a result of inflection contains only terminal symbols then no further parsing is possible. While inflecting a string, every inflection or derivation is considered to be a step in the parse-tree construction; then we get a graphical display of inflections of a string as a tree, called as *parse tree* of a string. It satisfies the following requirements:

1. All the leaf nodes of the tree labeled by terminals of the grammar.
2. The root node of the tree is labeled by the start symbol of the grammar.
3. A string whose inflection is represented by the given tree, is a string obtained by concatenating the labels of the leaf nodes of parse tree in a left to right order.

2. Parsing

The *Tamil Verb Parsing Engine (TVPE)* is a tool basically used to identify the morphemes or components of a complex or derived verb. It takes a derived verb as an input string and separates it into the root verb and the corresponding morphemes as tense marker and PNG marker. The function of each suffix or morpheme is labeled. It uses rule-based approach. The steps involved are as follows:

1. A tensed finite verb is given as an input. Then, the morphological parsing engine or analyzer starts scanning the string from right to left to looking for suffixes. A list of suffixes ruled out based on lexical morphology is maintained for this purpose.
2. It, then, searches for the longest match in the suffix list.
3. The morphological analyzer removes the last suffix; and then, determines its tag, finally adds it with the word's suffix list.
4. Then, it checks the remaining part of the verb in the lexicon of nearly 800 verb roots (for the purpose of this seminar article) and exits if the entry is found.
5. Based on the identified suffix, it generates the next possible suffix list.
6. The second step is repeated with the current suffix list.

Examples

Simple Present

1. பேசுகிறான் (peecukiRaan)

பேசு < Verb_Root >

-கிற < Present Tense >

-ஆன் < 3rd Person Singular Masculine >

2. பேசுகின்றான் (peecukinRaan)

1. பேசினான் (peecinaan)

பேசு < Verb Root >

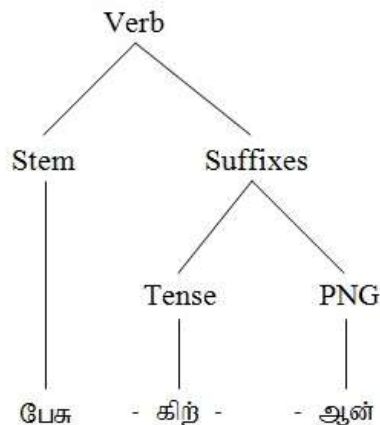
1. பேசுவான் (peecuvaan)

2. பேசும் (peecum)

பேசு < Verb_Root >

உம் < Future Tense & 3rd Person Singular Neuter > ie.

following parse tree.



Note that in the above examples, the suffixes italicized in the following set of grammar, are affixed with the verb stem *peecu* for denoting *tense* and *person-number-gender (PNG)*.

$V \rightarrow S_T T$ PNG (structure of finite verb - tensed)

$S_T \rightarrow peecu$ (verb_root)

$T \rightarrow P_R / P_A / F_U$

$P_R \rightarrow -kiR- / -kinR- / -kkiR- / -kkinR-$ (present tense markers)

$P_A \rightarrow -in- / -t- / -tt- / -nt- / -T- / -R- / -n- / -i-$ (past tense markers)

$F_U \rightarrow -v- / -pp- / -p-$ (future tense markers)

$-um-$ (future tense & 3rd person singular neuter)

$PNG \rightarrow -een / -oom / -iir / -irrkaL / -aay / -aan / -aaL / -aar /$

$/ -aarkaL / -anar / -atu / -ana$ (person-number-gender markers)

V stands for Verb S_T stands for STEM

T stands for TENSE PNG stands for PERSON-NUMBER- GENDER

P_R stands for PRESENT P_A stands for PAST

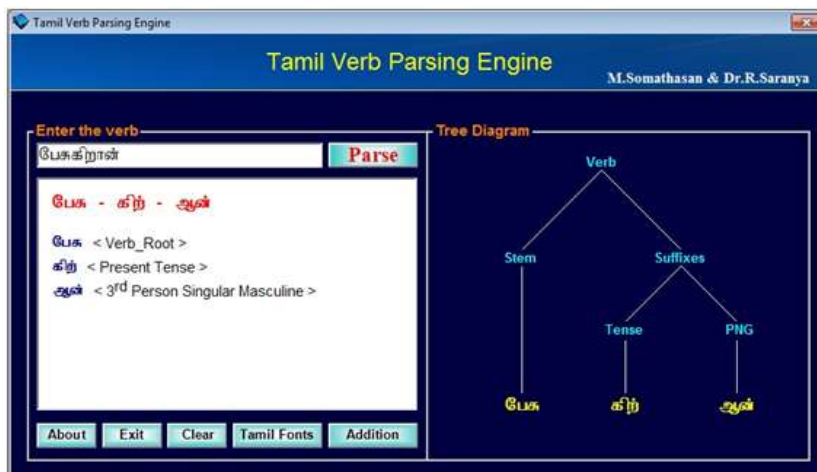
F_U stands for FUTURE

Implementation of Parsing Engine:

The parsing engine for the above grammar can be implemented by writing a program in VB6 to process the input string for performing the successful completion of the parsing. Thus, the parsing engine scans the characters of the input string, and whenever it gets the sufficient characters which constitute a token of the string, it reflects the morphologically parsed output. In order to perform this task, the parsing engine must know the tokens of the string to be implemented, i.e. it must know the keywords, identifiers, operators, delimiters, punctuation symbols, etc. Then, when it scans the source string, it will be able to return a suitable token, whenever it encounters a lexeme for a token. The lexeme is a term used to refer to a sequence of characters that expresses lexical as well as grammatical meanings in the given string that is matched by the pattern specifying the identifiers, operators, keywords, delimiters, punctuation symbols, etc. in the language. Therefore, the features involved in the design of the parsing engine are:

1. Specification of the token of the input string.
2. Designing a suitable recognizer for recognizing the tokens.

Sample Visuals:



Conclusion:

This attempt results the output as *a full-fledged efficient parser for Modern Tamil Verbs (tensed)* . The aim of this attempt is successfully attained, and it implies that the hypothesis is fully proved and accepted. Further, this tool will be very useful for all kinds of learners, language teachers, researchers, programmers, and scholars who are involved and work in the field of computational linguistics particularly in machine translation and Tamil verb (tensed) grammar analyzing and teaching, i.e. Tamil computation.

References:

1. Bellomo, T. (2009, April), *Morphological analysis and vocabulary development: Critical criteria*, Reading Matrix, 9 (1), 44-55
2. Anand Kumar M, Dhanalakshmi V, Soman K.P and Rajendran S. *A Sequence Labeling Approach to Morphological Analyzer for Tamil Language*, (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 06, 2010, 2201-2208
3. Joakim Nivre, *Dependency Grammar and Dependency Parsing*
4. E. Antworth. *PC-KIMMO: A two-level processor for morphological analysis*, Dallas, TX: Summer Institute of Linguistics.1990
5. D.J. Weber, H.A. Black & S.R. McConnel, *AMPLE: A tool for exploring morphology*, Dallas, TX: Summer Institute of Linguistics.1988.

Machine Learning of Sandhi Rules for Tamil

K.Rajan

Dr.V.Ramalingam

Dept. of Computer Science & Engineering

Dr.M.Ganesan

CAS in Linguistics

Annamalai University, Annamalainagar.

Abstract

This paper presents machine learning solutions to a practical problem of morphophonemic changes. The machine learning algorithms treat the morphology acquisition problem as one of '*learning to map*' one form of representation into another. For example, sequence of characters into morphemes, set of morphemes into a grammatical category.

The study which deals with all kinds of changes like addition, deletion, alternation etc., when two or more morphemes or words occur together is called '*Sandhi*'. In general the end of the first morpheme or word and the beginning of the following one are taken into account in *Sandhi*. The end and the beginning of morphemes that are added can be either a vowel or a consonant or consonant cluster. There are many combinations of such characters. These *Sandhi* rules should be explicitly specified for morphological analysis in a rule based system. These rules can be learned automatically by the system from the training samples and subsequently be applied for new inputs.

In this paper we discuss the machine learning model which learns the morphophonemic rules from the training data. When adding case suffixes to nouns, certain nouns undergo sound changes. The *Sandhi* rules describe these changes. In this model, we have trained the system to learn the noun inflectional patterns. The Decision tree algorithm is used for modeling.

Introduction

Many of the Indian languages are morphologically rich languages. Words in these languages are formed by combining meaningful constituents, called morphemes. Some morphemes are free morphemes and others are bound morphemes which join with free morphemes. The way in which these constituents are organised in word-forms follows morphological rules. Morphemes may have varied form in particular contexts, when we speak of allomorphic variations of that morphemes.

One important aspect of morphemes is that they change when grouped together. Actually, they influence each other phonologically. These changes are called morphophonemic changes so that we can say that they are changes in the phonological structure of words which occur when morphemes are grouped together.

When suffixes are added with words, a change in the word or addition of consonant or vowel takes place depending upon the nature of the suffix. The rules that make a change in words when adding suffixes are called sandhi rules. The word 'sandhi' is a Sanskrit word meaning 'meeting together'. Application of sandhi rules is necessary to correctly write and pronounce complex word forms.

Morphophonemics is the study which deals with the phonemic variations of a single morpheme in the given environment. It deals with all kinds of changes like addition, deletion, change etc when two or more morphemes occur together when they stand in conjunction, is called sandhi.

It is undeniable that the morphological aspect is indispensable for making sandhi rules. **Internal sandhi** occurs when two morphemes are added with one another and **external sandhi** occurs between two free-forms or two grammatical forms. The internal sandhi occurs in the morphological structure because, it relates to the sandhi of the morphemes within a word. So it is called "morphological sandhi". The latter one occurs in the syntactic structure of a language. So, it is called

compound or phrase or “syntactical sandhi”. A morphophonemic rule has the form of a phonological rule, but is restricted to a particular morphological environment.

An example of a morphophonological alternation in English is provided by the plural morpheme, written as “-s” or “-es”. Its pronunciation alternates between [s], [z], and [ɪz], as in *cats*, *dogs*, and *horses* respectively. A purely phonological analysis would likely assign to these three endings the phonemic representations /s/, /z/, /ɪz/. On a morphophonological level, however, they may all be considered to be forms of the underlying object //z//, which is a morphophoneme. The different forms it takes are dependent on the segment at the end of the morpheme to which it attaches – these dependencies are described by morphophonological rules. The alternative realisations of morphemes are called *allomorphs*.

Machine Learning

Machine learning is one of the important research and application areas of artificial intelligence (AI) . Machine learning is concerned with acquiring knowledge from an environment in a computational manner, in order to improve the performance. Machine learning is the capability of a computer to learn from experience (training data) and to extract knowledge from examples. A successful learner should be able to make general conclusions about the data it is trained on. This allows it to act appropriately in new situations. Many machine learning techniques have been applied to NLP tasks.

Most natural language processing (NLP) tasks require the translation of one level of representation to another. For example, in text to speech systems, the spelling representation of words is translated to a corresponding phonetic representation; in part-of-speech (POS) tagging, the words of a sentence are translated into their contextually appropriate POS tags. All these types of NLP tasks can be formulated as a *classification task*, and are therefore appropriate problems for machine learning methods. Classification-based learning starts from a set of instances (examples), each consisting of a set of input features (a feature vector) and an output class.

Decision Trees

Decision trees are one of the techniques for solving classification problems. Decision trees are easy to create, to understand, and to apply, and they are quite accurate. Other important methods are maximum entropy models, memory-based learning, neural networks and Genetic algorithms. Decision trees have been applied to a wide range of NLP problems including grapheme-to-phoneme conversion, part-of-speech tagging, tokenization, parsing, language modelling, classification of unknown proper names and spam detection. Many of these methods directly apply decision trees as classifiers [Black et al.].

Decision trees are learned from training data. Each data item consists of a set of features describing an object and the class of the object. Table 1 shows an example data items. Decision trees are recursively built beginning with the top most node by (1) computing the best test for the current node according to some *splitting criterion*, (2) creating a subnode for each possible outcome of the test, and (3) recursively expanding each subnode in the same way until a given *stopping criterion* is satisfied. Usually, the decision tree is simplified (*pruned*) in order to avoid overfitting of the training data[J.R.Quinlan].

Sandhi rules in Tamil

Tamil is an agglutinative and concatenative language, where morphemes are strung together to form long words. There are free morphemes and bound morphemes. The bound morphemes act as affixes which combine with other morphemes to form inflectional and derivational categories.

Affixes can be realised in many ways. For example, past tense is realized with *îç*, *îçîç*, *îçîç*, *øç*, and *îç* and plural is realized with *è÷ç*, *èçè÷ç*, *éçè÷ç*. These alternations are conditioned by the phonological properties of the surrounding sound segment of the affixes. However, morphotactic constraints which involve the ordering relations among morphemes are also effective in realisation of word forms. That is, affixes cannot be attached one after another in a free order. Thus in

designing language analysers, one should design and implement the mechanism that performs the phonological alternations and check the validity of ordering for the realisation of morphemes.

When adding a particular morpheme with another morpheme, the changes occurred depends not only on the characters but also on the type of the morpheme. In the following case, unless we know whether the first member of the given combination is a noun or a verb or something else, it may not be possible for us to predict the resultant form.

நாடு N + ஐ → நாட்டை

நாடு V + ஐ → Not applicable

படி N + ஆல் → படியால் (Instrumental case)

படி V + ஆல் → படித்தால் (Conditional verb)

Here, neither the phonemes nor the syllabic structure is responsible for the difference found in their sandhi behaviour. The factor which is responsible for the difference is the morphological structure, that is, the first one is of noun plus suffix type and the next one is verb plus suffix.

In Tamil, a noun is inflected for both number and case, the plural suffix is first added to the noun stem, optionally followed by the euphonic increment -in/-an, and then the case suffix is added. When a noun is inflected for case only, a case suffix is added to the stem or to the oblique stem, if any. When different types of suffixes are added to the noun, various morphophonemic rules operate.

When adding case suffixes to nouns, certain nouns undergo sound change. The Sandhi Rules describe which nouns undergo which changes. When a word that ends in a vowel and a suffix that begins in a vowel are added together, a glide - either *ய்* or *வ்* - is inserted in between. Selection of either *ய்* or *வ்* is determined based upon whether the final vowel is one of the 'front' vowels இ, ஈ, எ, ஏ, or ஐ one of the 'back' vowels அ, ஆ, ஒ, ஓ உ or ஊ respectively. Front vowels take the glide *ய்* and back vowels take the glide *வ்*.

Words with the syllable structure CVC, nouns ending in அம், டு and று undergo a number of different changes when a case suffix is added. Some case suffixes such as interrogative suffix, conjunctive suffix and the emphatic suffix (ஆ, உம், ஏ) do not follow these rules. Some nouns ending with consonants with a syllable structure CVC where the vowel is short is added with any suffix, the final consonant is doubled.

Developing a rule based system to model these sandhi changes involves many complex linguistic rules. But, a supervised machine learning techniques can be used to develop such models from the sample data. The morphophonemic rules can be learned automatically from the examples given to the system. The following table 1 shows the types of sandhi changes on the given noun stems when case suffixes, adjective marker, adverbial marker and plural suffix are added.

Table 1. Various noun inflections and plural forms

		Class 1	Class 2	Class 3	Class 4
Category	Suffixes	ய்-inserted	வ்-inserted	Doubling	த்த inserted (-ம்)
Nominative	NULL	படி	பசு	கண்	படம்
Accusative	-ஐ	படியை	பசுவை	கண்ணை	படத்தை

Instrumental	-ஆல்	படியால்	பசுவால்	கண்ணால்	படத்தால்
Dative	-உக்கு	படிக்கு	பசுவுக்கு	கண்ணுக்கு	படத்துக்கு
Locative	-இல்	படியில்	பசுவில்	கண்ணில்	படத்தில்
Ablative	-இலிருந்து	படியிலிருந்து	பசுவினிருந்து	கண்ணிலிருந்து	படத்திலிருந்து
Sociative	-உடன்/ ஓடு	படியுடன்	பசுவுடன்	கண்ணுடன்	படத்துடன்
Genitive	இன்/இனுடைய	படியினுடைய	பசுவினுடைய	கண்ணினுடைய	படத்தினுடைய
Adjective	-ஆன	படியான	பசுவான	கண்ணான	படமான (7)
Adverb	-ஆக	படியாக	பசுவாக	கண்ணாக	படமாக (7)
Plural	-கள்	படிகள் (7)	பசுக்கள் (9)	கண்கள் (7)	படங்கள் (8)

Sample data	Class
படி + ஐ → படியை	1 - ி-insertion
பசு + ஐ → பசுவை	2 - வ்-insertion
கல் + ஐ → கல்லை	3 Consonant Doubling
படம் + ஐ → படத்தை	4 -த்த் inserted (-ம் removed)
விறகு + ஐ → விறகை	5 - உ deletion
காடு + ஐ → காட்டை	6 - உ deletion, Doubling
கால் + ஐ → காலை	7 No change (Normal)
Various plural forms of the given noun stems	
படி + கள் → படிகள்	7 Plural form 1 (No change)
படம் + கள் → படங்கள்	8 Plural form 2
பசு + கள் → பசுக்கள்	9 Plural form 3

கல் + கள் → கற்கள்	10 Plural form 4 (Consonant change ல்-ற்)
சொல் + கள் → சொற்கள்	
ஆள் + கள் → ஆட்கள்	11 Plural form 5 (Consonant change ள்-ட்)
நாள் + கள் → நாட்கள்	

Feature representation

Different types of sandhi changes take place on different stem and suffix combinations. Training data is collected from the corpus for the above classes. The class numbers represent the type of sandhi changes. Maximum of 10 characters from the end of stem and 5 characters from suffix are used as nominal (symbolic) features. As the category of stem in this model is only noun, it is not specified as input explicitly. The class number is given as the output during training. The table 2 shows the feature vector for காடு + ஐ and படம் + கள் which belongs to output categories 6 and 8 respectively.

Table 2. Feature vector with input features and output class

Stem feature											Suffix Feature					Class
1	2	3	4	5	6	7	8	9	10		11	12	13	14	15	16
X	X	X	X	X	X	க்	ஆ	ட்	உ		ஐ	X	X	X	X	6
X	X	X	X	X	ப்	அ	ட்	அ	ம்		க்	அ	ள்	X	X	8

The training data in the specified format is supplied to the decision tree classifier that is used to classify feature vector of the new words. The classifier generalizes from the training data. The output classes match the output of the rule based system. The decision tree learning algorithm was tested on a Java based WEKA open source machine learning tool.

Conclusion

The sandhi rules are important in word formation. An automatic sandhi checker or sandhi generator can be used as a critical component in morphological analyser, generator and text-to-speech synthesizer. In this paper we have presented an approach for a machine learning model which acts as a classifier to classify sandhi changes on noun inflections. Studying the capabilities of these types of algorithms on various word categories with different representation schemes are encouraging.

References:

1. Aha, D. W., D. Kibler, & M. Albert (1991), Instance-based learning algorithms. *Machine Learning* 6:37–66.
2. Kienappel, A. K., & R. Kneser (2001), Designing very compact decision trees for grapheme-to-phoneme transcription. *Proceedings of the Eurospeech Conference* 1911–14.
3. Magerman, David (1995), Statistical decision-tree models for parsing. *Proceedings of the 33rd Meeting of the ACL*, 276–83.
4. Quinlan, J. Ross (1986), Induction of decision trees. *Machine Learning* 1:81–106.
5. Sproat, Richard (1992), *Morphology and Computation*. Cambridge, MA: MIT Press.
6. M.A.Nuhman (1999), Basic Tamil Grammar, University of Peradeniya, Sri Lanka.
7. Thomas Lehman (1989), A grammar of modern Tamil, Pondicherry Institute of Linguistics and culture.
8. Van den Bosch, A., Walter Daelemans, & T.Weijters (1996), Morphological analysis as classification: an inductive learning approach. *Proceedings of NEMLA*.

Dependency Parser for Tamil classical literature- *Kurunthokai*

Dr.V.Dhanalakshmi

Head of the Department, Department of Tamil, FSH, SRM University, Kattankulathur-603203

M.Anand Kumar

Assistant Professor, CEN, Amrita Vishwa Vidyapeetham, Coimbatore.

C.Murugesan

Research Associate, Department of Tamil, FSH, SRM University, Kattankulathur-603203

Introduction

Dependency parsing is a **form of syntactic parsing** of natural language based on the theoretical tradition of dependency grammar [4]. It is a process of analyzing the dependency structure of a sentence. Tamil is always head-final language. The verb comes at the end of the clause with a typical word order of Subject Object Verb (SOV) . However, Tamil language allows word order to be changed, making it a relatively word order free language. In Sangam literature poems the syntactic structure are more complex due to this relatively free word order nature of Tamil language. Dependency grammar is better suited for languages with free or flexible word order. Dependency parser for the contemporary Tamil language has been already developed [5]. This paper proposes a dependency parser for Tamil classical literature – kurunthokai.

Kurunthokai

The Sangam literature contains about 50, 000 lines of poetry contained in 2381 poems attributed to 473 poets including many women poets. '**Kurunthokai**' is one of the eight anthologies (Ettuth thokai), which belongs to the Sangam age. It has 401 short poems of 6 to 8 lines, and written by 205 poets. The syntactic structure of the poem is very complex since many clauses are embedded in it. So, developing an automatic dependency parser for the Tamil classical literature- Kurunthokai will be a most wanted work of the modern era.

Architecture of the Proposed System:

The architecture of the proposed system is given below:



Fig:1 Architecture of the Proposed System

The proposed system contains three modules:

1. POS Tagger
2. Dependency Parser

Part-of-Speech Tagger: The Part of Speech (POS) tagging is done to obtain the labeled Part of Speech or other lexical class marker such as noun, verb, adjective, etc in the given verse .

Dependency parser:. Dependency parsing is done to obtain the grammatical relation of the words in the verse. This decides which relations (subject, object, location, etc.) they have with the main verb or the root word. This is also used to obtain the types of clauses present in the verse.

Methodology of the proposed system

The methodology adopted to develop the modules is the Machine learning approach. Machine learning deals with techniques that allow computers automatically learn to make accurate predictions based on past observations. The major focus of machine learning research is to extract information from annotated data automatically, by computational and statistical methods. So the annotated data plays a key role in supervised machine learning techniques.

There are two main tasks involved in supervised machine learning; 1) . learning or training, 2) . Prediction. The system is trained with a set of examples called training data. The primary goal is to automatically acquire an effective and accurate model from the training data. This is a typical task for supervised learning and is usually called concept learning or learning from examples. The accuracy of the model will be better if the training data size is larger. The second phase of machine learning is the prediction, wherein a set of inputs is mapped into the corresponding target values, based on the generated model. The main challenge of machine learning is to create a model with high prediction accuracy. Here the training data will be generated for all the verses in the Kurunthokai, so the testing data will be of gold standard data.

POS Tagger

MBT (Memory Based Tagger) is an approach to POS tagging based on Memory-based learning. It is an extension of the classical k-Nearest Neighbor (k-NN) approach to statistical pattern classification. Here all instances are fully stored in memory and classification involves a pass along all stored instances. The approach is based on the assumption that reasoning is based on direct reuse of stored experiences rather than on the application of knowledge (such as rules or decision trees) abstracted from experience. Hence the tagging accuracy for known words will be cent percentage. As our test data will be of gold standard data and we use MBT for POS tagging the kurunthokai verse.

Customized POS tagset

The customized tagset contains 23 tags. We are not considering the inflections. The customized tagset is developed with the perspective of Tamil grammar Tholkaappiyam [6]. The customized tagset with examples is given below Table (1) .

S.NO	Main Category	Tag Category	Tag Set	Example
1.	பெயர் (Noun)	பெயர்ச்சொல் (Common Noun)	<NN>	மரம்
2.		சிறப்பு பெயர் (Proper Noun)	<NNP>	அதியமான் முசிறி
3.		பதிலிடு பெயர் (Pronoun)	<PRN>	சுட்டுப்பெயர்கள் அவன், அது

4.		வினா பதிலிடு பெயர் (Interrogative Pronoun)	<PNI>	யாவன், யாது, யார்
5.		பிரதி பதிலிடு பெயர் (Reflexive Pronoun)	<PNR>	தான்
6.	எண்ணுப் பெயர் (Numbers)	Cardinal	<CRD>	ஒன்று, 100, க
7.		Ordinal	<ORD>	ஒன்றாம், ஒன்றாவது
8.	வினை (Verb)	வினைமுற்று (Finite Verb)	<VF>	வந்தான்
9.		வினை எச்சம் (Adverbial Participle)	<AVP>	வந்து
10.		(Infinitive Participle)	<IVP>	வர
11.		நிபந்தனை எச்சம் (Conditional Participle)	<CVP>	வந்தால்
12.		பெயர் எச்சம் (Adjectival Participle)	<AJP>	வந்த
13.		தொழில் பெயர் (Gerundial Noun)	<GN>	வருதல்
14.		பெயரடைப் பெயர் (Adjectival Noun)	<AN>	நல்லவன்
15.		வினையாலணையும் பெயர் (Participle	<PN>	வந்தவன்

		Noun)		
16.	உரிச் சொல் Modifiers	பெயரடை (Adjective)	<ADJ>	நல்ல, அழகான
17.		வினையடை (Adverb)	<ADV>	அழகாக, நேற்று
18.		அளவடை (Quantifiers)	<QF>	பல, சில
19.	இடைச் சொல்	ஒட்டு (Clitic)	<CL>	உம், ஓ, ஏ, ஆ, தாம், தான்
20.		உவம உருபு (Comparision Word)	<CW>	ஒப்ப, உறழ
21.		வேற்றுமை உருபு (Case Marker)	<CM>	ஐ, ஆல், கு
22.		பின்னுருபு (Postpositional word)	<PPO>	போல, குறித்து
23.		அசை (Particle)	<PAR>	சின், மன், ஒரு

Table 1: Customized POS Tagset

Dependency Parser

Parsing is related to the automatic analysis of syntactic structure according to a grammar. It is usually performed after basic morpho-syntactic (POS) categories have been identified in a text. Dependency parsing is a form of syntactic parsing of natural language based on the theoretical tradition of dependency grammar. The dependency structure of the sentence is defined using dependency label and dependency head. Whenever two words are connected by a dependency relation, we say that one of them is the head and the other is the dependent, and that there is a link connecting them. In general, the dependent is in the form of modifier, object or complement. The head plays the larger role in determining the behavior of the pair.

In our proposed dependency parser system the representation the source of the edge represents the modifier and destination points to the head word. The customized dependency tagset contains 35 grammatical relations. MALT¹ parser tool is used to obtain the dependency parser for kunthokai verse. An example of the data set is given below.

S.no	Word	POS tag	Depn arc	Depn tag
1	யாயும்	<NN>	2	<CJ_SUB>
2	ஞாயும்	<NN>	3	<CL_SUB>
3	யார்	<PNI>	4	<COP>
4	ஆகியரோ	<VF>	23	<F_CL>
5	எந்தையும்	<NN>	6	<CJ_SUB>
6	நுந்தையும்	<NN>	9	<CL_SUB>
7	எம்	<PNI>	8	<ATT>
8	முறைக்	<NN>	9	<LOC_ATT>
9	கேளிர்	<NN>	23	<F_CL>
10	யானும்	<PN>	11	<ATT>
11	நீயும்	<PN>	14	<CL_SUB>
12	எவ்	<PNI>	13	<ATT>
13	வழி	<NN>	14	<LOC_ATT>
14	அறிதும்	<VF>	23	<F_CL>
15	செம்	<ADJ>	16	<ATT>
16	புலப்	<NN>	18	<ATT>
17	பெயல்	<VN>	18	<ATT>
18	நீர்	<NN>	19	<CL_SUB>
19	போல	<PPO>	23	<PP_CL>
20	அன்புடைய	<ADJ>	21	<ATT>
21	நெஞ்சம்	<NN>	22	<R_SUB>

¹<http://maltparser.org/userguide.html>

22	தாம்	<PNR>	23	<NSUB>
23	கலந்தனவே	<VF>	0	<ROOT>

Conclusion

MALT parser tool is language independent tool used for dependency parsing which is implemented for several languages. Using this tool the dependency labels and position of head in Tamil language is obtained.

References

1. Akshar Bharati, Rajeev Sangal, Dipti Misra Sharma and Lakshmi Bai, AnnCorra: Annotating Corpora Guidelines for POS and ChunkAnnotation for Indian Languages, Technical Report, LanguageTechnologies Research Centre IIIT, Hyderabad, 2006.
2. Dhanalakshmi V, Anand kumar M, Vijaya M.S, Loganathan R, Soman K.P and Rajendran S. (2008), "Tamil Part-of-Speech tagger based on SVMTool", Proceedings of the COLIPS International Conference on natural language processing (IALP-2008), Chiang Mai, Thailand.
3. Dhanalakshmi.V, Anand Kumar M, Soman K P and Rajendran S (2011), "Shallow Parser for Tamil" Tamil internet Conference (2011), University of Pennsylvania, Philadelphia, USA.
4. Nivre J (2005), "Dependency grammar and dependency parsing", Technical Report MSI report 05133, Vxj University: School of Mathematics and Systems Engineering.
5. Lehmann Thomas (1983), "A Grammar of Modern Tamil", Pondicherry: Pondicherry Institute of Linguistics and Culture.
6. Tholkaappiyam, sollathikaaram, kazaka veliyidu (1994) .

POS tagging for Classical Tamil Texts

G.Palanirajan

Associate, Central Institute of Classical Tamil, Chennai.
matrixpalani@gmail.com

Abstract

This paper presents a rule based model of Parts of Speech (POS) tagset for Classical Tamil Texts (CTT) . The noun forms are Type pattern (a word form), verb forms are Lemmatization (grouping together all of the different inflected forms of the same word) . This is based on form agreement method. This is a very efficient and novel approach because Tamil Language has a build-in system of agreement/concord in the sentence. Classical Tamil Tagset is divided into two basic classifications, namely noun morphology and verb morphology.

1. INTRODUCTION

There are a few POS Tagsets for Modern Tamil Language such as IL-ILMT, AU-KBC, IL-POST Microsoft, LDC-IL Mysore, etc. There are various methods in POS tagsets like linguistic rules, stochastic models and hybrid approaches, and each approach has its own merits and demerits (Rajendran. 2007) . In this context, Classical Tamil further presents a challenge in developing an automatic POS tagger as the language is highly inflectional and morphologically rich. Hence, it is essential to consider text processing prior to POS tagging in order to achieve high performance and more reliability.

2. POS TAGGER FOR CLASSICAL TAMIL

To assign an appropriate Parts of Speech to a word in a sentence automatically this is our fundamental objective. Input as a string of word and an output as a single best tag for each word is provided in this model. Words are grouped together into classes (sets) which show similar syntactic behavior, often typical semantic type. Word categories are systematically related to morphological process. Significant amount of information about the word and its neighbors. The tagset consists of two broad subcategories, namely 1) closed class types (relatively fixed membership involving Determiners, Pronouns, Postpositions, conjunctions, Auxiliaries, Particles and Numerals), 2) open class types (relatively unfixed membership connected with Nouns, Verbs, adjectives and Adverbs) .

3. BASIC IDEA OF THE POS TAGGING

The noun form carries plural marker, case markers, clitics, postposition, etc. The noun form may be a root or a stem or an oblique base form. The root form means that there is no future segmentation (kāl 'leg', kaṇ 'eye') and the base form means that the root word can be added with any suffix for example (kaṇṇaṇ (kaṇ+aṇ) 'male person', kuṇṇam (kuṇṇu+am) 'hill'), the oblique means the form that carries a suffix or an equivalent to another form kiṇṇaru (kiṇṇu) 'well', āṇṇu (āṇṇu) 'river', eṇ (nāṇ) 'I', taṇ (tāṇ)) 'my'.

CLASSICAL TAMIL POS TAGS (50)

1. ADJ - Adjective	பெயரடை
2. ADV - Adverb	வினையடை
3. CCD - Co-ordination	வாக்கிய இடைச்சொல்
4. CCS - Subordination	தொடர் இடைச்சொல்
5. DEA - Demonstrative Adjective	சுட்டுப்பெயரடை
6. DEM - Demonstrative	சுட்டுப்பெயர்
7. DIA - Interrogative Adjective	வினாப்பெயரடை
8. DIN - Interrogative	வினாப்பெயர்
9. DRF - Reflexive	தற்குட்டுப் பெயர்
10. MA – Accusative	செயப்படுப் பொருள் வேற்றுமை

11. MD - Dative	கொடை வேற்றுமை
12. MG - Genetive	உடைமை வேற்றுமை
13. MI – Instrumental	கருவி வேற்றுமை
14. ML - Locative	இட வேற்றுமை
15. MOB - Oblique	சாரியை
16. MP - Plural	பன்மை உருபு
17. MSO - Sociative	உடனிகழ்ச்சி வேற்றுமை
18. MS – Source	நீங்கல் வேற்றுமை
19. NC - Common Noun	பொதுப் பெயர்
20. NCR - Cardinal	எண்ணுப்பெயர்
21. ND - Derived Noun	ஆக்கப் பெயர்
22. NOR - Ordinal	எண்ணுப்பெயரடை
23. NP - Proper Noun	இயற்பெயர்
24. PADV - Adverbial Particle	வினையடை இடைச்சொல்
25. PAR - Particle	இடைச்சொல்
26. PCLI - Clitic	ஒட்டுகள்
27. PCOM - Comparative Particle	உவம உருபு
28. PEXP - Expletive	அசைச்சொல்
29. PINFR - Intensifier	வல்லடை
30. PINT - Interjection	வியப்பிடைச் சொல்
31. PONOM - Onomatopoeia	ஒலிக்குறிப்பு
32. POS - Postposition	பின்னுருபுகள்
33. PU - Punctuation	நிறுத்தக்குறியீடு
34. UNK -Others	பிற
35. VAX - Auxiliary verb	துணைவினை
36. VB - Verbal Base	வினையடி
37. VF - Verb Finite	வினை முற்று
38. VFA - Appellative Verb Finite	குறிப்பு வினை முற்று
39. VFP - Finite Participle	முற்றெச்சம்
40. VINF - Infinitive	செயவென் எச்சம்
41. VN - Verbal Noun (non-past)	இறப்பில்தொழிற் பெயர்
42. VNP - Verbal Noun (past)	தொழிற் பெயர்
43. VOP - Optative	வியங்கோள்
44. VP - Verbal Participle	வினையெச்சம்
45. VPA - Appellative Verbal Participle	குறிப்பு வினையெச்சம்
46. VPC - Conditional Participle	நிபந்தனை எச்சம்
47. VPN - Participial Noun	வினையாலணையும் பெயர்
48. VPNA- Appellative Participial Noun	குறிப்பு வினையாலணையும் பெயர்
49. VRP - Relative Participle	பெயர் எச்சம்
50. VRPA - Appellative Relative Participle	குறிப்புப் பெயரெச்சம்

3.1 Plural marker

Plural means a grammatical form that designates more than one of the things specified.

மலர்+கள் = மலர்கள் (NC) ; flower+MP = flowers

3.2 Case markers

A case marker is a suffix, it occurs after noun and pronoun.

Accusative Case marker –ஐ (ai), Instrumental case marker –ஆல் (āl), Sociative case marker– ஓடு (ōṭu), Dative case marker – க்கு (kku), Genitive case marker – இன் (in) and Locative case marker – கண் (kaṇ)

3.3 Postpositions

A postposition is a word that occurs after noun which may or may not be an appendage to the case marker.

மலர் + உள் (NC) = மலருள்; flower+postposition = 'inside flower'

பெயர்நிலை + முன்னர் (NC) = word + postposition = 'before word'

3.4 Vers

In general, a Tamil verb form takes tense marker along with person, number, gender markers (PNG) . There are no multiple meaning features of tense markers and PNG markers. But it provides many conjugated forms of the verbs. It is well known that in almost all natural languages, verbs are considered to be the most important part of speech. Verbs play a very important role in languages. As Tamil verbs are inflected to various grammatical categories as distribution of the Tamil verbal base is necessary. In such a way, the verb forms consist of verb root and conjugations are classified. There are sixteen major types of verb conjugations as follows 1) Verbal Base (Vb), 2) Infinitive (Vinf), 3) Verbal Noun (non-past) (Vn), 4) Verbal Noun (past) (Vnp), 5) Verbal Participle (Vp) 6) Appellative Verbal Participle (Vpa), 7) Relative Participle (Vrp), 8) Appellative Relative Participle (Vrpa) 9) Conditional Participle (Vpc), 10) Participial Noun (Vpn), 11) Appellative Participial Noun (Vpna), 12) Verb Finite (Vf), 13) Appellative Verb Finite (Vfa), 14) Finite Participle (Vfp), 15) Auxiliary verb (Vax) and 16) Optative (Vop)

3.5 Adjectives (ADJ)

A word or phrase naming an attribute, added to or grammatically related to a noun to modify or describe it.

There are two types of adjectives. They are inherent and derivative adjectives. Both are marked as single tag. Ex.

எல்லா, பல (ellā, pala)

3.6 Adverbs (ADV)

An adverb is a part of speech. It is a part of language verb, there are two types of adverbs. They are inherent and derivative adverbs. Both are marked as single tag (அங்கு, ஆங்கனம் (aṅku, āṅkanam)).

3.7 Pronouns (D)

A pronoun is a word that takes the place of a noun. (அவன், அவள் (avaṇ, aval)).

3.8 Particles (P)

A particle is a function word that does not belong to any of the inflected grammatical word classes (such as nouns, pronouns, verbs, or articles) . It is mostly used for words that help to encode grammatical categories (such as negation, mood or case), or fillers or discourse markers that facilitate discourse such as well, ah, anyway, etc. என்று, என்பது (eṇṇu, eṇṇatu)

3.9 Clitics (PCLI)

Clitics may belong to any grammatical category, though they are commonly determiners, or adpositions.

மலர் + ஏ = மலரே

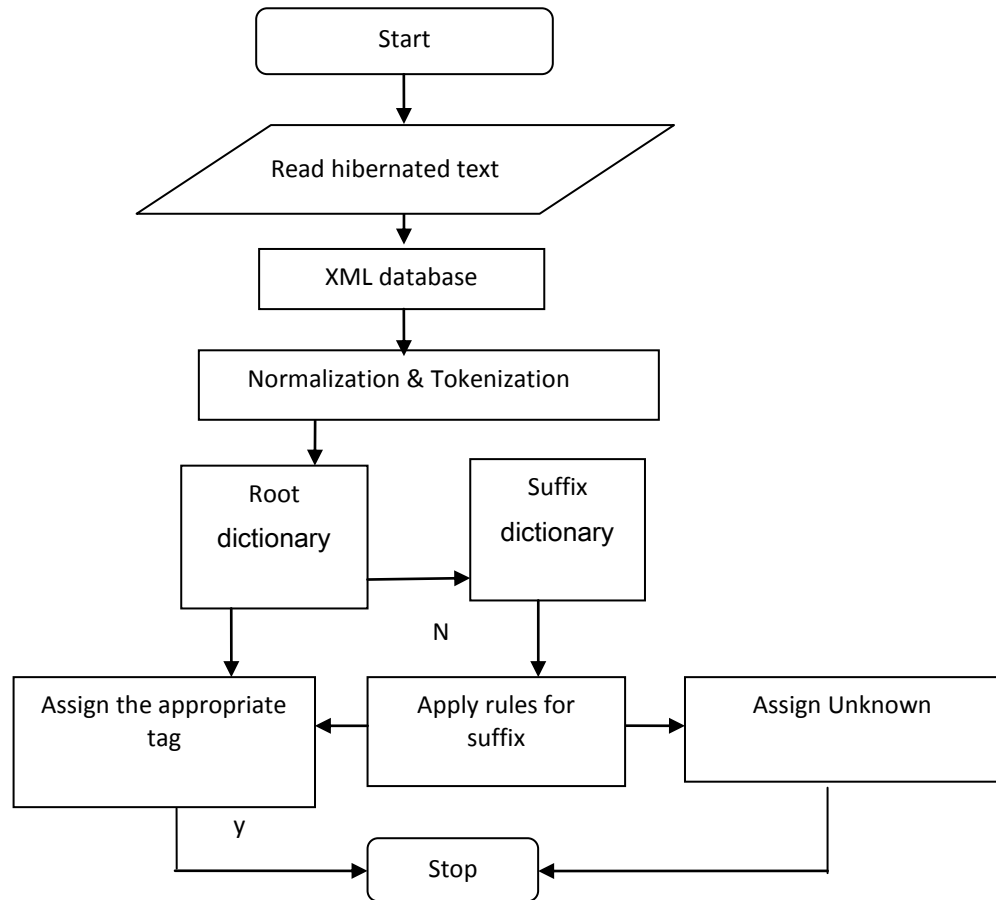
3.10 Numerals (N)

A numeral is a noun, divided into two types cardinal and ordinal. Both the types can be marked as a single tag (ஒன்று, இரண்டு, ஒன்றாம், ஒன்றாவது (oṇṇu, iraṇṇu, oṇṇām, oṇṇāvatu))

4. POS TAGGER PROCEDURE AND ARCHITECTURE

4.1 Procedure

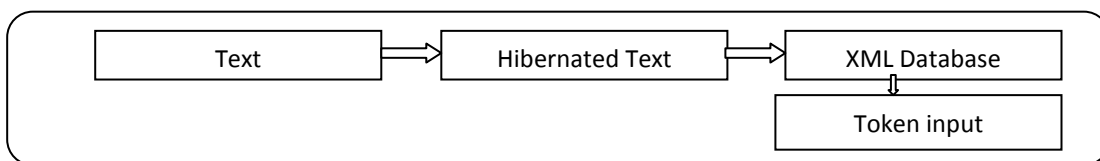
POS Tagger has two tasks: Training task and tagging task. In the training task we have trained the validated tagged data into base-level training module. This module generates a database that validates the root word. It searches the form of the word in the root dictionary if it can find out the root dictionary it assignd the appropriate tag, if it fails, it goes to the root dictionary in order to find out the rules of the suffix list to match the word. If the word is available, it assigns the appropriate tag and if it fails, it names it as an unknown category. The tagging process follows the following procedure.



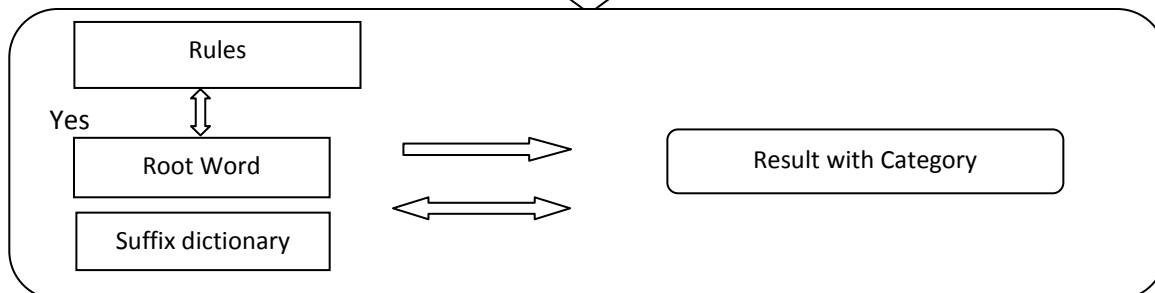
4.2 Architecture of POS

The architecture of POS Tagger consists of three layers such as Data Layer (DL), Process Layer (BL) and Presentation Layer (PL) . The system of architecture is shown below

Data Layer



Process Layer



The Data layer is developed at the time of Training; it encapsulates all the information related to data from the tagging module like Electronic Text, Hyphenated text and Extensible Markup Language Database. The Process layer contains logic

for retrieving persistent data from the DL and placing it into process objects. The PL gives graphical user interface (GUI) application for end user application

Input: A verse from the kuruntokai

யாய் உம் ஞாய் உம் யார் ஆகியர் ஓ
எந்தை உம் நுந்தை உம் எம் முறை கேளிர்
யான் உம் நீ உம் எ வழி அறிதும்
செம் புலம் பெயல் நீர் போல
அன்பு உடை நெஞ்சம் தாம் கலந்தன ஏ .

Output: An annotated text

யாய்/NC உம்/PCLI ஞாய்/NC உம்/PCLI யார்/DIN ஆகியர்/FV ஓ/CLI
எந்தை/NC உம்/PCLI நுந்தை/NC உம்/PCLI எம்/PIA முறை/NC கேளிர்/NC
யான்/PRO உம்/PCLI நீ/PRO உம்/PCLI எ/PIA வழி/NC அறிதும்/VFA
செம்/ADJ புலம்/NC பெயல்/RP நீர்/NC போல/POS
அன்பு/NC உடை/RPA நெஞ்சம்/NC தாம்/PRF கலந்தன/FV ஏ/PCLI ./PU

CONCLUSION

In significant part of the development of any Natural Language Processing system, Parts Of Speech system has to be framed. Tagged corpus is of basic resource for all NLP research. The Parts of Speech tagger is mainly used for Morphology analyzer, generator, Parser, etc., The tool provision category of the text is essential in this case. Ambiguities are common. To avoid such ambiguities, there is a need for syntactic and semantic information of a particular word. Hence POS for an analysis of Classical Tamil plays a major role in NLP.

REFERENCES

1. Andronov, M. 1969. The Standard Grammar of Modern and Classical Tamil. Madras: New Century Book House, Pvt. Ltd.
2. Arulmozhi. P, Sobha. L, Kumara Shanmugam. B. Parts of Speech Tagger for Tamil. In the Proceedings of the Symposium on Indian Morphology, Phonology & Language Engineering, Indian Institute of Technology, Kharagpur. pp 55- 57, 2004
3. Anandan, P. K. Saravanan, Ranjani Parthasarathi and T. V. Geetha. Morphological Analyzer for Tamil. In: International Conference on Natural language Processing, 2002.
4. Lakshmana Pandian S. and T.V. Geetha, "Morpheme Components based Part of Speech tagging", International Conference on Natural language processing, 2008
5. Rajendran S, (2007) Complexity of Tamil in POS tagging, Language in India, Jan 2007.
6. Rajan, K. Corpus analysis and tagging for Tamil. In: Proceeding of symposium on Translation support system STRANS-2002

POS TAGGING FOR WSD IN TAMIL FOR COMPUTATIONAL ANALYSIS

Dr.P.CHANDRAMOHAN

Assistant Professor, CAS in Linguistics, Annamalai University

Introduction

Each and every language in the world has its own structure which is well constructed and well principled in terms of application as well as formation. It should be noted that there is no direct relationship between word and sense assignment, but in between there is some object orientation which helps in designate the sense. The lexical units are arranged according to the language specific contexts demands. Of course the language carries some common features. i.e. called universal ones.

Many words have more than one possible meaning. For example the word **aTi** have number of meanings, when it comes with tense marker it is a verb which means 'to beat'. Similarly when it comes with case marker it is considered as a noun which means 'foot of the leg, line of verse' etc. After an infinitive of a verb, it serves as a causative marker and an auxiliary verb, used to indicate that the action in the main verb is executed with some harshness. Linguists discuss this phenomenon under multiple meaning and it proves to be complex. It can be seen that a word may have many meanings. Some of which are very different and some are related. Given these complications it is important for a computer which attempts language processing to be able to correctly determine the meaning in which a word is being used. The different meanings of polysemous and homophonous words are known as 'sense' and the process of deciding which is being used in a particular context "Word sense disambiguation". WSD is regarded as one of the most interesting and longest standing problems in the analysis of languages.

POS tagging and WSD

Part of speech and sense tagging are similar operations. Both perform a type of lexical ambiguity resolution and each assigns a single tag to all or most of the words in a text. Word types for which each polysemous word has a distinct grammatical category associated with it. These words will always be disambiguated if its part of speech in a text is known. For example the contexts free situations where the information obtained from the base / root word as well as the suffix or case ending used with a word. This becomes useful because many inflected words in a language are added with particular sense after using class – specific suffix or case ending. Each of these types of information is potentially useful for disambiguating meaning be it manually or by computer.

In Tamil the words like **paTi, o:tu, o:TTu, na:Tu** etc. Can either be a noun or a verb in context free situation.

Verb

paTi (to read)

paTitta:n

paTikkira:n

paTippa:n

Noun

paTi (foot step/ measuring vessel)

paTiyai

paTiya:l

paTikku etc.

This can be demonstrated by considering the following sentences.

- ❖ **Surya a:Ra:m vakuppu paTikkira:n** ("Surya is studying sixth standard")
- ❖ **ma:Tip paTiyil viLaiya:Da:te:** (Don't play on the step")
- ❖ **Sahana naTanam a:Tukira:L** ("Sahana is dancing")
- ❖ **avan a:TTaip pa:rtta:n** ("He saw the goat")

These words can be treated as a noun if a nominal case ending is tagged to it. Or they are verbs if a verbal suffix is tagged to it. Therefore these usages are disambiguated by their syntactic behaviour. This information source can be used for disambiguation when two meanings of a word have different parts of speech and the grammatical category of a particular usage can be determined.

It is difficult the words which are only ending with nominal form or verbal suffix. These words will be disambiguated by only some part of speech assignment, and others will not be disambiguating without identifying the sense relations. Consider the following examples:

In Tamil the words like *athikaaram*, *alaku*, *iTam* etc., are multiple meaning words which are accepted only case marker. Similarly the words like *kaTTu*, *koTTu*, *aTi* etc are the forms considered as only verbs. Here part of speech is not useful to disambiguate the words. To disambiguate the words first we have to identify the textual meaning. Further on the basis of this concept we have to identify the sense on the basis of context.

Words convey certain references by implying the sense representation associated with its semantic implications. In fact the nature of a word in terms of semantic implicational coverage can be considered as direct and transferred. The transferred concept here refers to the suggestive meaning of a word, where as the direct refers to the etymological meaning of a word. This distinction mainly depends upon the semantic coverage related to the possible context in which a word can be used. If a word occurs in major /dominant context it will be considered as direct semantic nature, where as in the case of secondary (idiomatic) concept the word it occurs in more than one context. More over the meaning will be considered as static in the entire context, when the word functions a direct concept. In contrast the meaning assigned through the secondary concept varies from context to context.

Context to refer to an immediate linguistic environment (rarely detached or isolated) in which a particular word occurs. Since it is not always explicit, it may be hidden within the neighbouring members of a word used in a piece of text. If we cannot extract the information relevant to the meaning of a word from its immediate linguistic environment, we need to take into account the topic of discussion as a sphere of necessary information. Taking these factors into consideration, Context may have classified into three types: (a) Micro context (b) Transferred context and (c) Topical context.

Micro Context

The micro context refers to the immediate environment of the MMW in a sentence where it has occurred, surrounding its immediately preceding and succeeding words. Conceptually, the immediately preceding (i.e., left) word (LW₁), the key word (MMW), and the immediately succeeding (i.e., right) word (RW₁) (= LW₁ + MMW + RW₁) constitute a lexical block, where the MMW is the main member while the LW₁ and the RW₁ are supporting members. Systematic interpretation of the lexical block will supply necessary information to retrieve the contextual meaning of the MMW. The members of the lexical block generate a network of semantic relationship from which the intended meaning of the MMW is derived by integrating meanings provided by the LW₁ and the RW₁. Thus, in majority of cases, proper importance to the micro context will help us to obtain the actual meaning of the MMW. Within the sphere of structural semantics, it is a unique network of syntactically related members within which each member derives its meaning from the interface of its semantico-syntactic relation with other members. To examine how the micro context supplies information to understand the contextual meaning of the MMW, on experimental basis, I have used 4 sets of lexical blocks where I have put the MMW in the middle and one word in each side (± 1) in the following manner (Figure 2) . I have distributed the sample data set to Tamil speakers to examine if they can understand the contextual meaning of the MMW by associating meanings of the neighboring words without referring to the sentences. The informant is provided with only one set of data at a time in sequential order: [Set 1] > [Set 2] > [Set 3] > [Set 4].

[Set 1]	LW ₁	MMW	RW ₁		
[Set 2]	LW ₁	MMW	RW ₁	RW ₂	
[Set 3]	LW ₂	LW ₁	MMW	RW ₁	
[Set 4]	LW ₂	LW ₁	MMW	RW ₁	RW ₂

Figure 2 Position of MMW in the local context (LW = Left Word, RW = Right Word)

The micro context provides us necessary information to know if the MMW holds relation with its neighboring members. A POS tagger assigns each word in a text to its word class. For example,

Puttakam paTikkira:n

“(He) reads book”

<s> <wt= “NP” >puttakam </w><wt=“VP” >paTi<mc</w><wt=“pt” >kkir</w><wt=“PNG” >a:n</w><c>.></c></ s >

Surya a:Ra:m vakuppu paTikkira:n

“Surya is studying sixth standard”

<s><wt=“NP”>surya</w>a:Ra:m</w>a:Ra:m<wt=“VP”>paTi<mc</w><wt=“pt”>kkir</w><wt=“PNG”>a:n</w><c>.></c></ s >

avaL paTiyil pa:l koTutta:L

“She gave the milk in the measuring vessel”

<s><wt=“NP”>avaL</w>paTi<mc</w><wt=CM>il</w><wt=“VP”>koTu</w><wt=“pt”>tt</w><wt=“PNG”>a:L</w><c>.></c></ s >

In the above sentences the occurrence of the MMW (i.e., **paTi**) and RW₁ (i.e., **pa:l**) as a lexical block within the micro context helps us to consider them together with a direct meaning, which is possible to derive if the words are treated separately in the sentence. Without further reference to any other context, we can understand that **paTi** is not used in general meaning.

Transferred context

The information obtained from this context is useful for understanding lexical collocation of words used in a lexical block. From here, we can know if co-occurrence of any two words is caused by choice (to evoke an intended sense) or by chance (having no special significance). I have found that association of two different words (W₁ and W₂) can denote a special meaning (idiomatic and/or metaphoric), which is not obtainable from the outline of individual literal meanings of the words. That means the co-occurrence of the W₁ and the W₂ in a particular lexical block can generate a special meaning, which is different from the literal meaning of each word. Moreover, collocation of the MMW with the new words generates new special meanings, which are different from the literal meaning of the words. Consider the following examples (Table 1)

Neighboring Words	Literal Meaning	Special Meaning
eTu	to take	To take something with hand
tuNi etu	to take cloth	To buy the dress material
paNam eTu	to take money	Draw money from bank
taNNi eTu	take water	Collect water
V e:laikku a:L eTu	take the man for work	Recruit
Kaiyai eTu	to take hand	Remove the hand

Table 1 Variation of meaning due to lexical collocation of words

Topical Context

The topical context refers to the topic of discussion and focuses on the content of a piece of text. Quite often, it is found that the actual meaning of the MMW depends heavily on the topic which has a strong role to alter etymological meaning of the MMW. The following examples can show that the event of meaning variation of the MMW takes place due to variation of topic or content— a regular phenomenon noted in all natural languages. It also implies that we should extract relevant information from the topic to trail the change of meaning of the MMW.

1. **caTTai paccai niRamaaka irukkiRatu** ("The shirt is green in colour")
2. **na:m paccai mi:nai ca:ppiTa muTiya:tu** ("we can't eat raw / uncooked fish")
3. **paccai maNNil pommai ceyyala:m** ("We can make the toy with wet clay")
4. **paccai piLLai pa:l kutikkitu** ("Young baby drinking milk")
5. **avan paccaiya:ka pe:cuva:n** ("He used unparliamentarily words")

If we analyze the above sentences independently, we shall find that the MMWs (paccai) do not have any notable variation in meaning. But if we combine all the sentences with context together and analyze, we can easily extract a special meaning of the MMW. Taken together, the sentences display a network of meanings, which is not obtainable from individual sentences. Here, special meaning is possible to extract only when we refer to the topic and interpret the sentences with close reference to the topic of the text.

Conclusion

Thus identification of suffix marker or case ending often helps to categories a word to particular lexical class and sense with reference to its context of use. But, if we find that information obtained from the context free situation is not sufficient for understanding sense variation. In linguistics a word has a bundle of information related to phonology, morphology, lexicology, semantics, syntax, text, grammar, etymology, metaphor, discourse, pragmatics and the world knowledge. It is not easy to capture all the information of a word just by looking at its surface form or to its orthography. We require a versatile system along with our native language intuition to interpret all the possible explicit and implicit meanings of a word used in a text for computational analysis.

References

1. ALLAN, Keith. 2001. *Natural Language Semantics*. Oxford: Blackwell Publishers.
2. DASH, Niladri Sekhar. 2002. Lexical polysemy in Bengali: a corpus-based study. *PILC Journal of Dravidic Studies*. 12 (1-2): 203-214.
3. DASH, Niladri Sekhar. 2005a. Role of context in word sense disambiguation. *Indian Linguistics*. 66 (1-4): 159-175.
4. MILLER, George A. and Leacock, Claudia (2000) Lexical representations for sentence processing. In RAVIN, Yael and LEACOCK, Claudia (eds.), *Ploysemy: Theoretical & Computational Approaches*. New York: Oxford University Press Inc. Pp. 151-160.

Morphological Analyzer for Classical Tamil Texts: A Rule-based approach for Case Marker

R.Akilan (akilan.rp@gmail.com)

Programmer, Central Institute of Classical Tamil, Chennai.

Prof E.R.Naganathan (ernindia@gmail.com)

HOD, Department of Computer science and Engineering, Hindustan University, Chennai.

Dr. G.Palanirajan (matrixpalani@gmail.com)

Associate, Central Institute of Classical Tamil, Chennai.

Abstract

This paper describes the works to build a Morphological Analyzer for Classical Tamil using Rule-based approach. Morphology is the study of internal structure of the word. Morphological analysis is a process of segmenting words into morphemes and a process of analyzing the word formation. Morphological analyzer is a tool for any type of Natural Language Processing work. It is a computer program which takes words as input and produces its grammatical structure as output. It identifies and segments the words and assigns the grammatical information. Capturing the agglutinative structure of Tamil words by an automatic system is a challenging job. This paper is going to reveal a rule-based approach for case marker.

Introduction

Natural Language Processing (NLP) is a computerized approach to analyze the text based on a set of theories and set of technologies. And, being a very active area of research and development, the basic objective of Natural Language Processing is to facilitate human-machine interaction through the means of natural human language.

Morphological analysis of a word is the process of segmenting the word into component morphemes and assigning the correct morphosyntactic information. For a given word, a morphological analyzer (MA) will return its word and the word class along with the other grammatical information depending upon its word class. MA returns all possible parse for a given word, without considering the context. MA is a very essential for languages having rich inflectional and derivational morphology such as morphologically rich languages like Dravidian languages.

Morphological Analyzer is a vital tool in NLP applications. In morphological rich languages, as there are multiple affixation, the finer grammatical information which helps in building efficient NLP applications, can be obtained only from Morphological Analyzer. Morphological Analyzer is required in most of the applications such as information extraction, QA system, machine translation and spell checker. There are several approaches attempted for Morphology for Tamil, We present a methodology for morphological analysis of Tamil, a morphologically rich, in this paper. We present a rule-based method for Morphology for Classical Tamil, particularly case marker.

Tamil morphology

Tamil belongs to the Dravidian family of languages. It is one of the Classical Languages. It is a verb-final language and has a relatively free word order; it is an inflectional language. Agglutination is another feature of the language. Tamil morphology is characterized as agglutinative or concatenative, i.e., Words are formed by successfully adding suffixes to the root word in series. When suffixes attach to the root several morphophonemic changes take place. The orders in which suffixes attach to a root form determine the morphosyntax of the language and the various changes that take place when a suffix attaches are called the morphophonemics.

Challenges in Morphological Analyzer for Classical Tamil

Tamil is a classical language which belongs to the Dravidian language family. Tamil literature has existed for over two-thousand years. The morphological structure of Classical Tamil is quite complex since it inflects to person, gender, and

number markings and also combines with auxiliaries that indicate aspect, mood, causation, attitude etc in verbs. A single verb root can inflect for more than two-thousand word forms including auxiliaries. Noun root inflects with plural, oblique, case, postpositions and clitics. A single noun root can inflect for more than five hundred word forms including postpositions. The root and morphemes have to be identified and tagged for further language processing at word level. The structure of verbal complex is unique and capturing this complexity in a machine analyzable and generatable format is a challenging job. The formation of the verbal complex involves arrangement of the verbal units and the interpretation of their combinatory meaning. Phonology also plays its part in the formation of verbal complex in terms of morphophonemic or *sandhi* rules which account for the shape changes due to inflection.

Methodology

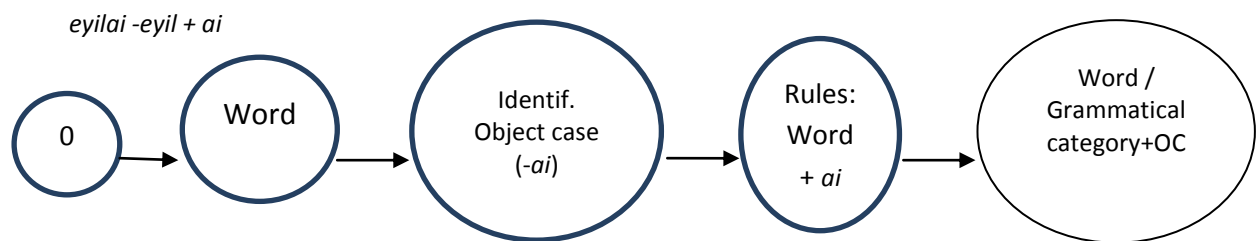
Finite State Automata (FSA)

FSA is a model of behavior composed of a finite number of states and transitions between these states. FSA is an abstract device used for recognizing simple syntactic structures or patterns. An automata is normally depicted by directed graph, called State Diagram and it is also represented in a tabular form as State Table. An FSA, as a string processing device, accepts strings as input and decides if the structure is correct, that is, it either accepts or rejects the string. From a mathematical perspective it is regarded as a function, mapping a set of string to the set {Accept, Reject}.

Case Marker

Case system links a noun phrase and other parts of the sentence through inflection markers, or a word which may be called as a adposition including preposition and postposition

1. Object case (-ai)



2. Instrumental case (-oṭu)

koṭiyōṭu-koṭi + oṭu

3. Associative case (ku, -kku, -akku, -ukku)

4. Dative case (iṇ)

karumpiṛku-karumpu + iṇ + ku, cāṭṭaṛku-cāṭṭaṇ + ku
maturaikku-maturai + (k) ku, tamakku-tam + akku
avaṇukku-avaṇ + ukku

5. Genitive case (-atu)

eḷḷatukuppai- eḷ + atu + kuppai, paṭaiyatukuḷām - paṭai + atu + kuḷām

6. Locative case (-kaṇ)

poruṭkaṇ - poruḷ + kaṇ, malarkkaṇ - malar + kaṇ

7. Nominative case (0)

Rules for Case Markers

Rule 1: Case suffix

The morphosyntax of case suffix may be summarized as

Root + {*ai*} {*āl*} {*iṇ*} {*il*} {*ku*} {*kku*} {*akku*} {*ukku*} {*atu*} {*kaṇ*}

a) After segment the case marker, if remaining word end with consonant later add ‘-u’.

The following example illustrate the inflection of a case suffix of ‘*iṇ*’

aṇpiṇ -> check the root word dictionary

aṇp + {*iṇ*} (if ‘no’ remove the suffix *iṇ*)

aṇp + {*iṇ*} (add ‘u’ at the end of the root word), *aṇpu* + *iṇ*

The following example illustrate the inflection of a case suffix of ‘*il*’

eḷuttil -> check the root word dictionary

eḷutt + {*il*} (if ‘no’ remove the suffix *il*),

eḷutt + {*il*} (add ‘u’ at the end of the root word), *eḷuttu* + *il*

b) Case suffix for *ai*, Remove the suffix ‘*ai*’, Check for root word dictionary

If yes print result, If no remove (y / n) case marker, Check for root word dictionary

If yes print result

Ex. *tōḷiyai* -> *tōḷiy ai* {remove the suffix *ai*}

tōḷiy {check dictionary / no word}

tōḷi y {remove y the last phoneme} = *tōḷi+y+ai* {Result}

c) To split a suffix from the word. The word consists of double letters between words and suffix namely *ṇ*, *l*, *l*, *ṇ* which are segmented separately.

(*kaṇ*, *kaḷ*, *col*, *taṇ*)

Ex. *kaṇṇukku* -> *kaṇ+ṇ+ukku*, *kaḷḷukku* -> *kaḷ+ḷ+ukku*, *collukku* -> *col+l+ukku*

taṇṇai -> *taṇ +ṇ+ai*

Analysis

The morphological analysis identifies root and suffixes of a word. Generally rule-based approaches are used for morphological analysis which are based on a set of rules and dictionary that contains root words and morphemes. In rule-based approach, a particular word is given as an input to the morphological analyzer and if that corresponding morpheme

or root word is missing in the dictionary then the rule-based system fails. Here each rule depended on the previous rule. So if one rule fails, it affects the entire rule that follows. In the course of testing of the rule, certain inconsistencies and lapses in recognizing certain word, First have been found nineteen thousand and nine hundred Classical Tamil root word corpus has been taken for analysis of that corpus is applied the case markers rules. The careful appraisal and study of the words is conducted to identify and overcome the lapses by incorporating certain amount of data into the root word to enhance the coverage and the overall performance of the morphological tools. The following problems are also well noted

1. Some words end with y (\dot{y}) and v (\dot{v}) which is a part of the word. Do not operate their *Sandhi* rules. For examples *añcāy*, *kāy*, *pāy*, *vāy* and *tev*.
2. In some words which require the doubling of the end consonant before add the suffix. For example
 $kaṇṇai$ ($kaṇ+ṇ+ai$), $maṇṇai$ ($maṇ+ṇ+ai$)
 $tammai$ ($tam+m+ai$), $emmai$ ($em+m+ai$)
 $collai$ ($col+l+ai$), $pallai$ ($pal+l+ai$)
 $neyyai$ ($ney+y+ai$), $meyyai$ ($mey+y+ai$)
 $poṇṇai$ ($poṇ+ṇ+ai$), $viṇṇai$ ($viṇ+ṇ+ai$)
3. Moreover -u ($\dot{t}u$ and ru) adding rule is a role between word and suffix. Here it is doubling the final consonant after removing the case marker. If the check before the dictionary For example
 $nāṭṭai$ ($nāṭu+-ṭṭ+ai$), $vīṭṭai$ ($vīṭu+-ṭṭ+ai$), $āṛṛai$ ($āṛu+-ṛṛ+ai$), $kūṛṛai$ ($kūṛu+-ṛṛ+ai$)
4. -u adding rule plays an important role between word and suffix. Here doubling of a variant consonant occurs after removing the case marker. For example
 $mārpai$ ($māṛpu+--rp+ai$), $cālpai$ ($cālpu+--lp+ai$)

Conclusion

This paper has described the Morphological Analyzer for Classical Tamil rule-based approach; in this paper rule-based approach is applied for Case marker. These rule-based approaches for case markers produce the result with more accuracy. In future, using the approach we can develop a rule-based approach for the analyzing not only case markers but also other markers and grammatical variations.

Reference

1. Vijay Sundar Ram R, Menaka S and Sobha Lalitha Devi, "Tamil Morphological Analyser", in "Morphological Analysers and Generators", (ed.) Mona Parakh, LDC-IL, Central Institute of Indian Languages, Mysore, pp. 1 –18. 2010
2. Anand Kumar M, Dhanalakshmi V, Soman K.P, "A Sequence Labeling Approach to Morphological Analyzer for Tamil Language", International Journal on Computer Science and Engineering Vol. 02, No. 06, 2010
3. Beesley, Kenneth R. Arabic, "Finite-State Morphological Analysis and generation", Proceedings of the 16th International Conference on Computational Linguistics, Vol.1.Copenhagen, Denmark. Page No: 89-94, 1996.
4. Viswanathan, S., Ramesh Kumar, S., Kumara Shanmugam, B., Arulmozi, S. and Vijay Shanker, K. "A Tamil Morphological Analyzer", Proceedings of the International Conference On Natural language processing ICON 2003, Central Institute of Indian Languages, Mysore, India. pp. 31–39, 2003.
5. Parameshwari K, "An Implementation of APERTIUM Morphological Analyzer and Generator for Tamil", an e journal of Language in India (www.languageinindia.com), May 2011 Special Volume: Problems of Parsing in Indian Languages, May 2011

AgriQC- A NOVEL QUESTION CLASSIFICATION SYSTEM FOR AGRICULTURAL KNOWLEDGE MANAGEMENT SYSTEM IN TAMIL

S. Augustina Mary (hepaug@gmail.com)

Teaching Assistant, Department of Mathematics

Dr. T. Mala (malanehru@annauniv.edu)

Assistant Professor, (Sr Gr), DIST

College of Engineering, Guindy, Anna University

ABSTRACT

The technologies of the “Knowledge Age” are transforming our focal point from personal, secluded information systems and repositories to a wide exchange and sharing of information in order to widen the size and depth of knowledge available to an individual.

The availability of such a kind of sharable knowledge is possible with the help of a Knowledge Management System (KMS). To access knowledge from the knowledgebase of a KMS, a user can query the system through a query processing subsystem of the KMS. The system in turn should provide accurate, relevant and consistent information to the user. The accuracy of the information depends not only on the technical intricacies in designing the knowledge retrieval subsystem but also in the intelligence the subsystem should possess in analysing the query submitted by the user, so as to retrieve relevant information from relevant sources. Thus it is very important that some machine learning techniques be adapted to analyse and classify the queries. A good classification system improves the overall performance of the KMS. This paper deals with the classification of queries for the Agricultural Knowledge Management System (AgriKMS) developed in Tamil.

Developing a question classification system for an AgriKMS in Tamil is a complex task. AgriQuestion Classification (AgriQC) helps in minimizing search space by analysing the Tamil language queries and generating the appropriate structured query to be submitted to the knowledge retrieval system. Machine learning approach is adopted for the AgriQC so that it provides important clues for answer selection and extraction. The design of a classifier and the extraction of the candidate features are the two key issues to be tackled in the AgriQC. Four features namely subordinate word category, question focus, syntactic and semantic structure are used for feature extraction.

The feature set is then used for classifying the questions given by the agriculturists to the AgriQC system. This system thereby acts as a sub system of an agricultural knowledge management system whose goal is to collect knowledge from experts in agricultural domain, unstructured and semi structured knowledge from www and social community sites and to disseminate knowledge to the end users by answering their queries using the AgriQC.

1. Introduction

Knowledge services can be otherwise called knowledge management activities are, the activities like creation, access, transfer and application of knowledge. From management perspective it is clear that these activities involve a heterogeneous group of people or entities working on the knowledge and therefore requires a consistent and reliable service. Therefore these activities must be governed and controlled by a management system.

A KMS (Knowledge Management System) refers to an information technology-based system to support and assist communities in knowledge management activities. To provide these services a basic Knowledge Management System consists of a knowledge retrieval module, knowledge extraction module, knowledge updation module and knowledge transfer module. This complicated task requires the storing of information in a structured Knowledge base which provides the needed infrastructure for a variety of knowledge services.

This knowledge base plays a vital role in the knowledge management system. It is a repository of domain specific information managed by the knowledge management system .

Thus a KMS is not a mere knowledge distribution system, but rather an interactive and open work environment. Of the number of user groups like knowledge creators, updaters, modifiers another major group interacting with the system is a group questioning the system for retrieval of information in their natural language. The question answering system, a module of the knowledge retrieval system should manage the user queries, select a correct answer from several candidates and provides exact answers for the query since a single query may be reformulated in different ways.

Question classification (QC) provides important clues for answer selection and extraction and hence a good QC system improves performance of the overall QA system. QC helps in minimizing search space by generating the appropriate queries to the information retrieval system. The natural reasoning system is capable of differentiating and indentifying queries posed to it and accordingly returns the expected answer. It handles the task with no difficulty and this process is termed as the human cognition. But it is difficult indeed to introduce this reasoning process within a machine which is still under research. Query framing differs from user to user. Processing queries of varied types appearing in different formats is a bottleneck to a machine. It is a hard task developing a question answering system which will produce exact answers in natural language. The difficulty of pinpointing and verifying the precise answer makes question answering more challenging than the common information retrieval task done by search engines.

This research attempts to provide insights into the following issue:

How should agricultural related questions in Tamil be classified in the question answering systems with relation to the type of answer sought?

The remaining part of this paper is organized as follows. Section 2 describes the different approaches being used for QC. Section 3 explains our original work for AgrilQC in Tamil . In Section 4 we conclude this work .

2. Literature Survey

The difficulty of Natural Language Processing (NLP) has limited the ability of a question answering system to give accurate answer to questions that are quite specific to a particular domain.

MULDER is claimed to be the first general-purpose, fully automated question-answering system available on the web. MULDER's architecture, relies on multiple search-engine queries, natural-language parsing, and a novel voting procedure to yield reliable answers. However, the difficulty of Natural Language Processing (NLP) has limited their ability to give accurate answer to questions that are quite specific to a domain. In addition to the traditional difficulties associated with syntactic analysis, there remains many other problems to be solved, e.g., semantic interpretation, ambiguity resolution, discourse modelling, inference, common sense, etc[14].

In order to avoid a number of difficulties in developing QA systems, a candidate set of question are generated automatically by analysing the sentences present in the selected documents. The question-answering system architecture basically uses sentences within a document as a source of question/answer[3].

Praveen Kumar et al [7] discuss about a question answering system that extracts domain specific keywords from the question and converts it into a query. This query is given as input to the search engine which retrieves relevant documents from the web after weighing and ranking them. Case based classification method is used to classify question types. They claim that this classification methodology helps in the later part of answer selection to put selectional restrictions to check which candidate answers satisfy the semantic constraints [7].

Q/A system based on knowledge base finds the most similar answer-sentence with the target question from the knowledge base and then return the corresponding answer to the user, the retrieval mode of which fully takes advantage of the large-scale of texts, and also avoids bringing the error resources in, which is a very good retrieval mode [5].

The prior knowledge of the estimated answer type helps the QAS to extract correct and precise answers from the document collection[2].

Emanuela Moreale and Maria Vargas [1] use argumentation for finding answers in the specific domain of student essays. The argumentation categories are obtained by means of bottom up approach: categories being definition, reporting, positioning, strategy, problem, link, content/expected, connectors and general. The query classification phase involves processing the query to identify the category of answer that the user is seeking[1].

Works (Hovy et al. 2001; Moldovan et al. 2002; Roth et al. 2002) have shown that locating an accurate answer hinges on first filtering out a wide range of candidates based on some categorization of answer types given a question. Specifically, this classification task has two purposes. First, it provides constraints on the answer types that allow further processing to precisely locate and verify the answer. Second, it provides information that downstream processes may use in determining answer selection strategies that may be answer type specific.[11].

3. AgriQC – A Novel Question Classification System

The AgriQA (Agricultural Question Answering System) requires both context based retrieval of information and inferencing of knowledge. People questioning the system may be of mixed type; both educated and uneducated. The research is oriented towards both the category taken into consideration.

In this section we concentrate on the Agricultural Query Processing Module, a submodule of the AgriQA, the technical details required for it's implemnetation. The work is organised and approached as follow: a. Challenges in Analysing a question in Tamil; b. question classification strategy; c. the features to be extracted discussed.

3.1 Agricultural Query Processing Module

The query processing module for the agricultural domain involves processing of simple and complex queries. The feature set includes namely subordinate word category, question focus, syntactic and semantic structure. Unlike simple questions, complex questions cannot be answered by simply extracting named entities. These questions require inferencing and synthesizing information from many sources.

3.2 Challenges in Analysing a question in Tamil

3.2.1 Position of the question tag

Usually when a question is raised, the question tag is analyzed first. Based on the question tag the next level of analysis continue. In English the question tag usually appears as the first word in the question. This helps in chunking the question for the question tag initially and based on this it is easier to classify the questions. But in Tamil the position of a question tag is usually present at the end or in the middle and may also vary. For eg

நாற்று நட்டல் என்றால் என்ன?
கம்பு கேழ்வரகு எவ்வளவு காலம் வளரும்?
உழுந்து துவரை எந்த காலத்தில் உற்பத்தி செய்யப்படுகிறது?
நெல்லில் உள்ள ஊட்ட சத்துக்கள் என்னென்ன ?
எந்த காலத்தில் நெல் உற்பத்தி செய்யலாம்?

And also the question tag helps for further identifying the answer type in the AgriQA system.

3.2.2 Question characteristics namely lexical, syntactical and semantics play a major role in the classification of questions.

Extraction of syntactic and semantic information usually requires language-specific tools. Tools, such as parsers, are not well suited for processing questions since they are trained on a corpora that don't contain many questions [4]. The relatively simple bag-of-word approaches that are successfully used in text retrieval are not sufficient for extracting specific, fact-based answers in an AgriQA system. It requires linguistic processing ranging from the relatively simple pattern matching and lexical look-up tables to the more difficult analysis of syntactic and semantic structure. Natural language processing techniques like shallow parsing, named entity recognition, part-of-speech tagging and logical transformations [9] are required.

3.3 Question Classification Strategy

A novel strategy, in addition to conventional search and NLP techniques, is used to construct the AgriQC system. We took a set of questions in Tamil in the domain. The questions are largely factual, requiring a listing, descriptive answers, . We manually classified them into four classes as listed below and are supported with suitable examples. Since no standard set of categories exists in the domain, we chose a rather coarse set that seemed appropriate to us.

Named Entity (animal, people, thing)

நெற்பயிரோடு வேறு என்ன பயிர்களை நான் ஊடு பயிராக வளர்க்கலாம்?
எந்த வகையான பூச்சிகள் நெற்பயிரை தாக்கக்கூடும்?

Location (city, country)

நெல் கொள்முதல் தமிழ்நாட்டில் எங்கு நடைபெறுகிறது?
நெற்பயிர் அதிகமாக பயிரிடப்படும் மாநிலங்கள் எவை?
நெற்பயிர் அதிகமாக எங்கு விளைவிக்கப்படுகிறது?

Description (definition, description, manner, reason)

➤ Definitions

The question tag what (என்ன) helps in retrieving a description of the entity searched for eg. ஊடு பயிர் என்ன? நாற்று நட்டல் என்றால் என்ன? நன்செய் நிலம் என்றால் என்ன? புன்செய் நிலம் என்றால் என்ன?

➤ Explanation

எந்த மாதிரியான மண் நெற்பயிருக்கு உகந்தது,

➤ Manner

நெற்பயிரை எவ்விதம் பயிரிடலாம்?

➤ List

நெல்லில் உள்ள ஊட்ட சத்துகள் என்னென்ன?
நெல் அமோக விளைச்சல் பெற எந்தெந்த உரங்களை பயன் படுத்தலாம்?

Numeric

➤ Measurement

செம்மை நெல் சாகுபடி முறைக்குத் தேவையான விதை அளவை என்ன?

➤ Seasonal

நெல் அறுவடைக்கு ஏற்ற காலம் எது?
நெற்பயிர் முதிர்ந்த உடன் எப்பொழுது அறுவடை செய்ய வேண்டும்?
உழுந்து துவரை எந்த காலதில் உற்பத்தி செய்யப்படுகிறது?

➤ Order

நெற்பயிருக்கு அடுத்து கரும்பை உற்பத்தி செய்யலாமா?

The question types can further be classified as low order and high order and can be either closed-ended (convergent) or open-ended (divergent) depending upon the type of answer they return [10]. A low order question requires the system to simply recall a simple single fact. On the other hand a high order question expects the system to recall facts and simultaneously show that they understand the topic, situation or solution to a stated problem. In otherwords, a high order question will expect that a system should understand the relationship between a fact or piece of knowledge within the greater context of the situation.

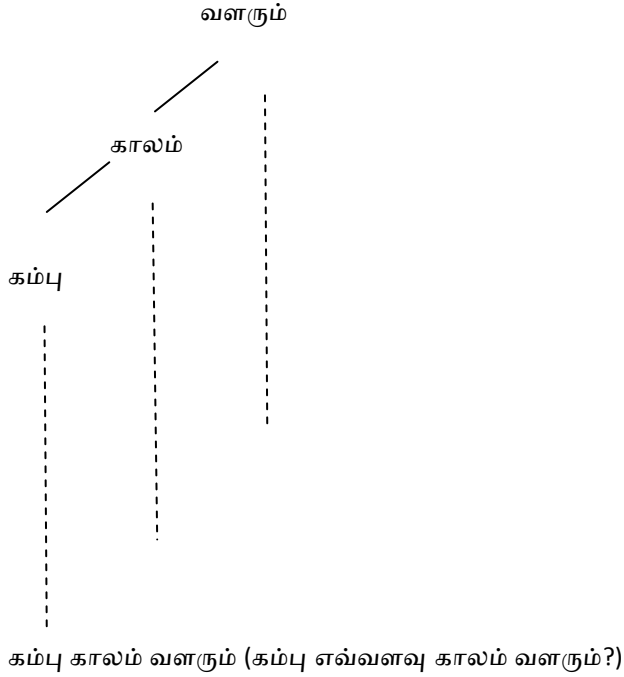
Closed-ended questions requires factual data that are generally short, while divergent type of questions requires an ability to recall some information from stored memory, but must apply that knowledge and other knowledge to explain, extrapolate or further analyze a topic, situation or problem[10].

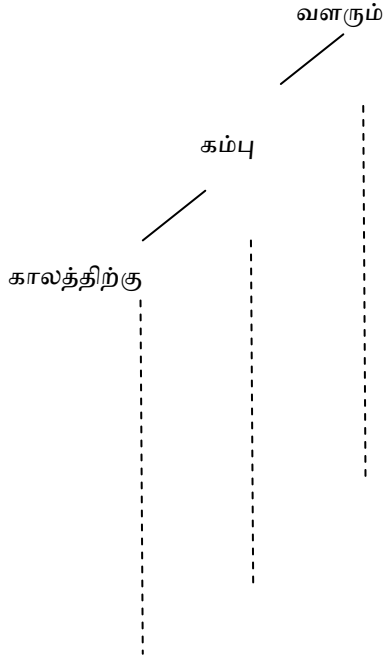
Feature to be extracted

In linguistics, subordination is a principle of the hierarchical organization of linguistic units. The principle is applicable in semantics, syntax, morphology, and phonology and most work in linguistics use the term "subordination" in the context of syntax. Thus to put it in a nutshell "Subordination is a relation existing between two syntactic units, whereby one unit is subordinate to the other and the latter is superordinate to the former"[13].

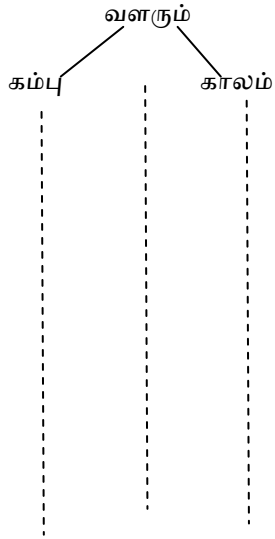
The relationship can be represented as a syntactic structure by trees. Trees are used for exhibiting constituency and dependency structure. Dependency grammar focuses on syntactic dependencies.

The analysis of an agricultural question posed by a user revealed the free word order. Thus dependency structures is well suited for the language. The example below shows a dependency tree, an ordered tree i.e. it shows actual word order.





எவ்வளவு காலத்திற்கு கம்பு வளரும்



கம்பு வளரும் காலம் (கம்பு வளரும் காலம் என்ன?)

Conclusion

The aim of the Agricultural question classification system is to analyse the query posed by an user to the system and retrieve the apt answer expected by the user. We discussed some of the methods used in the existing QA system and proposed a new strategy suitable for the agricultural domain. In the query processing module we have discussed about the difficulties involved in processing a query in tamil language, identification of question types, subordinate word feature extraction which will further help in the retrieval of the exact answer and the dependency structure suitable for the language.

References

1. Emanuela Moreale and Maria Vargas, "A Question-Answering System Using Argumentation", Third International Mexican Conference on Artificial Intelligence, 2004.
2. Muthukrishnan Ramprasath and Shanmugasundaram Hariharan, "A Survey on Question Answering System", International Journal of Research and Reviews in Information Sciences, Vol. 2, No. 1, March 2012.
3. Min-kyoung Kim, Han-joon Kim, "Design of Question Answering System with Automated Question Generation", Fourth International Conference on Networked Computing and Advanced Information Management, 2008.
4. Olga Feiguina and Balázs Kégl, "Learning to Classify Questions", Département d'informatique et de recherche opérationnelle Université de Montréal, 2005.
5. Wen-hua liu¹, hai-yan kang, "Study on question-answering system of restricted domain based on knowledge base", Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding, 12-15 July 2009.
6. Huan Cui, Dongfeng Cai, Xuelei Miao, "Research on Web-based Chinese Question Answering System and Answer Extraction", Journal of Chinese Information Processing. Vol 18, No.3, pp.24-31, 2004.
7. Praveen Kumar, Shrikant Kashyap, Ankush Mittal, "A Query Answering System For E-Learning Hindi Documents", South Asian Language Review, Vol.Xiii, Nos 1&2, January-June, 2003.
8. Yllias Chali and Shafiq R. Joty, "Answering Complex Questions Using Query-focused Summarization Technique", International Conference on Tools with Artificial Intelligence - ICTAI, vol. 2, pp. 131-134, 2008.
9. Stacey m. Bailey, b.s., "Providing question coverage through answer Type classification in natural language Question answering (nlqa) systems", 2001.
10. William F. McComas and Linda Abraham, "Asking more effective questions", Rossier School of Education
11. Xin Li and Dan Roth, "Learning Question Classifiers: The Role of Semantic Information, Journal Natural Language Engineering, Volume 12 Issue 3, September 2006 Pages 229 – 249.
12. Rusli Abdullah, Azmi Jaafar, Hamidah Ibrahim, Mohd.Hasan Selamat, Suhaimi Napis and Rodziah Atan, "Knowledge Management System Architecture in Agricultural Industries Environment", in the proceedings of IEEE, 2008.
13. [http://en.wikipedia.org/wiki/Subordination_\(linguistics\)](http://en.wikipedia.org/wiki/Subordination_(linguistics))
14. Cody Kwok Oren Etzioni Daniel S. Weld, "Scaling Question Answering to the Web", 2001.

Topic E
Natural Language Processing
இயற்கை மொழி ஆய்வு

TAMIL NOUN TO SIGN LANGUAGE A MACHINE TRANSLATION APPROACH

D.Narashiman, Teaching Fellow (dnarashiman@gmail.com)

S.Vidhya, Student, Multimedia Technology

Dr.T.Mala, Assistant Professor (Sr Gr) (malanehru@annauniv.edu)

DIST, CEG, Anna University, Chennai

ABSTRACT

This paper focuses on a novel automatic machine translation system for translating Tamil noun words into sign language variants. Noun words generally do not contain a generic meaning or predefined sign, so a special method has to be designed for translating Tamil noun words to sign gestures in sign language. Two methods are followed in representing nouns words in sign language. Namely finger spelling approach and Air writing approach. This work explores a novel approach of translating Tamil noun words to sign gestures using finger spelling approach to sign gesture. Finger spelling is a act of spelling out alphabets with the help of fingers and fingers displacement. Finger alignment for each letter is represented using a Dactylology chart. The final outcome of the proposed system is hand animated sequences of letters in Tamil noun words.

1. INTRODUCTION ON SIGN LANGUAGE

Sign language is the native language for hearing impaired people and is not universal. It varies slightly from region to region just like dialects of a spoken language. A single sign can have different meaning in different sign language. For example same sign is interpreted as “slow” in SriLankan sign language and “easy” in American sign language (ASL) .

To overcome the communication barrier between hearing impaired and normal hearing community an automatic machine translation will be very useful. There are various machine translation systems being developed to translate spoken language text to native sign language. Some of the systems are especially developed to be used in particular domains such as in banking hall, weather forecasting, in post offices by the hearing impaired community[3]. A complete and full fledge system that translate spoken text to Sign language is not yet developed and only little work is carried out in translating Tamil text to Tamil sign language.

A machine translation system has to bridge the structure of sign language ie the usage of the hand movements, palm orientation, fingers (manual) and body, facial expression (non manual) on one side to the grammar of spoken language on the other side, since the structure of Sign language is different from the spoken language and there is huge lexical gap between them. Therefore reordering of signs is needed to convey the correct meaning.

Usually the sentence of the spoken language may contain noun, verbs, adjectives, tenses, and numbers. More care should be taken to represent them in sign language. Tenses, temporal aspects and verbs are signed in different position in signing space and emotional words are signed through facial expressions[4][5]. Word sense disambiguation issues in sign translation systems can be overcome by contextual models. In the case of proper noun there is a need to apply a different approach in translation process. The objective of the proposed system is to translate nouns in Tamil text to Tamil sign language in an effective manner.

Generally in sign language, nouns and proper nouns are signed by Finger spelling and Air writing method. The Air Writing is an approach where with the help of the fingers the signer will draw the alphabetic representation in the air. This means that structure of the alphabets is shown in air. The important aspects in Air writing method is that the signer must correctly, perfectly and accurately show the alphabet orientation.

Though Air writing method is simple to represent the alphabets it has the following drawbacks: It depends on the signer speed in representing the alphabet, knowledge of the signer about the alphabet structure; as it vary among individual and focus of the recipient. These constraints/drawbacks can be overcome by a better and robust method namely Finger

spelling method. It is mostly used to represent the noun words in sign language. The proposed approach deals with the Finger spelling representation of Tamil noun words.

This paper is organized as follows section II discusses on various machine translation approaches used in sign language translation systems. Section III deals with the finger spelling concepts. Section IV dealt with system architecture and its working and Section V is about evaluations of the system and section IV about the conclusion and further enhancement of the system.

2. SIGN LANGUAGE TRANSLATION SYSTEM

Various methods are followed in general to translate spoken text to native sign language. The machine translation methods are classified as Example based method[6], direct transfer method and Interlingua method. In example based method initially a set of sentences in spoken language is translated manually into sign sequences and stored. With the help of this repository, the translation process is carried out. Direct translation method is based on word-byword translation[7]. Each spoken language word is mapped onto a sign. Simple reordering rule is applied to get the sign sequences. The system is limited to the availability of signs and sign-word mapping. In Interlingua Approach the source language text is initially analyzed and converted into some abstract meaning representation, called an Interlingua. Then sign language is then generated from the Interlingua representation. All these methods depend on the morphological aspects of the spoken language.

In these methods tagger and parser are used to identify the nouns, verbs, adjectives tenses, plurals, genders, named entities and other parts of speech. Apart from tagger and parser, word net dictionary is also used. Nouns and proper noun words will occur in the sentence in different forms. It is a challenge for a translation system to identify the noun words which occur in a sentence in different forms. The noun can be name of a person, name of a place or can be referred as he, him and by using POS tagger, dictionary and anaphora resolution techniques these words can be identified, then these words are tokenized into individual letters and these letters are represented using finger spelling method.

3. FINGER SPELLING APPROACH

One of the approaches that are used in sign language to communicate is the Finger Spelling which is the representation of letters by using only hands which are the basic units of sign language. As there are no proper and specific signs for noun words Finger spelling method is used to represent noun words in sign language. It also helps native Tamil speakers.

In the finger spelling approach the nouns are first tokenized into individual letters. Then based on the way the letter is written symbols are assigned to them using dactylology chart. Tamil letters have circular shapes; partially due to the fact that they were originally carved with needles on palm leaves, a technology that favored rounded shapes. The writing of a Tamil is a combination of alphabetical and syllabic systems. Tamil has twelve vowels, eighteen consonants, four Grantha consonants ('ஜ், ஷ், ஸ், ஹ்') and one special charter 'ஃ'. Table I shows an example the Tamil Finger Spelling based on dactylology chart for the word "உமா".

TABLE I Tamil letter mapping based on dactylology chart.

Word	Tamil Phonetic Spelling	Tamil Finger spelling
உமா	உ + ம + ஃ	  

Dactylogy chart gives gesture representation for each Tamil letter in sign language using hand movements. Figure 1.2 denotes the finger spelling which is representing letter 'ஏ', which is a vowel in Tamil language. In Tamil finger spelling, left hand is dominated by the right hand. Right hand makes alternative gestures to represent each Tamil letter.



Figure 3.1 Representation of letter 'ஏ'

Moreover, as shown in figure 3.1 right hand travels through each finger of left hand to represent vowels. Another most important point is that, these finger spellings should be made within the human body range, which means within a shoulder range. All Five Tamil vowel letters ('ஆ', 'ஈ', 'உ', 'ஏ', 'ஒ') have exploited finger movement.

4. SYSTEM ARCHITECTURE OVERVIEW

The proposed system architecture for translating Tamil noun words to Tamil sign sequences is shown in fig 4.1. The architecture of the Tamil noun words to Tamil sign language translation system is essentially a pipeline of three stages namely Reader, Planner and Hand Animator. Initially noun words are identified using tagger and Tamil dictionary after which the noun words are separated and are given as input to the reader. The reader along with help of the dactylogy chart[9] maps the signs and this is sent to the planner where the Hand movements are fixed and in the hand animator module the sign sequence for the noun word is generated and finally animated in 2D screen[2][8].

A. Text preprocessing

The given Tamil sentences are preprocessed with the help of Word net dictionary and anaphora resolution techniques to identify noun words. Then these words are given as input to the Reader.

B. Reader

The Reader takes Tamil Noun words as input. Using Tokenizer (uyirmei algorithm) the given noun word is separated into individual letters or Tokens. These tokens are classified into uyir (உ), mei (ம) and uyirmei (மு) letters. This algorithm takes into account the Unicode of the character and accordingly identifies the uyir and mei letters and with the help of this the uyirmei letters are generated. Then the symbols are generated from the symbol predefined chart namely Dactylogy chart.

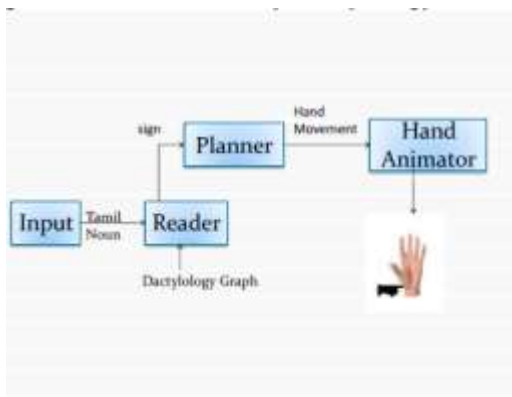


Figure 4.1 Architecture of translation system

The sign representation of each letter or alphabet of Tamil language set is available in the dactylology chart. Another important task performed in this module is the identification of the sign for uyirmei letters. These letters formed from both uyir and mei characters. Though there are two character representations for uyirmei while generating the signs for these letters it should be a single step and this issue is taken care by the planner.

C.Planner

In planner the characters are checked whether they are kuril or nedil character which is most important in determining the Hand animation. The tokens or letter along with their signs are fed into the planer. The kuril and nedil characters are identified and depending on this the hand gestures are generated. If there are kuril characters then the display is a static gesture and nedil characters then the display is dynamic gesture. In static gesture the right hand just touches the left hand. Whereas in dynamic gesture the right hand makes movement on the fingers of left hand. The output of this module is given to Hand animator to animate the sign sequence.

Hand Animator

This is the last module in the system where the signs for the noun words are animated. The animated signs are sequenced and are displayed on a GUI for effective understanding by hearing impaired people. Fig 4.2 shows the output screen for a word. In this depending on the kuril and nedil characters the gestures are generated.

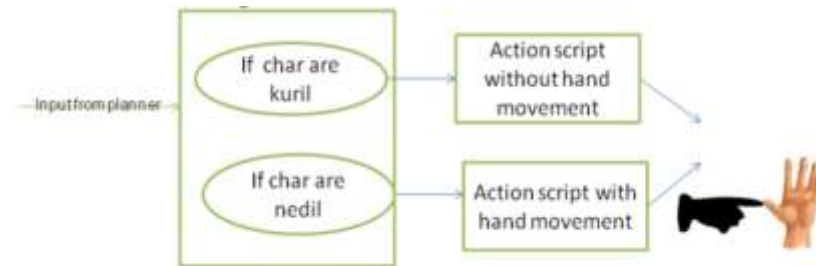


Figure 4.2 Hand Animator

5. EVALUTION CRITERIA

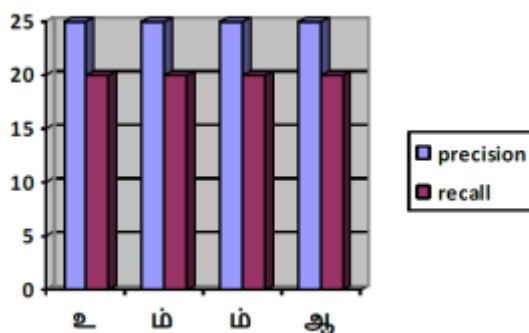


Figure 5.1 precision and recall graph

The elevation is done based on precision and recall. To the example presented the precision and recall graph is obtained as shown in the figure 5.1 both are calculated using the formula each letter $\text{Precision} = \frac{TP}{TP+FP}$, $\text{Recall} = \frac{TP}{TP+FN}$ For உ precision= $\frac{1}{1+3} = 0.25 = 25\%$, Recall= $\frac{1}{1+4} = 0.2 = 20\%$. Similarly for the remaining characters the precision and recall are calculated for our example.

6.CONCLUSION

The proposed work is synthesize of Tamil Sign Language gesture for Tamil noun words. The complexity is reducing in the process, instead of determining uyir-mei characters as uyir and mei separately, we can created a single image. This reduces both the time complexity and space complexity. At present sign gestures has created only for Tamil nouns words, Later it can be emphasized for the verb, tense, etc. Tamil Signs were developed by using the Dactylology chart. In future, individual character or written form of uyir mei characters can be synthesized based on Regular Expression and by generating the grammar. The system can also be developed to synthesize the common noun.

REFERENCES

1. Ameer M.Almasoud and Hend S. Al-Khalifa, " A Proposed Semantic Machine Translation System for Translating Arabic Text to Arabic Sign Language, " *second Kuwait conference on E-services and ESystems, April 5-7, 2011*.
2. Maria Papadogiorgaki, Nikos Grammalidis, Dimitrios Tzovaras, and Michael G.Strintzis, " Text-to-sign language synthesis tool, " in proceeding of EUSIPCO, 2005, p.130-134.
3. Mohamed Mohandes, "Automatic translation of Arabic text to Arabic sign language", *Artificial Intelligence and Machine Learning*, 2006, vol.6 (4), pp.15-19.
4. D.Narashiman, Dr.T.Mala „An interactive system for conversion of text to sign language, " in proceeding of International Conference on Information Systems & Software Engineering, 2009, PP.164-167.
5. D.Narashiman, Dr.T.Mala "A machine translation system for converting Tamil textto sign-language" in proceeding of 9th Tamil Internet Conference, 2010, PP.572-578.
6. Sara Morrissey and Andy Way, " An example based approach to translating sign language, " in proceedings of Example Based Machine Translation, 2005, vol.11, pp.106-116.
7. Tirthankar dasgupta, Sandipan dandpat and Anpam basu, "Prototype machine translation system from text-to-Indian sign language", in Proceedings of the International Joint Conference on Natural Language Processing Workshop for Less Privileged Languages, p.19–26, Jan 2008.
8. Tony Veale, Brona collins, and Alan Conway, " The challenge of cross-modal translation", proceedings of Association of Machine Translation, 2000, vol.13 (1), pp.81- 106.
9. S.Vanjikumaran, and G.Balachandran, "An Automated Vision Based Recognition System for Sri Lankan Tamil Sign Language Finger Spelling", in proceeding of International conferences on Advance in ICT for emerging regions, 2011.

Development of Telugu-Tamil Bidirectional Machine Translation System: A special focus on case divergence

Parameswari K (parameshkrishnaa@gmail.com)

Sreenivasulu N.V. (nv.sreenivasulu@gmail.com)

Uma Maheshwar Rao G (gurao@gmail.com)

Christopher M (efthachris@gmail.com)

Centre for Applied Linguistics and Translation Studies, University of Hyderabad

Abstract

Each case marker has a number of functions which lead to case mismatches in machine translation (MT). Hence, any effort in the direct mapping of case markings often results in non-standard or ungrammatical constructions in the target language. In detail, this paper attempts to find out the divergent functions of the nominative, accusative and dative case markers between Telugu and Tamil and provides various solutions to handle them in Transfer Based MT system.

1. Introduction

Building a machine translation (MT) system is one of the challenging tasks in Natural Language Processing. Even though Tamil and Telugu are two closely related languages belonging to Dravidian language family, still they exhibit a considerable amount of diversity at every level viz. morphological, syntactic, semantic and lexical levels. Keeping these in mind, building a MT system for this language pair can be not only non-trivial but also challenging. The paper deals with the issues in the development of an automatic Telugu-Tamil bidirectional MT systems.²

To build a more sophisticated and effective MT system, it is significant to identify divergences (i.e. cross linguistic differences) between the pair of languages. This paper focuses on case divergences i.e. the differences that occur due to case selection in Telugu and Tamil.

2. Case mismatches

Case is a system of marking dependent nouns for the type of relationship they bear to their heads (Blake, 1994:1). In certain cases, languages show a great mismatch in the patterning of the morphological case against the patterning of the syntactic case. It arises due to the distinction of a case (form) and its grammatical role (function) that it expresses.

Case marking with nominals in Telugu and Tamil is exhibited as suffixes to express the different syntactico-semantic functions. The inflection of case is overtly marked except in the case of Nominative which is \emptyset in both the languages as in Table-1.

Language	NOM	ACC	DAT	INST	ASS	LOC	ABL	GEN
Tamil	\emptyset	-ai-	(u) kku/-āl -ku	-ōṭu/-il/-iliruntu/ uṭaṇ	-iṇ/-uṭaiya/ -iṭam	-iṭamiruntu		-atu
Telugu	\emptyset	-ni/-ki/	-tō -tō -lō/ -nu -ku	-nuMḍi/ -yokka -daggara		-nuMci/		

Table-1: Case System of Telugu and Tamil

2. Machine Translation systems discussed here are being developed under the project of IL-IL MT at CALTS, University of Hyderabad, as a part of the Consortium of Indian Languages to Indian languages Machine Translation Systems funded by DIT, Ministry of Information Technology, Govt of India.

Here are some of the major reasons for case mismatches between Telugu and Tamil.

1. Case Syncretism
2. Case with Complements to Postposition

3. Case Syncretism

Case syncretism occurs when a single inflected form corresponds to two or more case functions (Comrie 1991: 44-47) . Distinct case values are determined on a language-specific basis, so that the case syncretism by the definition involves an observable asymmetry between paradigms within a language (Bearman, 2009:219) .

Both in Telugu and Tamil, a number of instances of case syncretism are found and that poses problems in MT. Here, we list the divergent behaviour of the use of the nominative, accusative and dative case markers between Telugu and Tamil.

3.1 Nominative Case Marker

Nouns in the nominative case have different functions such as (i) subject, (ii) predicate, (iii) subject complement, (iv) object complement and (v) object (Lehmann, 1989:25) .

3.1.1 Subject: The nominative case marked noun is used as a subject in both Telugu and Tamil. When a subject is in the nominative case, it controls verb agreement.

(1) Te.*rāmuḍu_i ikkaḍi-ki vacc-ā-ḍu_i* .

Ram.NOM here-DAT come-PST-3SGM

Ta.*rāmaṇ_i iṅkē va-nt-āṇ_i* .

Ram.NOM here come-PST-3SGM

'Ram came here'

However, when a predicate indicates the capability mood, Tamil optionally uses the subject inflected for the instrumental case marker or otherwise called as inflected for the lexical passive (Subbarao & Bhaskararao, 2004:163), whereas the subject in Telugu is in the nominative case.

(2) Te.*nēnu ī pani ceyy-a-gala-nu*

I.NOM this work-Ø (ACC) do-INF-can-1SG

Ta.*eṇṇ-āl/nāṇ inta vēlai.y-ai.c ceyy-a-muṭi-y.um*.

I-INST/I.NOM this work-ACC do-inf-can-3SGN (default)

'I can do this work'

3.1.2 Predicate: Noun phrases in their predicate position occur in the nominative case in Tamil. In Telugu, the predicate nominals are in the oblique case agreeing in *person and number* with the subject when it is in the first or second person singular or in the first person plural. Such agreement is manifested by the occurrence of a pronominal suffix which is co-indexed with the subject.³

(A) Predicate noun:

³The addition of pronominal suffix with a noun is otherwise called pronominalized noun (Krishnamurti and Gwynn, 1985:134)

With first person singular	With first person plural	With second person singular
<p>(3) Te. <i>nēnu yuvakuḍi-ni</i> .</p> <p>I.NOM young man.OBL-1SG.</p> <p>Ta. <i>nāṇ ilaiṇṇaṇ</i>.</p> <p>I.NOM young man.Ø</p> <p>'I am a young man'</p>	<p>(4) Te. <i>manamu yuvakula-mu</i>.</p> <p>we.incl.NOM young man.OBL-1PL.</p> <p>Ta. <i>nām ilaiṇṇar-kaḷ</i>.</p> <p>we.incl.NOM young man-PL.Ø</p> <p>'We are young men'</p>	<p>(5) Te. <i>nuvvu yuvakuḍi-vi</i>.</p> <p>You.NOM young man.OBL-2SG.</p> <p>Ta. <i>nī ilaiṇṇaṇ</i>.</p> <p>You.NOM young man.Ø</p> <p>'You are a young man'</p>

In Telugu, abstract nouns of quality with adjectival force occur in the position of a predicate in equative [NP-NP] construction [Krishnamurti and Gwynn, 1985:p125]. These nouns do not show agreement with subject and are in nominative form and similar to Tamil. The nouns like *poVdugu* 'tall', *eVwwwu* 'height', etc., do not agree with the subject.

(6) Te. *nēnu mī-kaMṭē poḍugu*

I.NOM you.PL-than tall (noun) .Ø

Ta. *nāṇ uṇkaḷ-ai viṭa uyaram*

I.NOM you.PL-ACC than tall (noun) .Ø

'I am taller than you'

(B) Predicate Adjective: The predicate adjective too behaves in a similar fashion as mentioned above in (3), (4) and (5) in Tamil and Telugu.

(7) Te. *nēnu maMcivāḍi-ni*.

I.NOM good person.OBL-1SG

Ta. *nāṇ nallavaṇ*.

I.NOM good person .Ø

'I am a good person'

(C) Predicate Numeral: A Predicate numeral is the number expression that occurs in the predicate position. The numerals in predicate position function exactly like predicate nouns in terms of agreement in Telugu. Hence, they appear in oblique form.

(8) Te. *mēmu muggura-mu*.

we.excl.NOM three person.OBL-1PL

Ta. *nāṇkaḷ mūvar*.

we.excl.NOM three person.Ø

'We are three persons.'

3.1.3 Subject Complement: The nominative case is used in Tamil when a noun functions as a complement to the subject with a verbal predicate. Similar to predicate nouns, the subject in Telugu agrees with the subject complement when it is in the first or second person singular or in the first person plural.

(9) *Te.nēnu_i maMciṽḍi-ni_i ; ayy-ā-nu_i*

'I.NOM good man.OBL-1SG become-PST-1SG'

Ta.nāṇ_i nallavaṇ_i ; āṇḍēṇ_i

'I.NOM good man.Ø become-PST-1, sg'

'I became a good man'

3.1.4 Object Complement: The noun must be in the nominative case when it functions as an object complement in Tamil. In such cases, Telugu noun is marked for accusative case. Irrespective of the ontology of object, the accusative case marker is mandatorily assigned to the object and its complement object in Telugu as in (10) .

(10) *Te.nēnu vāḍi-ni rāyi-ni/ maniRi-ni cēs-ā-nu.*

I he- ACC stone-ACC/ human- ACC do- PST- 1SG

Ta.nāṇ_i avaṇ-ai kal/ maṇitaṇ_i ākk-iṇ-ēṇ_i

I he- ACC stone.Ø / human.Ø make-PST- 1SG

'I made him into a stone/a human'

3.1.5 Object: The noun may be in the nominative case when it functions as an object which is [-animate] and [-specified] both in Telugu and Tamil.

(11) *Te.nēnu pāṭa pāḍ-ā-nu*

I.NOM song Ø (ACC) sing- PST- 1SG

Ta.nāṇ_i pāṭṭu pāṭ-iṇ-ēṇ_i.

I.NOM song.Ø (ACC) sing-PST-1SG

'I sang a song'

3.2 Accusative Case Marker The direct object of a transitive verb is accusative case marked either overtly or covertly. In both the languages, the accusative case marker shows syncretism with the nominative case marker (Ø) when the direct object marker is not overtly case marked.

The major divergence on marking accusative arises due to the following reasons between Telugu and Tamil:

1. Differential Object Marking
2. Accusative Case Marking of Theme in the Non-nominative Subject (NNS) Construction

3.2.1 Differential Object Marking

The variation in the occurrence of case marking with direct object is termed as Differential Object Marking (DOM) (Bossong 1991, Comrie 1989, Aissen 2003, Subbarao 2012 amongst others) . DOM depends on the direct object's ontological properties. The higher one in the scale is more prominent and tends to be marked by the accusative case. This can be shown in the following scalar dimensions.

Definiteness scale: [+pronoun] > [+proper noun] > [+definite] > [-definite, +specific] > [-specific]

Animacy scale: [+human] > [-human, +animate] > [-animate]

The animacy and definite hierarchy are applied to find out the divergent behaviour of DOM in Telugu and Tamil. The major divergence in the usage of the accusative case marker between Telugu and Tamil occurs with the object which is [+human], [+pronoun, -human, +/- animate] and [+definite, -animate].

(i) Nouns [+human]: Both in Telugu and Tamil, human nouns on the definiteness scale are obligatorily marked for the accusative case. However, in certain cases in Tamil, the accusative case suffix is used optionally with the object noun (O) [+human] as in (12) whereas Telugu uses it obligatorily.

(12) Ta.kumār māppi!lai- (y.ai.t) tēṭu-kir-āṇ.

KumAr son-in-law- (ACC) search- PRS- 3SGM. (Lehmann, 1988:p29)

Te.kumāru alluḍi-ni vetuku-tunn-āḍu.

KumAr son-in-law- ACC search- PROG- 3SGM

“Kumar is looking for a son-in-law.”

The absence of accusative case in (12) indicates that the object noun is incorporated into the verb and forms a complex predicate.

(ii) Nouns [+pronoun, -human, +/- animate]: In Tamil, the object [+pronoun, -human, +/-animate] is marked accusative obligatorily⁴, whereas in Telugu it is optional. Though pronouns [+deictic] and [+proper] are definite and referential [Bossong, p160:1991], Telugu has an option to mark the pronouns [-human, +/-animate] with the accusative case suffix as in (13) . The pronoun *idi* 'it.prox (imal) ' / *adi* 'it.dist (al) ' can be used as referential to nouns [-human, +animate] (eg. dog) and nouns [-animate] (eg. computer) in Telugu and correspondingly *itu/atu* in Tamil.

(13) Te.nēnu idi cūs-ā-nu.

I it see-PST-1SG

Ta.nāṇ it-ai.p pār-tt-ēṇ.

I it-ACC see-PST-1SG

'I saw this'

⁴ Our corpus data (CALTS Corpus with 3 million data) show that they are 84 cases in which the DO is accusative marked in contrast to 3 cases where it is not accusative marked; Since there are only 3 cases statistically it is considered as insignificant.

The pronouns *idi* 'it.prox' and *adi* 'it.dist' may also refer to nouns that are [+human, +female]. In such cases, intimacy or derogatoriness of the subject towards the direct object is indicated. Here, the accusative case suffix is used obligatorily as in (3), since the pronoun [+human] has a higher status in the animacy scale.

(14) Te. *nēnu dīni-ni cūs-ā-nu*

I she/it-ACC see-pst-1SG

'I saw her/this'

(iii) Nouns [+definite, -animate]: In Tamil, when a direct object[-ani] is modified by definite modifiers such as determiners (15) and possessives (16), the presence of accusative case is obligatory in contrast to Telugu where it is optional.

<p>(15) Ta. <i>vēlu eṇ *vīṭu /vīṭṭ-ai.p pār-tt-āṇ..</i></p> <p>Velu.NOM my *house.NOM/house- ACC see-PST-3SGM</p> <p>Te. <i>vēlu nā illu / iMṭi-ni cus-ā-ḍu</i></p> <p>Velu.NOM my house.NOM/house- ACC see- pst-3SGM</p> <p>'Velu saw my house'.</p>	<p>(16) Ta. <i>vēlu anta *vīṭu / vīṭṭaip pār-tt-āṇ.</i></p> <p>Velu that *house/house- ACC see-PST-3SGM</p> <p>Te. <i>vēlu ā illu / iMṭi-ni cūs-ā-ḍu.</i></p> <p>Velu that house/house- ACC see-PST-3SGM</p> <p>'Velu saw that house.'</p>
---	--

3.2.2 The Accusative Case Marking of Theme in Non-nominative Subject Construction (NNS)

Dative Subject Construction is the most widespread in Dravidian (Subbarao, 2012:p134) . The dative marked subject acts as an experiencer subject and the verb agrees with the object. In Tamil, stative predicates expressing the notion of mental and emotional experiences require the case marking pattern of DAT-ACC (Lehmann, 1989: p180) and DAT-NOM in Telugu. Verbs such as *weVri* 'know' and *puri* 'understand' and etc., in Tamil and *weVlusu* 'know' and *arWamavvu* 'understand' and etc., in Telugu refer to some mental experience as in (17) . Verbs such as *piti* 'like' and etc., in Tamil and *naccu* 'like' and etc., in Telugu express the emotional experience as in (18) .

<p>(17) Ta. <i>eṇ-akku avaṇ-ai.t teriyum.</i></p> <p>I- DAT he- ACC know.3SGN (default)</p> <p>Te. <i>nā-ku vāḍu telusu.</i></p> <p>I- DAT he.NOM know.3SGN (default)</p> <p>'I know him'</p>	<p>(18) Ta. <i>eṇ-akku uṇṇ-ai.p piṭi-kk-um</i></p> <p>I-DAT you-ACC like-FUT-3SGN (default)</p> <p>Te. <i>nā-ku nuvvu; nacc-ā-vu_i</i></p> <p>I- DAT you.NOM like-FUT-2SG</p> <p>'You pleased me (lit.) / I like you'</p>
---	---

3.3. Dative Case Marker:

The dative case marker in Telugu and Tamil has a wide range of functions and it is marked with the nouns expressing beneficiary, goal, the experiencer subject, purpose, possession, point in time and place, comparison, distribution and etc. (cf. Ramarao, 1975; Krishnamurti and Gwynn 1984; Ramanarasimham, 2006, Paramasivam, 1983:151 as quoted by

Lehmann, 1989:31; Arden 1942, Subbarao, 2012) . However, There are lots of instances where the dative has divergent functions and cannot be directly mapped between Telugu and Tamil.

3.3.1 Location: Nouns of inanimate category when inflected for the dative in Telugu express possession or appropriately the locative function. On the contrary, Tamil uses the locative marker. Here, the dative case marker relates two noun phrases which have holonymy-meronymy relationship. The word which is holonymy is marked for the dative case marker in Telugu and the locative in Tamil. Example:

(19) Te. *gōḍa-ku kiṭikī uM-di.*

wall-DAT window be-PRS.3SGN

Ta. *cuvarr-il jaṇṇal iru-kkiṛ-atu.*

wall-LOC window be-PRS.3SGN

The wall has a window.

3.3.2 Direction: The dative marked noun can express the direction in Telugu. The swapping of dative marker is found between Telugu and Tamil and thus named as dative swapping which is marked as a configurational difference between them. We use *dative swapping* to refer to the exchange of the dative between the postposition and their complement Ns in the translated structures. In Telugu, a noun (N) followed by a Spatial Noun (which functions as postposition (P)) inflects for dative, whereas in Tamil the complement noun receives dative. It can be illustrated as in the following:

Te. (([N][P]) +DAT) <=> Ta. (([N]+DAT/GEN) [P])

For instance, the following examples explicate the dative swapping between Telugu and Tamil.

(i) Direction towards inside/outside, up/down, forward/backward of a place: Place nouns occurring with the complements of P denoting the motion towards inside/outside, up/down and forward/backward inflects for the dative in Tamil and in the swapping order in Telugu.

(20) Te. *kumār iMṭi lō-ki/lōpali-ki veḷḷ-ā-ḍu.*

'Kumar house inside-DAT/inside-DAT go-PST-3SGM

Ta. *kumār viṭṭu-kku uḷḷē ce-ṇṛ-āṇ.*

'Kumar house-DAT inside go-PST-3SGM

Kumar went inside the house.

(ii) Direction towards left side/right side/ the point of compass: Place nouns occurring with the complements of P denoting the motion towards leftside/rightside (21) and the point of compass (22) inflect for the dative in Telugu and Tamil. Optionally, the dative swapping is also observed in Telugu.

<p>(21) Te. a. <i>rāmuḍu iMṭi-ki kuḍivaipu vēḷḷ-ā-ḍu.</i></p> <p>Ram house- DAT leftside go-PST-3SGM</p> <p>b. <i>rāmuḍu iMṭi kuḍivaipu-ku vēḷḷ-ā-ḍu.</i></p> <p>Ram house.OBL leftside-DAT go-PST-3SGM</p>	<p>(22) Te. a. <i>mā iMṭi uttarāni-ki/paḍamaṭi-ki veḷḷ-ā-nu</i></p> <p>our house.OBL north-DAT/west-DAT go-PST-1SG</p> <p>b. <i>mā iMṭi-ki uttaraM/paḍamaṭa veḷḷ-ā-nu.</i></p> <p>my house-DAT north / west go-PST-1SG</p>
---	--

Ta. <i>irāmaṇ vīṭṭu-kku valatupuram ceṇ-ṛ-āṇ.</i>	Ta. <i>eṇkaḷ vīṭṭu-kku vaṭakk-ē/mēṛk-ē ceṇ-ṛ-ēṇ.</i>
Ram house- DAT leftside go-PST-3SGM	our house-DAT north-LOC/west-DAT go-PST-3SGM
'Ram goes to the left side of the house'	'I went to the north/west side of our home'.

(iii) Direction towards a place: Both in Telugu and Tamil, nouns denoting place inflect for dative directly when they occur as complements of verbs of motion. However, non-place nouns when occur as complements of nouns denoting space i.e. locative nouns do not inflect for dative but the locative head inflects for dative. But in Tamil, both place (22) and non-place nouns (23) occurring as complements of locative nouns inflect for dative.

(22) Ta. <i>skūlu-kku.p pō.</i>	(23) Ta. <i>kumār uṭaṇē rāju.v-ukku/rāju.v-iṇ aruk-il ceṇ-ṛ-āṇ.</i>
school-DAT go-IMP	Kumar immediately Raj-DAT/Raj-GEN near-LOC go-PST-3SGM
Te. <i>skūlu-ku veḷḷu.</i>	Te. <i>kumār veMṭaṇē rāju daggari-ki veḷḷ-ā-ḍu.</i>
school-DAT go-IMP	Kumar immediately Raj near-DAT go-PST-3SGM
Go to School.	Immediately, Kumar went near Raj.

However, nouns inflected for dative can be complements of verbs of motion in Tamil, but these complements have postpositions as their heads. Whereas in the case of Telugu it permits nouns to be complements of postpositions inflected for dative.

(iv) Direction towards Nouns of space of generic type: Nouns of space of generic type like in Telugu *akkaḍa* 'there', *ikkaḍa* 'here', etc., when inflect for the dative indicate direction, whereas the corresponding forms in Tamil do not require the dative. In Tamil, these generic space nouns have the -e marker which shows the direction as in *aṅk-ē* 'there', *iṅk-ē* 'here', *uḷḷ-ē* 'inside', *veḷḷi.y-ē* 'outside', and etc., Example:

(24) Te. *rājā akkaḍi-ki veḷḷ-ā-ḍu.*

Raja there-DAT go-PST-3SGM

Ta. *rājā aṅkē ce-ṇṛ-āṇ.*

'Raja there go-PST-3SGM

Raja went there.

3.3.3 Duration: The dative here refers to the use of it as the expression of specific time. Sometimes it is realized as obligatorily in Tamil, whereas in Telugu, it is not.

(i) Duration of time (within X, where X is a temporal noun): The Tamil noun marked for dative case expresses a specific time limit or duration of time roughly equivalent to the English 'within'.

(25) Ta. *eṇ-akku pattu nāṭ-kaḷ-ukkuḷ puttakatt-ai.k koṭu-ṇkaḷ.*

I-DAT ten day-PL-DAT.inside book-ACC give-imp[hon]'

Te. *nā-ku padi rōju-la lōpu/logā pustakaM ivv-aMḍi*.

I-DAT ten day-PL inside book.Ø (ACC) give-imp[hon]'

'Give me the book within ten days.'

(ii) Temporal Specification (within X, where X is a temporal noun):: Certain nouns like *nēṛṛu* 'yesterday', *iṇṇu* 'today' and *nāḷai* 'tomorrow' take dative case in Tamil whereas Telugu does not (26) . Whenever Telugu uses the dative with these nouns (27) to indicate 'by X', where X is a temporal noun.

<p>(26) Ta. <i>nāṇ nāḷai/nāḷai-kku varu-v-ēṇ</i>.</p> <p>I tomorrow/tomorrow-DAT come-FUT-1SG</p> <p>Te. <i>nēnu rēpu vas-tā-nu</i>.</p> <p>I tomorrow come-FUT-1SG</p> <p>'I will come tomorrow.'</p>	<p>(27) Ta. <i>nāṇ nāḷai-kku varu-v-ēṇ</i>.</p> <p>I tomorrow-DAT come-FUT-1SG</p> <p>Te. <i>nēnu rēpaṭi-ki vas-tā-nu</i>. I tomorrow-DAT come-FUT-1SG</p> <p>'I will come by tomorrow.'</p>
--	--

(iii) The period of time: A noun of the period of time with the verb [+motion] takes the locative case marker to express the time in Tamil and Telugu. Telugu uses the dative with these nouns (29) to indicate 'by X', where X is a period of time.

<p>(28) Te. <i>nēnu udayaM/udayā-na vas-tā-nu</i>.</p> <p>I morning/morning-LOC come-FUT-1SG</p> <p>Ta. <i>nāṇ kālai.y-il varu-v-ēṇ</i>.</p> <p>I morning-LOC come-FUT-1SG</p> <p>'I will come in the morning'</p>	<p>(29) Te. <i>nēnu udayāni-ki vas-tā-nu</i>.</p> <p>I morning-DAT come-FUT-1SG</p> <p>Ta. <i>nāṇ kālai.y-il varu-v-ēṇ</i>.</p> <p>I morning-LOC come-FUT-1SG</p> <p>'I will come by morning'</p>
--	---

(iv) Limitation of time: To express the limitation of time, Tamil marks the noun with the dative marker and Telugu nouns is in the nominative case.

(30) Te. *nēnu ikkaḍa iMkā reMḍu saMvatsarā-lu uM-ṭā-nu*.

I here more two year-PL.NOM be-FUT-1SG

Ta. *nāṇ iṇku iṇṇum iraṇṭu varuṭattiṛ-ku iru-pp-ēṇ*.

I here more two year-DAT be-FUT-1SG

'I will be here for two more years.'

3.3.4 Distribution: The dative inflected noun in Telugu and Tamil exhibit distributive sense. Telugu inflects with the dative in the second part of the reduplicated noun to show the distributiveness. Whereas, in Tamil the first part of the noun is inflected for dative to express the same. The following examples illustrate the dative with noun[-animate] as in (31) and [+animate] as in (32) .

<p>(31) Te.<i>iMṭi iMṭi-ki ceṭ-lu peMcu-dā-M.</i></p> <p>house.OBL house-DAT tree-pl grow-FUT-2PL</p> <p>Ta.<i>vīṭṭ-ukku vīṭu maram vaḷar-pp-ōm.</i></p> <p>house-DAT house.Ø tree.Ø grow-FUT-2PL</p> <p>'We grow trees in each house.'</p>	<p>(32) Te.<i>maniSi maniSi-ki tēḍā uM-di.</i></p> <p>human.OBL human-dat difference be-PRS.3SGN</p> <p>Ta.<i>maṇitaṇ-ukku maṇitaṇ vērupāṭu uḷ-ḷ-atu.</i></p> <p>human-DAT human.Ø difference be-PRS-3SGN</p> <p>'Each human has a difference.'</p>
---	---

3.3.5 Standard of Comparison: While comparing two noun phrases, the standard of comparison in Tamil is expressed by the marking of the dative case, whereas the Telugu noun is marked for the associative and optionally the dative case.

(33) Te.*nuvvu nā-tō/ nā-ku samānaM*

you.NOM I-SOC/I-DAT equal

Ta.*nī eṇ-akku.c camam*

you.NOM I-DAT equal

'you are equal to me'

3.3.6 With verbs of *reach*, *belong to* and *touch*: The arguments of verbs such as reach (34), belong to (35) and touch (36) require the dative marking in Telugu and the accusative marking in Tamil.

<p>(34) Te. <i>iMṭi-ki cērukonn-ā-v-ā?</i></p> <p>house-DAT reach-PST-2SG-INTR</p> <p>Ta. <i>vīṭṭ-ai ceṇṇṇai-nt-āy-ā?</i></p> <p>house-ACC reach-PST-2SG-INTR</p> <p>'Did you reach home?'</p>	<p>(35) Te. <i>vāḍu japān-ku ceMd-ina-vāḍu.</i></p> <p>he.NOM Japan-DAT belong-PST-ADJ_PART-3SGM</p> <p>Ta. <i>avaṇ japāṇ-ai.c cēr-nta-vaṇ.</i></p> <p>he Japan-ACC belong-PST-ADJ_PART-3SGM</p> <p>'He belongs to Japan'</p>
--	---

(36) Te.*nā illu nadi-ki ānukon-i uM-di.*

my house.Ø river-DAT touch-CPM be.HAB-3SGN

Ta.*eṇ vīṭu nadi.y-ai oṭṭ-i iru-kkiṛ-ātu.*

my house.Ø river-ACC touch-CPM be.HAB-3SGN

'My house is located close to river'

4. Case Assigned By Postpositions:

There are postpositions in both languages, which assign case to their complements (noun phrase) which is morphologically manifested. For instance, consider the following example, where *nōkki* 'towards' in Tamil assigns

accusative to the right to the NP that is governed, whereas the corresponding PP in Telugu assigns an oblique case .
Example:

(37) Te.kumār nī kēsi ve||-ā-ḍu.

Kumar you.OBL towards-DAT go-PST-3SGM

Ta.kumār uṇṇ-ai nōkki.p pō-ṇ-āṇ.

Kumar you-ACC towards go-PST-3SGM

'Kumar went towards you.'

The other examples are,

TeluguTamilMeaning

N_NOM + tappaN_ACC + tavira'except'

N_OBL+ lāmṭiN_ACC + pōl/pōla'like'

N_OBL + guriMciN_ACC+ kuṛittu'about'

N_ACC + baṭṭiN_GEN + pati'accordingly' and etc.,

5. Conclusion

Though Telugu and Tamil are genetically related languages, the above case distributions demonstrate that there are lots of divergences between which poses problems in MT. The divergence due to case mismatches are the major ones that should be tackled systematically in any Machine Translation system. Our Machine Translation system handles these types of cross-linguistic variations by formulating appropriate rules in Transfer Grammar which ensures the better output in MT.

References

1. Aissen, Judith. 2003. *Differential object marking: iconicity vs. economy*. *Natural Language and Linguistic Theory* 21. 435-83.
2. Akshar Bharati and Vineet Chaitanya and Rajeev Sangal. 1995. *Natural Language Processing: A Paninian Perspective*, Prentice-Hall, New Delhi.
3. Arden, A. H. 1942. *A progressive grammar of the Tamil language*. Madras: Christian Literature Society.
4. Baerman, M. 2009. *Case syncretism*. The Oxford Handbook of Case, ed. by A. Malchukov and A. Spencer, 219-230. Oxford: Oxford University Press
5. Blake, Barry. 2001. *Case*. Cambridge: Cambridge University Press. Second Edition.
6. Bosson, Georg. 1991. *Differential object marking in Romance and beyond*. In *New analyses in Romance linguistics: selected papers from the XVIII Linguistic Symposium on Romance Languages*. Dieter Wanner and Douglas Kibbee eds. Amsterdam: John Benjamins. 143-70.
7. Butt, Miriam. 2006. *Theories of Case*. Cambridge: Cambridge University Press.
8. Comrie, B. 1989. *Language Universals and Linguistic Typology*. 2nd edn, Chicago: University of Chicago Press.
9. Comrie, B. 1991. *Form and Function in Identifying Cases*. In *Paradigms: the Economy of Inflection*. F. Plank (ed.), (Empirical Approaches to Language Typology 9), 41-55. Berlin: Mouton de Gruyter.
10. Dorr, Bonnie. 1993. *Machine Translation: A View from the Lexicon*. Cambridge, Mass: The MIT Press.
11. Krishnamurti, Bh and Gwynn, J.P.L. 1985. *A Grammar of Modern Telugu*. New Delhi: OUP.
12. Lehmann, Thomas.S. 1989. *A Grammar of Modern Tamil*. Pondicherry: Pondicherry Institute of Linguistics and culture.

13. Ramanarasimham, P. 2007. *An Intensive Course in Telugu*. Mysore: CIIL.
14. Ramarao, C. 1975. *telugu vākyam*. Hyderabad: APSA.
15. Subbarao, K.V. 2012. *South Asian Languages: A syntactic Typology*. Cambridge: Cambridge University Press.
16. Subbarao, K. V. and P. Bhaskararao. 2004. *Non-nominative Subjects in Telugu*, in P. Bhaskararao and K. V. Subbarao (eds.), *Non-nominative Subjects, vol. II*. (Amsterdam/Philadelphia: John Benjamins), 161–196.
17. Uma Maheshwar Rao, G. 2002. *A Computational Grammar of Telugu*. mimeo (350pages) . Hyderabad: University of Hyderabad.

QUALITY TRANSLATION USING THE VAUQUOIS TRIANGLE FOR ENGLISH TO TAMIL

M.Mayavathi (dm.maya05@gmail.com)

K. Arul Deepa (karuldeepa@gmail.com)

Bharath Niketan Engineering College, Theni, Tamilnadu, India

ABSTRACT

The aim of this work is handling complex sentences and alignments of words. Hybrid Machine Translation is automatically acquires knowledge from large amounts of training data at different languages. The system is to translate complex sentence structures to process able chunks and translating the text English to Tamil. The system is first separates the source text word by word with POS category and searches for their corresponding target words in the bilingual dictionary. Rule Based Reordering, Morphological Analyzing, and dictionary based translation to the Target language. The transfer rules for reordering from English parse tree with respect to Tamil help us to get the output in the syntactic pattern of target language. The reordered output after morphological generation of Tamil words is displayed as the final output of the machine translation system and then errors in the translated sentences are corrected by applying Statistical technique.

1. INTRODUCTION

Machine Translation is a process of translating the sentences from one language to the other based on the information in the Knowledge Base without human intervention. There are three approaches to machine translation: Statistical, Example based and Rule based machine translation systems. Synchronous Tree Adjoining Grammar associated aligned tree/string training data and a method of converting these grammars to a weakly equivalent tree transducer for decoding.

Natural Language Processing is a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications. Also, the contents of the documents that are being searched will be represented at all their levels of meaning so that a true match between need and response can be found, no matter how either are expressed in their surface form. There is typically a well-defined problem setting, a standard metric for evaluating the task, standard corpora on which the task can be evaluated, and competitions devoted to the specific task.

In earlier years the machine translation is done only at the word level i.e. word by word translation. This project is carried out at many places for years but still need of a good translation system. Any basic translation requires two main view points: First is the linguistic point of view and second is the mathematical point of view. The three major techniques involved machine translations are [1] Rule Based, Statistical, and Example Based Technique. The Statistical and Example Based Techniques needs parallel corpora for translation. In such cases adopting only the statistical technique will not result in proper translation to the target language.

2. SYSTEM FEATURES

A Hybrid technique is developed for a system that generates simple sentences translation with part of speech tagging, chunking and morphological generator, segmentation is done. The preprocessing tool for machine translation that simplifies the complex sentences into simple sentences. This system uses rule based technique for sentence simplification and uses characters such as („, „, „?“) as delimiters for sentence separation. They have designed this system as a preprocessing tool for English to Tamil translation.

Hybrid machine translation (HMT) is leverages the strengths of statistical and rule-based translation methodologies. Several MT companies (Asia Online and Systran) are claiming to have a hybrid approach using both rules and statistics. The approaches differ in a number of ways: **Rules post-processed by statics**: Translations are performed using a rules based engine. Statistics are then used in an attempt to adjust/correct the output from the rules engine. **Statistics guided by rules**: Rules are used to pre-process data in an attempt to better guide the statistical engine. Rules are also used to

post-process the statistical output to perform functions such as normalization. This approach has a lot more power, flexibility and control when translating.

3. CURRENT WORK

The given source sentence is parsed and tagged using POS tagger, the tagged information is stored in a separate file. The rule based reordering of the sentence has to be done in the above formulated order using the tagged information. Chunking of the source sentence has to be done using the bi-gram model, and the bi-grams are translated into Tamil language by means of a word dictionary file. Then the word by word translation has to be done with the bilingual dictionary and if a word does not exist in dictionary, it may be a proper noun which is to be transliterated to Tamil language. Then apply gender ending rules to get the target output sentence. The error in the target sentence is corrected statistically using the file which contains collection of Tamil verbs with proper tense and gender endings. Finally the Tamil sentence for the corresponding English sentence is generated.

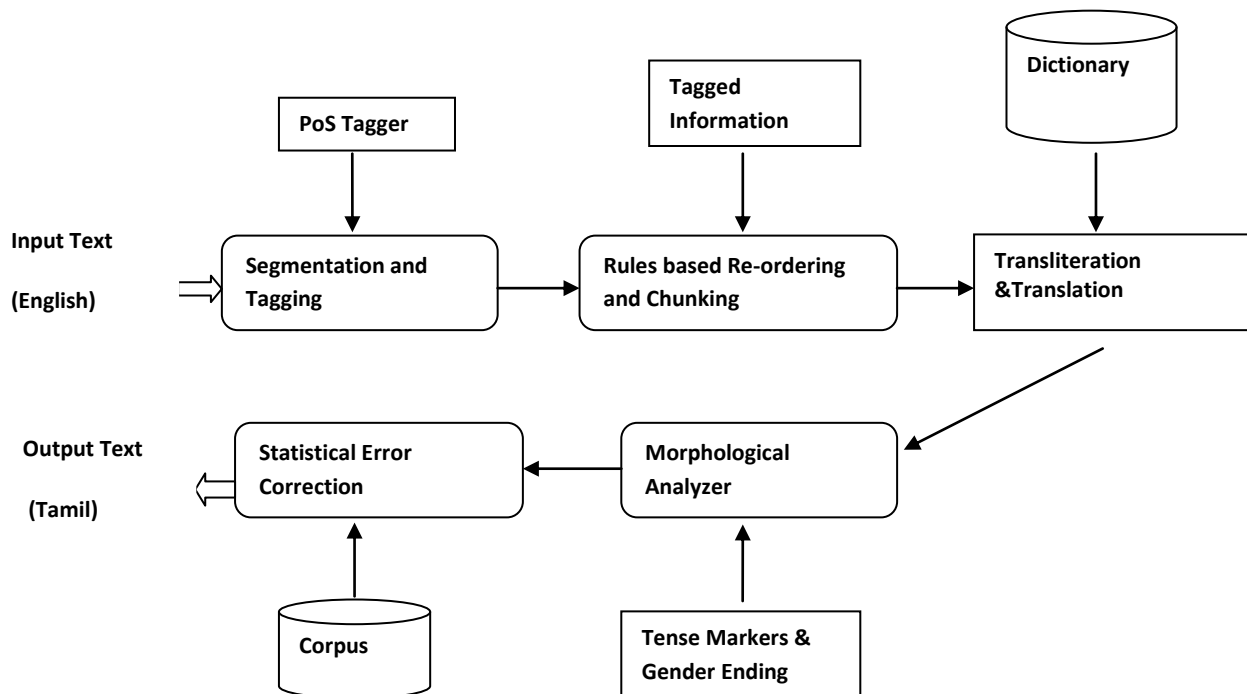


Figure 1: Overall System Architecture

4. SYSTEM METHODOLOGY

This paper presents an effective methodology for English to Tamil translation. Hybrid Machine Translation is handled by mapping from input to output sentence. Input is the English sentence which is enriched with segmentation, parsing and bilingual dictionary information. Output is a Tamil sentence with statistical error correction. The purpose is to group sequences of words are translated from source sentence to target sentence using hybrid techniques. The system can be translating complex sentences by creating new morphological reordering rules. Since a word in English has multiple meaning in Tamil, an effective word dictionary file (4500 words from English to Tamil) is used in order to achieve better results in translation. Gender ending verbs for all possible tenses were created for the purpose of statistically correcting the errors in the output sentences.

4.1 HMT Process

The process is acquires knowledge from training data and also enhance the input text with POS tagging and morphological information. After applying the local word grouping rules to the Tamil sentence (s), based on their four methods to process and align. (1) **Dictionary lookup approach (DL)** is used verbs and other groups are processed with DL

approach; HWGs with categories such as proper nouns, city, job-title, location, and country are processed with TS approach. **(2) Transliteration Similarity** is transliteration system maintains a consistent correspondence between the alphabets of two languages, irrespective of sound. Given two words, each from a different language, we define “transliteration similarity” as the measure of likeness between them. This could exist due to the word in one language being inherited or adopted by the other language, or because the word is a proper noun. Named entities such as city, job-title, location, country and proper nouns, all recognized by the local word grouping algorithm are compared using a transliteration similarity approach.

Neighbors approach works on this principle and aligns one or more words with one of the English words. Considering one HWG at a time, we find the nearest Tamil word that is already aligned with one or more English word (s) . We assume that the words in English-Tamil phrases follow a similar order and align the rest words in that group accordingly. The algorithm retrieves expected English word (s) from the HWGs and tries to locate them in the English sentence. This approach can be useful to locate one or more English words that align with one or more Tamil words.

INPUT

Ravi waited for the train but the train was late.

OUTPUT

ரவி ரயிலுக்காக காத்திருந்தான் ஆனால் ரயில் தாமதமாக வந்தது

Segmentation and Tagging

The segmentation and tagging of the source sentence is done using the Parts Of speech tagger. In our work we use Stanford POS tagger for the tagging purpose. The English sentences are taken as an input to the parts of speech tagger. The tagger tokenizes each word in a sentence and identifies the parts of speech information such as verb, noun, adjective etc. of that word. Then the words and their tagged information are stored in a separate file which is used for reordering of sentences. Characters are arranged in document lines following some type setting conventions which we can use to locate characters and find their style. The above complex sentence can be split into simple sentences.

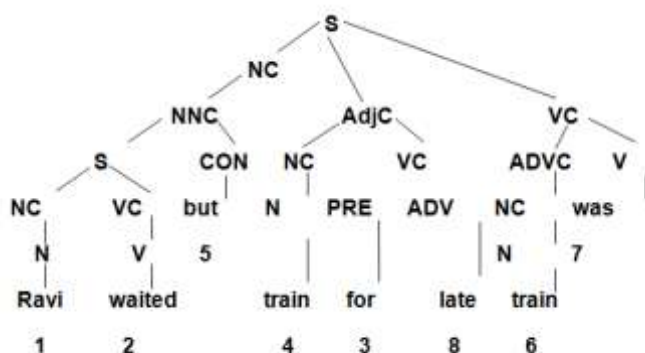
Segmentation: 1. Ravi waited for the train, 2. but the train was late

Tagging: Ravi // waited // for // the // train // but // the // train // was // late.

NNP VBD IN DT NN CC DT NN VBD JJ

Rule Based Reordering

The tagged words are stored separately for the purpose of reordering according to the morphological structure of the Tamil language. The tagged words are arranged according to the order which is mentioned below UH/ PP/ WP/WRB/WDT/ NNP/ PRP/ RB/ DT /CC/ JJ/ PP\$/ WP/JJR/ JJS/ IN/ NN/ NNS/TO/ VB/ VBD/ VBG/ VBN/ VBP/ VBZ/ MD. The above mentioned order suits all most all the types of simple sentence when reordering it from English to Tamil language.



Handling Complex sentence:

Noun, adjective and adverb clauses are considered.

Step 1: Conversion of complex to minimal sentence by grouping the clauses

Step 2: Minimal simple sentence can be analyzed as mentioned earlier

Step 3: Integration of clauses into the minimal simple sentence

4.2. Morph Analyzer/ Morph generator:

The source text is passed to the morphological analyzer. Morphological analyzer extracts the root word and its feature equations. These feature equations will be used in the later part to generate or add proper inflections to the target language. The sole purpose of this module is to handle the morphology of the target language. Features stored in the target structure might be needed for producing the proper inflected target lexical form.

4.3. Transliteration

The transliteration is the process of labeling the text in one language with other. In English to Tamil transliteration, the English text is replaced with the Tamil text by preserving the spell. The SVM based Multilingual Amrita English- Tamil Transliteration tool is developed by Amrita – CEN and we use the same in the machine translation system. First the corpus of English words are collected and preprocessed. The preprocessing involves two level Romanization, segmentation and alignment. The English words are romanized into Tamil words, by English- Tamil mapping. The romanized Tamil words are again romanized back to English, by Tamil - English mapping.

4.4. Statistical Error Correction Method

Even though we write gender ending rules, in some cases accurate verb with proper gender ending cannot be obtained for Tamil language. Particularly when writing rules for past tense sentences many contradictions arises. In such cases there occurs an error in the target sentence. For example consider a source sentence “Ravi waited for the train but the train was late” for that we may get the target sentence .we have around 70 base verbs in Tamil language with all possible gender ending and tenses.

Wrong Sentence

ரவி காத்திருந்தார் ரயில் ஆனால் தாமதமாக இருந்தது ரயில்

Correct Sentence

ரவி ரயிலுக்காக காத்திருந்தான் ஆனால் ரயில் தாமதமாக வந்தது

5. APPLICATIONS

The list of some of the most commonly researched applications of machine translation. There is typically a well-defined problem setting, a standard metric for evaluating the task, standard corpora on which the task can be evaluated, and competitions devoted to the specific task. The problem of sentence understanding deals with understanding individual sentences, and determining their meaning in the context of preceding sentences. The problem is divided into three stages: semantic parsing, semantic classification, and discourse modeling. Information retrieval (IR) is the area of study concerned with searching for documents, for information within documents, and for metadata about documents, as well as that of searching relational databases and the World Wide Web. The FST based morphological analyzer and generators are widely implemented for many languages. Automatic summarization involves reducing a text document or a larger corpus of multiple documents into a short set of words or paragraph that conveys the main meaning of the text. Summarizations of multimedia documents are used in education, website and etc. Greatly speed ups syntactic analysis Tagging is local and No need to process the whole sentence to find that a certain tag is incorrect. The application of POS Tagging is Senses depend on semantic context and less structured, longer distance dependency.

6. CONCLUSIONS

The overall design, architecture, functions, and translation methodologies are presented and reviewed in details. The system applies Translation Corresponding Tree structure for annotating bilingual and Constraint Synchronous Grammar for analyzing the syntax of bilingual texts. A major drawback with the statistical model is that it presupposes the existence of an aligned parallel corpus. The work done based on limits to translation of complex sentences from English to Tamil. The sentences are sub divided into words using Word-Based Translation models and words are aligned corresponding to translation models. Tree can be constructed based on word alignment. 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. If tokenization creates a one-to-one mapping, the number of tokens in both languages should be the same by adjusting this parameter. The system with all the necessary modules are in place, scalability is a key to improve its performance. Transliteration, Morph-synthesizer and extracting features are on its own a big task and these have to be enhanced as well to improve the overall performance of the system along with the root word lexicon and the reordering rules. The bilingual dictionary lacks the word sense information, so the semantic ambiguity arises in the system for many words.

7. FUTURE ENHANCEMENTS

In future works increasing the re-ordering rules; increase the database entries, fine tuning the Morph generator, and scalability. The system can be translating complex sentences by creating new morphological reordering rules. An effective word dictionary file (4500 words from English to Tamil) is used in order to achieve better results in translation. Gender ending verbs for all possible tenses were created for the purpose of statistically correcting the errors in the output sentences. The system is implemented using java codes. Multiple parse trees are used handled by the Stanford parser and the dependency parser is also used in the translation system. To handling the verbal phrases is possible by the system. The transliterator is limited to the Indian place names so performance of transliterator is very low when it's used for vocabulary words which are not present in the database. The morph generator is implemented for certain cases but the dependency information of many inflectional categories is given by the parser, such cases works well in morph generator and translation of sentences. The reordering rules are confined to the nodes of the branches and same rule could be handled for different cases with same syntactic structure. Question type of sentences are handled it is also one of the limitation of system.

REFERENCES

1. "A weighted tree automata toolkit", (May, J. & Knight, K -2006)
2. "An overview of probabilistic tree transducers for NLP", (Knight, K. & Graehl, J - 2005)
3. "Comparing Evaluation Metrics for Sentence Boundary Detection", (Yang Liu, -2007)
4. "Extending BLEU Evaluation Method with Linguistic Weight", (Lixin Wang Haoliang Qi, Sheng Li, Liu Daxin - 2008)
5. "Improving Statistical Machine Translation using Lexicalized Rules selection", (Zhongjun He1, Qun Liu1 -2008)

6. "Is Word Error Rate a Good Indicator for Spoken Language Understanding Accuracy", (Wang, Y., Acero, A., and Chelba, C -2003)
7. "Machine Translation System for Indian Languages", (Latha R. Nair, David Peter S- 2012)
8. "Phrase based English – Tamil Translation System by Concept Labeling using Translation Memory", (R. Harshawardhan, Mridula Sara Augustine and K.P. Soman-2011)
9. "Rule based Sentence Simplification for English to Tamil Machine Translation System", (Poornima.C Dhanalakshmi. V, Anand Kumar. M-2011)
10. "Segmentation and Alignment of parallel text for statistical machine translation", (Yonggangdeng.-2007)
11. "Synchronous Tree Adjoining Machine Translation", (Steve DeNeefe and Kevin Knight-2009)
12. "Semantic Role based Tamil Sentence Generator", (S. Lakshmana Pandian, V.Geetha-2010)
13. "Semantic translation error rate for evaluating translation systems", (Subramanian, K. Stallard.D, Prasad.R. S, Natarajan, P -2007)

DOSA 2020: THE FUTURE OF CELL PHONE AND TABLET PC IN TAMIL INTERNET USING *protégé*

A.M.PONRAJ

ASSOCIATE PROFESSOR, DEPARTMENT OF COMPUTER SCIENCE, MADURAI KAMARAJ UNIVERSITY

ponrajcomputer@gmail.com

ABSTRACT

Dosa Ontology is the very first multilingual ontology in Indian languages developed first in tamil letters using protégé . Dosa Ontology will be useful for all who are interested in developing semantic based tamil internet using protégé which was developed by Stanford Center for Biomedical Informatics Research (CBIR) . This ontology was developed to demonstrate various capabilities of protégé to execute queries in the semantic web. It describes the different types of dosa and its ingredients, Nutritive values for each dosa. People can know all the details of dosa by having look at the object properties such as Ingredient (சேர்வையுறுப்புக்கள்), spiciness (காரமதிப்பு), dosa size (அளவுஅடிப்படை), etc... The user also knows all the details of **Carbohydrates, Proteins, Lipids and Vitamins by having look at the data properties.**

Introduction

Aadhaar's guarantee of uniqueness and centralised, online identity verification would be the basis for building these multiple services and applications, and facilitating greater connectivity to markets Aadhaar would also give any resident the ability to access these services and resources, anytime, anywhere in the country Aadhaar can for example, provide the identity infrastructure for ensuring financial inclusion across the country – banks can link the unique number to a bank account for every resident, and use the online identity authentication to allow residents to access the account from anywhere in the country Aadhaar would also be a foundation for the effective enforcement of individual rights. A clear registration and recognition of the individual's identity with the state is necessary to implement their rights –to employment, education, food, etc. The number, by ensuring such registration and recognition of individuals, would help the state deliver these rights.

Let us lister to a sample imaginary dialogue that will take place in 2020 using cell phone and aadhaar.

ஆபரேட்டர்: இராயப்பர் உணவு விடுதியை அழைத்ததற்கு நன்றி. நான் உங்கள் தேசிய அடையாள அட்டையில் இருக்கும் எண்ணை தங்களிடமிருந்து பெறலாமா ?AADHAAR.....

வாடிக்கையாளர்: வணக்கம், நான் ஒரு கோரிக்கை வைக்க விரும்புகிறேன் .

ஆபரேட்டர்: முதலில் உங்கள் AADHAAR எண் வேண்டும், ஐயா .

வாடிக்கையாளர்: என் AADHAAR எண்... ம்ம், அது 6102049998-45-54610.

ஆபரேட்டர்: திரு மணி அவர்களே! உங்கள் தொலைபேசி எண் 494-2366. சரிதானே. உங்கள் அலுவலக எண் 65234167764-45-23234, உங்கள் செல் எண் 266-2566. மின்னஞ்சல் முகவரி mani@home.net <mailto: mani@home.net>.

வாடிக்கையாளர்: ஆ! நீங்கள் இந்த தகவலை பெற?

ஆபரேட்டர்: AADHAARல் உள்நாட்டு பாதுகாப்பு அமைப்பில் இந்த தகவலை கிடைத்தது, சார்.

வாடிக்கையாளர்: அது என்ன?

ஆபரேட்டர்: நாங்கள் உள்நாட்டு பாதுகாப்பு அமைப்பில் உள்ளோம் . இந்த உங்கள் ஆர்டர் நேரம் 15 விநாடிகள் மட்டுமே.

வாடிக்கையாளர்: (sighs) ஓ, நான் உங்கள் இறைச்சி சிறப்பு தோசை ஒரு ஜோடி சாப்பிட விரும்புகிறேன்.

ஆபரேட்டர்: அது ஒரு நல்ல ஆர்டர் என்று நினைக்கவில்லை, சார்.

வாடிக்கையாளர்: என்ன?

ஆபரேட்டர்: சார், உங்கள் மருத்துவ பதிவேடுகள், நீங்கள் அதிக கொழுப்பு மற்றும் உயர் இரத்த அழுத்தம் மிக்கவர் என்று குறிப்பிடுகின்றன. உங்கள் தேசிய உடல்நலம் வழங்குநர் பரிந்துரைக்கு மாறானா.... அது போன்ற ஒரு ஆரோக்கியமற்ற உணவை அனுமதிக்க மாட்டேன்.

வாடிக்கையாளர்: என்ன?!?! நீங்க....

ஆபரேட்டர்: நீங்கள் எங்கள் குறைந்த கொழுப்பு சோயாபீன்ஸ் தோசை முயற்சி செய்யலாம் .

வாடிக்கையாளர்: நான் அந்த மாதிரி ஏதாவது விரும்புகிறேன் என்று உங்களுக்கு எப்படித் தெரியும்

ஆபரேட்டர்: நீங்கள் கடந்த வாரம் உங்கள் உள்ளூர் நூலகத்தில் இருந்து இதைத்தானே சாப்பிட்டீர்கள்?

வாடிக்கையாளர்: சரி, சரி. எனக்கு இரண்டு குடும்ப அளவிலான சோயாபீன்ஸ் தோசை கொடுங்கள் .

ஆபரேட்டர்: அது உங்கள் மனைவி உங்கள் நான்கு குழந்தைகள் , உங்கள் 2 நாய்கள்.... எல்லாமே சாப்பிடலாம்

வாடிக்கையாளர்: நீங்கள் என் கடன் அட்டை என்னை வாங்கிக் கொள்ளலாமே...

ஆபரேட்டர்: மன்னித்துக்கொள்ளுங்கள், ஆனால் உங்கள் கடன் அளவு அதிகமாக உள்ளது .

வாடிக்கையாளர்: நான் ஏடிஎம் வழியாக, உங்கள் டிரைவரிடம் பணம் கொடுக்கலாமா?.

ஆபரேட்டர்: சார். உங்கள் சோதனை கணக்கு isoverdrawn.

வாடிக்கையாளர்: பரவாயில்லை! தோசை அனுப்புங்கள். நான் பணத்தை தயார் செய்ய வேண்டும். அது எவ்வளவு நேரம் ஆகும்?

ஆபரேட்டர்: இது 45 minutes ஆகும். அவசரம் என்றால் ..நீங்கள் ஒரு மோட்டார் சைக்கிளில் சென்று வெளியே பணம் பெற்று வரலாம்.

வாடிக்கையாளர்: எப்படி நான் ஒரு பைக் சவாரி என்று உங்களுக்குத் தெரியும்?

ஆபரேட்டர்: நீங்கள் கார் கடன் செலுத்தத் தவறினீர்கள், அதனால் உங்கள் கார் வங்கியால் பறிமுதல் செய்ய பட்டு விட்டது.

வாடிக்கையாளர்: சரி, நான் ஒரு # % # ^ & \$ % ^ \$ @ # வேண்டும்

ஆபரேட்டர்: நான் உங்கள் மொழி பார்த்து ஆலோசனை செல்ல விரும்புகிறேன், ஐயா.

வாடிக்கையாளர்: (பேச்சு)

ஆபரேட்டர்: வேறு எதுவும் இருக்க முடியுமா சார்?

வாடிக்கையாளர்: ஆமாம்..... கோக் ஒரு இலவச 2 லிட்டர் ஒரு கூப்பன் வேண்டும் ..

ஆபரேட்டர்: மன்னித்துக்கொள்ளுங்கள், ஆனால் எங்கள் விளம்பரத்தின் பிரத்தியேக பிரிவு (நீரிழிவு) இலவச சோடா வழங்குவதிலிருந்து இருந்து எங்களை தடுக்கிறது. இராயப்பர் உணவு விடுதியை அழைத்ததற்கு நன்றி!

In this paper, different types of dosa are collected for the ontology creation. The ontology is classified into Vegetarian Dosa, non-vegetarian dosa and fruit dosa etc... The ingredients for the dosa are inserted into the individuals. These classifications is used for user can view the data in meaningful way. This ontology is developed using protégé 4.2 Alpha.

Protégé 4.2 Alpha is a free and open -source platform. It used to developed a meaningful ontology. This editor has many tools to view and search the data. The tools available are DL-Query tab, owl2Query tab, ACE view tab, onto graph, Spreadsheet importer etc.

Using query tools, the user can retrieve the information related to the types of dosa, such as Nutritive value, ingredients, etc. Using Onto graph tool, the user can view the structure of the ontology, like class structure, properties and individuals. Using ACE view tool, the user can view the many different views to the ACE text.

1.2.1 What Is The Semantic Web?

The Semantic web is introduced by Tim Berners-Lee. ஜெனீவா நகரில் உள்ள "செர்ன்"/CERN (Conseil European pour la Recherche Nuclaire (as it called in French), or European Council for Nuclear Research (in English) குழுவைச் சேர்ந்த கணினி விஞ்ஞானி "டிம் பெர்னர்ஸ்- லீ" (Tim Berners-Lee) . இணையத்தின் ஆதாரமான "hypertext" ஐ

அறிமுகப்படுத்தி அதை அனைவரையும் முயன்று பார்க்க வெளியிட்ட அந்தக் கோரிக்கையே இணையத்தின் பிறப்பாக அனைவராலும் கருதப்படுகிறது .The Semantic Web is an extension of the current web that provides an easier way to find, share, reuse and combined information. It is based on machine readable information and builds on XML technologies and RDF flexible approach representing data. The semantic web provides common formats for the interchange of data.

Purpose of ontology

“People can’t share knowledge if they do not speak a common language” [Davenport & Prusak, 1998]. “An Ontology is an explicit specification of a conceptualization” [Gruber, 1993]. Ontology’s enable a better communication between Humans/Machines Ontology’s standardize and formalize the meaning of words through concepts.

ABOUT PROTÉGÉ

Protégé is a free, open-source ontology editor and knowledge-based framework. Protégé platform supports two modeling ontologies editor. That’s protégé-frames and protégé-owl editors. Protégé ontologies can be exported into a variety of formats including RDF (S), OWL and XML schema.

Protégé can be extended by way of a plug-in Architecture and a Java-based Application Programming Interface for building knowledge-based tools and applications.

Protégé-OWL editor enables users to:

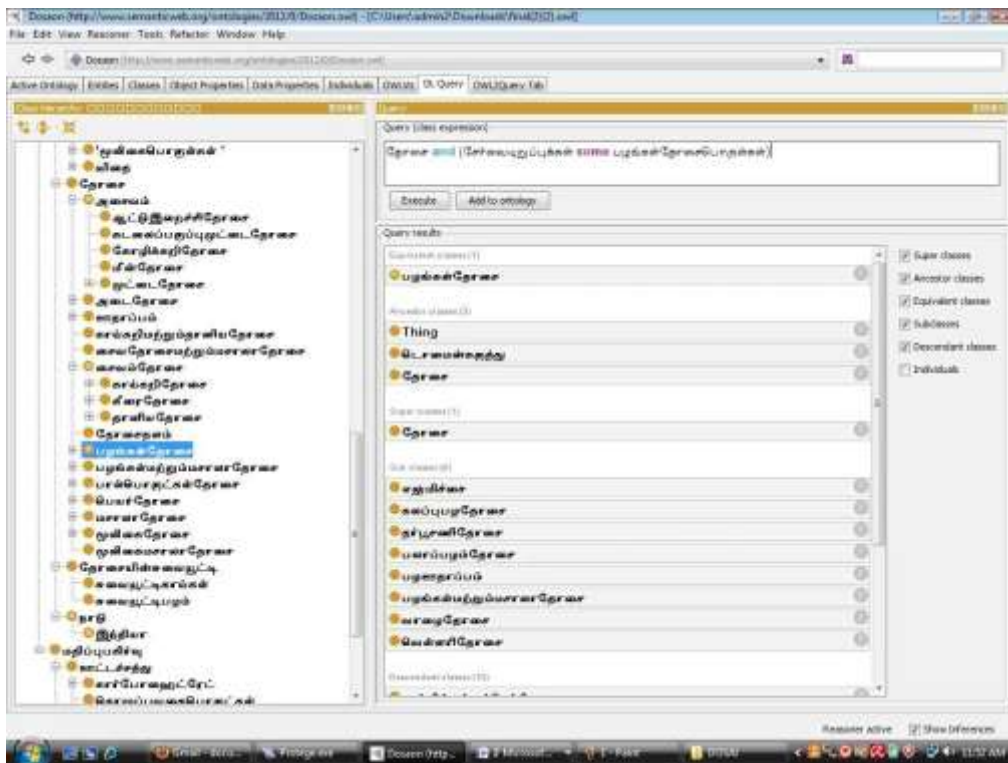
- Load and save OWL and RDF ontologies.
- Edit and visualize classes, properties, and Rules.
- Define logical class characteristics as OWL expressions.
- Execute reasoners such as description logic classifiers.
- Edit OWL individuals for Semantic Web markup.

PROTÉGÉ 4.2 alpha

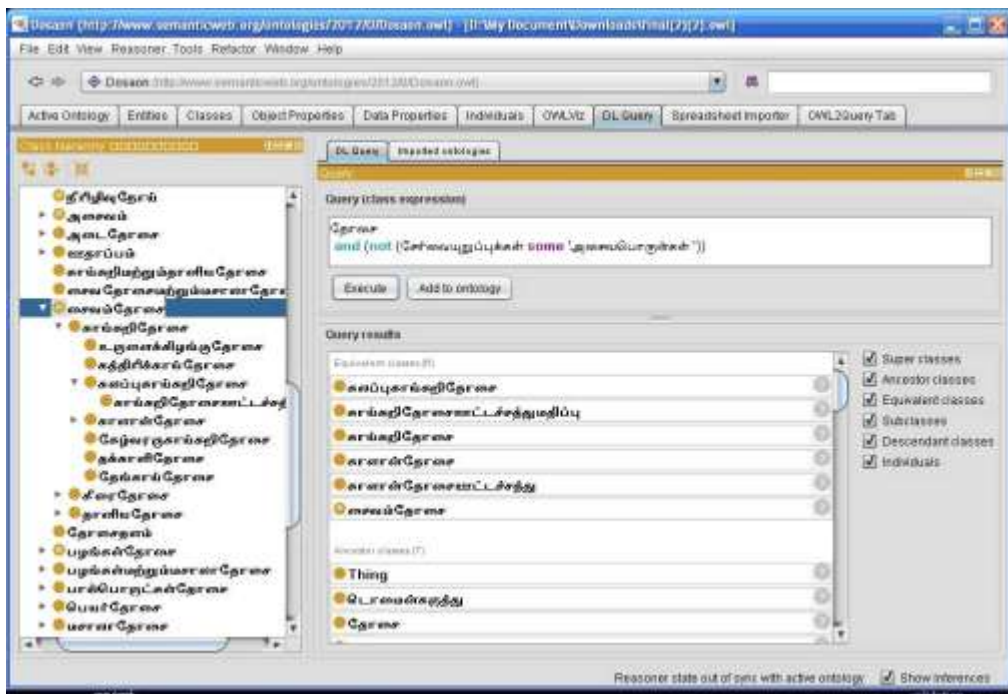
Protégé 4.2 Alpha is a free, open-source platform that provides a growing user community with a suite of tools to construct domain models and knowledge-based applications with ontologies At its core, Protégé implements a rich set of knowledge-modeling structures and actions that support the creation, visualization, and manipulation of ontologies in various representation formats. Protégé can be customized to provide domain-friendly support for creating knowledge models and entering data.

Conclusion

The semantic web is an extension of the current web in which information is given well defined meaning, so that the machines can understand the content of the web and provide the relevant information for the user’s query. The dosa ontology is created with the tool protege4.2 alpha. This helps the user to fetch the relevant information by using this in semantic web applications. We have demonstrated the usage of protégé for creating the ontology in tamil fonts. Protégé can create the java scripts which will run in selected hand held devices. This attempt is a pilot study for tamil e-commerce and e-business.



Appendix 2: A DL query executed in Protégé



Sense Mining for Tamil

W. Aisha Banu, B. Shamreen, S. Sasikala and B. Nimala Devi

B.S. Abdur Rahman University, Chennai

Abstract

Sense mining is classifying a sentence as subjective or objective and expressed it as Positive, Negative or Neutral. As the web has proliferated, opinions about various entities have also proliferated. There is a need to understand the sense of the contents expressed. The domain of sense mining focuses on this aspect. Sense mining for Tamil is a non-trivial task. The early work in the domain has focused on aspects such as bag of words, support vectors or rating systems. Natural languages like tamil need more sophisticated approaches which focus on the granularity of the grammar and richness of the lexicon. In this work, a simple sense mining approach that takes into account the capability of the lexicon is outlined. This work will form the base for more sophisticated approaches in the future.

1. Introduction

Sense mining is an unique research in text and information retrieval in that the focus is on aggregating the contents and deriving an overall foci. This foci is finally expressed in the form of a classification of word tense. This itself is tough as there are a lot of combinations that are possible in natural languages. Double negative, positive negatives are some of the combinations which can test even experts. The work in the domain can be classified into machine learning, lexicon based and grammar based systems. The challenge in machine learning is the adaptability to new untrained text. The lexicon based statistical works focus on the advances in the clustering systems to identify the cluster in which the smoothened output resides. This cluster based approaches work with lexicons and are effective for untrained text. But, classification is difficult. Grammar based approaches rely on the codification of the text. The issue here is that lots of opinions need not follow the subject verb object structure of the grammar.

The seminal work in this domain is [1] which provides a framework and methods for identifying the sentiments and a rich set of rules for the analysis of subjectivity. This is the basis for this work. [2] provided a framework for the architecture and implementation of the system. They have given a direction for the work to be extended in the future. The approach used in [3] to classify the words into the categories (Nouns, verb, adverb) are used in this work.

2. Proposed work

The stages in Sentiment mining for Tamil are given in Figure 1. The key stages are as follows.

- Linguistic processing
- Parts of speech and semantic tagging
- Opinion identification

Linguistic preprocessing

The two steps involved are the initial splitting and form reduction. The text in the opinion is split into sentence level structures. This is done with a parser and the use of syntactical constructs like ".".

Parts of speech and semantic tagging

In this stage, the sentences are tagged with the use cases such as

- JJ: adjective
- CD: cardinal and
- NN/NNS, VB/VBD/VBN/VBG, RB/RBR/RBS represent different forms of noun, verb and adverb.

The critical aspect is that lexicons must be used for this stage.

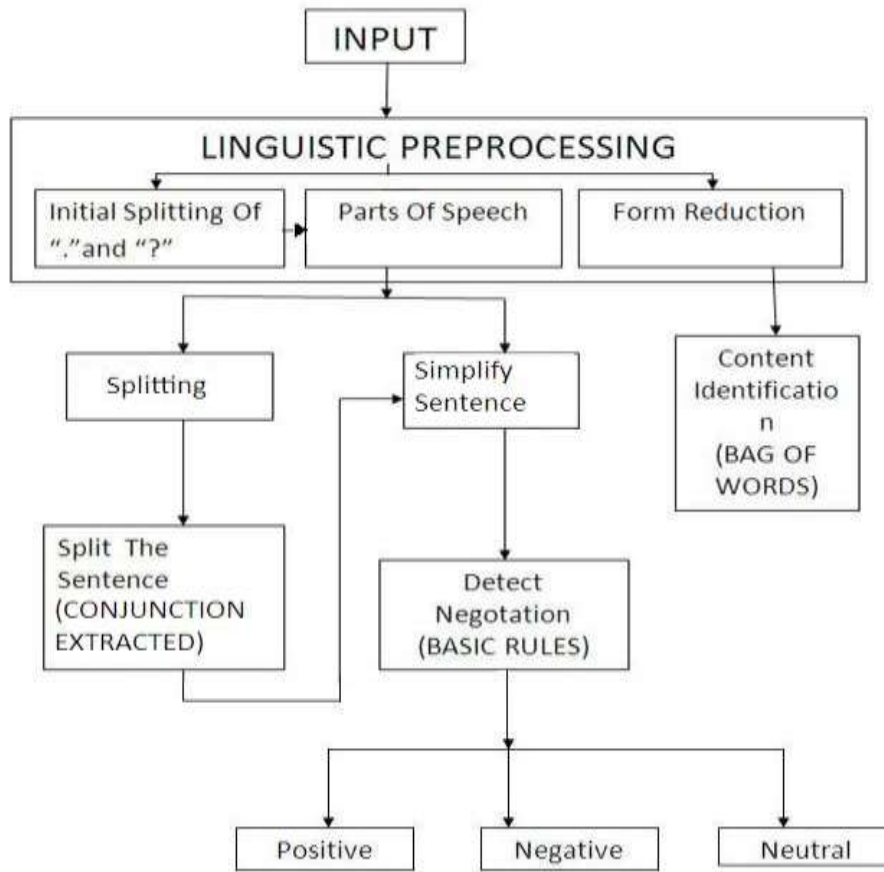


Figure 1: Overall architecture

Opinion identification

The lexicons can give the sense of the words. But, a classification system is needed to identify the proper sense of the terms. The following rules are used as a guideline.

- Negative --> Negative
- Positive --> Positive
- Negative Negative --> Positive
- Negative Positive --> Negative
- Decreased Negative --> Positive
- Decreased Positive --> Negative

The output for various use cases are shown below. If the sentence is neither positive or negative it is classified as neutral.

4. Methodology and use cases

The following examples explain the thrust of this work.

EXAMPLE 1

நாங்க நல்ல நண்பர்கள் எங்களுக்குள் போட்டியோ பொறாமையோ இல்லை என்று சிம்பு, தனுஷ் கூட்டாக அறிவித்தும் அவர்களை நண்பர்களாகப் பார்க்க சமூகம் மறுக்கிறது. எப்போதாவது இவர்கள் இருவரும் சேர்ந்து மேடையில் தோன்றுவதுண்டு ஆனால் அடுத்தநாளே இந்த நட்பு நாடகம் வெளுத்துவிடும்.

LINGUISTIC PREPROCESSING:

SIMPLE SPLITTING

1. நாங்க நல்ல நண்பர்கள் எங்களுக்குள் போட்டியோ பொறாமையோ இல்லை என்று சிம்பு, தனுஷ் கூட்டாக அறிவித்தும் அவர்களை நண்பர்களாகப் பார்க்க சமூகம் மறுக்கிறது
2. எப்போதாவது இவர்கள் இருவரும் சேர்ந்து மேடையில் தோன்றுவதுண்டு ஆனால் அடுத்தநாளே இந்த நட்பு நாடகம் வெளுத்துவிடும்

CONJUNCTION

1. நாங்க நல்ல நண்பர்கள் எங்களுக்குள் போட்டியோ பொறாமையோ இல்லை என்று சிம்பு, தனுஷ் கூட்டாக அறிவித்தும் அவர்களை நண்பர்களாகப் பார்க்க சமூகம் மறுக்கிறது
2. எப்போதாவது இவர்கள் இருவரும் சேர்ந்து மேடையில் தோன்றுவதுண்டு
3. ஆனால் அடுத்தநாளே இந்த நட்பு நாடகம் வெளுத்துவிடும்

PARTS OF SPEECH

1. நாங்க) NNS) நல்ல (JJ) நண்பர்கள் (NNS) எங்களுக்குள் (NNS) போட்டியோ (NN) பொறாமையோ (JJ) இல்லை (RB) என்று சிம்பு) NN), தனுஷ் (NN) கூட்டாக (RB) அறிவித்தும் (NN) அவர்களை) NNS) நண்பர்களாகப் (NNS) பார்க்க (VB) சமூகம் (NNS) மறுக்கிறது (VB)
2. எப்போதாவது) RB) இவர்கள்) NNS) இருவரும்) NNS) சேர்ந்து) RB) மேடையில் (NN) தோன்றுவதுண்டு (VB)
3. ஆனால்) IN) அடுத்தநாளே) JJ) இந்த) JJ) நட்பு) JJ) நாடகம்) NN) வெளுத்துவிடும். (RB)

Output

1. நல்ல, போட்டியோ, பொறாமையோ are POSITIVE. மறுக்கிறது is NEGATIVE
2. No Polarity is found. Its NEUTRAL.
3. வெளுத்துவிடும் is negative.

So the given paragraphs polarity is NEGATIVE.

EXAMPLE 2

மங்காத்தாவின் இந்திப் பதிப்பில் அஜித்தை நடிக்க வைக்க டைரக்டர் வெங்கட்பிரபு திட்டமிட்டுள்ளார். அஜித் நெகட்டிவ் ஹீரோ பாத்திரத்தில் கலக்கியிருந்தார் ஆனால் வெங்கட் பிரபுவு் பிரியாணி படத்தில் பிஸியாக இருப்பதால் இப்போதைக்கு இந்தி மங்காத்தா தாமதம் ஆகும் எனத் தெரிகிறது.

LINGUISTIC PREPROCESSING

SIMPLE SPLITTING

1. மங்காத்தாவின் இந்திப் பதிப்பில் அஜித்தை நடிக்க வைக்க டைரக்டர் வெங்கட்பிரபு திட்டமிட்டுள்ளார்
2. அஜித் நெகட்டிவ் ஹீரோ பாத்திரத்தில் கலக்கியிருந்தார் ஆனால் வெங்கட் பிரபுவு் பிரியாணி படத்தில் பிஸியாக இருப்பதால் இப்போதைக்கு இந்தி மங்காத்தா தாமதம் ஆகும் எனத் தெரிகிறது.

CONJUNCTION

1. மங்காத்தாவின் இந்திப் பதிப்பில் அஜித்தை நடிக்க வைக்க டைரக்டர் வெங்கட்பிரபு திட்டமிட்டுள்ளார்
2. அஜித் நெகட்டிவ் ஹீரோ பாத்திரத்தில் கலக்கியிருந்தார்
3. ஆனால் வெங்கட் பிரபுவு் பிரியாணி படத்தில் பிஸியாக இருப்பதால் இப்போதைக்கு இந்தி மங்காத்தா தாமதம் ஆகும் எனத் தெரிகிறது

PARTS OF SPEECH

1. மங்காத்தாவின் (NN) இந்திப் (NN) பதிப்பில் (JJ) அஜித்தை (NN) நடிக்க (VB) வைக்க (VB) டைரக்டர் (NN) வெங்கட்பிரபு (NN) திட்டமிட்டுள்ளார் (VB) .
2. அஜித் (NN) நெகட்டிவ் (JJ) ஹீரோ (RB) பாத்திரத்தில் (NN) கலக்கியிருந்தார் (JJ)
3. ஆனால் (CC) வெங்கட் பிரபுவு் (NN) பிரியாணி (NN) படத்தில் (NN) பிஸியாக (JJ) இருப்பதால் (CC) இப்போதைக்கு (NN) இந்தி (NN) மங்காத்தா (NN) தாமதம் (JJ) ஆகும் (VB) எனத் (DT) தெரிகிறது. (JJ)

OUTPUT

1. திட்டமிட்டுள்ளார் **evitisoP** si
2. நெகட்டிவ் ஹீரோ, கலக்கியிருந்தார் **evitisop** era
3. தாமதம் is **negative**. so the given paragraph is **NEUTRAL**

4. Results

So far we have progressed in identifying the overall sense of reviews. The work has been done by hand and the algorithm and architecture are under development in the software system now. The results of the work over 18 use cases show that the results obtained in this work are consistent with the overall expert opinions. For this, the approach was applied to the select reviews. An expert well versed in movies was asked to give the tense of the work without knowing the results of the proposed approach. The results were compared.

5. Conclusion and future work

At present the algorithm is at a developmental stage, but shows a lot of promise for Tamil. The algorithm will be customized for a large set of use cases and combinations and implemented in the software in future and the results shared. The proposed approach focuses only on three sense: Positive, negative and neutral. In future this will be expanded with more sense words being used.

References

1. B. Liu, Sentiment Analysis and Subjectivity, Handbook of Natural Language Processing, Second Edition, 2010.
2. C. Rohrdantz, M. C. Hao, U. Dayal, L-E Haug and D. A. Keim, Feature-Based Visual Sentiment Analysis of Text Document Streams. ACM Transactions on Intelligent Ssystems and Technology, 3 (2): 26 (2012)
3. K.M. A. Kumar and Suresha, Lexical analysis to effectively detect Users' opinion, International Journal of Web & Semantic Technology, 2 (4), pp. 105-113, October 2011.

Sense Mining for Natural Language Processing using Collaborative Filtering Approach

Dr. R. Shriram (shriram@bsauniv.ac.in)

B.Manikandan (bmanibala@gmail.com)

Dept of Computer Science and Engg, BS Abdur Rahman University, Chennai

Abstract

Text mining refers to the process of deriving high-quality information from text. Analysis of sentiments of the content for natural languages is a non-trivial task. Existing methods consider only the keywords or ontology based 'related words'. The sentiments expressed are difficult to express and manipulate. The objective of this work is to propose a methodology for sentiment mining of natural language content (Tamil) for a specific domain. A specific domain (movie reviews) is taken and the sentiment mining system is applied for this system. The novelty of this work is in the methodology of the sentiment mining system for Tamil.

1. Introduction

Reading and understanding a sentence could be easy for humans even though there are lots of ambiguities. To overcome this problem Word Sense Disambiguation (WSD) task is to be done which assigns the appropriate sense for all the occurrences for ambiguous words. Sentimental Analysis is a complicated field which involves the process and interpretation of Natural Languages. It does not depend on specific sentence or text. The first work in sentiment mining involved the use of NLP techniques. [1] used part of speech tagging to estimate a 'semantic orientation' of meaningful phrases in Opinions reviews. The entire review was then classified as either positive or negative by averaging the semantic orientation of the phrases within. The application of a number of different machine learning techniques to sentiment classification. A Naive Bayes classifier, maximum entropy and support vector machines were used classify movie reviews as either positive or negative. Using unigrams, the support vector machine gave higher levels of accuracy, but needed trained data sets[2]. Part of speech tagging is the most commonly encountered, though there are also papers detailing classifiers using co reference resolution[3], and even the use of a full syntactic parse tree. The Stanford POS tagger uses a maximum entropy technique [4]. To avoid constantly referencing a subject by name, natural languages usually contain alternative words that can be used when referring to a previously mentioned subject.

2. Proposed System

Our proposed approach consists of the following stages: Parsing, Parts of speech tagging and Emotion classification. In our work, the crucial contribution is in the emotion classification process. The emotion classification system considers the entire document as a source and tries to make many passes over the given text to identify the proper word senses and disambiguate it. The lexicon is modified to account for

- similarity to the classifiers
- encode domain knowledge
- encode the knowledge about the structure of reviews.

In movies, the following classifiers: action, thriller, sentiment, sadness, love and comedy are used. Hence not only is each term/Part of speech parsed for its meaning, but its relevance to the classifier. In our experiments we have used the existing movie review to determine the emotion. Movie review contains positive and negative sentences of the movie. We experimented the result using Proposed approach, Statistical approach and Expert View. The statistical view comprises of counting the occurrence of terms to the classifiers and categorizing it. Experts views are taken as the final judgment. The following table (Table 1) shows evaluation of result using various methods.

Table 1: Comparison of methods

Name of the Movie	Proposed Approach	Statistical Approach	Expert View
Thupakki	Action	Action	Action
Pizza	Thriller	Action	Thriller
Naduvuala konjam Pakktha Kanam	Comedy	Comedy	Comedy
Thandavam	Action + Love	Action	Action + Love
Poda Podi	Love	Love	Love
Neerparavai	Sadness	Sadness	Sadness

Statistical Approach: Based on statistical approach (Figure 1) thupakki movie falls under action movie.

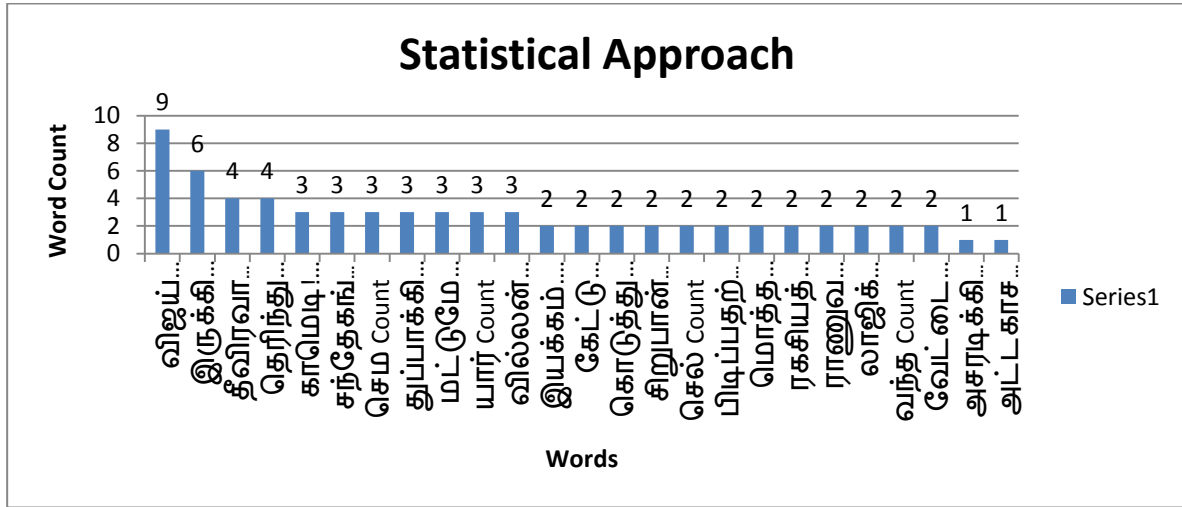


Figure 1: Statistical approach for Thupakki

Proposed Approach: By doing proposed approach we have chosen the words which describes about the movie exactly. The below mentioned are the words taken form the thupakki movie review.

தீவிரவாதிகளை வேட்டையாடும் ராணுவத் 'துப்பாக்கி, தீவிரவாதியைப், குண்டுவெடிப்புத், வேட்டையாடுகிறார், சகாக்களைக் கொன்ற , டிக்டிக் தித்திக் க்ளைமாக்ஸ் , அனைவரையும் சுட்டே சாகடிக்கிறார் , காமெடிக்கு, செம சீரியஸ் காமெடி.

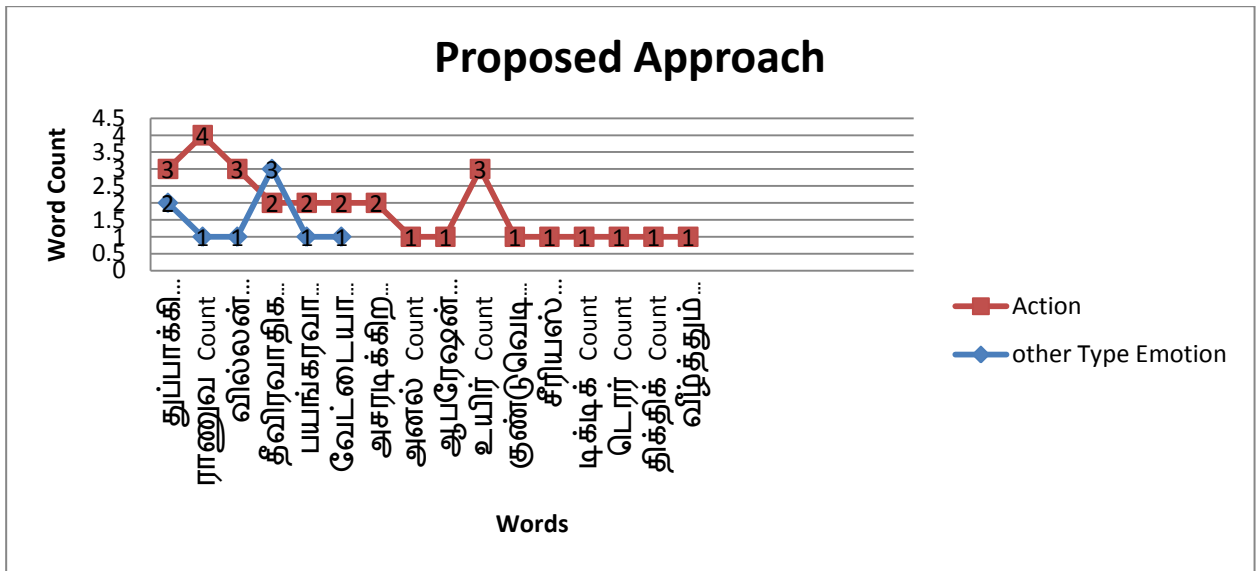


Figure 2: Proposed Approach for Thupakki

Based on proposed approach thupakki movie comes under action movie.

Statistical Approach: Based on statistical approach pizza movie falls under action movie.

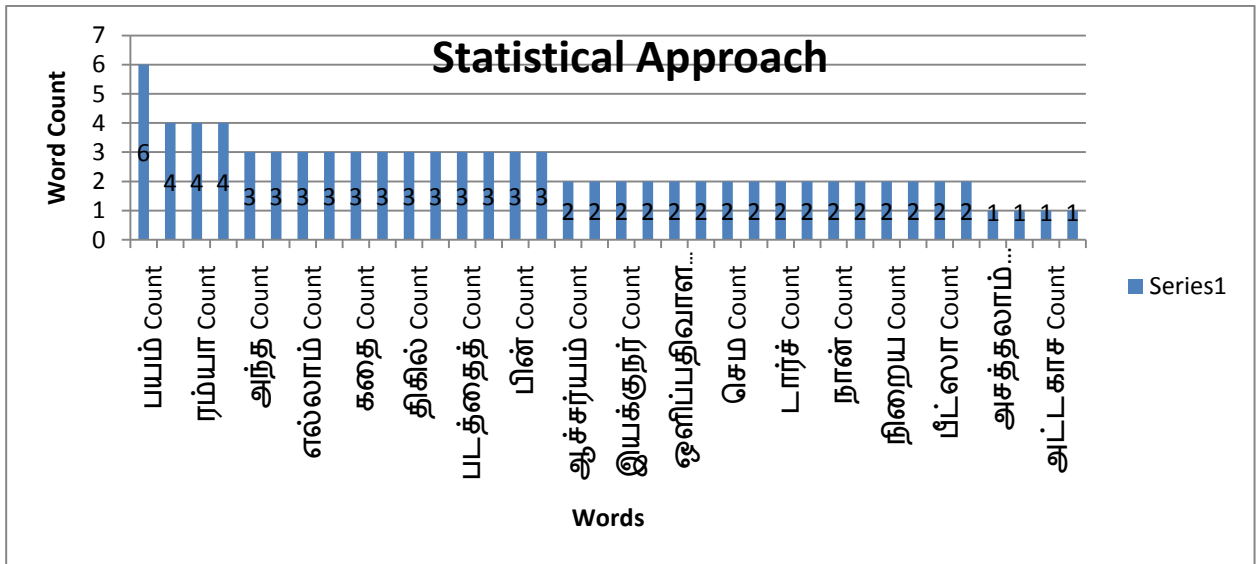


Figure 3: Statistical approach for Pizza

Proposed Approach: The below mentioned are the words taken form the pizza movie review.

ஜில்' ரொமான்ஸ், 'திக்' திகில், 'அட' ஆச்சர்யம் கலந்துகட்டிய செம ஹாட் தரில், திகில் படத்துக்குக் , இருட்டு, நிறைய நிறையத் திகில் , 'மிரட்டி', முதல் பாதி முழுக்கப் பயத்துடன் திகில், கதறுவதுமாக, இரவில் இருளில் , பயத்தில் உதறி வெலவெலத்து, டெர்ர அத்தியாயம் பயமுறுத்தும், பேய் நாவல்.

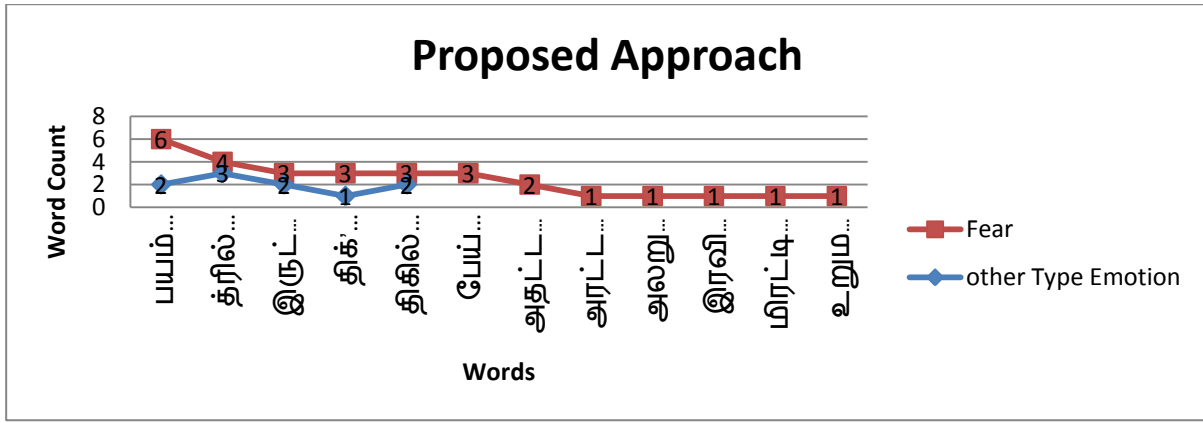


Figure 4: Proposed Approach for Pizza

Based on proposed approach pizza movie comes under thriller movie. In all our use cases so far done, the proposed approach is always equal to or better than the performance of the statistical approach. This is because our approach not only counts the occurrence but defines the co-occurrence based on the lexicon as well.

3. Conclusion

The main technique used to sentiment mining was the emotion classification system. This paper gives an outline for the sentiment mining for Tamil movie reviews in Tamil. Further development of framework could be applying technique can also be implemented using the software. So far, we have implemented and tested this by hand only. In future we can explore the possibility of dynamic online based systems.

4. References

1. Turney, P., (2002) "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews". Proceedings of the 40th Annual Meeting on Association For Computational Linguistics. New Jersey, p. 417-424.
2. Pang, B., Lee, L., and Vaithyanathan, S. (2002) "Thumbs up?: sentiment classification using machine learning techniques." Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10, New Jersey, p. 79-86.
3. Nicolov, N.; Salvetti, F. and Ivanova, S. (2008) . "Sentiment Analysis: Does Coreference Matter?" AISB 2008 Convention Communication, Interaction and Social Intelligence.
4. <http://nlp.stanford.edu/software/tagger.shtml>
5. Fernando, P, Naftali, T., Lillian, L. (1993) "Distributional Clustering of English Words ". (1993) Proceedings of ACL '93, pp. 183-190.

SEMANTIC ROLE LABELING OF TAMIL DOCUMENTS

AN AID FOR CONCEPT - BASED CLASSIFICATION

V. Janani (jananise@gmail.com)

K. Arul Deepa (karuldeepa@gmail.com)

Bharath Niketan Engineering College, Theni, Tamilnadu, India.

ABSTRACT

The aim of this work is to design and implement a system to identify, analyze and tag the constituents in the sentence which fill a semantic role expressed by some target verbs of a sentence in Tamil. The system reads a Tamil text document and performs tagging of semantic roles associated with a given target verb within such a document. Any two terms can have the same frequency in the documents, but one term contributes more to the meaning of its sentences than the other term. Thus, the underlying model should indicate terms that capture the semantics of text. In this case, the model can capture terms that present the concepts of the sentence, which leads to discover the topic of the document.

1. INTRODUCTION

Natural Language Processing (NLP) is the computerized approach to analyzing text that is based on both a set of theories and a set of technologies. Natural Language Processing is a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications.

A new concept-based mining model that relies on the analysis of both the sentence and the document is introduced. The concept-based model can effectively discriminate between non-important terms with respect to sentence semantics and terms which hold the concepts that represent the sentence meaning. It performs the mining process based on lists of words and stems. It uses a concept-based approach to add semantics into the mining process. Thus the system proposes an effective and simple approach of combining different semantic role labeling systems which significantly improves its performance.

In the few last years there has been an increasing interest in shallow semantic parsing of natural language, which is becoming an important component in all kind of NLP applications. As a particular case, Semantic Role Labeling (SRL) is currently a well defined task with a substantial body of work and comparative evaluation. Given a sentence, the task consists of analyzing the propositions expressed by some target verbs of the sentence [2]. In particular, for each target verb all the constituents in the sentence which fill a semantic role of the verb have to be recognized. Typical semantic arguments include Agent, Patient, Instrument, etc. and also adjuncts such as Locative, Temporal, Manner, Cause, etc.

2. SYSTEM FEATURES

From the set of concept terms and their semantic role, a machine learning technique is developed for a system that generates simple sentences with labels. By annotating the corpus using part of speech tagging, chunking and semantic role labeling processes, preprocessing work is done. These are capable of producing the appropriate sequence of semantic roles of the concept term for realizing sentence.

SRL (Semantic Role Labeling) is the process of detecting basic event structures such as who did what to whom, when and where. It identifies the arguments of a given verb and assigns them semantic labels describing the roles that play in the predicate i.e., identify predicate argument structures. There are many syntactic variations possible in the understanding semantics of the Tamil sentence. The structural view of mapping from input to output structures involves the input is to be text enriched with morpho-syntactic information and output is to be a sequence of labeled arguments[7]. The linguistic nature of the problem is handled by the argument identification which is strongly related to syntax.

3. CURRENT WORK

POS tagging tool for Tamil has been used for POS tagging of Tamil document corpus. The POS tagged corpus is input for the Chunk parsing. Chunking tool for Tamil has been used for Chunking of POS tagged corpus. This Phrase chunked corpus is input to the Semantic role labeling process [3]. Automatic generation of natural language text is an essential task in many language processing like summarization, automatic document generation, question answering system, translator etc. This learner is to be designed as a statistical model that is formulated from a preprocessed corpus of sentences. The semantic role labeling is also a tagging process by which the semantic roles like subject, action, agent, co agent, location, time, manner, direction, cause, extent, purpose, negation and modal of each concept in a sentence[10].

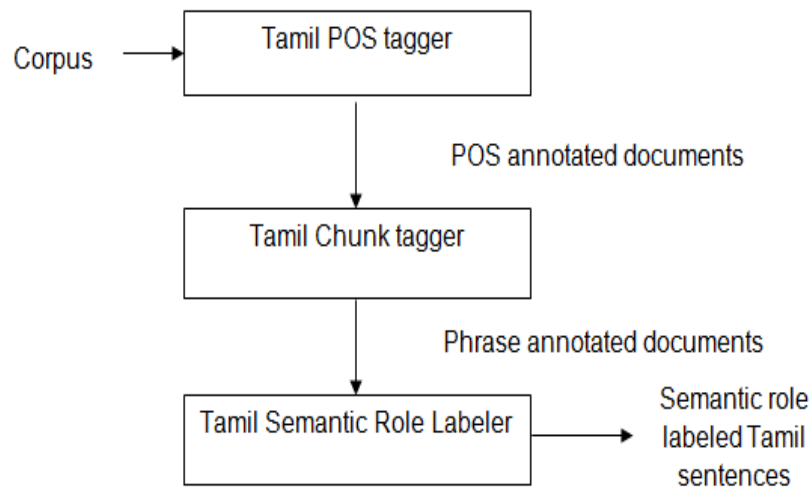


Figure 1: Overall System Architecture

4. SYSTEM METHODOLOGY

In this paper, syntactic variations are handled by mapping from input to output structures. Input is the text which is enriched with morpho - syntactic information. Output is a sequence of labeled arguments. In semantic role labeling the goal is to group sequences of words together and classify them by using semantic labels. For meaning representation the predicate-argument structure that exists in most languages is used. In this structure a word (most frequently a verb) is specified as a predicate, and a number of word groups are considered as arguments accompanying the word (or predicate) .

4.1. SRL – Process

The process involves the identification the semantic relationships, or semantic roles, filled by constituents of a sentence within a semantic frame. Given an input sentence, the system labels constituents with either abstract semantic roles such as AGENT or PATIENT, or more domain-specific semantic roles such as SPEAKER, MESSAGE, and TOPIC[1]. In terms of processing a sentence to extract its meaning, it corresponds to the idea that the analysis can be decomposed into two stages. A few NLP programs perform the input translation in a single stage called as "conceptual" or "semantic parsing, but more often the task is split into two phases "Syntactic Analysis" (or "Parsing") and "Semantic Interpretation".

The first stage uses grammatical (syntactic) information to perform some structural preprocessing on the input, to simplify the task of the rules which compute a symbolic representation of the meaning. This preprocessing stage is usually known as parsing, and could be roughly defined as "grouping and labeling the parts of a sentence in a way that displays their relationships to each other in a useful way". The structure built by the parser should be a suitable input to the semantic interpretive rules which will compute the "meaning" of the sentence.

Natural Language Understanding Question Answering

Input

ராம் நேற்று ஒரு பந்தைக் கொண்டு ராஜாவை அடித்தான்.

Output

[Agent ராம்] [am-tmp நேற்று] [instrument ஒரு பந்தைக் கொண்டு] [patient ராஜாவை] அடித்தான்.

- யார் ராஜாவை பந்தைக் கொண்டு அடித்தது?
- யாரை ராம் பந்தைக் கொண்டு அடித்தான்?
- எதைக் கொண்டு ராம் ராஜாவை அடித்தான்?
- எப்பொழுது ராம் ராஜாவை பந்தைக் கொண்டு அடித்தான்?

Syntactic Analysis

In linguistics, syntax are the rules of a language that shows how the words of that language are to be arranged to make a sentence of that language. Hence it is the study of words that are combined together to form a sentences. The purpose of syntactic analysis is to determine the structure of the input text. This structure consists of a hierarchy of *phrases*, the smallest of which are the *basic symbols* and the largest of which is the *sentence*. It can be described by a tree with one node for each phrase. Basic symbols are represented by leaf nodes and other phrases by interior nodes. The root of the tree represents the sentence. The goal of syntactic analysis is to parse the lexemes into a parse tree. According to syntax the component parts of a sentence are called constituents which are the natural groupings of a sentence .Each of them has:

- a Grammatical Category (Whether it is a noun phrase, a verb phrase, etc)
- a Grammatical Function (whether it is the subject, or object, or predicate, etc)

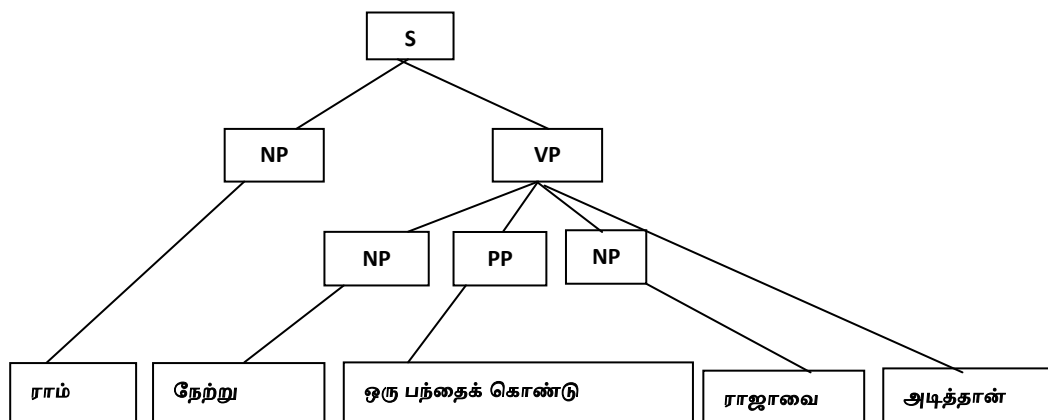


Figure 2: Syntactic Analysis

Semantic Analysis

This is the level at which most people think meaning is determined, however, as one can see in the above defining of the levels, it is all the levels that contribute to meaning. Semantic processing determines the possible meanings of a sentence by focusing on the interactions among word-level meanings in the sentence. This level of processing can include the semantic disambiguation of words with multiple senses; in an analogous way to how syntactic disambiguation of words that can functions as multiple parts-of-speech is accomplished at the syntactic level. Semantic disambiguation permits one and only one sense of polysemous words to be selected and included in the semantic representation of the sentence. A wide range of methods can be implemented to accomplish the disambiguation, some which require information as to the frequency with which each sense occurs in a particular corpus of interest, or in general usage, some which require consideration of the local context, and others which utilize pragmatic knowledge of the domain of the document.

- Agent – Who did? - ராம்
- Patient –To whom? – ராஜா
- Instrument – With what? – பந்து
- Temporal Adjunct – When? – நேற்று

4.2. Word Sense Disambiguation

The purpose is to mark each word with a syntactic tag. It involves the following methodology such as assigning each word all the possible syntactic tags and disambiguate words as much as possible using linguistic information (hand-coded rules) and risks are avoided by leaving the words ambiguous than guess wrong. Treat sense disambiguation like POS tagging, just with “semantic tags”

The problems differ as follows:

- POS tags depend on specific structural cues (mostly neighboring tags)
- Senses depend on semantic context – less structured, longer distance dependency

4.3. Handling Complex Sentences

Complex sentences with noun clause, adjective clause and adverb clause are considered. Grouping the clauses is the main process of the parser. The clauses are generally indicated by special cue suffixes / cue phrases. Grouping is done by the position of cues. These complex sentences are parsed by the following steps,

- Conversion of complex sentences into simple sentence by grouping the three clauses with their corresponding words and forming Noun Clause and Verb Clause.
- Parsing the simple sentence can be done.

5. APPLICATION OF SRL – CONCEPT BASED CLASSIFICATION

Usually the category profile preparation and test document ruling feature extraction process were done based on statistical analysis of one or more terms. i.e., literal keyword investigation is done to extract features. The profile or feature set of such systems would be a collection of keywords. The problem here is that, when more than two terms have the same frequency in a document, one may contribute more meaning than the other. The procedures of identifying such terms which capture the meaning of the sentence are known as the concept based mining approach. Here we come-up with a system that prepares profiles and feature set as a collection of concepts instead of keywords. This concept filled profile will explain the intension of the text in a much better manner. The frequency of the term of concept is known as concept factor. The concept factor is identified at different levels of the text including sentence, document and corpus.

The concept of the sentence is analysed with the aid of the Semantic Role Labeller. It is the process of labelling the sentences according to its Part-Of-Speech (POS) . In shallow semantic parsing, semantic tags are assigned to the arguments, or case roles, associated with each predicate in the sentence. Initially each labelled part of the sentence is considered as a concept. The term frequency of each concept is analysed by the labels assigned at the sentence level but not by the literal count of keyword occurrences. Then the concept investigation is extended to the document level. Document level concepts are identified by surveying the relationship between the concepts of sentences. Thus here the content of a text is described by a set of concepts. [11]

Each sentence of the document is labeled by a semantic role labeler that determines the terms which contribute to the theme of that sentence. Each term that has a semantic role in the sentence is called a concept. Shallow semantic parsing is a process to annotate naturally occurring text with semantic roles that can facilitate the discovery of patterns of information in large text collections producing such a mark-up. Who is doing What to Whom (WWW) classifies the contribution of each term in a sentence to the main topic of that sentence [12].

6. CONCLUSION

To date, SRL systems have been shown to perform reasonably well in some controlled experiments, with standard test collections for Tamil. Still, a number of important challenges exist for future research on SRL. It remains unclear what is the appropriate level of syntax needed to support robust analysis of semantic roles, and to what degree improved performance in SRL is constrained by the state-of-the-art in tagging and parsing. Beyond syntax, the relation of semantic roles to other semantic knowledge (such as WordNet, named entities, or even a catalogue of frames) has scarcely been addressed in the design of current SRL models. A deeper understanding of these questions could help in developing methods that yield improved generalization, and that are less dependent on large quantities of role-annotated training data.

7. FUTURE WORK

Indeed, the requirement of most SRL approaches for such training data, which is both difficult and highly expensive to produce, is the major obstacle to the widespread application of SRL across different genres and different languages. Given the degradation of performance when a supervised system is faced with unseen events or a testing corpus different from training, this is a major impediment to increasing the application of SRL even within Tamil, a language for which two major annotated corpora are available. It is critical for the future of SRL that research broadens to include wider investigation of unsupervised and minimally supervised learning methods. In addition to these open research problems, there are also methodological issues that need to be addressed regarding how research is conducted and evaluated. Shared task frameworks have been crucial in SRL development by supporting explicit comparisons of approaches, but such benchmark testing can also overly focus research efforts on small improvements in particular evaluation measures. Improving the entire SRL approach in a significant way may require more open-ended investigation and more qualitative analysis.

REFERENCES

1. Automatic Labeling of Semantic Roles (Daniel Gildea & Daniel Jurafsky-2002)
2. A Generative Model for Semantic Role Labeling (Cynthia A. Thompson¹, Roger Levy², and Christopher D. Manning-2003)
3. FrameNet-based Semantic Parsing using Maximum Entropy Models (Namhee Kwon, Michael Fleischman & Eduard Hovy-2004)
4. Maximum Entropy Markov Models for Semantic Role Labeling (Phil Blunsom-2004)
5. Semantic Parsing Based on FrameNet (Cosmin Adrian Bejan, Alessandro Moschitti, Paul Morărescu, Gabriel Nicolae & Sanda Harabagiu-2004)
6. Semantic Role Labeling Using Different Syntactic Views (Sameer Pradhan, Wayne Ward, Kadri Hacioglu, James H. Martin & Daniel Jurafsky-2005)
7. Shared Task: Semantic Role Labeling (Xavier Carreras & Lluís M`arquez-2005)
8. Towards Free-text Semantic Parsing: A Unified Framework Based on FrameNet, VerbNet and PropBank (Ana-Maria Giuglea & Alessandro Moschitti-2006)
9. Semantic Role Labeling Using a Grammar-Driven Convolution Tree Kernel (Min Zhang, Wanxiang Che, GuoDong Zhou, Aiti Aw, Chew Lim Tan, Ting Liu, & Sheng Li-2008)
10. Semantic Role based Tamil Sentence Generator (S. Lakshmana Pandian, T.V.Geetha-2009)
11. A Self Appreciating Approach of Text Classifier Based on Concept Mining (Proc. IEEE Int'l Conf-K.Arul Deepa, C.Deisy -2012)
12. Semantic Role Parsing: Adding Semantic Structure to Unstructured Text (S. Pradhan, K. Hacioglu, W. Ward, J.H. Martin, and D. Jurafsky – 2003)

Application of Natural Language Processing Techniques to Augmentative Communication Systems

Aravind Kannan (vu2brains@gmail.com)

Consultant Speech Language Pathologist, Department of Speech, Hearing & Communication, National Institute for Empowerment of Persons with Multiple Disabilities (NIEPMD), Chennai

Dr. V. Thiruvalluvan (vtvalluvan@yahoo.co.in)

Professor in Linguistics, Center for Advanced Study in Linguistics, Faculty of Indian Languages, Annamalai University

Natural language processing is an efficient example to denote the interaction between humans and computers. The computer fetches the input; and transforms the meaningful information into natural language; and finally, produces natural language as the output. NLP falls under the category of computational linguistics. There are several forms of techniques and applications of the natural language processing. The goal of this paper is to identify the potential applications of natural language processing techniques that can be incorporated in the augmentative communication systems. This research can extend the communication rate of the children / individuals who are with physical disabilities.

Augmentative and alternate communication (AAC) system refers to an integration of components that include symbols, strategies and aids that enhance the communication abilities. This is commonly used to semantically process the contents and communicate in an effective manner with the recipients. In an AAC system, the main purpose of communication is to convey information, monitor and control the social interactions or control the environment. In the process of employing the natural language processing techniques in the AAC system, the semantic interpretations tend to conflict. The input to the system is a set of content words – “*Car See Pinky Now*”. The regular system implementation does parse the content semantically to classify the grammar of the words and relate the term with language specific instantiations.

The final sentence processing yields the meaningful content as “*The Car Is Seen By Pinky Now*”. The paper here presents the general processing phases that take place within an AAC system and the various techniques that could be employed for better communication and understanding of the contents of the sources. These techniques might help in attaining the granularity of the representation of the knowledge and delivering the right content as a result of the language processing within the system.

Natural Language Processing

Natural language processing has evolved with multiple techniques that would simplify the source text and present the right components as the result. The techniques of NLP enable the users to accomplish the purpose of the language analysis. Naturally occurring texts might refer to any form of language, genre and mode; and these might be in oral or written forms. The texts generally evolve with no intended purpose (Liddy, 2003) . According to the human-like language processing, NLP might even be regarded as a category of artificial intelligence.

The model of NLP involves the seven major components namely phonology, morphology, lexical, syntactic, semantic, discourse and pragmatic. These components have modified the basic approaches that were followed to implement natural language processing in the experiments and real life scenarios. The approaches are classified into four broad categories namely connectionist, symbolic, statistical and hybrid. Each of these approaches follows similar processes. However, the purpose and activities involved in every process differed to a greater extent (See Figure 1) .

Having known the richness of natural language processing in transforming the undefined and meaningless information to clear and meaningful information, the applications have grown enormous. The commonly used applications where the NLP is significantly used include information retrieval, information extraction, dialogue systems, summarization, question-answering and machine translation. The machine translation application is a conventional one and it simplifies even higher levels of analysis. However, the other applications are embedded in the commonly used systems for various lexical analyses and uses under different environments.

Stages of NLP

NLP is basically a three stage process that involves parsing, translating and generating the input contents. The output of every stage is fed as the input to the next stage of the system. The natural language processing works based on semantic structures and generations. Hence, the first step is to find out the possible interpretations for the input through the semantic parser (Liddy, 2003) . The problem here is that a word can fall under several semantic classifications. With a sub frame designed for every word as in input, the semantic type is designed and is substituted to the word to develop the interpretations. For instance, 'take' can be replaced by 'get' and 'from'.

The translator serves as the middleware to the system. The main use of this stage is to create placeholders for the words in the translator's dictionary and covert the words with probable structural translations. Syntactic realization takes place at the second stage of the natural language processing. The generator phase is primarily responsible for the sentence generation based on the knowledge of specific language. The result becomes the final component of the entire system. Lexical classification and generation take place at this final stage. The output of the NLP system directs the patient with the right options and translations for their doubts (Liddy, 2003) .

Augmentative and Alternate Communication

Communication generally requires a mutual understanding between the speaker (s) and receiver (s) . The dynamic process becomes complex when the communication disorders arise. For instance, in autism which is a cognitive language disorder has presented numerous hindrances to the type of communication being held between people. The augmentative and alternative communication systems (AACs) mainly aim at eliminating the communication disorders of people by introducing new substitutes for communication (Figure 2) . Some of the speech pathologists make use of paper based PECS to effectively improve communication with people having disorders. On the other hand, technical solution like plasma picture communication table is also used to interact in a creative and fun-filled method (Vanderheiden, 2002).

The paper based communication system is effective even when it is embedded in the mobile devices so that children who are experiencing severe autism might use it as a tool to convey information. There are also AAC systems to address complex communication needs. Some people might experience physical and sensory issues that might restrict them from communicating with people. For these groups of people, communication is facilitated with advanced techniques along with training to explain the movement patterns and enhance the participation of the patient in the society.

In today's trend, speech pathologists and linguistic systems support AAC technologies and concepts mainly because the service delivery is effective and the regular human interventions are overcome in that case (Leroy & De, 2010) . The application of AAC systems is widened to people of any sort of communication needs and to those who are in need of additional aids. Knowledge transmission and reception processes become easier and are carried with relevant evidences in case of AAC systems.

NLP in AAC

The individual benefits of NLP and AAC systems are enormous. When these concepts are grouped together, it delivers a better space for communication with no evidence of disorders. For people suffering from autism, the concept of NLP in AAC would be fruitful. NLP helps in the development and delivery of semantic structure whereas AAC helps in presenting new substitutes to the ease communication. In this scenario, it shows that when NLP is used in AAC, it is feasible to ease communication by clearing the semantic ambiguities and similarities; and providing a clear text as a result to the recipient (McCoy et al, 1990) .

Language disorders might be experienced either in the reception type or expression methods. As mentioned earlier, a NLP in AAC system might simplify the communication by breaking the words, matching the words with its context and grammar and reframing the sentence in a meaningful manner. For instance, the sentence "*Work I Done*" is presented as follows:

Step 1: The words in the sentence are broken separately. The result is: "*Work, I, Done*".

Step 2: The semantic parser parses the words and matches the words with the appropriate grammar. Subsequently, the context is also analyzed.

Step 3: The words are also compared with relevant aids such as pictures to identify the relations and patterns of them.

Step 4: The semantic strategy gives the result as *"Work Is Done by Me"*.

Step 5: The results of the strategy are considered and the individual's cognitive level is evaluated.

Step 6: If the cognitive level is very low, the picture-based communication is made. If it is medium, then paper based or even speech-based communication can be made to convey the end result.

In this process, the steps 1, 2 and 4 are carried out by NLP and the purpose of the process is accomplished when NLP is involved in AAC to deliver the result based on the intended recipient's cognitive level.

NLP in AAC is often considered as the best interface to communicate with the patients (McCoy et al, 1990) . As the mere aim of the system is to relieve the person with disorders and enhance the efficiency, the machine translation is used in case of NLP; and the AAC system covers the phonetic and lexical ambiguities faced by the patient. In most of the cases, the patients feel the linguistic ambiguity either due to the context or phonetics used by the speaker in the oral communication. When written communication is taken into consideration, grammar, lexical levels and type of narration might confuse the readers.

People with speech and language disorders expect a high level of clarity in the information presented and the AAC system does assure clarity through different means. When the patient accesses the AAC system, he / she tends to receive multiple formats of a sentence so that there is a consistency in the information delivered and relevance of the words used. The communication aids are generally chosen depending on the patient's abilities. The application of NLP in AAC is expected to develop a new world of communication in terms of clarity and ease in understanding and capabilities.

APPLICATIONS OF TECHNIQUES OF NLP

NLP in Just-In-Time Vocabulary Support

Lack of knowledge of Vocabulary is a serious problem for most of the patients. The problem with the interpretation is that a sentence might, at times, match with multiple meanings and there develops a serious issue in choosing the right one among them. Just-in-time vocabulary support is primarily designed with an aim of assisting the patients with context-based vocabulary support so that the complexity in understanding the language is minimized. The vocabulary coverage is vast as it identifies and presents all possible meanings for a particular word. When a sentence is analyzed with the help of this vocabulary support, it adds clarity to the resultant meaning (Epp et al, 2012) .

When NLP is employed in AAC system, it breaks the sentence into its set of words and compares each one of them with the vocabulary support system. The result that indicates the highest match is thus presented to the patient. When the patient enters three different keywords namely 'paper', 'paper please' or 'just one paper', they all denote the same context that the person needs paper. So, as a result the device gives the output to the patient. This is certainly a new experience and an advantageous one to the patients who are experiencing problems with vocabulary.

NLP in Language Tasks

The commonly performed language tasks are vocabulary development, daily activity on studying word lists and regular activities on acquisition of knowledge with the help of tools. These seem to take place in classrooms as well as at home. The connectionist approach of NLP is used here to monitor and sequence the language tasks (Conover et al, 2012) .

Generalized models of the words are developed and connectionist representations are employed. These would eliminate the constraints in the manipulation of sentences. Moreover, when this approach is used in an AAC system, it certainly improves the connectedness of the children with the society and the local interacts are greatly enhanced. These language tasks might develop ambiguities or issues related to inference. However, these are easily resolved with the connectionist approach of NLP as it employs advanced mechanisms to eliminate word – context ambiguities, inference issues and language generation issues (Conover et al, 2012) .

Whenever there is a query raised by the student / patient, the AAC system delivers the task result with relevant vocabulary terms as well as the semantic representations. These representations could be in the form of images that indicate the connectivity between the terms. By this deployment, the understanding skill is improved and the students tend to learn new terms at a rapid pace more than as expected. The speech language pathologist might witness the progress of the patient in terms of communication and minimization of the language disorders.

NLP Translator for Speech Generating Device

Speech generating devices (SGD) are the commonly known voice output communication aids that were particularly used for people with linguistic or phonetic disorders. These devices contain the contents and vocabulary that are varied depending on the user's requirements and the contexts. The patient operated selector mechanism works on the symbols that are equipped with photoelectric cell to enable the patients to read the characters at an ease. The system is connected with a typewriter so that the patients can type the contents with proper aid available to recognize and distinguish the letters (Vanderheiden, 2002).

The speech generating devices are said to be easier modes to imbibe knowledge on the language. NLP has a potential application in these speech generating devices. As already mentioned, the context of the input might influence the output obtained from the device. In case of fixed display speech generating device, the symbols and items are stationary whereas dynamic display device tends to generate visual symbols in a dynamic way where the symbols can even be changed by the user. In such a situation, the NLP is to be employed in the selection set and vocabulary of the speech generating device. The selection set is actually a collection of messages and codes that a person can use.

When the content is chosen, NLP's translator would produce a better result. It can automatically translate the inputs as per the language definition and produce the right language outputs to the user. The automatic maintenance is only an advantage to the user as the contents are translated and stored in an up-to-date manner. Whenever the user accesses the system to learn the language, the translated outputs are also displayed as options. This process is certainly expected to improve the vocabulary of the patient. The introduction of NLP to this particular type of AAC can increase the rate of action and performance.

Symbolic Approach for User-Centered Design System

The main purpose of the user-centered design system is to ensure universal usability for the people with disabilities in accessing the system. The system is developed with the notion of "design for all". This augmentative and alternate communication system involves flexible interfaces, common dimensions appropriate for educational purposes, good presentations and high dynamism (Newell & Gregor, 2000) . The presentation of information is the most important of all. People with disabilities switch to these types of systems as aids and so, there is a need for better functionality. The user-centered design system involves both mainstream as well as supportive systems that covered the entire aspects of the learning process.

The natural language processing to this particular type of augmentative and alternate communication is certainly supposed to increase the overall efficiency. In the interface and user-centered designs, there are even options for the patients to enter the inputs as it is basically a dynamic system. In such a case, the introduction of natural language processing can eliminate the difficulty of the users in finding the right choice (Newell & Gregor, 2000) .

Firstly, the semantic parser can be used to process the input from the user panel. As the sensitivity is generally high in these systems, the sensitivity can be controlled with the semantic parser operation. Subsequently, the results of the middleware (translator) can be useful for the patients as they tend to get several choices. The final step of generation will produce the exact result that is highly suitable for the system. In these phases, the symbolic approach of NLP is being used as the user-centered system involves several symbols and dialects to improve the overall interactivity.

The symbolic approach for this system will benefit in a better way as the rules are human developed. The representation schemes are primarily set with the motive of applying in user-centric system. The implementation of this NLP approach to AAC system will work based on the semantic networks. It works as follows. Whenever the user raises a query, the decision trees and explanation-based learning are invoked. So, the query undergoes the three phases of NLP and the

output will be the result of explanation-based learning or decision tree. This might improve the reasoning skill as well as the learning process is incorporated with efficient strategies that can assure fruitful results.

A general AAC system for user-centered requirement is not as efficient mainly because the language learning process does not include all the essential strategies (phonetic, lexical and linguistic) . When NLP is included to this system, then the information extraction application ensures greater efficiency. Most of the speech pathologists prefer NLP to this system due to the quality outputs.

Parser for Emotion Detection System

There are several systems meant to detect the emotions based on the linguistic content present. The emotions are regarded as compound and they were analyzed further. The emotional language understanding is targeted towards the language learners who are more dependent on the expression through emotions. The emologus system works by fixing a specific emotion measure to every word in a sentence (Tallec et al, 2011) . This system presents the overall architecture of this system. Depending on the total value, the patient's emotion is measured. This process of sentiment analysis is functional in the augmentative and alternate communication methodology.

The parser of NLP system is going to even more benefit the entire emotional analysis process. The application of NLP to this system is basically to control and enhance the evaluation of semantic dependencies between the inputs. The application actually works as follows. The patient gives the input to the system. The words are individually broken and the emotion content percentage is applied to every word. Subsequently, the semantic linkages and dependencies between the words are evaluated through the parser of NLP (Tallec et al, 2011) . The parser would classify the semantic functions of the words and then, the frames help in a better and accurate evaluation of the emotions. The emotions are evaluated with respect to the semantic structure they belong to.

This NLP concept is going to better the predictions of the emotions. Semantic parser of NLP helps the patients to see the preferences and the right structures that the sentences belong to. This is a way to improve the linguistic knowledge and an easier way to determine the emotion with just parsing of the individual words.

Pictographic System

Pictographic writing is said to be a new and potential application of natural language processing. The cognitive design that the pictograms support is highly interactive and might quickly let the people to write the text. However, the writing process becomes slow when patients are not given assistance. Therefore, the NLP would suggest sentence fragments. There are basically two major applications of NLP in this domain and these are based on the morpho-syntactic structures. The advantage of introducing this application of NLP to AAC system is that the disabled persons can enhance their domain knowledge in addition to alphabetic writing (Abraham, 2011) .

For a handicapped person who wishes to write but fails to use alphabetic writing, NLP presents the pictograms where the person can simply choose and form the word. This is rather interesting; and the disabled person can understand the concept. These would improve the overall interactivity; and the linguistic and cognitive level of the person gets improved. This application is not only proposed for the people with disabilities but also to the people who are on the initial stages of learning and are finding difficulties in following it. The application area is definitely broad although disabled persons are to be benefited at the maximum. When pictograms are already built in AAC, the NLP system can associate the pictograms to relevant semantic structures.

Assessment of NLP to AAC from Success Perspective

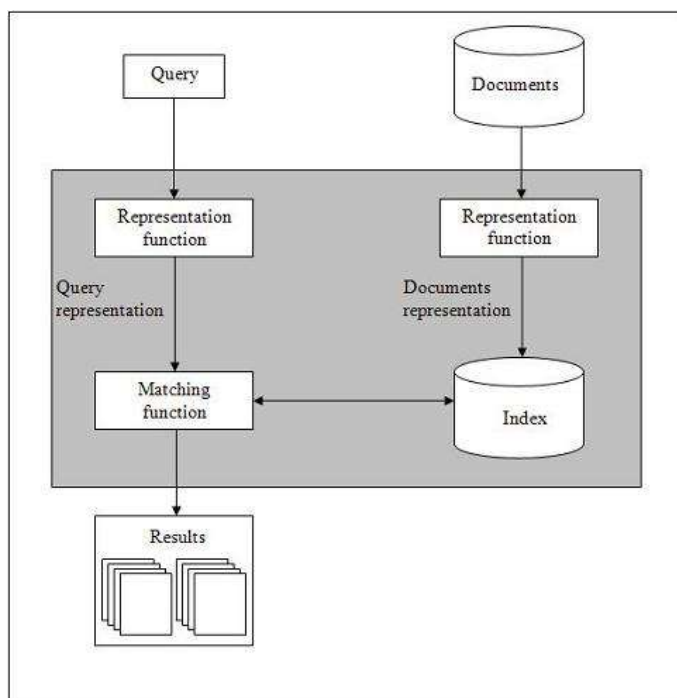
AAC is a widespread system that is commonly used among the patients with autism, language understanding issues and apraxia (Cumley, 2005) . Children with serious disorders associated with phonology fail to have proper communication interaction. So, the AAC system tends not only to serve as an aid for communication but also to improve the language intelligibility. As language influences the overall development and cognitive understanding of a child, the AAC systems are developed with this perspective.

There are certain issues faced by the people in the case of AAC. There are no pre-stored samples to the patients in the AAC systems whereas the application of NLP to AAC can enable better interpretation and automatic content maintenance, which resolves the above problem. The other point is that there is no option for dynamic choice of the vocabulary in an AAC system. The language representation of NLP appears to be easier, no matter whether it is a picture-based system or alphabet-based system or voice based system (Hill, 2011) .

There are two major fields where NLP to AAC system can prove to be high; and they include interface design and the word prediction. The future applications are more relied on NLP due to the main reason that the word prediction adds better sense to the outputs of the AAC system. This implies that the success rate is high and there is a huge potential for these applications in the near future. The evaluations that have been conducted so far on these applications have clearly denoted that the next generation of technology is to rely on these applications and there are greater chances for the affected people / patients to overcome the disorders at a rapid pace (Hill, 2011) . So, it is highly recommended to utilize these applications and also spread the idea to those patients with language / communication disorders. The authors of this paper are positive that the speech language pathologist can obtain a better result.

CONCLUSION

NLP and AAC are two different techniques and both of them go together on the semantic and language terms. AAC system serves as an aid to the patients with autism and any other form of speech and language disorders whereas NLP techniques when applied in AAC serves as an information retriever, dialogue systems that primarily ease the communication and provide instant solutions to the communication disabled. The step by step process of NLP ensures accurate outputs at the end and the output of every process is individually used for several applications. For instance, the parser might be highly appropriate for the emotion detection system. The picture-based system might involve a translator or so. However, the whole process of AAC is to support and clear the doubts of patients with relevant interface support. Some of the aspects, where AAC fails, are all overcome by the introduction of NLP to AAC. The individual applications that were stated above indicate the effectiveness of NLP techniques and the appropriate influence on the outcomes of AAC systems. As most of the patients prefer highly interactive systems to develop their language-based knowledge, NLP fills them with the maximum features. NLP provides explanations, options and trees that can enable the persons with communication disorder to improve their vocabulary and obtain a clear knowledge on the output for their respective inputs. So, the applications of NLP to AAC systems are highly efficient and mandatory for the elimination of communication disorders among people.



I CAN HEAR PERFECTLY	PLEASE REPEAT AS I TALK (THIS IS HOW I TALK BY SPELLING OUT THE WORDS)	WOULD YOU PLEASE CALL
A AM HE	AM ARE ASK BE BEEN BRING CAN	ABOUT ALL
HER I IT ME	COME COULD DID DO DOES DON'T	AND ALWAYS
MY HIM SHE	DRINK GET GIVE GO HAD HAS HAVE	ALMOST AS
THAT THE THERE	IS KEEP KNOW LET LIKE MAKE MAY	AT BECAUSE
THEY THIS WHOSE	PUT SAY SAID SEE SEEN SEND SHOULD	BUT FOR FROM
WHAT WHEN WHERE	TAKE TELL THINK THOUGHT WANT	HOW IF IN
WHICH WHO WHY	WAS WERE WILL WITH WON'T WOULD YET	OF ON OF
YOU WE YOUR	ER EST ING LY N'T S TION	TO UP WITH
A B C D E F G	AFTER AGAIN	
H I J K L M N	ANY EVERY	
O P Q R S T U	HERE JUST	
V W X Y Z 1	MORE ONLY	
2 3 4 5 6 7	SOME SOON	
8 9 10 11 12 13	THERE VERY	
SUN MON TUES	PLEASE THANK YOU	GOING OUT
WED THUR FRI	MR. MRS. MISS	START OVER
SAT BATHROOM	MOTHER DAD DOCTOR	END OF WORD

REFERENCES

1. Abraham, M. (2011) . From a Cognitive Model Towards an Assistive and Augmentative Written Language. In C. R. Murray, & P. M. McCarthy (Eds.), *Proceedings of the Twenty-Fourth International Florida Artificial Intelligence Research Society Conference* (pp. 132 - 137) . Palm Beach: AAAI Press.
2. Conover, A., Packer, J., & Mills, D. (2012) . Implementation in the classroom. Retrieved May 5, 2012, from AAC Language Lab: <http://www.aaclanguagelab.com/materials/implementation.pdf>
3. Cumley, G. (2005, 2 24) . Other Disabilities and Apraxia. Retrieved 2 10, 2012, from The Childhood Apraxia of Speech Association of North America (CASANA): <http://www.apraxia-kids.org/site/apps/nlnet/content3.aspx?c=chKMI0PIIsE&b=788457&ct=464167>
4. Epp, C. D., Djordjevic, J., Wu, S., Moffatt, K., & Baecker, R. M. (2012) . Towards providing just-in-time vocabulary support for assistive and augmentative communication. *IUI '12 Proceedings of the 2012 ACM international conference on Intelligent User Interfaces* (pp. 33 - 36) . New York: Association for Computing Machinery.
5. Hill, K. (2011, May 23) . Achieving Success in AAC: Assessment and Intervention. Retrieved 2 10, 2012, from AAC Institute: <http://www.aac institute.org/Resources/Press/AssessmentModelpaper/aacsucc.html>
6. Leroy, G., & De, L. G. (2010) . Augmentative and Alternative Communication Technologies. In M. M. Cruz-Cunha, A. J. Tavares, & R. Simoes, *Handbook of Research on Developments in E-Health and Telemedicine: Technological and Social Perspectives* (1 ed., pp. 1070-1087) . Medical Information Science Reference.
7. Liddy, E. D. (2003) . Natural Language Processing. In M. Drake (Ed.), *Encyclopedia of Library and Information Science* (2 ed., pp. 2126 - 2136) . New York: Marcel Decker, Inc.
8. McCoy, K., Demasco, P., Jones, M., Pennington, C., & Rowe, C. (1990) . Applying Natural Language Processing Techniques to Augmentative Communication Systems. In H. Karlgren (Ed.), *13th International Conference on Computational Linguistics*. 1, pp. 413 - 415. Helsinki: International Committee on Computational Linguistics.
9. Newell, A. F., & Gregor, P. (2000) . "User sensitive inclusive design" — in search of a new paradigm. *CUU '00 Proceedings on the 2000 conference on Universal Usability* (pp. 39 - 44) . Arlington: Association for Computing Machinery.
10. Tallec, M. L., Antoine, J.-Y., Villaneau, J., & Duhaut, D. (2011) . Affective Interaction with a Companion Robot for Hospitalized Children: a Linguistically based Model for Emotion Detection. *5th Language and Technology Conference* (pp. 445 - 450) . Poznan: Springer series Lecture Notes in Artificial Intelligence, SpringerLink.
11. Vanderheiden, G. C. (2002, November/December) . A journey through early augmentative communication and computer access. *Journal of Rehabilitation Research and Development*, 39 (6), 39 - 53.

Topic F
Education
கல்வி

Kalvi: An Adaptive Tamil mLearning System

Keshava Rangarajan, Chief Architect, Landmark (Halliburton) Corporation

Jayaradha Natarajan, Software consultant at TIBCO

Arivoli Tirouvingadame, Principal Member of Technical Staff, Oracle America, Inc.

Abstract

Learning Management Systems (LMS) in Tamil are at an early stage today. They typically model users as a homogenous group with content representing a static structural organization of the course material determined by the educator. In most cases, LMS users who are the primary consumers, represent heterogeneous groups with diverse learning characteristics, needs and goals. This lack of sophistication in LMS is particularly noticeable when they are used to deliver course content that are typically taught in languages other than Tamil. The *adaptation* of e/m-education systems to an individual or to a group based on their characteristics, expectations, knowledge, background and preferences of the students is understood as critical but is under served today. Emphasis is moving slowly towards learner-oriented platforms and putting the learner's expectations, motivations, habits, learning styles, needs, etc. as the focus of interest. This paper delves into an LMS for Tamil language education. It also explores the mobility aspect of the LMS systems, which makes the e-learning systems available via mobile devices like iPad, iPhone, Android based devices, etc., thus making the LMS much more accessible and in line with current computing trends.

Introduction

This paper proposes *Kalvi*, an adaptive Tamil mLearning System, which is based on the Sakai project. Sakai is a community of academic institutions, commercial organizations and individuals who work together to develop a common Collaboration and Learning Environment (CLE) . The Sakai CLE is used for teaching, research and collaboration. It is a free, community source, educational software platform distributed under the Educational Community License. Sakai is a Java-based, service-oriented application suite that is designed to be scalable, reliable, interoperable and extensible.

Anatomy of an academic course

Before we begin, let's first take a closer look at the structure of a typical educational course. A course typically consists of an ordered sequence of learning modules. Each of these learning modules consists of a sequence of topics/lessons that introduce and illustrate concepts. The lessons could include quizzes at various points that test the learner's grasp of concepts and reinforces previously learned concepts. Such a course can be viewed as a directed, typically acyclic *graph* where modules and concepts (and potentially quizzes) form the *nodes*. These nodes are connected by *links* that represent the learner's transitions from topic to topic. Educators and students progress via links across the vertices (modules and concepts) from start to end during the life cycle of a course.

Non-adaptive and adaptive courses

In a *non-adaptive* course, the connecting links/arcs are static, pre-determined globally and follow a pre-determined path. In an *adaptive* course, firstly, the links are initially configured based on the information (descriptive attributes) available about the learner. Additionally, there are many possible link flow paths. These paths are conditional i.e. based on an ongoing evaluation/scoring of the learner's progress through the topics over a given period. Additional nodes/topics may be brought in dynamically based on a dynamic evaluation of the learner's level of knowledge as she/he progresses through the course. The topics introduced are driven by analytical insight gained from community use.

Existing problems

To summarize, existing problems in Tamil LMSs are as follows:

1. There are very few modern Learning Management Systems for education via Tamil language especially ones that deliver content typically taught in other languages (like English)

2. Even if they do exist, these LMS systems deliver content in a static fashion; they do not take into account the user's preferences, level of skill, learning goals and other factors explicitly into account and use this as the basis for learning content delivery and learn from user activity

Data mining, Machine learning and Analytics in LMS

Data mining, Machine learning and Analytics forms the core of LMS systems.

- **Data mining and Machine learning:** Learning management systems and Learning Content management systems deal with volumes of data. Users consuming the course material leave a trail of data while performing their activities. These data can and needs to be mined to extract insight into learning patterns, learner groupings, Topic classifications (eg: easy, difficult, etc.) . Machine learning techniques like Dynamic Regression, Support Vector Machines (SVM), Neural Net engines, etc. can be employed to mine the data to extract insight
- **Analytics:** Analytics plays a big role in LMSs. The broad promise of analytics is that new insights can be gained from in-depth analysis of the data trails left by individuals in their interactions with others, with information, with technology, and with organizations. At a high level, the following are the types of analytics of interest here:
 - **Learning Analytics:** Wikipedia defines Learning Analytics as the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs. Learning analytics are largely concerned with improving learner success.
 - **Academic Analytics:** Wikipedia describes Academic analytics as the term for Business Intelligence used in an academic setting. Academic analytics is the improvement of organizational processes, workflows, resource allocation, and institutional measurement through the use of learner, academic, and institutional data. Academic analytics, akin to business analytics, are concerned with improving organizational effectiveness.

Adaptive e-learning systems

Now let's take a look at adaptive e-learning systems. An e-learning system should be designed to match students' needs and desires as closely as possible, and adapt during course progression. It is considered to be adaptive if it is capable of:

1. Modeling users, monitoring the activities of its users;
2. Interpreting these on the basis of domain-specific models;
3. Inferring user requirements and preferences out of the interpreted activities, appropriately representing these in associated models; and
4. Acting upon the available knowledge on its users and the subject matter at hand, to dynamically facilitate the learning process.

Thus, adaptive e-learning system can be described as a personalized system, which is able to:

1. Perform content discovery and assembly,
2. Provide an adaptive course delivery, an adaptive interaction, and adaptive collaboration support

Architecture of KALVI system

The Kalvi system proposed in the paper is built on Sakai LMS platform. At a high level, the Kalvi system has two parts to it:

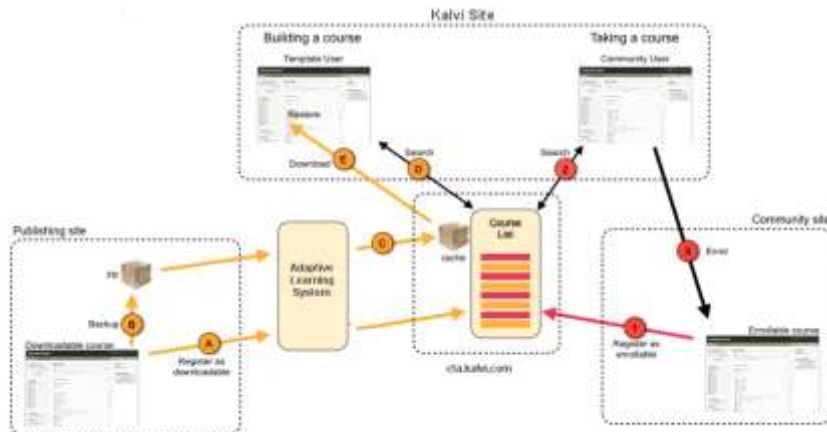
1. Kalvi server
2. Kalvi client

Kalvi server

The Kalvi server is the backend module. It supports all the full-fledged features of a typical LMS. There is a central repository of the offered Course list. Educators can build and publish new courses via the publishing site. The students

can search the course list and select their courses of interest and take them via the community site. The server has the Adaptive Learning system as well, which is responsible for making the LMS adaptive. All data is persisted in a central backend database.

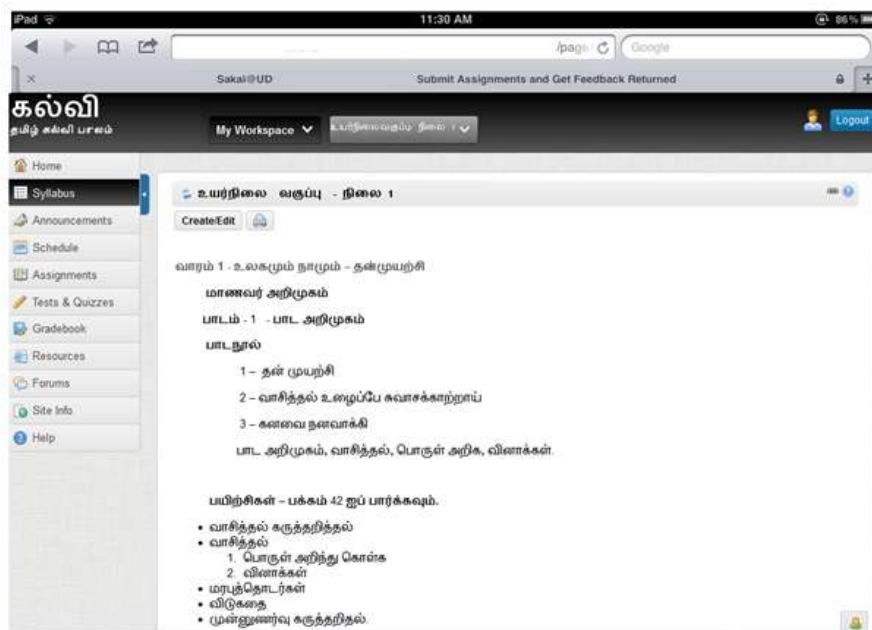
Here is the architecture diagram of Kalvi:



Kalvi client

Kalvi supports both web based and mobile clients. Students can take a course via mobile devices like iPad, iPhone, Android based devices, etc. The mobile client downloads the course from the server and saves it locally. Along with the course, the client piece of the Adaptive learning system pertinent to the course is also downloaded to the mobile device. The student then takes the course in the mobile device. While taking a course from the mobile device, it is not required to stay connected to the server. That is, courses can be taken from the mobile devices both in *online* and *offline* modes. All the data obtained by monitoring and recording student activities during the course life cycle are persisted in a local database in the mobile device. When they are connected, the Kalvi server and client can sync up periodically.

Here is proposed sample screen shot of a typical Tamil course taken from iPad:



Here is proposed sample screen shot of a typical Tamil course taken from iPhone:



Concluding thoughts and future work

The key barrier here is not the veracity of the concept or the implementation of the LMS but it is their incorporation into the current educational processes and culture which is a rather static. This is more so in Tamil LMSs and Tamil educational systems today. This requires evangelization as well as a high level of engagement from all participants in the education process to effect a change. But this is the clear trend forward. Irrespective of the subjects and courses offered, the demography served, and the medium of languages delivered to, the learning methodologies and techniques are the same as they broadly rely on data mining, machine learning and analytics to deliver adaptive learner-centric content in mobile form factors for the current and next generation of learners. The promising aspect of this paper is that the proposed adaptive LMS system could be applied ubiquitously.

Acknowledgement

The proposed Kalvi LMS is based on the Sakai project. A free trial hosted instance of the Sakai CLE from Longsight (<https://trysakai.longsight.com/portal>) was used during this research. The authors would like to acknowledge the contributions of all the people involved in Sakai project and Longsight, and their numerous colleagues.

References

- U.S. Department of Education - "Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics", an Issue brief. <http://www.ed.gov/edblogs/technology/files/2012/03/edm-la-brief.pdf>
- George Siemens & Dragan Gasevic, Caroline Haythornthwaite & Shane Dawson, Simon Buckingham Shum & Rebecca Ferguson, Erik Duval & Katrien Verbert, Ryan S. J. d. Baker - Society for Learning and Analytics Research - "Open Learning Analytics: an integrated & modularized platform", July 28, 2011 <http://solaresearch.org/OpenLearningAnalytics.pdf>
- <http://www.sakaiproject.org/>
- http://en.wikipedia.org/wiki/Sakai_Project
- http://en.wikipedia.org/wiki/Machine_learning
- <http://www.longsight.com/>
- <https://trysakai.longsight.com/portal>
- <https://moodle.org/>
- www.apple.com/ipad
- www.apple.com/iphone

E- Learning module of Imparting Medical and Health Education Using Classical Tamil Literature via the Internet

Dr.Semmal

Associate Professor, Department of Physiology, Sri Ramachandra University
Managing Director, Manavai Mustafa Scientific Tamil Foundation

Scientific Tamil is the need of the hour. It is an essential requirement, not an option. The fortification of the scientific armamentarium of Tamil will gift self-sustenance to the language in the future. Any misleads and self-centered approaches will only pull back the wagon – a move which should not be allowed by the intellectuals of the present day population gifted with Tamil as their mother tongue. Waiting for governments and well-established organizations to do the good job is a nicer option, but not the only one. Most of the bodies do develop a slog and working difficulties as they grow, for the very rules and guidelines laid by them will slow down the progress of their work at the ground level. Individuals can make the difference; they have made it and will make it in the future as well.

Medical Education is a rapidly evolving specialty in Science. It has been duly increasingly appreciated that the concept of Medical Education is not just teaching information to the medical students but to shape them as more responsible and more intellectual professionals best suited to pursue the profession as a medical person amidst an ever increasing population of mankind in which knowledge base is maximally open even to the common man.

Medical Education units have evolved in various educational institutions across the globe and the trend has successfully spilled over into the Indian Subcontinent. The view shared by the opponents of changes in Medical Education originates from the fear that the radical changes may alienate the students from the traditional roots unique to the Indian culture, because many still believe that preparing students with global outlook can actually make them belittle their local and traditional roots. A comprehensive method of teaching is seriously required incorporating the modern and the traditional inputs cloned in elegance. The greatest advantage unique to the Indian students is their access to the classical pieces of literature left for them by the greatest Indian civilizations of the ancient World.

The Manavai Mustafa Scientific Tamil Foundation (MMSTF), a non profit, non government organization devoted to promote scientific Tamil development has crafted an unique formula that blends Modern Medical and Health Education concepts with the Classical Indian Literature - the **Philocine module of Education**, a process by which the students can be educated about the modern Neurophysiological concepts based on literary lines crafted some 2000 years before. The Philocine approach will be a unique venture in Tamil literature for it focuses on the concept of dissemination of knowledge more than the usual formula of praising and endowing exaggerated acknowledgements for the literary works and their authors.

The MMSTF has been successful in making many publications in National and International journals as well as in making oral presentations in the World Classical Tamil Conference, Alagappa University International Conference, and Sekkizhlar International Conference. All the presentations were done using the Philocine formula for deciphering the ancient Indian literature. The 21st century approach to ancient Tamil literature using the Philocine formula will have its ramifications well beyond the preview of Tamil Literature itself for it will enhance the human understanding of the fields of Neurophysiology and Neurobiology. In actuality the reach and importance that ancient Tamil literature would gain due to the Philocine formula will be unprecedented in the history of Tamil Literature.

After proper initial planning and groundwork the revolutionary concept of Philocine, an innovative technique that conglomerates the thought process of ancient Tamils and that of the great thinkers of the western world like Charles Darwin and Sigmund Freud was unveiled by Dr.M.Semmal on the eve of the World Classical Tamil Conference during 2010.



Enumerating the Psychophysiological Insights in the ancient Indian literature is a highly potential area for research, as a sample when considering Sangam Tamil alone - a long list of classical literatures extending up to 2800 separate works by more than 470 separate authors are present, all of them are cherished for their literary excellence for more than 2000 years were actually carrying varying grades of Scientific information's hidden and embedded.

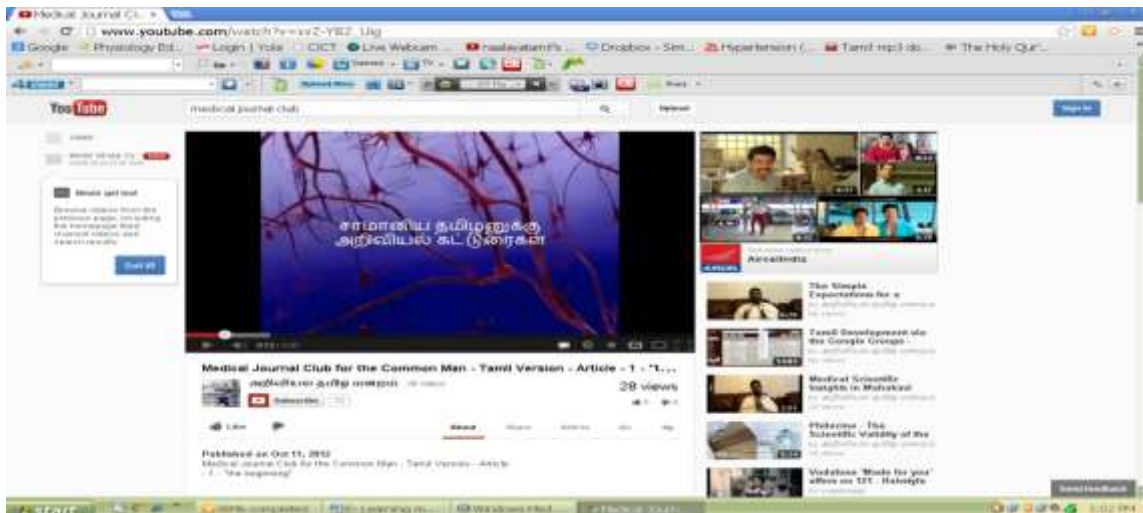
Able to fully appreciate the importance of E- learning for the Tamils, MMSTF has designed various innovative E – Learning modules for the development of scientific Tamil, on 15th June 2010 the **“Ariviyal Tamil Mandram You Tube channel”** was launched by MMSTF. This digital wing of the MMSTF is designed to impart Medical and Health Education using classical Indian Literature which would enhance the Self esteem levels of the students, it is needles to explain how greater it will enhance the process of learning among the students in the coming decades. This channel is the worlds first you tube channel exclusive for scientific Tamil. Till date 128 students have been processed by the MMSTF to present papers in various National and International conferences.



MMSTF has also introduced the concept of Medical Journal Club in Tamil and has released multiple videos in the concept of Integrating Medical, Engineering and Higher secondary school students with Physical Education students. Dissemination of recent advances in medical sciences to the common man directly from open access journals was

launched on experimental basis on the Ariviyal Tamil Mandram You Tube channel; it is designed to impart the latest of the information in simple language to the school children and the common man.

There is a strong ideology amongst many that Tamil development can be accomplished only by using the medium of Tamil, to deviate from the mainstream becomes a necessity as many second generation of Tamils have already deviated well away from Tamil, to tune them back into Tamil requires to reach them via the language well known to them and this is yet another innovative E – learning exercise by the MMSTF, all the contents in this section will be 100 percent in English, still this also means Tamil development, this is not a compromise but an innovation and an outreach, a necessity at least until the end of the next decade – purely to bring back the population from distant nations, an eventuality possible due to their extreme proximity to languages other than Tamil.



As of date the Ariviyal Tamil Mandram has processed 126 Tamil students to present papers at various National and International conferences using the E – Learning module specially designed by MMSTF, the process by which students from a rural background will be trained and processed via the internet and will be primed for the events.

The first virtual paper presentation by a team of students integrating Medical, Engineering and Sports fields were performed in the premises of the prestigious Alagappa University in 2012, It was made possible due to the E – learning strategy developed by MMSTF.

Learner Support Services and Online Distance Education: A Case Study on Bachelor of Business Management (Online) Degree Programme

Rathirane, Y

Senior Lecturer, Department of Financial Management, University of Jaffna, Sri Lanka
rathi.yogen@yahoo.com

Abstract

Distance education has evolved from a teacher centered model to a learner centered model, the roles and responsibilities of instructors and other learner support providers have changed to being more practical than imprudent. In 1990's a new generation of distance learning emerged in the form of a computer based instruction delivered over networks. The internet would begin to serve as the main medium of delivery. This type of Distance Learning course is often referred to as online learning. Bachelor of Business Management (BBM) Online Degree Programme is the first online Tamil Degree Programme in Jaffna University and which had been introduced in 2009. Due to the conflict situation in Jaffna it was commenced on 2009 however the fund was granted by the Asian Development Bank (ADB) through the Distance Education Modernization Project (DEMP) in Sri Lanka in 2006. When the programme was initiated, it faced challenges and difficulties without awareness regarding the online mode of education among the staff of the Jaffna University. Most of the students who work in the private and the public organizations follow the degree and they depend on learner support services to enhance their knowledge and skill and qualify their examination. A critical element of an effective succession program for online students is Learner Support Services (LSS). Learner support comprises all the assistance provided by a distance education or an online mode system which matches the facilities which a face to face system provides for the success of its students. LSS include tutoring mentoring, content development, Technical support services, interaction with students and administrative activities. This study examines the first time online learners' perceptions of the learner support services provided by the faculty of Management Studies & Commerce, University of Jaffna, Sri Lanka. The study suitably used both primary and secondary data. In this case, the researcher selected 96 students from BBM online degree programme conveniently. Qualitative and quantitative techniques have been used to survey the study. The data was gathered from the forums, discussions and interview with the students with the help of structured questions. Descriptive analysis and frequency table have been used to find out the results. While many factors contribute to the learner support, the level of interaction and support are at the top of the list. It is concluded that sixty percentage (60%) of the students in Online learning program report feelings of isolation, lack of self-direction and management, and finally decrease in motivation levels. This study identifies the types of LSS strategies that can effectively address the challenges. Suggestions from the study by getting feedback from the students that the facilities such as usage of modern technology, ie using video conferencing, uploading PowerPoint presentation with sound recording and the advance instructions regarding examination and other face to face activities and the effective administrative procedures should be enhanced regarding the LSS. In order to provide LSS that help students successfully participate in online degree programme at University of Jaffna interactive strategies are needed to enhance students' self directed learning skills and reduce their feelings of isolation. Outcomes of the study would benefit the academicians, researchers, policy makers, and practitioners of Sri Lanka and other similar countries.

1. Background of the study

In recent times the distance education mode has been developed in the developed and developing countries as one of the mode of education model which helps to all students to study without physical appearance. More universities and organizations are starting online programs. Much of this increase is due to the demands of the students who are conspired by distance education, mostly because they face a number of obstacles that make conventional method of teaching.

In Sri Lanka the Open University is an important institution providing the distance education mode system and the online learning system has been introduced by some of the Sri Lankan Universities and other institutions. Online learning system uses the web based education system and the study guidance is provided by Learning Management System (LMS). The Jaffna University provides Bachelor of Business Management (BBM) Degree programme in Tamil medium as an external degree under the faculty of Management Studies which was the first degree programme in Jaffna University.

Learner support activities include tutoring and teaching; counselling and advising including such services as orientation, learning and study skills assistance, academic advising, and career and personal counselling; and administrative activities such as admission and registration, library and information systems, and infrastructure support for activities such as peer tutoring and alumni organization. In other words, learner support activities are all those interactive processes that are intended to support and facilitate the learning process. This research explores the Students' perception on learner support services for the BBM Online students in Jaffna University.

Traditionally learner support in distance education has been identified as being a completely different set of activities. However, with the implementation of online learning, this distinction does not always hold and the line between the two sets of activities has become much more distorted (cf. Thorpe, 2003). The BBM Online Degree programme provides the variety of LSS to the students and who are facilitated to interact with other students and teachers. No opportunities to find out the feelings of the students whether they are satisfied or not with the services. This study tries to answer the following question: "Does the first time online students perceive the learner support services of BBM Online Degree is remarkable?"

The main objective of this study is to identify the online students' perception regarding the learner support services provided by the faculty of Management Studies and Commerce. In addition to this the following objectives also are incorporated in this study.

- To find out the obstacles when the LSS are provided
- To suggest the possible solutions to overcome the obstacles and
- To recommend the strategies for learner support services to be satisfied the students.

2. Theory and Hypothesis

Learner support services play a key role (Ludwig-Hardman & Dunlap, 2003) since they serve as a vitally important and effective mechanism to counter challenges of attrition related to online learners' sense of isolation and inability to function as self-directed learners (Keegan, 2003). In addition, knowing who the learners are and using this knowledge to inform the design and provision of learner support services increases the likelihood that they meet learner needs as well as inform best practices in this area (Hughes, 2004; Moisey & Hughes, 2008).

According to Abrami and Bures (1996) some students in distance learning programs and courses report feelings of isolation, lack of self-direction and management, and eventual decreases in motivation levels. These factors are supported by Tinto's (1997) model of institutional departure with its central notion that student persistence is strongly predicted by their degree of academic integration (e.g., performance, academic self esteem, identity as a student, etc.) and social integration (e.g., personal interaction, connection to academic community, etc.). One of the most reliable problems associated with distance learning environments is logic of isolation due to lack of interaction (Bennett, Priest and Macpherson, 1999; Harasim, Hiltz, Teles and Turoff, 1995). This logic of isolation is linked with attrition, instructional ineffectiveness, failing academic achievement (Booher and Seiler, 1982), and negative attitudes and overall dissatisfaction with the learning experience (Thompson, 1990). In this case the following assumptions have been established as hypothesis:

- H₁: Learner support services provided is remarkable in the BBM Online Programme.
H₂: Students' perception on learner support services avoids the feelings of isolation.

3. Methods

The study suitably used both primary and secondary data.. There are two batches are going on smoothly. This is a three year degree programme and two semesters per year. Altogether 6 semesters with 96 credits must be completed by the students. First batch is in the fifth semester and the second batch is in the 2nd semester. In this case, the researchers selected ninety six (96) students from BBM online degree programme conveniently. The data was gathered from the forums, discussions and interview with the respondents. The Learner support services have been categorized as tutoring – mentoring, content development, Technical support services, interaction with students and administrative activities. Feeling of isolation and lack of self direction and management are included under the Students' perception which have

been measured by three point likert scale from strongly agree to strongly disagree that is related to learner support services and revealed by the percentage of respondents. In this case above the percentage of 50 has been accepted as a remarkable learner support services have been provided in this programme. The linkage between students' perception and learner support services have been illustrated by using correlation analysis.

4. Results

Students' perception on learner support services has been illustrated by using percentage of the respondents' responses as follows:

Table 1: Percentage of Students' perception on Learner Support Services:

Learner Support Services	Strongly Agree or Agree %	Neutral or No Opinion %	Strongly Disagree or Disagree %
Tutoring – Mentoring	45	37	18
Content Development	48	38	14
Technical Support Services	40	37	23
Interaction With Students	47	28	25
Administrative Activities	32	38	30

Source: Survey results

Table 1 shows that the perception on learner support services reveals below 50% agreed with the learner support services which can't be said as remarkable. Therefore, the first hypothesis has been disproved from this result.

Table 2: Relationship between Learner support services and students' perception:

	Tutoring mentoring	Content development	Technical Support Services	Interaction with students	Administra tive activities	No Feeling of Isolation	self direction and mgt.
Tutoring- mentoring	1	0.668**	0.576**	0.437**	0.576**	0.413**	0.437**
Content development		1	0.466**	0.472**	0.657**	0.478**	0.468**
Technical Support Services			1	0.474**	0.658**	0.415**	0.428**
Interaction with students				1	0.676**	0.497**	0.478**
Administrative activities					1	0.416**	0.437**
No Feeling of Isolation						1	0.578**

Self direction and mgt.							1
-------------------------	--	--	--	--	--	--	---

** - Correlation is significant at the 0.01 level

Correlation test was made to examine the relationship between variables and results are tabulated. From the Table 2 the results revealed that there is a significant relationship between the Learner Support Services (LSS) and perception of the students at 0.01 significant levels. From this result the hypothesis 2 has been disproved because the relationship between the perception and learner support has the weak moderate relationship and it could be proved that the students feel isolation and lack of self direction.

5. Conclusion

From the above results it can be concluded students dissatisfy with their learner support services because they feel isolation, lack of self direction and management, unethical submission of assignments and lack of technical facilities etc. Insufficient academic support leads to them to withdraw from their course. Further, students' performance depends on effective LSS and the possible suggestions from the study by getting feedback from the students that the facilities such as usages of modern technology i.e., using video conferencing, uploading power point presentation with sound recording and the advance instructions regarding examination and other face to face activities and the effective administrative procedures should be enhanced as LSS in order to improve the performance. And some strategies are highlighted as suggestions to improve the quality of the LSS in future.

When this delivery mode is new, student's lack of access to support services such as advising, tutors, libraries and technical support, placed them at a disadvantage. It should be concentrated carefully by the faculty. Designing support services should be available for the online learning students, including support for admissions, registration, career advice, tutoring, academic advising, library, etc., will result in a more personalized approach for online students. This also includes groups formed between librarians, advisors, and technical support with teaching faculty, which can be essential in helping students complete their courses successfully. Technical support by telephone, email, online chat, or web-based tutorials, for both students and tutor-mentors, are crucial in limiting the extra burden of managing the technology along with mastering the course content. Students can refer to this support when they have difficulties on completing exams or submitting assignments, rather than failing to complete the required work. Failure of internet access and power failure should be eliminated in the Jaffna region.

References

1. Bennett, R. (1986) Through the open door - today's revolution in open access and distance learning, *Journal of European Industrial Training*, vol. 10, no. 6, pp. 1
2. Cartwright, W. and Miller, S. (2000) Setting a research agenda for multimedia cartography, *Proceedings of Mapping Sciences Institute Australia National Conference - Beyond 2000 Spatial Business Daring to Change*, December, 2000, Sydney, pp. 333-343.
3. Daniel, J. (2001) Distance education and the developing world, *Association of Commonwealth Universities Bulletin*, no. 147.
4. Keremidchiev, D. & Yankov, P. (2001), *Challenges and Advantages of Distance Learning system*, Information & Security, Volume 6, Pp 111-121.
5. Stacey L, Hardman, C. (2003) Learner Support Services for Online Students: Scaffolding for success, *The International Review of Research in Open and Distance Learning*, 4 (1) 1492-3831
6. McDougall, K. and Dowling, D. G. (1996) Education, communication and integration: a look at the future of distance education in surveying, *Proceedings of 37th Australian Surveyors Congress*, Institution of Surveyors Australia, April 1996, Perth, Australia, pp. 435-444.
7. Nyirenda, J.E. and Kakanda, A., (1982), *Distance Education at the University of Zambia. Media in Education and Development*, 15, (1) 2225, 1982.
8. Rumble, G. (2000) Student support in distance education in the 21st century: Learning from service management, *Distance Education*, vol. 21, no. 2, pp. 216-235.

9. Thomas, W. R. (1999) Electronic Delivery of High School Courses: Status, Trends and Issues. Southern Regional Education Board .
10. Nanayakara, G.L.S. George, W, (2007) Research Studies on Tertiary Education Sector, National Education Commission, Sri Lanka.
11. Lakshman, J, Hettiaratchi, K, (2009) National Online Distance Education Service - NODES Handbook for School Libraries and use of Career Guidance Teachers.
12. BBM Online Degree Proposal submitted to the DEMP Office, Memorandum of Understanding and other Faculty board, Senate Reports and documents.
13. Direct interviews with related staff and direct observations and experiences regarding the BBM Online Degree Programme.

COMPUTER AIDED TAMIL LANGUAGE TEACHING TO CHILDREN WITH LEARNING DISABILITIES

Dr. R. Lalitha Raja

Assistant Professor, CAS in Linguistics, Annamalai University

M. Ramana Kumar

Student, BASLP, SRM Medical College Hospital and Research Center.

1. Introduction

Teaching language to children with learning disability is a challenging area that always focuses on learner-centered approach. Many projects have been introduced (throughout different parts of the world) and innovations have brought changes in materials (supplementary, self-study, authentic, etc) and technology devices (language laboratories, tape recorders, video recorders, computers and software, etc) . More recently, the use of technology as a tool to develop the different language skills has received great attention so that special educators as well as speech-language pathologists are frequently exposed to new practices.

1.1 Learning Disability

A learning disability (LD) is in most situations a "hidden disability" with no outward signs of a disability. Children with LD are those who have average/high intellectual ability but struggle in their academics. A learning disability is unique to the individual and can appear in a variety of ways. It may be difficult to diagnose, to determine impact, and to accommodate.

Learning disabilities usually fall within four broad categories:

- **Spoken language** - listening and speaking
- **Written language** - reading, writing, and spelling
- **Arithmetic** - calculation and concepts
- **Reasoning** - organization and integration of ideas and thoughts.

A person with a learning disability may have discrepancies in one or all of these categories. Often people with Attention Deficit Disorder/Attention Deficit Hyperactive Disorder (ADD/ADHD) also have learning disabilities.

People with an LD are often neglected when considering assistive computer technology. Regarding Tamil language no specialized software are available for language teaching for these children. Many people with learning disabilities can benefit from mainstream and specialized hardware and software to operate a computer and further their academic and career goals. As computer based teaching/learning involves animated features and interactive sessions with feedbacks it helps the children to hold in learning process for long time and rehearsal makes them to store the linguistic feature in their long term memory.

2. Aim

The present study aims to frame a methodology for developing software to teach various linguistic features in Tamil which are considered to be a problematic area for the children with LD. The linguistic features important for reading and its main problematic areas of these children are; sound recognition, phonological awareness, morphological awareness and vocabulary. So the present study frames a methodology to build software for above said features.

3. Methodology

The methodology to build software for teaching sound recognition, phonological awareness, morphological awareness and vocabulary in Tamil are as follows.

3.1. Sound recognition

a. Teaching to Recognize Rhyme

Teaching children to click appropriate buttons to locate sounds.

- One sound - will need to click the buttons with one picture.
- Two sounds - will need to click the buttons with two pictures (both the same picture to show that the two sounds were the same) .
- Three sounds - will need to click the buttons with three pictures (all the same picture to show that the sounds were the same).

Teacher says	Child shows the picture
k . . . k	two picture, the same one
n . . . p	two picture, the different ones
r . . . r . . . r	three beads, the same one
t . . . r . . . l	three beads, all different pictures

b. Teaching to Recognize Rhyme

Giving tasks like;

"Do these words rhyme? / paTam/, /maTam/

"Which of these words rhyme? /vaTTi/, /vaLLi/, /aLLi/

and the child has to click the buttons with words that sounds rhyming.

3.2 Phonological Awareness

Phonemic awareness is "the ability to hear, identify, and manipulate the individual sounds, or phonemes, in spoken words" a key for successful phonological processing. Phonemic awareness gets much attention in the field of literacy, in particular as an essential skill needed in the development of print literacy (Adams, 1990) . The following are the phonological awareness skills.

- **Phoneme Blending** - to find the word made by blending the phonemes
The picture of tree sounds five phonemes - "p-a-T-a-m"
The fruit in the ground has words with 4, 3 and 5 phonemes –
"p - a-m" "p-a-T-a" "p-a-T-a-m"
Click on the fruit that has the word made by blending these sounds.
- **Phoneme Segmentation** - to find number of phonemes in a word
The picture of a flower sounds a word with 4 phonemes - "k-o-T-u"
The leaves in the ground has numbers 3, 4, and 2
Click on the leaf that has the number of phonemes by counting the phonemes in the word
- **Phoneme Identification** – to identify the sound you hear at the beginning of the word
The picture of a ball sounds a word - "karam"
The bats in the ground has 3 different words –

“ram” “aram” “kal”

Click on the bat that has the sound you hear at the beginning of this word.

- **Phoneme Addition** - the word you get when you add a sound
The picture of cup sounds a word - *“eru”*
The saucers in the ground has 3 different words –
“peru” “etu” “aru”
Click on the saucer that has the word you get when you add a sound.
- **Phoneme Deletion** - the word you get when you remove a sound
The picture of bag sounds a word - *“naaRu”*
The books in the ground has 3 different words –
“naaRRu” “nuuRu” “aaRu”
Click on the book that has the word you get when you remove a sound.
- **Phoneme Manipulation** - the word you get when you change a sound
The picture of cup sounds a word - *“aru”*
The saucers in the ground has 3 different words –
“paru” “eru” “arul”
Click on the ball that has the word you get when you change a sound.
- **Syllable Blending** – to find the word made by blending the syllables
The picture of balloons sounds three syllable - *“va” - “ru” - “vaar”*
The balls in the ground has words with 2, 3, 4 syllables –
“varaar” “varukiraar” “varuvaar”
Click on the ball that has the word made by blending these syllables.
- **Syllable Segmentation** - to find number of syllables in a word
The picture of a flower sounds a word with 4 syllables - *“ko – Tu – kiR – aar”*
The leaves in the ground has numbers 3, 4, and 2
Click on the leaf that has the number of syllables by counting the syllables in the word.
- **Sentence Segmentation** – to find number of words in a sentence
The picture with 3 balloons sounds three words - *“itu en peenaa”*
The balls in the ground has 4 words, 3 words and 2 words –
“itu en peenaa” “itu peenaa” “itu en putiya peenaa”
Click on the ball that sounds three words.

3.3. Morphological Awareness

Morphological awareness is defined as the ability to use the knowledge of word formation rules and the pairings between sounds and meanings (Kuo & Anderson, 2006) . With morphological awareness, learners are able to learn morphemes and morphemic boundaries by disassembling complex words into meaningful parts (e.g. *kuḷantaikaLukku* = *kuḷantai* + *-kaL* + *-kku*), learning the meanings of roots, affixes (*kuḷantai*= baby, *-kaL*= indicate plural nouns, *-kku*= to), and reassembling the meaningful parts into new meanings (*peNkaLukku*, *iLainkarkaLukku*, *muthiyavarkaLukku*) . The following are the phonological awareness skills.

- **morpheme identification** – find the segment of the words into meaningful chunks, we feed words (*kuḷantaikaLukkukaga*) which can be divide into free and bound morphemes with function/meaning of each morphemes, first we teach free morpheme (*kuḷantai*) with a picture, then with bound it (*kuḷantai – kaL*) we have to add a picture of two or more children. Then children are taught to identify morphemes by giving the function/meaning.
- to identify the sound you hear – *kiR* (or function of morpheme)
The picture of a ball sounds a word - *“paaTukiRaar”*

The bats in the ground has 3 different morphemes –
 “paaTu” “kiR” “aan”

Click on the bat that has the sound you hear at the beginning of this word.

- **Synthesis (morphological structure)** - find the word made by attaching the suitable morphemes. A root was given and there will be few markers are given as a choice, the child has to click the button with right marker. Pictures will be given as clue.

example:

itu oru pantu, mejaiyil iraNDu pantu ____ irukiRatu a) kaL b) ai c) aal

aNNan paTam varai ____a) kiRaana b) ntaaL c) kiRaAL

3.4. Vocabulary



a. Homophones

All homophones are presented with the context and with the help of pictures.

palli ‘lizard’ - cuvaRRil palli irukkiRatu.

paLLi ‘school’- avaL paLLi celkiRaaL.

Tool tip can given to each word highlighting the meaning.

<p>palli ‘lizard’ – cuvaRRil palli irukkiRatu.</p> 	<p>paLLi ‘school’- avaL paLLi celkiRaaL.</p> 
---	---

Kalvi kaRkum iTam

Exercise:

1. Matching context based picture.

2. Sentence filling.

For example: avaL _____ celkiRaaL.

i. paLLi

ii. palli

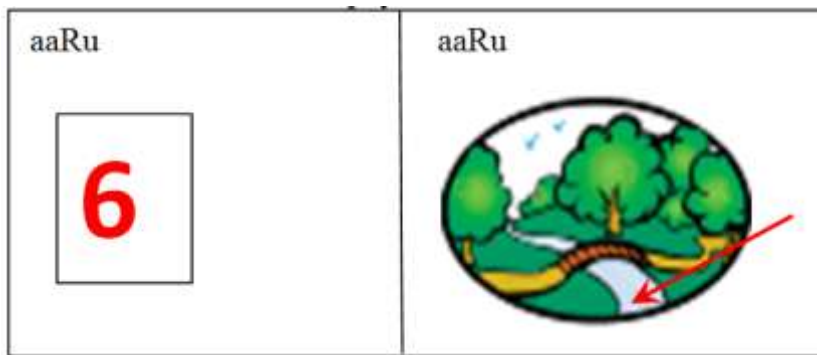


b. Homographs

All the homographs are fed to the computer. The context is provided with the help of pictures.

aaRu- muruganukku mugam aaRu.

aaRu- inta aaRRin peyar enna?

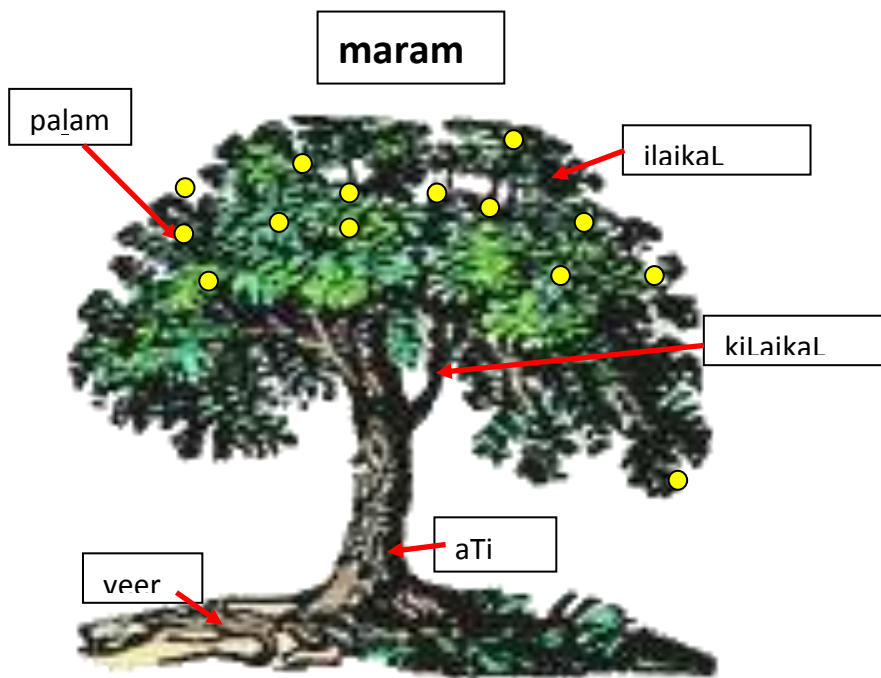


Exercise:

1. Context based picture
2. Matching the picture according to the context

c.Meronyms

All head words and related words are presented with picture. For example the picture of tree with a word 'tree' and its parts, are marked as branch, stem, fruit, leaves and root in the same picture. Zooming is provided to go further to view the parts.



Exercise:

1. Labeling the pictures.
2. Matching the words with the pictures.
3. Treasure hunt.

4. Concluding Remarks

The methodology provided above is a sample for each feature. As the problem of the dyslexics is to read normal text in printed form and as they have lot of distractions, computer aided interactive visualized teaching will be very effective to teach vocabulary to these children.

As these children have memory problems, this helps to store in long term memory as they are presented with pictures or animations. Also any amount of words can be given in each section and according to the time available to the students they can go through the entire list.

5. References

1. Adams, M.J. (1990) . Beginning to read: Thinking and learning about print. Cambridge, MA: MIT Press.
2. Ball, E.W., & Blachman, B.A. (1991) . Does phoneme segmentation training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly*, 26, 49–66.
3. Kuo, L. - J. & Anderson, R. C. (2006) . Morphological awareness and learning to read: A cross- language perspective. *Educational Psychologist*, 41, 161-180.
4. Scott, J., & Nagy, W. E. (2004) . Developing word consciousness. In J. F. Baumann & E.J. Kame'enui (Eds.), *Vocabulary instruction: Research to practice* (pp. 218-238) . New York: Guilford.
5. White, T.G., Power, M.A. & White, S. (1989) . Morphological analysis: Implications for teaching and understanding vocabulary growth. *Reading Research Quarterly*, 24, 283-304.

Challenges and Health Risk Faced by Tamil School Students in Malaysia with Immediate Need for ICT Interference

Saravanan Mariappan

Nexus NGN Sdn. Bhd., Malaysia
nexusngn@gmail.com

ABSTRACT

Primary schools are the first institution where children receive their early stage of compulsory education. Children in Malaysia require to attend primary school, from the age of seven until the age of twelve. During this period the bag weight and the body weight ratio is likely to be high as some students are relatively small to carry loads similar to older children. In recent years, this issue caught the attention of few parents but the problem of proposition of bag weight to weight and height of children continuously haunts the parent's mind without any form of solution to the problem. Hence the role of ICT are seek in order to shoulder the responsibility of enhancing the teaching engine which depend on traditional books to digital books. Modernizing the current Tamil school infrastructure should also be part of mainstream overhaul in order to create much healthier, safer and exciting environment to study Tamil. This paper address our research finding on possibility of long term health deterioration under gone by Malaysian Tamil school students. Recommendation will also be inked within this paper for better and safer studying environment for our students with improvement to current infrastructure.

1. INTRODUCTION

Primary schools are the first institution where children receive their first stage of compulsory education. Children in Malaysia require to attend primary school from the age of seven until the age of twelve. During this period the bag weight and the body weight ratio is likely to be high as some students are relatively small to carry loads similar to older children. In recent years, this issue caught the attention of few parents but the problem of proposition of bag weight to weight and height of children continuously haunts the parent's mind without any form of solution to the problem.

2. PROBLEM STATEMENT

The Times Of India (2010) have reported that a scientific study has confirmed that heavy back packs carried by children aged 10-13 years can caused disability at later age. The Times Of India also reported that the study conducted by the Movement Analysis Lab of Srinivas College of Physiotherapy and Research Center in association with Kasturba Medical College in Mangalore on 200 school boys has found that bags weighing 5% of the body weight of child could affect the trunk and lower limb angle and those bags weighing 15% could change all angles pertaining to head, neck, trunk, and lower limb changing the overall posture of the child.

Couple to Times Of India report, article released by "Towards Queensland Tomorrow" says that students who carry heavy school bags may experience fatigue, muscle strain, back pain, distortion of the spine's natural curves and rounding of the shoulders. This fact has also been highlighted by S.Dockrellin in the same report that musculoskeletal symptoms are believed to be multifactorial and Whitfield (2001) states that the carriage of heavy schoolbags is a suspected "contributory factor" and therefore represents an "overlooked daily physical stress" for school children. Risk factors like musculoskeletal discomfort associated with school bag carriage include the combined effects of heavy loads, load shape and size, time spent carrying the load, position of the load on the body and walking up staircases with heavy loads.

A general guideline of 10% body weight, initially proposed by Voll and Klimt in 1997 continues to be the recommended guideline when carrying a back pack style schoolbag. The findings of more recent studies may challenge this guideline. The weight carried by students varies from day to day and studies have reported different results in regards to the average school bag weight. The different age range for students included in the studies may explain some of this variation, but irrespective of age there is a considerable number of school children carrying in excess of 10% of their body weight. Field data sample survey was conducted at a Malaysian school by surveying about 150 students from standard 1 to 6 of all available 5 classes at each standard. Following information were extracted from the survey:

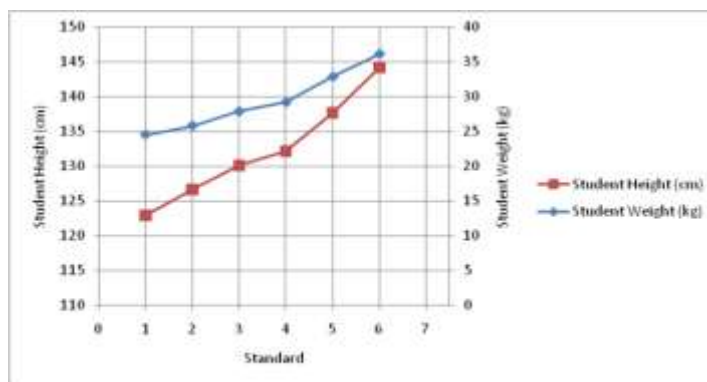
1. Bag weight (kg)
2. Student weight (kg)
3. Student height (cm)
4. Interview student health condition
5. Book sizes (biggest to average book size)
6. Interview teacher on how to reduce bag weight
7. Measure classroom dimensions to identify suitable location to place the locker
8. Measure current table sizes to identify the method to retrofit the table

Scales and measuring tape were using as tools for measurements. An average of 5- 10 students were randomly selected for the study from each class and standard.

3. SUMMARY OF FINDINGS

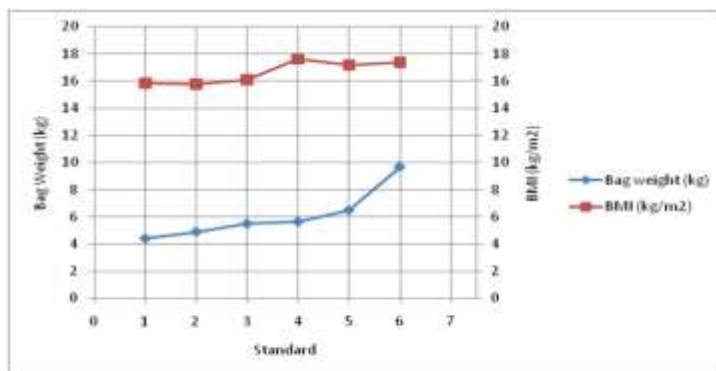
Few critical plots were made in order to simplify the data collected.

Plot 1: Student height and weight for all the standards.



Plot 1: Student height and weight for all the standards

Plot 2: Student BMI, bag weight for all the standard

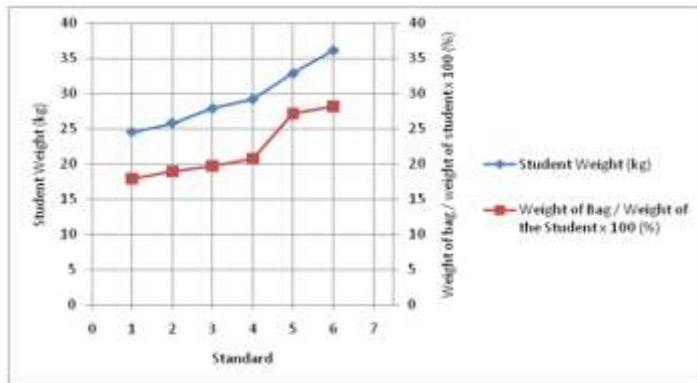


Plot 2: Student BMI, bag weight for the entire standard

With the plots and analysis made the students found to be carrying much heavier weight than recommended in the studies reported in "The Time of India". The bag weight should only be about 5% of the student weight, however the students are found to be carrying their school bag which is about 22.13% of their body weight much higher than recommended by Voll and Klimt (1997) . Hence there is a good possibility of the students to have long term health problem related to spine due to heavy school bags. By analyzing the school classroom sizes and possibility of adopting locker system for each student, we found the following problem:

- I. The classrooms are too small to place lockers of about 80 units (2 section, 40 students per section). Refer Figure 1.0.
- II. Most existing furniture are desk without pocket.
- III. It is unsafe to leave behind the books at the desk pocket without locker, as the next section student might displace the books.

Plot 3: Student weight, Percentage of weight of bag/ weight of student for all the standard.



Plot 3: Student weight, Percentage of weight of bag / weight of student for the entire standard

4. DISCUSSION

Hence it is proposed to have segmented lockers of 5.3 feet long by 1.2 feet width and 4 feet high with rollers for easy movement within the class room. After taking into account average book size refer Figure 2.0, the proposed locker dimension with improved desk suitable for Malaysian school environment are shown in:

- I. Figure 3.0 shows the stand alone locker system with wheels attached-suitable to move about. However not suitable for classrooms with space constrain.
- II. Figure 4.0 shows the improved desk system with build in two slim lockers. This approach can be retrofitted into existing table and suitable for class rooms with space constrain.

A combination of the above two system can be adopted with a possibility of deducting 30% -50% of the student bag weight. Bringing down to tolerable bag weight of less than 10% of student body weight.

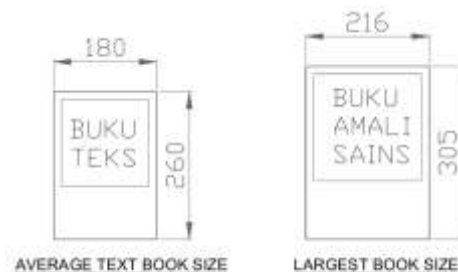


Figure 2.0: Average book size

5. HEALTH CHECK AND INTERVIEW WITH TEACHERS

Based on the study it is found that the students are carrying bag weight 2 to 3 times heavier than recommended 5-10% of body weight. The ideal weight should be only 5% of body weight. The students at this early stage found to be not able to report fatigue, exhaustion and muscle weakness due to carry heavy bags. They are silently tolerating the mass in the excitement of going to school and to avoid being punished by teachers. In depth interview and health check is required to

be conducted to identify the long term health problem to the children. When a child complains of leg or hand pain or neck pain, the child or parents most likely advice the client to rest, without realizing it is the early signs of spine distortion.

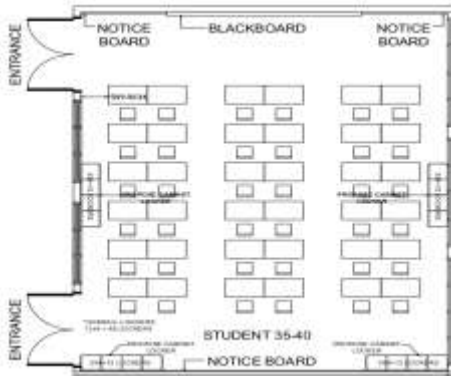


Figure 1.0: The classrooms are too small to place lockers of about 80 units
(2 section, 40 students per section)

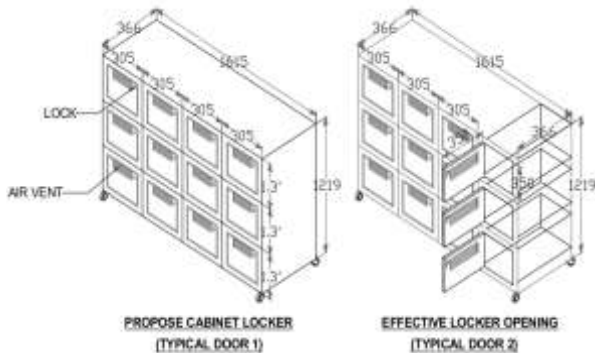


Figure 3.0 shows the stand alone locker system with wheels attached-suitable to move about.



Figure 4.0 shows the improved desk system with build in two slim lockers.

As for the teachers, reduction in bag weight is well received with the use of locker system. This could ensure the references or text books is alw ays available for the teacher to utilize for teaching.



Plate 1 – Weight of students with their bags are scaled



Plate 2 – Size of bag is measured



Plate 3 – Total numbers of bag carried by a student



Plate 4 – Height of student is measured



Plate 5 – Shows the health consequences of carrying heavy school bags.

6. RECOMMENDATION AND CONCLUSION

Based on the study it is found that the students are carrying bag weight 2 to 3 times heavier than recommended 5-10 % of body weight. The ideal weight should be only 5 % of body weight. To prevent long term health damage of the children, it is recommended carrying out the following measures:

- I. Teachers are recommended to guide the children what books are required to bring for the day or next lesson.
- II. Parents are recommended to assist the students in preparing their school bags according to time table and teacher's guide.
- III. Heavy text and reference books, it is recommended to be left behind in school lockers.
- IV. Propose to introduce lockers for each student within the classroom for reduce the bag weight
- V. Introduce new standalone locker system as shown in Figure 3.0.
- VI. Improve or retrofit the existing desk with pockets as shown in Figure 4.0.
- VII. Perform combination of standalone locker and retrofit the existing desk in order to bring down the weight of school bag.
- VIII. Avoid piggy riding work books – old completed work books are paste together with new work books to ensure continuity of notes, however this cause children to carry redundant weight causing the bags to be more heavier.
- IX. Install access ramp, similar to OKU's wheel chairs ramp in order for children to pull the bags on wheels over split platforms to avoid being carried over the staircases or walk long distances carrying the bags.
- X. Introduce single section school in order for students to safely leave their books within their class rooms and reduce bag weight.
- XI. Provide sufficient space at car parks or drop off area for bus and cars to to stop next to an access ramp to minimize bags being carried.
- XII. Install easy to deploy and retrofit hydraulic lift for building blocks higher than 2 floors (ground floor and first floor) .
- XIII. Recommend school building not higher than 2 floors (ground floor and first floor)
- XIV. Allow for larger classrooms to fit in built in lockers
- XV. Implement teaching system using ICT.
- XVI. Digital notes and exercise books using tablets will be able to reduce bag weight.

With all the above recommendation we believe the long term health of the children can be safe guarded. Plate 5 shows the health consequences of carrying heavy school bags.

REFERENCES

1. Health & Safe Fact Sheet, Towards Tomorrow's Queensland (Internet) 27th July 2012. Available at http://education.qld.gov.au/health/pdfs/healthsafety/factsheet_hevbags.pdf (Accessed 27th July 2012).
2. Time Of India, Heavy school bags can cause disability in children, (Internet) 27th July 2012. Available at http://articles.timesofindia.indiatimes.com/ 2010-07 -23/ chennai/ 28305425 _1_ heavy- school-school-bags-heavy-back packs (Accessed 27th July 2012)
3. Voll H-J and Klimt F (1977) translated by Theodridis D (1977): On strain in children caused by Carrying schoolbags; 39:369-378
4. Whittfield JK, Legg SJ, Hedderley DL. The weight and use of schoolbags in New Zealand secondary schools. Ergonomics 2001; 44: 819-824.

Developing a Linguistic-based Tool for Teaching and Learning of Tamil Verbs

Saranya, R

CAS in Linguistics, Annamalai University

M. Somathasan

Sri Lanka

Introduction

The things referred to by language are never merely things of the phenomenal world which can be demonstrated with the help of technology i.e. *science*. A large number of things are feelings, ideas, and values. In fact, the words in a language are so loaded with feelings and values that social scientists feel, sometimes, the necessity to devise a science, like mathematical science, which has signs with no such load to carry. Social phenomena like group solidarity, reference groups, and host of others are fostered through the help of a language though other group distinguishing signs are also often used. The load of meaning and emotion carried by words is sometimes made very explicit in the grammar of language. Thus, this paper tries to focus on teaching and learning of Tamil verbs which express the different moods of the speaker or writer.

Aim

The aim of this paper is to develop a linguistic based tool for teaching and learning of strong verbs in modern Tamil.

Objectives

The objectives of the developed tool are to make the students or learners:

- understand the concept of Tamil verbs and their structures.
- learn the different inflected forms of a verb such as finite and non-finite forms.
- know the various functions of verbs.
- prepare themselves at their earliest stage to work with computers and to compete with the technological challenges in the electronic educational set up.
- enhance their knowledge on verb paradigms in Tamil.
- stimulate them to learn eagerly by the immediate feedback of their learning activities.

Methodology

For the present study, about 40 strong verbs in modern Tamil have randomly been collected from the school text books as source data. The collected data have been analyzed linguistically to find out their components which are internally organized systematically for framing their structures. Finally, the linguistic information on verbs have been computed by applying the programming frame work namely VB.Net.

Discussion

Verbs in modern linguistic world are defined as one of the grammatical elements of speech that are used to express conceptualizations of activities, events, processes or states. In Tamil, they function morphologically and syntactically in three ways as follows:

- **As finite verbs:** As far as Tamil language is concerned, there are three types of finite verbs namely *imperative* (i.e. vaa / vaarungkaL 'come', *indicative* (i.e. vantaan 'came-he') and *optative* (i.e. varuka 'welcome/come') . They are distinguished according to the category of mood.

- **As non-finite verbs:** They function syntactically adverbs, adjectives. E.g, here the verbal participle *vantu* in *vantu poo* and *cenRa* in *cenRa paiyan* can not be syntactically considered as adverbs and adjective, respectively. But *cenRa* in *cenRa aaNTu* and *taLLi* in *taLLi nil* are considered as pure adjectives and adverbs respectively, since they qualify the noun as well as the verb which follow them.
- **As derived verbs:** The transitive verb *naTattu* 'conduct' and the causative verb *naTappi* 'make to walk' are derived from the intransitive verb *naTa* 'walk'. Similarly, the causative verb *tooRRuvi* 'cause to appear' is also derived from the inherent transitive verb *tooRRu* 'make to appear'.

The principal inflectional categories occurring with verbs appearing in the finite forms are tense, aspect, mood, gender, number and person, whereas in non finite forms are infinitive, verbal and adjectival participles, etc.; and they employ representations for syntactic category and semantic- syntactic function. For instance,

avan puttakam vaangka kaTaikkuc cenRaan.

'He went to the shop to buy a book'.

The non finite verb *vaangka* 'to buy' which is morphologically treated as infinitive but it is syntactically used for expressing the meaning of *purposive* that is known as semantic-syntactic function. But the developed tool highlights the inflectional forms of the verbs which express not only various functions but also the different moods such as:

- *Indicative:* It, however, is not completely devoid of emotional load as the context in which it occurs the tone in which it is uttered, might give it this weight.
- *Imperative:* The imperative as the name suggests, calls before our eyes a social situation of superiority and subordination, one of the most fundamental of social situations.
- *Optative:* This suggests an option and therefore a freedom to do or not to do a certain act. When the words *talaivar avarkaLee varuka!* are uttered, the form *varuka* 'come - optative' brings about a desire in the heart of the speaker to welcome the *talaivar* 'a leader'. The form *talaivar avarkaLee* 'third person epicene most honorific singular' which is loaded with social values, all that is sectioned and, do a certain extent, sanctified by the society, because the form *avarkaL* 'they' does normally not refer to singular. This form represents the whole value framework of a society. That the language should have recognized the various forms of verbs shows the range of social function which language fulfills.

The developed tool consists of four things such as,

1. Verb Classification
2. Concept of Verbs
3. Verb Classification with examples
4. Exercise

The visual is given below.



1. **Verb Classification:** It deals with how the verbs are structurally as well as functionally differed from one another. It simply shows the traditional classification of the Tamil verb. The illustration of the program has various states which are visualized in the following screenshots.

Visual-1



State-1

This is the starting state at which the teaching and learning process of the verb classification begins. When the term *verb* given in the visual -1 is clicked, it moves to next state that performs the results displayed in visual-2.

Visual-2



State-2

In this state, the basic classification of verb such as finite, non finite and defective verbs is depicted as it is given in the visual - 2. When the terms displayed in this visual are clicked, it moves to the next state where the results are displayed.

Visual-3



State-3

This state shows the results of the previous state, that is classification of finite verbs namely tensed and non –tensed which come under the term *Finite*. In this state, when the term *Tensed* displayed in the visual -3 is clicked, it goes to the next state.

Visual-4



State-4

This state depicts the results of the previous state that is the classification of the Tensed (finite) verbs which refer to the three tenses such as present, past and future. When the term *Non-Tensed* visualized in the visual -4 is clicked, it is moved to next state.

Visual-5



State-5

In this state, the classification of the Tensed (finite) verbs such as imperative and optative is visualized as the results of clicking made in state-4. The term *Non-Finite* displayed in the visual-5 is clicked to move to next state and perform its classification.

Visual-6



State-6

It visualizes the various forms of non-finite verbs namely infinitive, participle and conditional. Among them the term *participle* displayed in the visual-6 is clicked to go next, and display the sub-division of participle namely verbal and adjective participles.

Visual-7



State-7

The result of the previous state is visualized here. In this state, the term *Adjective (participle)* displayed in the visual-7 is clicked to go to the end state and perform the present, past and future Adjective participles.

Visual-8



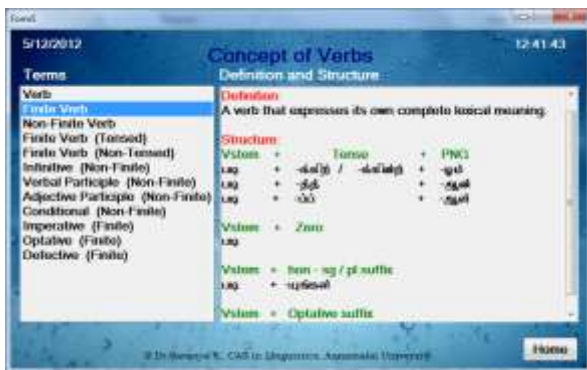
State-8

This state ends the program successfully with the results of the classified verb based on the structural aspect. And then click the *Home* button displayed in the visual -8 to go *HOME* page of this tool.

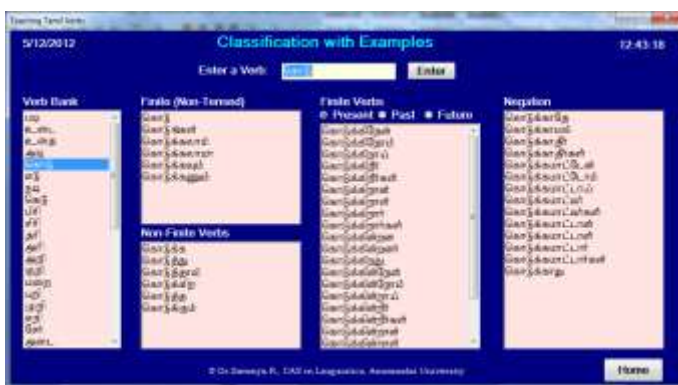
2. Concept of verbs

This part is divided into two sub-parts. The first one consists of the terms of the verbs. The second one contains the definition and the structure of the various verb terms given at the first part. This part describes the components of the verb, and then how the components are internally well organized as well as structured to form different forms of a single

verb. The terms which have to be defined are given in the left side. When the term is clicked, the definition, and its structure with example are displayed in the right side. The sample visual is as follows:



This part highlights what is the maximum quantity of the inflected forms of a single verb according to the above mentioned classification. The root verb bank is placed at the left side. While clicking any of the root verbs given in the bank, the maximum quantity of the inflected forms of that particular verb is displayed. It is shown in the following sample visual.



In this part, a few of the inflected forms are randomly given for identification of the term of form. If the verb form given in the list is clicked, multiple choices are displayed. When the correct answer is clicked, the mark is automatically scored. When the wrong answer is clicked, it shows the correct answer. This is simply for testing and evaluating the students or learners, and also for having feedback.

It has three states; the first state contains a few of the inflected forms which are randomly chosen from the database developed for this tool. When any of the forms is clicked, the choices are displayed. If the correct answer is clicked, the marks are scored. If the wrong answer is clicked, the statement 'It's wrong' is displayed with answer. The visuals of these three states are as follows:





Conclusion

This simple tool will definitely stimulate the cognitive power of students; and will be not only a virtual learning but also a teaching tool in promoting up a student's or learner's knowledge on Tamil verbs with their functions and structures.

References

1. Lehmann, T (1993) '*A Grammar of Modern Tamil*' Pondicherry Institute of Linguistics and Culture, Pondicheery
2. Saranya, R. & Mubarak Ali, A. (2007) '*Tamil Morphology*' (in Tamil) Annamalai University publication, Annamalaiagar.

Mobile Learning as support to Tamil Language Teaching

Dr. P. KUMARESAN

Dr. P. VIJAYA

Assistant Professors, CAS in Linguistics, Annamalai University

Introduction

One of the potential technologies deemed suitable to play a fruitful role in this regard airmobile ones. Portability and accessibility of mobile devices in this digital era have attracted many scholars to apply them in the educational settings. Furthermore, Several researchers have attempted to prove applicability of mobile learning as modern ways of teaching and learning (Naismith, 2004) . Moreover, applying portable technologies have been demanded by most of the modern learners who oftentimes are forced to study anywhere and anytime, for example, at work, in the bus or at weekends (Evans, 2008) .The wireless portable devices such as iPods, MP3 players, smart phones (like Blackberry, phone), and Personal Digital Assistants (PDAs) could provide opportunities to respond to the need of this generation. Evans (2008) believes that a distinguishable feature of mobile learning or M-learning is the potential to study when traveling on transport. Language learning is also expected to benefit from the extensive possession of mobile devices such as phones and media players (Kukulska-Hulme, 2006) . In addition to the above mentioned profits of M-learning for learners that can be attributed to TLT learners too, teachers may benefit from applying portable wireless devices through their teaching process. To increase access to authentic teaching and learning subjects, mobile technology can be used mainly by teachers even when they are t traveling to schools or arrange lesson plans.

Technology in Language Teaching and Learning

"Technology has accompanied the process of language teaching and learning for many years. Cassette players and television were traditional primitive technological tools, which were used in language classes as pedagogical aids. Today, when we think about technology, the first teaching aid that appears in our vision is computer. Likewise, in the field of Tamil language teaching, computer is a good teaching aid, especially since it has been complemented with the connection to the Internet. That is why a great deal of studies has been carried out to investigate the effect of computer-based or web-based language learning in the educational environments. Educators have also recognized the potential of technology as an instructional tool in language teaching and its application is increasing too. They asserted that technology is able to generate either collaborative or independent learning environment in which learners can practice and learn a new language. Combination of technology into language education has become a daily event, and the educational multimedia courseware is produced largely as reference subjects to promote Tamil language teaching and learning Moreover, usage of the Internet in language teaching has been considered a serious methodology. Hismanoglua (2010) emphasized that besides its technological function, the Internet can also be used as a pedagogical tool for improving language learning and teaching. The recent Internet- based technologies employed in foreign language instruction is Web 2.0 tools. The most common tools of Web 2.0 include wiki, blog, podcast, social network and video conferencing have demonstrated the capability of the current technology in language teaching and learning. Studies have revealed that wikis are useful tools for learning and teaching as they provide collaborative writing (Cress & Kimmerle, 2008) . Likewise, blogs or text formatted journal entries by users, can improve writing skills, promote active learning, and provide feedback for students and teachers (Alexander, 2006; Seitzinger, 2006) . Language learners normally use blogs in their classes to enhance both writing and reading skills (Sarica & Cavus, 2009) .Apart from wikis and Hogging, social networking is a good opportunity for language learners to improve their writing and reading ability especially when they type messages or read them (Sarica & Cavus, 2009) . Lam (2000) confirmed the potential of online messages to boost the writing ability of language learners and stated that online exchanges and discussions via the web and email messages may enhance the students' writing skill. Language teachers were the first to acknowledge the benefits of the application of social networking tools in language learning.

Mobile Learning

The creation of mobile devices such as cell phones, PDAs, iPods and pod cast has demonstrated that technology-based pedagogy is employed rather extensively in some academic environment. Although mobile learning is gradually being accepted in not many educational settings, its advantages cannot be overlooked. Mobile learning is defined as any service or facility that contributes to acquisition of knowledge regardless of time and location. According to Vavoula and Sharples (2002) learning can be considered mobile in three different contexts: learning is mobile in regard to space, it is mobile due to the different places, and it is mobile in terms of time. Hence, mobile learning system can deliver education to learners anytime and anywhere they need it. M-learning is limitless in terms of the content and geographical extent, so, this offers dispersed virtual classrooms accessible any time (Jalalyazdi, et al, 2009) . Another variety of M-learning which is applied exclusively for language learning is called Mobile assisted language learning (MALL) . Although, this is an illustration of technology-based language learning, it is different than computer assisted language learning (CALL) because it focuses on the "continuity or spontaneity of access and interaction across different contexts of use (Kukulska-Hulme, 2009, p. 162) . The novel model of mobile learning creates various learning environment since students can download applications synchronously or asynchronously. They can also access notifications, weekly activities, feedbacks, assignments, their courses, online libraries, grading reports and these have increased their interest in studies (Kristoffersen & Ljungberg, 1998) . Individual learners who are engaged in this type of learning can personalize their learning environments by deciding where and when to learn. Furthermore, to develop mobile learning activities, instructional designers should pay special attention in creating and managing the knowledge database such as the vocabulary databases, reading materials, and learning materials including audio or video files. In the meantime, accessibility and technical connection problems are the most important considerations (Park, 2011) . Chang (2010) claimed that mobile learning's an audio-based learning project that allows learners to participate in an asynchronous learning discussion on mobile devices instead of the text-based discussion. In other words, learners can download audio files recorded by their peers and listen to these recordings while on the move. Since multimedia message services (MMS), an evolutionary form of short message services (SMS), can send not only text but also graphics,

PDAs

Personal Digital assistants (PDAs) are pocket-sized computers that are expandable with some hardware components like keyboards and wireless networks and can be equipped with software programs such as word processors, flash-cards, databases, and bilingual dictionaries (Houser, et al., 2002) . Chinnery (2006) asserted that one of the primary roles of PDAs has been as a translator in a language-learning classroom. Other than that, software programs such as 'MobiLearn' have managed to convert PDAs into 'talking phrase books. In addition, various projects have been implemented for using PDAs in language learning environments.

Cell phones

Short Message Service (SMS), voice-messaging, cameras, video-recording and even Internet access for cell phone users are practical for language learning. Chinnery (2006) believed that all of these features allow language teachers to offer access to authentic content, communicative language practice, as well as completion of tasks to the students. A cell phone is the most popular and accessible mobile device in language learning as it is widely used by individuals regardless of their age and gender. Houser, et al. (2002) quoted the results of a study performed by Stanford Learning Lab on learning language via mobile phones. They provided some programs including translation of words and phrases, vocabulary practice, access to live talking tutors and quizzes. The findings of the study revealed that mobile phones were effective for quiz delivery carried out in small segments. It also concluded that quizzes and voice vocabulary lessons had great potential in the teaching and learning of language.

Conclusion

This paper has highlighted how to support a variety of mobile devices in the educational delivery, particularly in the field of TLT. Mobile devices like other technologies, at first appeared peculiar for pedagogical use but slowly, they have become a part of our life. Great changes in utilizing PDA, and cell phone for the teaching and learning of languages have proven the potential of mobile technologies. Furthermore, related literature has identified the adoption of this technology by language teachers. Portability and wide access to mobile phones have made it more popular in education.

A computer is perhaps more excellent than a mobile phone "for handling various types of information such as visual, sound, and textual information, but mobile phone is superior to a computer in portability (Yamaguchi, 2005) . Hence, the integration of mobile learning with Tamil teaching and learning may offer vast innovations in the coming days. Even though the utilization of mobile learning in TLT is not common in many countries, such educational setting seems a fashionable path in language learning. Mobile learning applications in language learning has its advantages and its potential should not be overlooked, for, the future holds great possibilities for this type of technological device for pedagogical use.

References

1. Alexander, B. (2006) . A new way of innovation for teaching and learning- *Educause Review*. 41 (2), 32-44.
2. Cress, U. & Kimmerfe, J. (2008) ., A systemic and cognitive view OB collaborative knowledge building with wikis,
3. Lee, M. & McLoughlin C. (2010) . Beyond distance and time constraints: Applying social networking toolsand Web 2.0 approaches in distance education. In G. Veletsianos (Ed.), *Emerging technologies in distance education* (pp). 61-87. Athabasca: Athabasca University Press.

Computer Assisted Tamil Language Teaching Tool

“Tamil Learning Mantra”

A. Reshma Parveen (reshmanasser@gmail.com)

Master of Computer Applications, N.G.M College of Arts and Science, Pollachi

M. Rajasenathipathi (senathi_pathi@yahoo.co.in)

Head, PG Department of Computer Applications, N.G.M College of Arts and Science, Pollachi

ABSTRACT

Language is the soul of communication and Computer is the heart of Technology. The borders of the world can be broken by combining the computer and language together using the help of computer assisted language teaching. The interest in learning a new language became less among the people in this world due to not awareness about that new language and no proper translation through a known language. Recent years, people has shown a boom of interest in using computer for language learning and teaching, but the roles of computer in teaching Indian languages is very less, that also its too less especially for Tamil language comparing to other Indian languages. In this paper we present the basic Tamil language teaching tool “TAMIL LEARNING MANTRA”. This Tamil teaching tool (TLM) is used to help the people across the foreign countries and also in other states of India to learn the basics of Tamil language. It is used as an aid to learn Tamil very easily and quickly for all categories of people.

INTRODUCTION

Tamil Language teaching tool is mainly used to perform a mode of teaching the Indian language Tamil to people of other states and countries around the world. It assists the people to learn the Tamil through the global language English. The teaching of language begins with the basic concepts such as, letters of the language from vowels, consonants and combined vowel consonants. The words which are commonly used in Tamil language are categorized into different groups and each word is pictorially represented for easy identification. The sentences are given as combination of basic phrases and words. It is simplified for better understandability. More interactive teaching techniques have been handled in this tool. it plays a major role in spreading the wings of Tamil language across worldwide.

EXISTING SYSTEM STUDY

The existing systems available for teaching Tamil are very less comparing to other languages in India. Among them few tools are available for Tamil conversion and typing alone. There are also some mobile applications available for Tamil learning. Many online websites are also helpful in teaching Tamil.

DRAWBACKS IN EXISTING SYSTEM

Offline systems

The offline tools are only capable of teaching the Tamil language generally without the deep knowledge of its concepts. These tools help people to learn Tamil language letters and words through the Tamil only, it does not use any other language to give its equivalent meaning. This does not help the people who are unaware of Tamil. The interface used is not very friendly for all categories of users. It just helps to navigate from first word/letter to last in sequence order, which cannot be flexible in picking a intermediate word/letter. The most important drawback is these tools do not help in providing the direction and method of writing the letters. They provide only the whole letter as it is. Only the basic letters are covered with limited examples. No steps taken to test the knowledge of the language for people who completed learning it. Some tools used to teach Tamil through other Indian languages which is useful only for a group of people belong to that particular language limit so it is also not useful worldwide. Only minimum numbers of examples are given for letters so the learner does not know the proper usage of letters in creating words. The another main drawback is the learner suffers in pronunciation of letters and words, which is very important to learn a language.

Online systems

These types of tools are only used to learn Tamil by online only. It needs an internet connection for learning. Some of the websites provide only lessons of Tamil language concepts rather than interactive teaching. Most of websites are under some universities and it is only adaptable for adults and not for children for learning. Although some websites have good teaching methodology, most of the times it generates error due to server problem in loading the animations and becomes very slow. Additional plug-in and add-ons are required to load the multimedia components in the web pages. Most of websites concentrate in teaching the stories and rhymes only, not on the language concepts.

4. PROPOSED SYSTEM

The propose Tamil teaching tool is used to teach the Tamil language from the basic level. It provides more flexibility to learners. This tool teaches the language to people with more examples and clear pictures for each. This is basic level teaching version which covers the basic letters, words and sentences. It concentrates on giving the best to learners by proper step by step process of teaching. The letters are taught at first and the words are developed by letters, then sentences are taught finally. It attracts the user with its colorful interface and animated components.

ADVANTAGES IN PROPOSED SYSTEM

This Tamil teaching tool is used to teach in attractive manner by using the writing methods. Each letter is animated to show the direction of writing it. This helps the learner to write the letter very easily. The pronunciation of the letters and words are given in English. Along with that, it provides the equivalent English word for the given Tamil word. The audio is played to pronounce the given word in Tamil as well as in English. This helps the users to speak the words easily. The learner has a facility to hover on the letter or word to hear the pronunciation of it. A huge variety of topic is covered in words such as animals, birds, colors, and all other normal things around us. The sentence teaching consists of the basic phrases, two words and three words sentence and finally simple sentences. The listening skill is enhanced by using the audio which can be heard by means of speaker or microphone to listen to the words and sentences. The writing skills are also enhanced by this tool. It provides a wide range of printable materials for letters, words and sentences in the form of pdfs. The learners can do writing practice by writing over the printed patterns of each letter. The Quiz section is another interesting part of this tool. It allows the learner to take a test in categories such as verbal quiz and image quiz. The verbal quiz has verbal question with multiple choices and the learner should select the correct answer. The image quiz consists of multiple choices as images and the correct image for given Tamil word should be selected by the learner. This tool provides a user friendly and interactive environment for learning. It overcomes all the drawbacks of existing system by providing audio for pronunciation problem, examples for easy learning and quiz for evaluating learner's knowledge. It also help in learning the writing methods which help learners to write by themselves without any help from others.

COMPONENTS DESCRIPTION

The Tamil teaching tool (TLM) consist of the components for teaching letters, words and sentences. It has additional components for enhancement of listening skills and writing practice. The description of each component is given as follows.

BASIC LETTERS

The component is used to teach the basic letters of the Tamil language. It consists of teaching the vowels, consonants and combination of vowels with consonants. The Tamil language consists of 12 vowels, 18 consonants and 216 vowel consonants. The component has letters with the example in Tamil and in English. The pronunciation of the letter and its example is displayed in English language. The learner can hover on the list of letters to display it animated writing method, example and pronunciation in both written and audio play back format. The learner can use the operational buttons to hear the audio of pronunciation, to print the material and to take writing practice from the single component. The GUI for teaching consonants is given in fig.1

BASIC WORDS

The words which are used commonly and the important for communication is categorized into animals, birds, food, fruits, vegetables and all basic categories. The learner can select any one category at a time and can learn the wide range of words in that particular category. The teaching is made interesting by using the colorful images related to the words and audio of the pronunciation. The word is written in Tamil as well as in English for easy understanding. The GUI for word categories is given in fig.2 and teaching under one category is shown in fig.3, 4.

USER INTERFACE DESIGN



Fig.1 GUI for Teaching Consonants



Fig.2 GUI for word categories



Fig.3 GUI for teaching in each category



Fig.4 GUI for teaching in each category

SENTENCES

The sentences are taught from the basic phrases of language and then towards the simple sentences. The basic phrases are covered from the common used phrases and relationship words. The two words sentences are formed from the simple words. The three word sentences are also similar to two word sentence. Each sentence has the audio format to give the pronunciation of it in Tamil. The GUI for sentence teaching is given in fig.5.

LISTENING COMPREHENSION

The listening is very important to carry out a proper communication with a language. The learner must have capability to listen to the pronunciation to improve the speaking skill of a language. This teaching tool for Tamil language provides a wide range of words and sentences to help the learners to enhance the listening skill. This tool gives a very user friendly interaction to listen the word and sentences just by hovering over it. The GUI for listening skill enhancement is given in fig.6.



Fig.5 GUI for teaching sentences Fig.6 GUI for Listening skill enhancement

WRITING COMPREHENSION

The writing practice is very essential one to learn a language completely. The Tamil learning mantra allows the learner not only learn using the application but also using the hardcopy. It has the Tamil language concept in the form of printable materials. All the letters such as vowels, consonants, vowel consonants, words and sentences are created in the form of handwriting material with the sample patterns. The material is in the form of pdf. The learner can just select the related concept and take printouts. Now the learner can refer the original letters and overwrite on shaded letters for improving the writing skill. The alternative method to make the children who learn Tamil is to make them interesting with the help of writing using mouse. The set of colors is given and one can select any color and use mouse to write over the given shaded letter in the screen. This makes children to learn more enthusiastically. The GUI of writing practice is shown in fig.7, 8.



Fig.7 GUI for writing practice Fig.8 GUI for printable pdf material

QUIZ FOR EVALUATION

The final component of this tool is the quiz which is used to evaluate the knowledge of Tamil language the learner gained from learning it using this tool. The quiz is separated into two type's namely verbal quiz and image quiz. The verbal quiz it deals with the verbal questions given in English and a related image is displayed for identification of the given question. The multiple choices of Tamil words are given. The learner has to choose the correct answer. The answer should be the Tamil word for given English word in question. This component is not only for evaluating the learner but also help them. When the learner select an answer from given choice it immediately intimate them whether it is correct or wrong. If the answer selected is wrong then it displays the correct answer near the intimation. This helps the learner to understand the meaning and correct their mistakes. The GUI for verbal quiz is given in fig.9.



Fig.9 GUI for verbal quiz Fig.10 GUI for image quiz

The image is quiz is similar to the verbal quiz but it has the difference in the choice. The question is given in the Tamil and three images are given as the multiple choice. The learner has to select the correct image for the given Tamil word in question. This also helps the learner when he/she select an answer from given choice it immediately intimate them whether it is correct or wrong. The GUI for image quiz is given in fig.10.

CONCLUSION AND FUTURE ENHANCEMENT

The Tamil teaching tool Tamil learning mantra is an interactive tool for learners to learn Tamil with the help of animation of writing method, voice playback for improving pronunciation and wide range of sample words. This will give a good experience for users to learn Tamil very quickly and easily with the Quality of the original Tamil. This tool is a basic version which covers only basics of Tamil language. In future it can be enhanced to teach the next levels of teaching Tamil such as complex sentence, passages, essays and conversations and finally the full fledged Tamil reading and speaking skills. It may also be used with machine learning and natural language processing to facilitate users to directly interact with computer for communication. This tool can also be enhanced to touch technology and used for the touch screen devices to make Tamil learning more effective and easy. Let this tool be one of the stepping stone for the oldest language in this world, to reach the people through the newest technology.

Developing Tamil Writing Skills: A Computational Approach

P. VIJAYA

P. KUMARESAN

Assistant Professors, CAS in Linguistics, Annamalai University

Introduction

Today, computers are extensively being utilized for natural language processing particularly for data storage, computation, and communication. It is a tool for solving a wide range of problems. It is obvious that students have difficulties in understanding the concepts of various functions of words at sentence level. The difficulties vary from individual to individual. In order to make teaching and learning process effective, it is essential to find out learning difficulties of students during instructions.

This paper provides a brief overview of how computers have been used and are being used for developing writing skills. It focuses not only on a technical description of hardware and software, but rather on the pedagogical questions that teachers have considered in using computers in the classroom for developing writing skills. There are four language skills, namely, Listening, Speaking, Reading and Writing. But this paper mainly concentrates on developing writing skill only.

Objectives

- To improve teaching and learning by using the computer as a tool that can enrich existing teaching practices and open up new ways of presenting linguistic information.
- To promote a self-determined learning instrument that can be used collectively at school, as well as individually at home.
- To increase motivation by supporting a decision making and problem solving activities.

Thus this paper aims to develop a computerized learning and teaching materials for improving five different writing skills as follows:

- Narrating*
- Describing*
- Interpreting*
- Form filling*
- Preparing applications and bio-data*

a. Narrating

We narrate ***past events***. This skill gives us the ways to share our experiences. This leads to the development of stories. By developing our ***narrating skill*** one can become a good ***story writer***. Most of the narrations are in past tense which indicates past events.

Activity 1

nii ippadippil evvaaRu ceerntaay enpataik kuuRa mudiyumaa?

Can you narrate, how did you join this course?

Or

unatu naNpan kalluuriyil evvaaRu ceerntaan? (maruttuvam, poRiyiyal, piRa kalluurikaL

How did your friend get admission to the college: Medical or Engineering or other colleges?

Help !

Write down first all the steps you have taken to get admission in simple sentences and then write connected account of the same.

Activity 2

(meelee kuuRiyataiyee) nii kalluriyil mihavum kasTappaTTEE ceeraveenTiyiruntatu enpatai iRantakaala etirmaRai
vinaikaLaip payanpaTuttik kuuRuka.

Can you convert your write-up into a difficult process of getting admission by using past negative constructions?

Activity 3

Dr. Rajkumarai viirappan kaTattiya nikaLcci allatu atu poonRa veeRu kuRRaviyal nikaLvinai vivarikka mudiyumaa?
Can you narrate the incident of abduction of Dr.Rajkumar by Veerappan or any other recent interesting incident that attracted you?

Help !

Collect the stories from newspapers and then try to narrate the incidents.

Here are a few questions you may try to answer them.

1. When did the abduction take place?
2. enkee? Where?
3. At what part of the time?
4. What happened immediately the next day?
5. What followed next?
6. Why did the Chief Minister meet?
7. What were the demands made?
8. What were the reactions followed then?
9. How about Dr.Rajkumar's life in the jungle?
10. What was Nakkeeran Gopal's role in this episode?
11. When was Dr. Rajkumar released?
12. How was he released?
13. What happened then?

Activity 4

vaaLkkaic caritam vivarittal. itu oru inRiyamaiyaata moLittiRan. nam ovvoruvariTattum, (ulaka aLavi) varalaaRU/ aracyal/
aRiviyal caarnta aRinartam vaaLkkaiyee vivarittukkuuRum paTTaRivu uLLatu. Anta aLaviRiku naam
vicuvaacamuLLavarkaLee. meeRkaNTa takavalkaLai evvaaRu ceekarikkiRoom?

Narrating life history. This is an important skill. Every one of us has this experience of narrating the life of some historical/political/scientific leaders of the world. How sincere we are? How do we collect the information?

Step 1

Collect all events in the life of that particular leader.

Step 2

Rearrange the events in the correct chronological order.

Step 3

Put your sentences together to make **connected accounts**. Convert your simple sentences into compound and complex ones using connectives, conjunctions, etc.

Step 4

Does your narration give a **complete outline** of the leader's life? If not, what do you require?

Step 5

Discuss with your friends.

Read their creations. Rewrite your own, if required.

Jawaharlal Nehru paRRi cila vaakkiyankaLai eLuta muTiyumaa?
Can you write a few sentences on the life of Jawaharlal Nehru?

Activity 5

toTar nikaLvukaLai kaaTTum paTankaLaip paarttu campantappaTTa katai/nikaLcciyee vivarittuk kuuRa mudiyumaa?
Can you narrate the story / incident with the given picture sequence? Picture enclosed

Help !

Name the boy. Before, narrating the incidents, describe each picture. (Picture is provided)

2. Describing

Describing simple tasks like preparing coffee/tea/sweets, cooking, making a telephone call, describing an action, etc. are very important for developing written communication skill. Sometimes, giving instructions to do certain things also lead to description. For example, in a chemistry practical class we learn and practice preparing some salt etc. for which instructions are always given.

Activity 6

teeniir/kaapi/ inippu tayaarippatai vivarittuk kuuRa mudiyumaa?

Can you describe 'how to prepare tea/coffee/sweets'?

Help !

1. Note down the different items necessary for the preparations.
2. Write down the steps – how it can be prepared.
3. Are there different steps or put everything together and prepare coffee/tea/sweets.
4. You may even discuss this with your friend/friends.

Activity 7

piRarukkuc ciRu veelaikoTuppatu enpatum atanai curukkamaay, teLivaay colvatenpatum oru moLittiRanee. unatu kuLuvinaruTan kalantu peeci kuRippeTuttukkonTu, iru cakkaravaNTiyeep paLutupaarkka viiTtin mun kuuTattaic cuttappaTutta piRarkku nii kuuRum kaTTaLaikaLai eLutuka.

Giving instructions precisely, clearly, is another skill that we have to develop.

Discuss now with your group and note down the steps that you have to take to give service to your scooter, or clear the hall of your house.

Activity 8

teecappaTam paarttup paTittaRital- taRaippaTattaik koNTu oru kuRippiTTa iTattai aTaivataRkaana vaLimuRaiyee naNpanukkuk kuuRutal

Map reading, giving instructions to your friends to reach a place using a map.

The road map of Chidambaram is given to you. You give instructions to your 3 friends to reach their destinations –

Friend 1 in the **railway station** wants to reach **Natarajar Temple**.

Friend 2 in the **Rajendran statues** wants to reach **Pichavaram**.

Friend 3 in the **Gandi status** wants to reach Dept. of **CAS in Linguistics**.

Help !

You may start like this

When you come out of the Railway station you will find a statues to your left and a diverson Annamalainagar Road; take the road when you start walking, to your right will be railway office and to your left will be a Rajendran statues so on.

Activity 9

Describing maps is an important task. Not only they give knowledge of the location, but also make you familiar with different places.

oree nilappakuticaarnta iru taraippaTankaLaip paarttu puRavaLiccaalai, vaLaiyappaatai aakiyavaRRai vivarikka muTiyumaa? avaRRiRkiTaiyee uLLa veeRupaaTukaL enna? ivaRRil etu ciRantatu?
Can you describe the two maps given to you on the same point, namely, outer ring Road? What are all the differences?
Which of the two is good?

Activity 10

When we describe a building, we normally do so in relation to the things near it, or the building/familiar to us.

The following words are often used.

in front of adjacent to there is opposite on the right of
there are by the side of across behind on the left of

meeRkaNTa coRkalaip poola meelum cila coRkaLaic ceerttu CAS in Linguistics kaTTiTavaLaakattai vivarikka.
You can add more such words and try to use them to describe the CAS in Linguistics building.

Activity 11

Like building, when you want to describe a house, how will you do that?

koTukkappaTTuLLa coRkkalaipayanpaTuttutal- meelum un conta coRkaLaic ceerttal- ataRku aLaveetum illai- un viiTTap
paRRi vivarikka.

Using the words given – add your own also, there is no limitation for that – describe your house.

Help !

Here is a list of words useful for you.

ground floor bed room garden dining room
first floor toilet compound wall floor area
second floor bath room hall/drawing puja room
stair case room balcony wardrobe
kitchen cupboard showcasesit-out
closet

Activity 12

Oru ceymuRai/ ceyalmuRai vivarippatu

How to describe the process?

aRiviyal caarnta ceymuraikalil ceyyappaaTTu vinai payanpaTuttappaTTa coRRoTarkaLaiyee kaaNkiRoom. yaar ceytaar
enpatai vita etu ceyyappaTTatu, evvaaRu ceyyappaTTaTu enpavaRRaiyee (ataavatu ceyapaTu porulaiyee)
mukkiyakaruttaaka ennumpootu ceyyappaaTTu vinaiyee payanpaTuttukiRoom.

In most of the scientific writings this is very common. In most of the Chemistry, Physics classes, one may notice the use of passive sentences. We may use passive sentences when we are interested in what is done, how it is done than who did it, in other words, it is impersonal.

Take for example

Almost all grains are cultivated in villages only.
The grains are procured from villages.
The grains are cleaned.
The grains are bundled.
The grains are stored in protected places.
The grains are distributed to the people through PDS.
The grains are purchased by people

Observe the sentences, which may be listed as below:

The grains -----

are cultivated

are procured
are cleaned
are bundled
are stored
are distributed
are purchases

All are passive sentences. We understand who does all these things: Yes! The Government. We do not bother much about who does it but bother about what is done. Whenever you describe processes and whenever you are concerned with the thing done, you can use passive sentences.

Observe the active-passive sentences.

- a) The people cultivate grains in the villages.
- b) Grains are cultivated by the people in the villages.

Note:

There are a few steps to transform *an active sentence into a passive*.

Step 1

The object of the active sentence becomes the subject of the passive sentence.

Step 2

The subject of active sentence becomes the object of the preposition “by”

Step 3

The verb is transformed as

“BE” + past participle form of the verb

(The verb has to agree with the new subject)

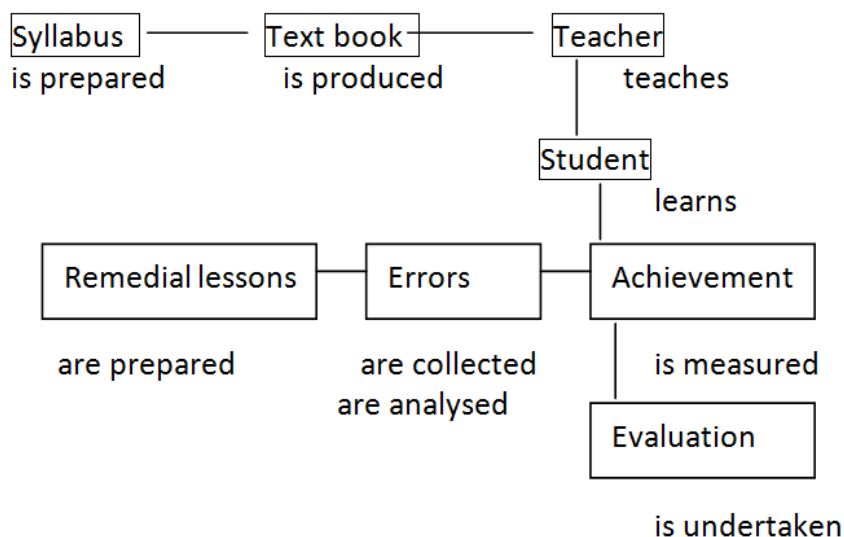
kiraamattu makkaLaal taaniyangkal kiraamangkaLil payiridappadukinRana.
Grains are cultivated by people in the villages.

Step 4

Delete the agent if not necessary.

Activity 13

Here is a flow chart of Teaching – learning process



kaRRaI- kaRpittal va_limuRai muLuvaakkiiyankalil vivarikkamuTiyumaa? inappuccoRkaLai, karuttut taTarcciyeyum maraveel.

Can you now describe the teaching- learning process in full sentences. Do not forget to use connectives & sequences.

Activity 14

Take any process of your choice and describe it.

Help !

- How is a T.V. serial made.
- How is water purified and destributed.
- How is milk supplied.
- How do letters reach the destination.
- Select a candidate for a post etc.

1)

Ettanai ceyappaaTTu vinai vaakkiiyankaLaip payanpaTuttamuTiyumaa attanai vaakkiiyankalaiyum eLutuka
Try to use as many passive sentences as possible.

2)

eLutap poovataic caarntu manattiraiyil oru kaaTciyee uruvaakkikkoLka. piRarkaruttaik keeTTu, naNparkaLitam kalantu
peeciyum pinnar eLutuka

Make a mental picture of what you are going to write. Take prompts and discuss with your friends and write.

Conclusion

The attempt made is very essential as well as important to note that 'linguistics' represents only one of the parameters and its use is indirect way of teaching language. It is supposed to give the language teachers, course designers, and producers of materials an inside view of the organization of language teaching and learning.

Reference

1. Aslam, Rekha. 1992 *Aspects of language teaching*. New Delhi.
2. Cook, Vivian, 1991. *Second language learning and language teaching*. London Arnold.
3. Ellis, Rod, 1994. *The study of second language acquisition*. Oxford.
4. Lado, R. 1964. *Language Testing*. Longmans.
5. Richards, J.C. (ed) 1977. *Error Analysis*. London: Longman.
6. Sam Mohan lal. 2000. *Aspects of Language Teaching*. Mysore: CIIL.
7. Ur, Penny, 1996. *A course in Language Teaching*, Cambridge.
8. Nadaraja pillai, N. (1986) *A Guide for Advanced Learners of Tamil*. Mysore: Central Institute of Indian Languages.
9. Wier, J. C. (1993) . *Understanding and developing language tests*. Eaglewood Cliffs, NJ: Prentice Hall.

ROLE OF ELECTRONIC DICTIONARIES IN TAMIL LANGUAGE TEACHING AND LEARNING

S. Raja

Annamalai University

Introduction

Lexicography is the discipline concerned with the principles and methods of preparing dictionaries. It is an applied field, whose theoretical background is partly utilized by linguistics. Almost all the branches of linguistics provide information in making or writing a dictionary. Dictionary is a simple tool which helps us to pronounce, respell and check grammar. It gives meanings to words which cannot be understood or words that have more than one meaning. Besides meaning, it also provides information on the reader syllables, intonation and pronunciation of words. Dictionaries are considered practical sources of information that learners and teachers can depend on. Dictionaries could take different formats like traditional paper dictionaries, handheld electronic dictionaries, CD ROM dictionaries, online dictionaries depending on the purpose of their required usage. Dictionaries have become very important tools in language learning especially the new forms of electronic dictionaries. A good dictionary and dictionary skills are the pillars of language learning. Several technological developments have led to the invention of electronic dictionaries in the operational processes of teaching and learning in TTL. Careful consideration should be paid to the rules and precautions of using them and how to avoid any possible drawbacks. Teachers need to play a considerable role in facilitating dictionary usage in class activities to enhance the teaching/learning process.

Electronic Dictionary

The term electronic dictionary can be used to refer to any reference material stored in electronic form that gives information about the spelling, meaning, or use of words. Thus a spell-checker in a word-processing program, a device that scans and translates printed words, a glossary for on-line teaching materials, or an electronic version of a respected hard-copy dictionary are all electronic dictionaries characterized by the same system of storage and retrieval. It is the retrieval system, rather than the information content, which makes electronic dictionary use such a revolutionary experience compared to the consultation of a hard-copy dictionary. Books organize information in a primarily linear way, which is appropriate for the listing of a succession of separate entries, but inadequate as a means of grouping and regrouping words according to their semantic and pragmatic similarities, their valency and collocational patterning, or simply their letter combinations. The A-Z sequence places headwords in an order which is virtually meaningless, shedding no light on the relationships between words that are alphabetically distant, and complicating searches for phrases and idioms. A text search of an electronic dictionary, on the other hand, provides almost instant access to groups of words in any information field created during its development.

Thus the Oxford English Dictionary on CD-ROM can act, among other things, as a synonym finder, a dictionary of quotations, or a terminology bank, simply through the rearrangement of the information which is provided in the original paper-based text. Similarly, several learners' dictionaries in electronic form offer filtered searches through examples, definitions and grammar codes, while Word Net, a dictionary purpose-built for electronic retrieval, offers a full range of links from the search word to its synonyms, co-ordinate terms, hypernyms and hyponyms. Pocket electronic dictionaries permit poor spellers to type in a "sound alike" version of their search word and select from a range of near matches the dictionary suggests, while cd-rom dictionary users can "hyperlink" and "interface" to call up entries for unknown words within the entry they are consulting, or in an unrelated on-line text (Nesi, 1999) .

As electronic dictionaries become more widely used, and users become more discerning, lexicographers and dictionary designers are refining those features that take best advantage of the medium, and are producing more lexicographical material for the Internet and cd-rom, without any hardcopy equivalent. Field searches, too complex for all but the most skilful of first generation electronic dictionary users (Nesi 1996), now tend to be much simpler to conduct; one click from an entry in the Longman Web Dictionary, for example, leads straight to all "related entries" (those containing the search word), while Collins E-Dictionary (1998) offers easily activated searches through words with similar grammatical behaviour, but has dispensed with the Word Bank and searches of semantic relations (antonym, synonym, hyponym and

superordinate) offered by Collins Cobuild on CD-ROM (1995) .

Skills of the Electronic Dictionary users

The first, and most fundamental dictionary skill is that of selecting an appropriate dictionary for a given task, and in this respect the advent of electronic dictionaries has placed much greater demands on dictionary users. Different dictionary formats suit different language activities, so now the choice is not just between a learners' dictionary or a native speakers' dictionary, general or specialist, bi- or monolingual, but also between a dictionary that requires access to a networked computer, a dictionary with links to a word-processing program, or a portable dictionary to consult while speaking and listening, or while reading hardcopy text. Nevertheless it is not hard to find comparative information about hardcopy dictionaries, as they are well known to teachers, reviewed in teaching journals, and available for inspection in bookshops and libraries. In contrast there are very few sources of unbiased advice available to the language learner who needs to select the right electronic dictionary for the job.

Dictionary skills also involve an understanding of the micro and macrostructure of the dictionary, its cross-referencing system and the contents of appendices and study pages. This kind of knowledge is particularly difficult for an electronic dictionary user to acquire because only one screen page can be viewed at a time, and the spread of contents can only be visualized, rather than viewed and handled as with a hardcopy book. The ability to identify the appropriate form of the word may not matter much either, as many electronic dictionaries accept "searches by letter" (type in the first letter (s), and select from a list of lemmas), or "sound alike" searches, and some automatically display the lemma of whatever inflected form the user types in. Instead, learners need to acquire the skills to use these dictionary functions correctly, as they are not always obvious.

Electronic Dictionaries are user-friendly, very fast and easier to carry. Students relate better to it and are more enthusiastic and sounds can be heard. They are better for class based activities, while paper dictionaries would be better for homework based tasks. When learners use an electronic dictionary their word search becomes faster. This allows learners to spend more time on reading comprehension or on searching word meanings.

A comparison between electronic and paper dictionary revealed that a lot of differences exist between them in some features. The vocabulary items covered in the electronic dictionary exceeds the limited number in the paper one. Some features do not exist in the paper dictionary such as databank, reference book, voice recorder, calculator, MP3. In addition, it does not include speech features, updates and interactive learning functions like irregular verbs, idioms, dialogues, sentence structure, accent correction, and grammar explanations. Obviously, the electronic dictionary is faster, lighter and more mobile than the paper one.

Learner's Dictionary

Dictionaries for foreign students are usually called Learner's Dictionaries or MLD (for Monolingual Learners Dictionaries) . Their most interesting features are their efforts to develop new ways of defining words and to provide information necessary for encoding. It is the latter feature that explains most of their characteristics as given below:

1. The language described is the usual everyday variety, not the formal or literary;
2. Preference is given to the quality of the treatment of words rather than to the quantity of words;
3. Particular care is taken in the treatment of the more or frequent words;
4. The obligatory or preferred syntactic patterns of verbs (at least) and sometimes nouns and adjectives, the countability or uncountability of nouns, the irregular forms of plurals, comparatives, conjugations, etc., are all indicated;
5. Pronunciation is indicated by means of the IPA;
6. Many examples, sometimes authentic but always representative of normal, everyday usage, are given.

Finally learner's dictionaries are also innovative in the sense that they have extensive introduction on "How to use the dictionary".

While compiling a learner's dictionary, three factors viz. user's linguistic difficulties, user's reference needs and user's reference skills should be taken into consideration. The first factor is the relative learning difficulty of various kinds of lexical and grammatical items in English. It is commonly agreed, that even advanced students have great difficulty in understanding, and in particular, have difficulty in using high-frequency words, because of the multiple meanings, derivatives, compounds and idioms which they give rise to. This is the area to which the lexicographer's principal efforts must be directed. The second factor is the learner's reference needs. Some learners may be more concerned with interpretation than production (i.e) they may use the dictionary primarily for quick retrieval of individual items or meanings when reading. The third factor is readers' reference skill, which the user is assumed to possess or can be expected to acquire, in handling a dictionary and making effective use of the information it contains.

Electronic and Online Dictionaries

Several websites are available for Tamil Dictionaries which can be utilized for teaching learning purpose. For example, www.sol.com.sg/classroom/dictionaries/html is a website for accessing the Tamil lexicon. Another website for an online web based English Tamil dictionary is www.geocities.com/Athens/Acropolis/8780/. The University of Chicago has developed a website to access the Cre-A: Dictionary of Contemporary Tamil (on line version) - www.lib.Uchicago.edu/LibInfo/Subjects/SouthAsia/. The National University of Singapore has developed a Tamil English Dictionary which is available online in the website www.irdu.nus.edu.sg/tamilweb/. Another website for Tamil- English, English-Tamil and Tamil-Tamil Dictionary www.murasu.com/akaram will be useful for learning Tamil language through online. Pocket Electronic dictionaries along with text translators like *T-9 Tamil*, *M-815* and *M-833* which displays Tamil in both English phonetic and Tamil characters are available. Apart from Tamil they also include 12 and 32 language translations. They include over 25000 translations which will be useful for teaching and learning Tamil language.

In the electronic dictionary, the given English word or sentences will be translated into Tamil along with the pronunciation through an audio mode. For example, the sentence *I missed my train, can you issue another ticket instead of this one?* has been translated as *naan enatu rayilai tavaravituviteen, inta payanacciTikku patil veRu payanaciTTu taruveerkaLaa?*

Phrasal Verbs

Phrasal Verb is a multi – word expression comprising a verb and one or more particles. (Svensen 1993) . A Phrasal verb consists of a “root verb” such as *go*, *put* or *set* and a “particle” (an adverb or preposition such as *away*, *on* or *out*) . Some phrasal verbs are used in idioms or other fixed expressions. Many phrasal verbs have more than one meaning. Every meaning of a phrasal verb has grammar patterns that show whether the verb takes an object, whether the object is a person or thing and where that object can go. Transitive phrasal verbs are those where the object or pronoun can go between the verb and particle or after the particle, while intransitive phrasal verbs are those verbs that never take an object.

Phrasal Verbs Dictionary

It is a dictionary that has been specially designed to help learners of English deal confidently with phrasal verbs. The dictionary provides all the information that is needed to understand phrasal verbs and to use them well. Learners often think that phrasal verbs are difficult; some students try to avoid using them altogether. But phrasal verbs are not only a central feature of the English language, they are also extremely useful. Expert speakers use phrasal verbs in all kinds of contexts-not just in informal situations such as conversations or emails, but quite often in formal and technical writing too. There are many contexts where a phrasal verb is simply the best, most natural-sounding way of expressing an idea, and so students should be encouraged to use them.

Comparison of Phrasal Verbs in dictionaries

The Phrasal Verbs of most frequently used lexical items such as *bring*, *send*, and *take* given in three commonly used learners dictionaries such as English-Tamil- Dictionary (1963) and Pals dictionary - English - English - Tamil (1994) are discussed in detail. The micro and macro structure of these dictionaries have been compared and their structural developments are also discussed.

Sustainability of Teaching Information and Communication Technology (ICT) Skills in Malaysia Tamil Schools using Tamil Open Source Computing

Gunasegaran Kandaswamy (kkvguna@yahoo.co.uk)

C.M. Elantamil

Program Titian Digital, Hindu Youth Organisation Port Klang, Malaysian Community Education Foundation, Malaysia

Program Titian Digital – *Digital Bridge Programme*

Introduction

Information and Communication Technologies (ICT) provide a great development opportunity by contributing to information dissemination, providing an array of communication capabilities, increasing access to technology and knowledge and making teaching and learning more interactive and effective.

However, access to and the cost of ICT continues to be a major obstacle in schools, particularly in the developing countries. In Malaysia, most of the Tamil schools are partially government aided schools and they are located at rural and peri-urban areas, where these areas are home to a great majority of underserved communities. Here, the problems are even bleaker, where the infrastructure is either old or non-existent.

As ICT becoming the crucial means for improved economic conditions, a holistic ICT programme is most needed particularly in tackling those young children who are neglected by macro policy who reside mostly at rural and peri-urban areas. Understanding the need, the **Hindu Youth Organisation Port Klang, Malaysia (HYO Port Klang) & Malaysian Community Education Foundation launched the Program Titian Digital (PTD)** in April 2009 by setting up sustainable ICT Labs in Tamil Schools, using open source computing in Tamil and thin client technology which are reliable and cost effective.

Implementation

(i) Identify potential participating schools or receive applications from potential schools.

- after identifying or receiving applications, meetings are held with the School Administration, Parent Teacher Association (PTA), School Board of Governors (SBG) to explain the guidelines and conditions to be followed by the school.

(ii) Setup ICT Labs

- Labs are set according to the highest number of students in a class. The lab with tiles flooring includes server, terminal, networking, LCD projector, screen, internet service, tables, chairs & air-condition.

(iii) Formation of ICT Development Committee

- This Committee consists of Headmaster, Senior assistance of Curriculum, ICT Coordinator, PTA Representative, SBG Representative and PTD Representative. They will ensure the successful implementation of the programme.

(iv) Conduct ICT Classes

- This classes are taught by a tutor employed by the PTA. Students will be attending the ICT Literacy classes minimum once a week (1 hour) . This lab is also used to conduct ICT based teaching (eg: teaching of Science)

(v) Training of Tutors

- Tutors attend compulsory training once a month. The training includes Module Content, Pedagogy and Technical.

(vi) Mentoring of Tutors

- Tutors will be observed and guided regularly by the school administration and every 2 months once by pedagogy experts.

(vii) Development of Curriculum, Tutor's Guide & Students Module

- A team of experts consist of IT experts, Senior ICT tutors, Educationist and Textbook Officers are involve in the above production.

Besides the above tasks, in order to ensure sustainability of the Programme, the PTD team also carries out the following:

1. Quarterly meeting with the ICT Coordinators to discuss on the performance of each ICT Lab & issues related to the implementation of the programme.
2. Annual centralise ICT Development Committee meeting. This meeting is to share experience, issues and give feedback and suggestions for the development of PTD.
3. Conduct hardware auditing quarterly to ensure the labs are functioning well.
4. Conduct annual accounts auditing to ensure the financial account are kept well.
5. Ensure that the PTA is financially sound. This is to make sure that the PTA is able to pay the salary of the tutor, maintenance are done regularly and upgradements are done when necessary.
6. Empower the local community to support the school to sustain the ICT Labs.

Program Titian Digital is proud to announce that it has unleash the ICT potential of 23, 000 students that have undergone a minimum of an hour ICT training every week besides teaching and learning via ICT at 60 Program Titian Digital ICT Labs throughout West coast of Peninsular Malaysia presently. The numbers are expected to reach over 35, 000 by end of 2013.

Note: There are 523 Tamil Schools (of which 371 are partially government aided schools) with 103, 000 students in Malaysia

Keeping cost low while maintaining cutting edge performance, the thin client systems at Program Titian Digital ICT labs are capable of providing affordable server-based open source computing solutions. ***Program Titian Digital holds The National Record in Malaysian book of Record as Largest Open Source User.***

Program Titian Digital works with subject experts to develop ICT teaching guide and student guide in Tamil. We have successfully published level 1 & 2 Student module (guide book) while work in progress on level 3 guide book.

Students undergoing PTD-ICT literacy programme will be evaluated at end of each level and a certificate of achievement awarded jointly by University of Science Malaysia - National Advanced IPv6 (USM NaV6) Centre, a premier centre in the area of Next Generation Internet.

Join us to reach out to the remaining **300** over Partially Aided Tamil schools.

மாற்றுத் திறனாளிகளுக்கான மின் உள்ளடக்கங்கள்

முனைவர் ப. டேவிட் பிரபாகர்

சென்னைக் கிறித்தவக் கல்லூரி
tamilprofessor@gmail.com

மரபு சார்ந்த கற்றல் - கற்பித்தல் முறைகளில் கல்வித் தொழில்நுட்பம் புதுமைகளை ஏற்படுத்தியுள்ளது . இந்நிலையில், மாற்றுத் திறனாளிகளையும் உள்படுத்தி அவர்களும் எளிதாக அணுகிப் பயன்கொள்ளும்வண்ணம் மின் உள்ளடக்கங்களை உருவாக்கவேண்டியுள்ளது . தமிழகச் சூழலில் , உயர்கல்வி பெற விரும்பும் பார்வை குறைபாடு கொண்ட மாற்றுத் திறனாளிகள் பலரும் தமிழ் , ஆங்கிலம், வரலாறு, தத்துவம் போன்ற குறிப்பிட்ட துறைகளையே தெரிவு செய்கின்றனர் . இத்தகைய துறைகளுக்கான மின் உள்ளடக்கங்களைத் திட்டமிட்டு உருவாக்குவது உடனடி தேவையாகும் .

எழுத்துக் கோப்புகளாக அமையும் உள்ளடக்கங்களை அணுகுவதில் , திரை வாசிப்பான், திரை உருப்பெருக்கி , சிறப்பு விசைப்பலகைகள் போன்ற துணைத் தொழில்நுட்பங்கள் பயன்படுகின்றன . ஒலி / ஒளிக் கோப்புகள் , படங்கள், அட்டவணைகள் போன்றவற்றையும் மின் உள்ளடக்கங்கள் கொண்டிருப்பதால் , அவற்றை எளிதில் அணுகுவதற்கான தொழில்நுட்பங்களைப் பயன்கொள்ள வேண்டியுள்ளது . மேலும், இணைய வழி உள்ளடக்கங்களை உருவாக்கும்போது பின்பற்ற வேண்டிய உலகளாவிய தரக்கட்டுப்பாட்டு வழிகாட்டு நெறிகளையும் (WCAG 2.0, BS 8878: 2010) கருத்தில் கொள்ளவேண்டியுள்ளது .

இணையப் பக்கங்களின் வடிவமைப்புகள் எளிதில் அணுகும் வண்ணமும் ஊட்டாட்டத் தன்மைகொண்டதாகவும் அமைதல் வேண்டும் . மேலும், மொழி சாராத கூறுகளை அணுகுவதற்கு வசதியாக தலைப்புகள் , விளக்கங்கள் இடம் பெறுதல் வேண்டும் . தாவிச் செல்லல் , இணைப்புகளுக்குச் செல்லல் போன்றவையும் மாற்றுத்திறனாளிகளைக் கவனத்தில் கொண்டதாக அமைதல் வேண்டும் . மூடில் போன்ற கற்றல் மேலாண்மை சார்ந்த மென்பொருள்களையும் பயன்படுத்தும்முறையிலும் மின் உள்ளடக்கங்களை உருவாக்க திட்டமிட வேண்டும் .

இக்கட்டுரை, தமிழகப் பல்கலைக்கழகங்கள் மற்றும் கல்லூரிகளில் வழங்கப்படும் தமிழ்க் கல்விக்குத் தேவையான மின் உள்ளடக்கங்களை மாற்றுத் திறனாளிகளும் அணுகும் வகையில் உருவாக்கவேண்டிய தேவைகள் , முறைகள், அணுகுமுறைகள், தொழில்நுட்பங்கள் ஆகியவற்றை விவாதிக்கிறது .

இளங்கலைத் தமிழ்ப்பாடத்தில் இலக்கியம் , இலக்கணம் ஆகியன முதன்மைப் பாடங்களாக விளங்குகின்றன . இலக்கியத் தாள்கள் பெரும்பான்மையும் செய்யுள் தொகுப்பாக அமைபவை . புனைகதை, தன்வரலாறு, கடிதம் போன்ற உரைநடை இலக்கிய வகைகள் , தினாய்வு, படைப்புக் கலை, இதழியல், தமிழக வரலாறும் பண்பாடும், தமிழ் இலக்கிய வரலாறு முதலிய தாள்கள் உரைநடையில் அமைகின்றன . இவற்றோடு மொழியியல் , நாட்டுப்புறவியல் போன்ற தாள்களும் முதுகலைப் பாடத்திட்டத்தில் இடம் பெறுகின்றன .

மேற்கூறிய பாடங்களுக்கான மின் உள்ளடக்கங்களை மாற்றுத்திறனாளிகளும் பயன்கொள்ளத்தக்கதாக வடிவமைப்பதில் அதிக சிக்கல்கள் எழ வாய்ப்பில்லை . ஏனெனில், இத்தகைய பாடங்களில் எழுத்துப் பனுவல் அல்லாத பிற ஊடகங்களைத் தவிர்ப்பதன் மூலம் பாடங்களை எளிதில் அணுகும்வண்ணம் உருவாக்கமுடியும் . அட்டவணை, கிளைப்படங்கள் ஆகியவற்றுக்குத் தலைப்புகள் , விளக்கங்கள் ஆகியவற்றை இணைப்பதன் வாயிலாக கருத்துத்தொடர்பை ஏற்படுத்த இயலும் .

முகப்புப் பக்கத்தில் , உள்ளே இடம்பெறும் தாள்கள் , பாடப்பகுதிகள், பக்க விவரங்களைத் தெளிவாகவும் முறைப்படுத்தியும் குறிப்பிடல் வேண்டும் . மாற்றுத்திறனாளிகளுக்காக உருவாக்கப்படும் பக்க வடிவமைப்பில் சிறப்புக் கவனத்தை மேற்கொள்ளவேண்டியுள்ளது . திரையில் உள்ள பல்வேறு கூறுகளின் இடம் தெளிவாக வரையறுக்கப்பட வேண்டும் . பி. டி. எஃ. கோப்புகளைத் திரை வாசிப்பான் படித்துக்காட்டாது . உலகளாவிய இணைய வழிகாட்டுக் குறிப்புகளுக்கேற்ப , தரத்துடன் உள்ளடக்கங்களை உருவாக்க வேண்டியுள்ளது . பக்க வடிவமைப்பில் எழுத்துப் பனுவல்களுக்கு மட்டுமின்றி , பக்க வடிவமைப்பில் இடம் பெறும் பிற கூறுகள் அனைத்திற்கும் வழிகாட்டுக் குறிப்புகள் வழங்கப்படவேண்டும் . கட்டளைகள், இணைப்புகள், பொத்தான்கள் போன்றவற்றை இவ்வகையில் தெளிவாக அமைத்தல் வேண்டும் . பக்க அமைவில் உள்கூறுகளை எல்லா பக்கங்களிலும் ஒத்த இடத்தில் அமைப்பது கற்றலை எளிமைப்படுத்தும் .

செய்யுள் பகுதியாயின் செய்யுள் பகுதி , அதன் ஆசிரியர் பற்றிய குறிப்பு , நூல் பற்றிய குறிப்பு , அருஞ்சொற்பொருள், இலக்கணக் குறிப்பு , பா வகை பற்றிய குறிப்பு , சங்க இலக்கியமாக இருப்பின் திணை , துறை பற்றிய குறிப்பு , பயிற்சி வினாக்கள் போன்றவற்றைக் குறிப்பிட்ட இடத்தில் முறையாக அமைத்தல் கற்றலில் தடுமாற்றத்தைத் தவிர்க்கும் .

மேற்கூட்டிய குறிப்புகளுடன் திரை வாசிப்பான் , திரை உருப்பெருக்கி , சிறப்பு விசைப்பலகைகள் போன்ற துணைத் தொழில்நுட்பங்கள் போன்ற தொழில்நுட்பங்களைப் பயன்படுத்தத்தக்கதாக மின் உள்ளடக்கங்களை உருவாக்குவதன் மூலம் மாற்றுத்திறனாளிகளின் கற்றலை எளிமைப்படுத்த இயலும் .

21 நூற்றாண்டில் புதிய ஊடகம் வழி தமிழ் - புத்தாக்க வளர்ச்சி

சி.குணசேகரன்,

புதிய ஊடகப் பயிற்றுவிப்பாளர், மென்பொருள் உருவாக்குநர், சிங்கப்பூர்
powerguna@gmail.com

நீண்ட நெடிய பாரம்பரியமும் தொன்மைச்சிறப்பும் மொழிவளமும் இலக்கியவளமும் கொண்டு சீரிளமைத் திறத்தோடு உயர்தனிச்செம்மொழியாய் உலகை வலம் வந்துகொண்டிருப்பது நம் தமிழ்மொழி . காலம் காலமாகப் பல்வேறு மாற்றங்களை அடைந்து இன்று தொழில்நுட்பத்தின் வழியாகத் தமிழை இனிமையாகவும் எளிமையாகவும் கற்றுக்கொடுப்பதற்குரிய சூழல் உருவாகி உள்ளது . சிங்கப்பூர் போன்ற புலம் பெயர்ந்த நாடுகளில் தமிழ் மொழி இரண்டாம் மொழிக்கு மூலில் கற்பிக்கப்பட்டாலும் தகவல் தொழில் நுட்பத்தைப் பயன்படுத்துவதில் சிறந்து விளங்குகின்ற நாடாகத் திகழ்ந்து கொண்டிருக்கிறது .

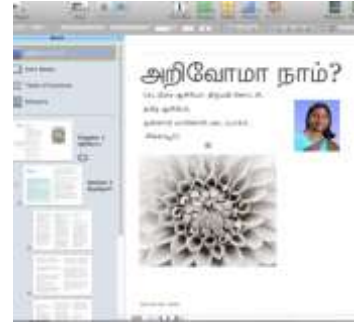
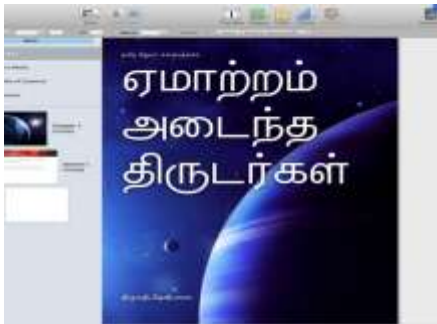
பன்மொழிச் சூழல் வழங்கும் சிங்கப்பூரில் எதிர் வீட்டிலும் அண்டை வீட்டிலும் சீனம் மற்றும் மலாய் மொழி பேசுவோர் வாழும் சூழலில் குழந்தைகள் இயல்பாகத் தமிழ் மொழியைக் கற்றுக்கொள்ள முடியாது . வகுப்பறைச் சூழலிலும் குறைவான மணி நேரமே தமிழைக் கற்றுக் கொள்வதால் அதனைச் சரளமாகப் பயன்படுத்துவதற்கு வாய்ப்புகள் குறைவு.

இப்படிப்பட்ட சூழலில் தகவல் தொழில்நுட்பத்தின் மூலம் பல்வேறு அணுகுமுறைகளைப் பயன்படுத்திக் கற்பிக்கும் பொழுது தமிழ் மாணவர்கள் மிகுந்த ஆர்வத்துடன் தமிழ் மொழியைக் கற்கின்றனர் என்பது ஆய்வின் வழி அறியப்படுகிறது . தொழில் நுட்பங்கள் பல்கிப்பெருகிக் கொண்டிருந்தாலும் நான்கு வகையான தொழில் நுட்பங்களைத் தேர்வு செய்து அவை 21 நூற்றாண்டில் தமிழ் வளர்ச்சிக்கு எவ்வகையில் உறுதுணையாக இருக்கும் என்பதை விளக்குகிறது இந்த ஆய்வுக்கட்டுரை .

நான்கு வகைத் தொழில் நுட்பங்கள்

ஒரு மொழியை அனைவரும் சிறப்பாகக் கற்றுக்கொள்ள தொழில் நுட்பம் பல்லாற்றானும் பயன்பட்டு வருகிறது . தொழில் நுட்பங்களின் எண்ணிக்கையும் நாளுக்கு நாள் உயர்ந்து கொண்டே செல்கிறது . அவற்றுள் குறிப்பிடத்தக்கனவாக விளங்குவது:

1. மின்னூல் வழிக் கற்றல் (Learn Through E-Book)
2. குறும்படம் வழி கற்றல் (Learn Through Digital Video)
3. கைத்தொலைபேசி வழி கற்றல் (Learn Through Smartphone)
4. மின்வழிக் கற்றல் (Learn Through LMS E-Learning)



மின்னூல் வழிக் கற்றல் (Learn Through E-Book)

காலம் காலமாகக் கல்வெட்டுக்களிலும் ஓலைச்சுவடிகளிலும் காகிதங்களிலும் நூல்களிலும் தமிழ்மொழியைக் கற்று வந்த காலம் மாறி இன்று அறிவியலின் அதீத வளர்ச்சியில் மின்னூல் உருவாக்கப் பட்டிருக்கிறது .

கணினி, மடிக் கணினி, தொட்டுணர் கருவி (ipad) போன்றவற்றில் தமிழ் நூல்களைப் பதிவேற்றம் செய்து , இருந்த இடத்தில் இருந்துகொண்டு வாசிப்பதற்கு ஏற்புடையதாக அமைவதே மின்னூல் எனப்படும் .

இன்றைய சூழலில் கணினி நம் அன்றாட வாழ்வில் இன்றியமையாத ஒன்று . அதன் வழி பதிவேற்றம் காணும் மின்னூல்கள், நூலகம் சென்று புத்தகம் தேடும் நேரத்தைக் குறைத்து கற்கும் வாய்ப்புகளை அதிகரிக்கின்றன . தமிழில் உள்ள பல்வேறு வகையான நூல்களை நினைத்த நேரத்தில் கற்பதற்கும் , கற்கும் போது ஏற்படும் ஐயப்பாடுகளை உடனுக்குடன் தீர்ப்பதற்கும் , நூலில் உள்ள கருத்துக்களைப் பற்றிக் கலந்துரையாடுவதற்கும் உரிய வழி முறைகள் மின்னூலில் அதிகமாக உள்ளன . அனைவரின் விருப்பத்துக்கு ஏற்ப நூல்களை தேர்வு செய்து கற்பதற்குரிய வாய்ப்பு இருப்பதால் அவர்களுக்குத் தமிழ் மொழியைக் கற்க வேண்டும் என்ற ஆர்வம் அதிகரிக்கிறது. மேலும், தனக்குள் மறைந்து இருக்கும் எழுத்துப் படைப்புக்களை உலகோருக்கு எடுத்து இயம்பும் வாய்ப்பும் நமக்கு ஏற்பட்டுள்ளது .பொருளாதார ரீதியாக தரமானத் தமிழ்ப் படைப்புகள் படிக்கும் வாசகர்களை உடனுக்குடன் சென்றடைவதுடன் , அதனைச் சார்ந்த நல்ல கருத்துகளும் , இணைய கருத்துப் பரிமாற்றங்களும் தரமான தமிழ் மின்னூல்கள் படைக்க அபரீதமான வாய்ப்புக்களை ஏற்படுத்தியுள்ளது .

மின்னூல்வழி கற்பதால் பல புத்தாக்கச்சிந்தனைகளும் மாற்றங்களும் உருவாகும் வாய்ப்புகள் அதிகரித்துள்ளன . பசுமை நிறைந்த உலகை உருவாக்குவதற்குக் கணினி சார்ந்த பயன்பாடு மூலம் காகிதங்களின் தேவை பெரிதும் குறைவதால், மரங்களின் ஆயுள்காலம் அதிகரிக்கப்பட்டு வருங்காலச் சமுதாயம் செழுமையுடன் வாழ வாய்ப்பு ஏற்பட்டுள்ளது. தெளிவான எழுத்துகள் , உடனுக்குடன் மின்னூல்கில் உலா வருவதன் மூலம் அதிகமான செய்திகளைக் கற்று அறியும் வாய்ப்பு, நிறைய நூல்களைத் திரட்டிப் படிக்கும் எளிய முறை போன்றவை , மின்னூல் வசம் நம்மைச் சுண்டி இழுக்கும் சக்திகளாக மாறி உள்ளன .

காணொளி, இசை, வண்ணமயமான விளக்கப்படங்கள் , இணையப்பக்கம்வழி இருவழி தகவல் பறிமாற்றங்கள் மூலம் தமிழில் எழுதவும், படிக்கவும் எண்ணற்ற வாய்ப்புகளை மின்புத்தகங்கள் வழங்குகின்றன .

குறும்படம் வழி கற்றல் (Learn Through Digital Video)



தமிழ் கற்றல் கற்பித்தல் மேம்பாட்டிற்குப் பெரிதும் பயனுள்ளத்தாக அமையும் தொழில்நுட்பங்களில் , மின்வழி தயாரிக்கும் குறும்படமும் ஒன்று .

காலம் காலமாக கற்ற பாடங்களைச் சோதனை செய்வதற்கு , தேர்வு முறை ஒப்படைப்பு , வினாடிவினா போன்ற முறைகளையே பின்பற்றிக்கொண்டிருக்கிறோம். அதிவேகமாக வளர்ந்து வரும் அறிவியல் தொழில் நுட்பம் , கற்றல் சோதனைகளை மாற்றி அமைத்து வருகிறது . பாடத்தலைப்புகளை ஒரு நிமிடம் முதல் இரண்டு நிமிடம் வரை இணையத்தின் வழி தொடர்புடைய அனைத்துச் செய்திகளையும் சேகரித்து விறுவிறுப்பான குறும்படம் தயாரித்து ஒப்படைக்கும் சூழலில் அதிக நேரம் செலவு செய்யாமல் சுருங்கக் கூறி விளங்க வைக்கும் திறனை வளர்த்துக் கொண்டு தன்னம்பிக்கையுடன் அப்படைப்பை உருவாக்கவும் செய்கிறார்கள் . பல நாட்கள் பல இடங்களுக்குச் சென்று பல நூல்களைக் கற்றுப் படைத்த நிலை மாறி குறும்படத்தி ன்வழி, குறுகிய நேரத்தில் பலவிதமான செய்திகளை ஒன்று திரட்டி உருவாக்கும் படைப்பு தமிழ் மொழியைக் கற்போருக்குச் சுமையாக இல்லாமல் சுகமாகவே அமைகிறது .

குறும்படத்தால் விளையும் பயன்கள்

கற்பனை ஆற்றலையும் , குழு உணர்வையும் சீரிய சிந்தனையையும் வளர்ப்பதற்கு குறும்படம் தயாரித்தல், தாராளமாக வாய்ப்புகளை வழங்குகிறது . தமிழில் உரையாடி உருவாக்கப்படும் குறும்படங்கள் , திட்டமிடும் திறன், தயாரிப்பு உத்திகள் , படைக்கும் ஆற்றல், புத்தாக்கச் சிந்தனை, தொழில்நுட்ப நுணுக்கங்களை அறிந்து கொள்ளும் ஆற்றல் போன்றவற்றை வளர்க்கின்றன .

தமிழுக்கு அணி சேர்க்கும் வகையில் தயாரிப்பாளர் , எழுத்தாளர், படைப்பாளர், நடிகர் என பல கோணங்களில் வகைபிரித்துப் பார்த்துத் தரமான படைப்புக்களை உருவாக்குவது தமிழில் புத்தாக்கச் சிந்தனைகளை ஒன்று சேர்ந்து வெளிக்கொணரும் நிலையை ஏற்படுகிறது .

இணைய உலகில் தமிழ்க் குறும்படங்களைப் பதிவேற்றம் செய்து அதைக் காண்பவர்கள் வழங்கும் சிறந்த மதிப்பீடுகளும் பயனுள்ள கருத்துக்களும்

மிகுந்த மனநிறைவைத் தரும் பொழுது படைப்பாளிகள் மேன்மேலும் தமிழில் புதைந்து கிடக்கும் எண்ணற்ற கருத்துக்களை ஒலி.ஒளி காட்சிகளாக உருவாக்க வேண்டும் என்ற சிறந்த சிந்தனை ஏற்படுகிறது .



கைத்தொலைபேசி வழி கற்றல் (Learn Through Smartphone)

இன்றைய சூழலில் நம்மிடையே கைத்தொலைபேசி என்பது உயிரும் உடலும் போல இரண்டறக் கலந்த ஒன்றாக அமைந்திருக்கிறது. அதன் வழியாகத் தமிழ்மொழியை மிகுந்த ஆர்வத்துடன் பயன்படுத்த முயல்வதைக் காண முடிகிறது. அனைவரிடமும் கைத்தொலைபேசி பயன்பாட்டில் இருப்பதாலும் அடிப்படையில் தமிழ் எழுத்துகள் அதனுள் இணைக்கப் பட்டுள்ளதாலும் கற்றல் கற்பித்தல் எளிமையாக நடைபெற வழி வகுக்கிறது .

உலக நாடுகளில் அறிமுகப் படுத்தப்பட்டுள்ள தடையில்லா இணையச்சேவை , கைத்தொலைபேசி கல்வி முறை , தமிழ் மொழி வளர்ச்சிக்கு புதுப்பாதை அமைத்துத் தந்துள்ளது .

எளிமையான கட்டமைப்பு மூலம் தமிழ்ப் பயனீட்டாளர்களை ஒருமுகப்படுத்துவதற்கு குறுந்தகவல் அனுப்புவதன் மூலம் தமிழில் கலந்துரையாட முடிகிறது . பாடத்திட்டங்கள், வாசிக்கும் வழி முறைகள் , கணினி, இணையம், கைத்தொலைபேசி இம்மூன்றையும் ஒருங்கிணைப்பதன் மூலம் தமிழ்மொழியை எளிமையாகவும் இனிமையாகவும் பயன்படுத்த முற்படுகின்றனர் .

தமிழ் கற்றல் கற்பித்தலில் கைத்தொலைபேசிகளின் பயன்கள்

தமிழ் மொழித் பயன்பாட்டை ஊக்குவிப்பதோடு , மொழித் திறன்களில் ஏற்படுகின்ற படிப்படியான வளர்ச்சி நிலையைக் கண்காணிப்பதற்கும் ஏதுவாக அமைகிறது . கற்றல் தலைப்பிற்குத் தொடர்புடைய செய்திகளைப்

பதிவிறக்கம் செய்து முழுமையான கற்றலில் ஈடுபடுத்த , இக்கருவி பெருந்துணை புரிகிறது . எந்நேரமும் கைகளில் தவழ்ந்து கொண்டிருப்பதால் நினைத்த நேரத்தில் அதனை காலவிரயம் ஏற்படாமல் பயன்படுத்த முடிகிறது .

இருவழிக் கலந்துரையாடலுக்கு இக்கருவி பயன்படுவதோடு , தயக்கமின்றித் தங்களுடைய பதில்களை வழங்குவதற்கு ஏதுவாக அமைகிறது . விளையாட்டு உத்தி முறைகளின் வழி , கற்றல் கற்பித்தலை நடத்தும் போது தமிழை எளிமையாகச் சென்றடைகிறது. பல்வேறு வகையான விளையாட்டு உத்திகளைக் கைத்தொலைபேசிகளில் பதிவிறக்கம் செய்து அதன்வழி புத்தாக்கமும் ஊக்கமும் ஏற்பட வழி வகை செய்கிறது.

கைத்தொலைபேசியில் ஜிபிஎஸ் (GPS), வீடியோ, புகைப்படக்கருவி, 4ஜி கட்டமைப்பு போன்ற வன்பொருள்கள் கற்றல் கற்பித்தலுக்கு மிகவும் பயனுள்ளதாக அமைந்திருப்பது அனைவரும் அறிந்த ஒன்றாகும் .



மின்வழிக் கற்றல் (Learn Through LMS E-Learning)

கணினி உலகில் ஒருவழித்தகவல் பரிமாற்றமே அதிகமாக நடைபெற்ற நிலை கடந்த பத்தாண்டுகளில் இருவழித்தகவல் பரிமாற்றமாக மாறும் முறையே மின்வழிக் கற்றல் என்பதாகும் .

இம்முறை தமிழ்மொழியின் உயர்வுக்கும் அதிகப் பயன்பாட்டிற்கும் வழி வகுத்துள்ளது . தாமாகவே அறிந்து உள்வாங்கிக் கற்கும் வகையில் கற்றல் கற்பித்தல் மறு உருவாக்கம் கண்டுள்ளது .

இம்முறையின் வழி ஒரே நேரத்தில் உலகின் பல்வேறு பகுதிகளில் வாழ்பவர்கள் தமிழ்க் கற்றல் கற்பித் தலைத் திறம்படச் செய்ய முடியும்.

குறிப்பிட்ட நேரத்தில் குறிப்பிட்ட இடத்தில் அமர்ந்து கற்க வேண்டும் என்ற நிலையை இம் முறை மாற்றி அமைத்துள்ளது. விரும்பிய நேரமெல்லாம் மின்வழிக் கற்றல் வழி கற்றுப் பயனடைவதற்கு ஏதுவாகத் தொழில்நுட்பம் அமைகிறது .செய்யும் தவறுகள் மற்ற வர்கள் முன்னிலையில் சரி செய்யப்படும் அவல நிலை இம்முறையில் இல்லை . தவறுகள் அப்பொழுதே திருத்தப் படுவதற்கும் சரியான உறுதியான விடைகளை இதன்வழி பெறமுடிகிறது . சுயமாகச் சிந்திப்பதற்கும் கருத்துக்களைப் பரிமாறிக்கொள்வதற்கும் மின்வழி கற்கும் வகுப்பறை (Virtual Classroom) பேருதவியாக அமைகிறது . தகவல்களைப் பரிமாறிக் கொள்வதற்கும் கற்றல் கற்பித்தலில் மிகுந்த ஆர்வம் ஏற்படுவதற்கும் மின்வழிக் கற்றல் உதவியாக அமைகிறது . நூல்களையோ ஏடுகளையோ சுமந்து சென்று கற்க வேண்டும் என்ற நிலை இம்முறையில் இல்லை .

கற்றல் மேலாண்மை அமைப்பு Learning Management System (LMS) எனும் சிறப்பு மென்பொருள் கற்பித்தல் மற்றும் கற்றல் நோக்கத்திற்காக உருவகப்படுத்துதல் பல்லுடகம், ஊடாடல், உருவாக்கம், நிர்வாகம் மற்றும் விநியோகம் , கருவிகளின் தொகுப்பு மின்கற்றல் படிப்புகள் மற்றும் தகவல்தொடர்பு மற்றும் கருத்துக்களை , இணைய அடிப்படையான பயிற்சிகள் மூலம் தமிழ்க்கல்வியை உலகளாவிய முறையில் அறிமுகப்படுத்த அடித்தளம் இட்டுள்ளது.

மின் கற்றல் நன்மைகள்

தகவல் மேம்படுத்தும் எளியமுறை மின் கற்றல் வழி சாத்தியமாகிறது . கற்கும் பயணத்தில் புதிய செய்திகளும் தகவல்களும் உலகளாவிய முறையில் உடனுக்குடன் மற்றும் செய்யப்பட்டு புதிய கோணங்களில் பற்பல வாய்ப்புக்களை வழங்குகின்றன .

இதன் மூலம் , தமிழ் மொழிப் பயிற்சித் திறன் கணிசமாக அதிகரிக்கிறது. எப்போது வேண்டுமானாலும், எங்கு வேண்டுமானாலும் மின் கற்றல் வழி தைரியமாக அணுகவும் , மற்றும் வேண்டிய அளவு கற்றுக்கொள்ளவும் முடிகிறது. சுயமாகப் பயில்வதைவிட , தமிழைத் தரமாகக் கற்க வாய்ப்புகள் அதிகம் உள்ளன . புவியியல் எல்லைகளைக் கடந்து தமிழின் இனிமை குன்றாது தமிழைக் கற்றுத்தர பற்பல செயலிகள் மின்வழிக் கற்றலுக்குப் பயன் படுத்தப்படுகின்றன. தமிழ் இணையம் மூலம் கற்கும் வாய்ப்பு அனைவரையும் எளிதாகச் சென்றடைகிறது .

பரிந்துரைகள்

மாறி வரும் தொழில் நுட்பத்திற்கு ஏற்பத் தமிழை எடுத்துச்செல்லும் வன்பொருட்கள் (Hardware) மற்றும் மென்பொருட்கள் (softwares), செயலிகள் (applications) பலரால் பல்வேறு மாறுபட்ட தொழில் நுட்பங்களில் வெளியிடப்படுகின்றன. இவை ஒருமுகப்படுத்தப்பட்டு , ஒரே சீரான தரக் குறியீடு கொண்டவையாக வர வேண்டும். ஆங்கிலத்தைப் போல , சீன மொழி கணினி உலகில் தனித்து வளர்வதற்கு இந்த ஒருமைப்பாடு பெரிதும் உதவுகிறது . தமிழுக்கு 31 குறியீடுகள், ஆங்கிலத்துக்கு 26. இடைவெளி 5 மட்டுமே. ஆனால் தட்டச்சு முறையில் இதுவரை தெளிவான சீரமைப்பு இல்லாததால் , தமிழ் பயனீட்டாளர்கள் பெரிதும் குழம்பிய நிலையில் இணையத் தமிழை பயன்படுத்தும் நிலையில் இருக்கிறோம் . யூனிகோட் முறை இந்த இடர்பாடுகளை தீர்க்கும் என நம்புவோம்.

விண்டோஸ்8 முறையில் தமிழ் பயன்பாடு இப்போது முன்னேற்றம் கண்டிருந்தாலும் , ஐ-பட் போன்ற மடிக் கணினிகளில் இன்னும் தமிழ் தாராளமயப்படுத்தப்படவில்லை . தமிழ் தட்டச்சு முறை இன்னும் வழங்கப்படாத நிலையில் தமிழை முழுமையாகப் பயன்படுத்த முடியாத நிலையில் இருக்கிறோம் . எதிர்காலத்தில் ஆப்பிள் நிறுவனம் அனைவரும் பயன்படுத்தும் பொதுவான ஒரு முறையை அறிமுகப்படுத்தும் என நம்புவோம் .

தமிழில் முப்பரிமான கணினி விளையாட்டுகளை உருவாக்குவதற்கு முயற்சிகள் மேற்கொள்ளப்பட வேண்டும் நமது தமிழ் இசைக்கு நடமாடும் கணினி மாந்தர்களோடு நாமும் சேர்ந்து விளையாடும் முப்பரிணாம விளையாட்டு , MICROSOFT, Apple, Unity போன்ற உலகளாவிய நிறுவனங்களுக்குத் தமிழ் விளையாட்டுக்களை அறிமுகம் செய்து வெற்றி பெறுவதற்கு , முதலீட்டாளர்களும், பயனீட்டாளர்களும் மனமுவந்து முன்வர வேண்டும் . எதிர் காலத்தில் சேர, சோழ, பாண்டியர்களின் வீரசாகசங்களையும் , நமது தமிழ் பண்பாட்டுக்கூறுகளையும் உள்ளடக்கிய கணினி விளையாட்டுகளை உருவாக்குவதற்கு இதுவே நல்ல தொடக்கமாக அமையலாம் .

தமிழில், இதுவரை பொருளாதார அடிப்படையில் பயன் உண்டாகும் வகையில் முயற்சிகள் மேற்கொள்ளப்படாததால் பல அரிய செயல் பாடுகள் முளையிலேயே கருகிப்போய் விடுகின்றன . அந்த நிலை தற்போது ஆக்ககரமான மாற்றங்களை பரவலாகக் காண முடிகிறது . 21 நூற்றாண்டில் தமிழ் சார்த்த தொழில் தொடங்கவும் தமிழை வாழ்க்கை மொழியாகவும் மாற்றுவதற்கு வாழ்நாள் பயிற்சிகள் , உதவித்திட்டங்கள், வேலை வாய்ப்புகள் போன்றவை புத்தாக்கத் தொழில் முனைவர்களை உருவாக்குதற்குப் பேருதவியாக அமைவதற்கு உலகளாவிய முறையில் முயற்சிகள் மேற்கொள்ளப்பட வேண்டும் .

முடிவுரை

தொன்மை மிக்க தமிழ் , தனக்கே உரிய அழகை , செழுமையை நம்மிடம் அடையாளம் காட்ட இதுவரை இணையம் ஒரு வழித்தடமாக அமைந்துள்ளது . இனி வரும் காலம் , இளைய சமுதாயம் தமிழை சுவாசிக்க , நேசிக்க புதுவகையான சிந்தனைகளை உள்ளடக்கிப் புத்தாக்க எழுச்சியோடு செம்மொழியை உலகோர் அனைவரும் கசடறக் கற்கவும் 21 நூற்றாண்டையும் கடந்து அதன்படி நிற்கவும் , புதிய ஊடகமும் இணையமும் புத்தாக்கச் சிந்தனையும் பேருதவி புரியும் என மனமார நம்புவோம் .

References Websites

http://www.eonreality.com/news_releases.php?ref=news/news_releases&sid=475
<http://tamilelibrary.org/>
http://www.tamilcube.com/res/tamil_ebooks.html
<http://www.projectmadurai.org/>
www.coursesites.com
<http://ta.wikipedia.org/wiki>
<http://en.wikipedia.org/wiki/Tamil>
www.microsoft.com
www.apple.com
<http://developer.apple.com/>
<http://www.ulakathamizhchemmozhi.org/en/>
www.infitt.org
<http://www.hindu.com/2010/06/24/stories/2010062454530400.htm>
www.infitt.org/...%20The%20Use%20of%20Multi%20Media%20in%20Teaching%20Tamil-English.pptx
<http://tamilchannels-online.blogspot.com/2011/12/get-bold-using-social-media-tocreate.Html>
<http://tamilelibrary.org/teli/sintalk1.html>
conference.nie.edu.sg/2007/paper/papers/ITE283.pdf
https://wiki.smu.edu.sg/digitalmediaasia/Digital_Media_in_Singapore
www.ida.gov.sg
www.mda.gov.sg

Reference Books

1. Salmon, Gilly (2002) E-tivities: The Key to Active Online Learning, London, Kogan Page
2. Paul Chin, Using C&IT to Support Teaching
3. Slater, Paul and Varney-Burch, Srah (2001) Multimedia in Language
4. Susan Ko, Steve Rossen: Teaching Online -A Practical Guide, 2nd Edition
5. Reynol Junco& Jeanna Mastrodicasa:Connecting to the net.generation
6. Les Lloyd, Editor: Technology and Teaching
7. Kristen J. Nelson;2nd Edition: Teaching in the Digital Age
8. Shirley Bennett, Debra Marsh and Calre Killen: Handbook of Online Education

தொடக்க/நடுநிலைப் பள்ளிகளில் தமிழ் இணையப் பயன்பாடு

கவி.செங்குட்டுவன் (எ) செ.இராஜேந்திரன், எம்.ஏ, எம்.பில், பி.எட்

பட்டதாரித் தலைமையாசிரியர், ஊ.ஒ.ந.நி.பள்ளி, ஜோதிநகர், ஊத்தங்கரை – 635207

rajendrankavi@yahoo.co.in

முன்னுரை

நேற்றைய உலகம் கணினி உலகம், இன்றைய உலகம் இணைய உலகம். அன்று நிலவைக் காட்டி குழந்தைக்கு சோறு ஊட்டினோம். இன்று நிலவுக்கேச் சென்று சோறு ஊட்டலாம், நாளை நிலவிலேயே சோறு சமைக்கலாம். இத்தகைய நவீனம் நாளும் நிகழ்ந்துக்கொண்டிருக்கும் வேளையில் தமிழ் இணையத்தின் தன்னிகரில்லாத் சேவைகள் மற்றும் தேவைகள் பற்றியும் அதை கல்விக்கு எவ்வகையில் பயன்படுத்தலாமென்பது பற்றியும் குறிப்பாக துவக்க/நடுநிலைப் பள்ளிகளில் எவ்வகையில் பயன்படுத்தலாம் என்பது குறித்தும் விரிவாக அலசுவதே இக்கட்டுரையின் நோக்கமாகும்.

கணினியில் தமிழ் மொழியின் தாக்கம் மற்றும் ஆளுமை

கல்வெட்டுக்கள், ஓலைச்சுவடிகள், செப்பேடுகள், காகிதம் என பல்வேறு வழிகளில் வளர்ச்சி பெற்ற தமிழ் தற்போது இணையத்தில் உலா வருகிறது. இதன் விளைவுதான் இன்று நாம் இங்கு ஒன்று கூடி விவாதிக்கும் வாய்ப்பைத் தந்துள்ளது. 1980ம் ஆண்டு முதல் கணினியில் இடம் பெறத் துவங்கியது தமிழ். 1995ம் ஆண்டு முதல் தமிழ் இணையத்தில் உலாவர ஆரம்பித்தது. இணையத்தில் தமிழ் இடம் பெற்று ஏறக்குறைய 17 ஆண்டுகள் ஆகிவிட்ட இந்தச் சூழலில் அதன் வளர்ச்சி குறிப்பிடத்தக்க இடத்தைப் பெற்றுள்ளது.

பத்து ஆண்டுகளுக்கு முன்னர் வரை கணினி அல்லது இணையம் தமிழை வளர்க்க ஒரு துணைக் கருவியாகப் பயன்பட்டு வந்தது. ஆனால் தற்போது கணினித் தமிழே தனித் துறையாக வளர்ந்து வருகிறது. ஆங்கில வழியாகக் கற்ற கணினித் தொழில் நுட்ப வல்லுநர்கள் மெல்லத் தமிழில் நுழைந்து தொழில் நுட்பங்களைத் தமிழ்ப் படுத்தித் தமிழ் இணையத்தை வலுப்பெறச் செய்து வருகின்றனர்.

இணையத்தில் தமிழ் இன்று ஆட்சி செய்து வருவதால் எளியவரும், வறியவரும், மொழி ஆளுமை அற்றவரும் கூட இன்று உலகம் முழுமைக்கும் இருந்த இடத்திலிருந்தே பயணிக்க முடிகிறது. இன்றைய இணைய உலகால் தமிழ்ப் பயன்பாடு, கலாச்சாரம் உலகம் முழுதும் மிக விரைவாக பரவி வருகிறது. தமிழில் உள்ள வலைத் தளங்கள் வாயிலாக தமிழர் செய்திகளை அறிந்து கொள்ளவும், பல்லாயிரக் கணக்கான தமிழ் நூல்களைப் பெறவும், இணையத் தமிழ் இதழ்களைப் படிக்கவும், தமிழிசையைக் கேட்கவும், தமிழ்க் காணொலிகளைக் காணவும் முடிகின்றது. மொத்தத்தில் தற்கால தகவல் தொழில் நுட்ப உலகில் தமிழைத் தழைக்கச் செய்ய வேண்டிய முயற்சிகள், செயல்பாடுகள் மிக வேகமாக நடைபெற்று வருகின்றன என்பதே உண்மை நிலை.

கணினி மூலம் கல்வி வளர்ச்சி

மற்ற எல்லாத் துறைகளையும் போலவே இணையம் கற்றல் கற்பித்தலிலும் மிகப் பெரிய மாற்றங்களை ஏற்படுத்தியுள்ளது. பழைய மரபு வழி சார்ந்த கற்றல் கற்பித்தல் முறைகள் பலவும் தற்போதைய வகுப்பறைகளில் மாறி நவீனமான புதிய முறைகள் இடம் பிடித்துள்ளன. இணையம் வழியான கல்வி முறை ஆசிரியர்களின் கற்பித்தலிலும் மாணவர்களின் கற்றலிலும் புதிய வகைப் பரிணாமங்களை உண்டாக்கிக் கொடுத்திருக்கின்றன. எளிமை, விரைவு, விரிவு, பயன்விளைவு, கவர்ச்சி, மனமகிழ்வு, பல்லுடகம் முதலான தன்மைகளைக் கொண்டிருப்பதால், இணையம் வழியான கல்வி முறை இன்றைய காலத்திற்கு மிகவும் ஏற்றதாகவும் தவிர்க்க இயலாத ஒன்றாகவும் ஆகிவிட்டது. மேலும் கற்றல் கற்பித்தலுக்கு உதவும் வகையில் பல வகையான இணைய வசதிகள், செயலிகள், மென்பொருள்கள், வலைத்தளங்கள் முதலியவை தற்போது நடைமுறைக்கு வந்துவிட்டன.

தகவல் தொடர்பு தொழில் நுட்பத்தின் மிக விரைவான வளர்ச்சி கல்வித் துறையிலும் மிகப் பெரிய தாக்கத்தை ஏற்படுத்தியுள்ளது. கணினியையும் இணையத்தையும் பயன்படுத்தாத கல்வி நிறுவனங்களே இல்லை என்கின்ற

அளவுக்குத் தொடக்கப்பள்ளி தொடங்கி உயர் கல்விக் நிறுவனங்கள் வரையில் எல்லா இடங்களிலும் இதன் தாக்கம் அதிகரித்து வருகின்றது. இத்தகையச் சூழலில் கற்றல் கற்பித்தலில் இணையமும், இணையம் சார்ந்த வசதிகளும் முக்கிய இடத்தைப் பிடித்து விட்டன எனலாம். கோலின்சு என்பார் 1996 - 97 ஆம் ஆண்டுகளிலேயே கற்றல் கற்பித்தலில் இணையம் தடம் பதித்து விட்டது என்கிறார். இதன் அடிப்படையில், இணையம் வழியான கற்றல் கற்பித்தல் இன்று மிகவும் புகழ் பெற்றதாகவும் எல்லோராலும் விரும்பப்படுவதாகவும் இருக்கின்றது.

இணையத்தில் தமிழ்

தமிழ் மொழி காலத்திற்கு ஏற்ற வகையில் தன்னை வளர்த்துக் கொண்டேவருகிறது. இன்றைய மிகப் பெரிய தகவல் ஊடகமாக விளங்குவது இணையம் ஆகும். இந்தத் தகவல் சாதனத்திலும் தமிழ் அதிக அளவில் பயன்படுத்தப் பெற்று வருகின்றது. ஏறக்குறைய இணையப் பயன்பாட்டில் தமிழ் மொழி குறித்த தகவல் பரிமாற்றம் உலக அளவில் இரண்டாம் இடத்தில் இருப்பதாக புள்ளி விபரங்கள் தெரிவிக்கின்றன.

தமிழ் மொழிக் கற்றல் கற்பித்தலுக்கு உதவும் இணைய வசதிகள் தற்போது அதிகரித்து வருகின்றன. மழலையர் கல்வி தொடங்கி முனைவர் படிப்பு வரையில் தமிழில் படிப்பதற்குரிய வாய்ப்புகளை இன்றைய இணைய உலகம் ஏற்படுத்திக் கொடுத்திருக்கிறது. கேட்டல், பேசுதல், வாசித்தல், எழுதுதல் முதலான அடிப்படைக் கல்வி முதல் ஆராய்ச்சிப் கல்வி வரையில் இணையத்தில் படிக்கக் கூடிய வசதிகள் தற்போது உள்ளது. இலக்கணத்தையும் இலக்கியத்தையும் புதிய ஊடகத்தின் வழியாக புதிய அணுகு முறைகளோடு கற்பதற்கும் கற்பிப்பதற்கும் வாய்ப்புகள் பெருகி வருகின்றன. அவ்வகையில், தமிழ் கற்றல் கற்பித்தலுக்குப் பங்களிக்கும் இணைய வசதிகள் குறித்து பார்த்தோமேயானால், இணையம் இன்று ஒரு பல்லாடகக் கருவியாக வளர்ச்சி பெற்று இருக்கின்றது. எழுத்துரு, படங்கள், அசைவுப் படங்கள், நிகழ்ப்படங்கள், ஒலி-ஒளி காட்சிகள், இணைய உரையாடல், நிகழ்ப்படக் கருத்தரங்கு, வானொலி, தொலைக்காட்சி எனப் பன்முகப் பயன்பாடுகளைக் கொண்டதாக விளங்குகின்றது. இவை அனைத்தும் தற்கால நவீன முறையிலான கற்றல் கற்பித்தலுக்கு மிகவும் ஏற்றவை மட்டுமல்ல எளிமையாகப் பயன்படுத்தக் கூடியவை. இவ்வாறு பல்லாடகமாகச் செயல்படும் இணையத்தில் உள்ளவற்றை கணினியில் சேமித்து வைத்தோ அல்லது பதிவிறக்கம் செய்தோ தேவையான நேரத்தில் பயன்படுத்திக் கொள்ள வாய்ப்பாக இருக்கின்றது. அதே போல், நம்மிடம் உள்ளவற்றை இணையத்தில் பதிவேற்றம் செய்து உலகம் முழுவதும் உள்ள கணினிப் பயனாளர்களோடு பகிர்ந்து கொள்ளவும் முடிகிறது. இப்படியான, அரிய வாய்ப்பினைத் தமிழ் மொழிக் கற்றல் கற்பித்தலுக்கும் பயன்படுத்திக்கொள்ள வேண்டியது நம்முடைய ஆற்றலும் திறமும் ஆகும்.

இணைய வழிக் கல்வி

இணையம் வழி கற்றல் கற்பித்தல் மூலம் குறிப்பிடத்தக்க நன்மைகள் விளைகின்றன. பல்வேறு விதமான கல்வி மேம்பாட்டு செயல்கள் நிறைவு பெற இயலும் அவற்றுள் குறிப்பிடத்தக்கவை :

- பாடங்களை மாணவர்கள் எளிதில் புரிந்துக் கொள்ள முடிகிறது .
- ஓர் குறிப்பிட்ட தலைப்பிலான பாடத்தை பற்றிய கூடுதல் தகவலை அறிந்து கொள்ளும் வகையில் இணைப்புகளைப் பயன்படுத்தலாம்.
- எல்லா மாணவர்களும் ஒரே விதமான பாடத்தை படிக்க வேண்டிய கட்டாம் இல்லாமல் அவர்களின் விருப்பத்திற்கேற்ப தேர்வு செய்து படிக்கும் வாய்ப்பு கிடைக்கிறது
- மாணவரை மையமாகக் கொண்ட பாடத்திட்டத்தை வகுக்கலாம் .
- மாணவர்கள் மகிழ்ச்சியுடன் கற்கவும் பல புதிய சிந்தனைகளை வெளிக்காட்டவும் பயன்படுகிறது.
- எழுத்து (text), ஒலி (sound), காட்சி (visual), அசைவுப்படம் (graphics), நிகழ்ப்படம் (video), உடலியக்கம் (psychomotor), இருவழித்தொடர்பு (interactive) எனப் பல வகையில் கற்பதற்குரிய வாய்ப்பு கிடைக்கின்றது.
- மாணவர்களின் கற்றல் திறனையும் தன்னம்பிக்கையையும் வளர்த்திடவும் அவற்றை அளவிடவும் இம்முறை சிறப்பானது.
- மாணவர்களின் சிந்தனையைத் தூண்டி ஆய்வு மனப்பான்மையை வளர்க்கிறது .
- கற்பதற்குரிய குறிப்பிட்ட காலமோ, இடமோ, தக்க சூழலோ தேவையில்லை.
- தேவையான அனைத்து தகவல்களையும் விரைவாகவும், விரிவாகவும் பெற முடிகின்றது.

- கற்றலில் ஏற்படும் சந்தேகங்களை பல முனைகளிலிருந்தும் ஆய்ந்து , கேட்டறிந்து தாமே தீர்வு காணும் வாய்ப்பை ஏற்படுத்தித் தருகிறது.
- தாம் கற்றதை தாமே மதிப்பீடு செய்துக்கொள்ளும் வாய்ப்பும் கிடைக்கிறது .

துவக்க/நடுநிலைப் பள்ளிகளில் இணைய அறிமுகம்

கடந்த பத்தாண்டுகளுக்கு முன்பு வரையில் கணினியை கண்ணால் காண்பதே மிகப்பெரிய அதிசயம். ஆனால் இன்றைய காலகட்டத்தில் தமிழகத்தைப் பொருத்தமட்டில் அனைத்து நடுநிலை, உயர்நிலை, மேநிலைப் பள்ளிகளிலும் கணிப்பொறிகள் பயன்படுத்தும் வசதி கிடைத்து விட்டது. ஆம், அனைத்து பள்ளிகளிலும் மேசைக் கணினிகளும், மடிக் கணினிகளும் உள்ளது. அவற்றைக்கொண்டு இன்று இணைய அறிமுகத்தை மேற்கொண்டால் நாளைய உலகம் நம்கையில் என்பதற்கேற்ப இன்றைய இணைய தலைமுறை மிக வேகமாக இத்துறையில் முன்னேற்றம் காண வழி கிடைக்கும். இதன்மூலம் உலக அரங்கில் இணைய பயன்பாடு என்பதே கோலோச்சும்.

துவக்க/நடுநிலைப் பள்ளிக் கல்வி வளர்ச்சியில் இணையத்தின் பங்கு

இன்று கல்வி என்பது வெறும் பாட அறிவு என்பதில் இருந்து வேறுபட்டு உலகியல் அறிவை வளர்ப்பதோடு அன்றாட நடைமுறைச் செய்திகளையும் புதுப்பித்துக்கொள்ளத்தக்க வகையில் அமைய வேண்டும் என அனைவரும் எதிர் நோக்குகின்றனர். இதனைக் கருத்தில் கொண்டே இன்று அரசாங்க துவக்க/நடுநிலைக் கல்வி கற்பித்தல் முறையை 1 முதல் 4 வகுப்பு வரையில் எளிமைப்படுத்தப்பட்ட செயல்வழிக் கற்றல் (SIMPLIFIED ACTIVITY BASED LEARNING) என்றும், 5 ஆம் வகுப்பு எளிய படைப்பாற்றல் கல்வி (SIMPLE ACTIVE LEARNING METHODOLOGY) என்றும், 6 முதல் 8 வகுப்பு வரையில் படைப்பாற்றல் கல்வி (ACTIVE LEARNING METHODOLOGY) என்றும் மாற்றம் செய்து அறிமுகப்படுத்தி உள்ளது. இச் சூழலில் வெறும் பாடபுத்தகத்தை மட்டும் வைத்துக்கொண்டு ஆசிரியர்கள் மாணவர்களுக்கு பாடம் நடத்தினால் அது அனைவரின் எதிர்பார்ப்புகளையும் பொய்த்துப்போகச் செய்துவிடும்.

இச்சூழலில்தான், நமக்கு கைகொடுக்கும் புதியமுறை இணைய வழிக் கற்றல்/கற்பித்தல் ஆகும். அனைத்து பள்ளிகளிலும் இணைய வழிக் கல்வியை நடைமுறைப்படுத்தும் போது ஆசிரியர்கள் தமது அறிவாற்றலை உயர்த்திக்கொள்ளும் அதே வேளையில் மாணவர்கள் நாளும் பல புதியச் செய்திகளையும், கற்றல் அனுபவங்களையும் பெற வாய்ப்பு ஏற்படும். மேலும் நடுநிலைக் கல்வி பெறும் வயது என்பது (6 முதல் 8 வகுப்புகள்) 10 முதல் 14 வரையானது இக்காலக்கட்டம் மாணவர்கள் எதையும் விளையாட்டு வழியில் எளிமையாகக் கற்கும் காலம் இக்காலகட்டத்தில் கற்கும் எதையும் தம் வாழ்நாளில் மறத்தல் இயலாது. எனவே இப்பருவத்தில் இணையவழிக் கல்வியை அறிமுகம் செய்யும் போது அது மிகு பயனை விளைவிக்கும். மேலும் இக்கல்வி முறையை தொடர்ந்து பின்பற்றும் மாணவன் ஓர் குறிப்பிட்ட காலகட்டத்தில், "தானே கற்றல்" (SELF LEARNING) முறைக்கு மாறிவிடுவான்.

தமிழ் இணையக் கல்விக் கழகம் அறிமுகம்

மேற்காண் இணைய கற்றல் முறையை ஆசிரியர்களும், மாணவர்களும் பயன்படுத்திட பெரும் உதவியாக அமைந்ததுதான் தமிழ் இணையக் கல்விக்கழகம். இதில் உள்ள வசதிகளை பயன்படுத்திக்கொள்ள ஓர் மாணவன் துவங்கிவிட்டாலே அவனுக்கு கற்பித்தலுக்கு ஆசிரியர் துணை தேவைப்படா. ஆம், காரணம் அதில் உள்ள நூலகம், இணைய வகுப்பறை, அகராதிகள், கலைச் சொற்கள், சுவடி காட்சியகம், பண்பாட்டுக் காட்சியகம் என உள்ள பல்வேறு வசதிகள் இணையத்தை பயன்படுத்தும் அனைவரையும் கவரும் .

தமிழ் இணையப் கல்விக் கழகம் மூலம் உலகின் எந்த மூலையில் உள்ள ஒருவரும் எளிதில் தமிழை அடிப்படை அரிச்சுவடியில் துவங்கி முதுநிலை பட்டப் படிப்பு வரையில் பயில இயலும் . இத்தளத்தில் காணும் நூலகம் சென்றால் அங்கு காணும் ஆயிரக்கணக்கான நூல்கள் மிக அரிதானவை இந்த நூலகத்தை இதுவரையில் பார்வையிட்டவர்கள் 2, 78, 145 பேர் (04.1.2012 வரையில்) ஆவர்.

கல்விக்கான சில இணைய தளங்களும் அவற்றின் சேவைகளும்

தமிழ் மொழிக் கற்றல் கற்பித்தலுக்குப் பயன்படும் பல இணையத் தளங்கள் உள்ளன. தமிழை முதல் மொழியாகக் கற்பதற்கும், இரண்டாம் மொழியாகக் கற்பதற்கும் இணையத் தளங்கள் வந்துவிட்டன. அதுமட்டுமல்லாது, ஆங்கிலம் உள்ளிட்ட பல அயல் மொழிகள் சார்ந்தோரும் தத்தம் மொழி வழியாகத் தமிழ் கற்றல் என்பது இன்று மிக எளிதானதாக உள்ளது . முறையாக வடிவமைக்கப்பட்ட கலைத் திட்டங்கள், பாடங்கள் ஆகியவற்றை இணையத்திலேயே படிப்பதன் வழியாகத் தமிழ் மொழியைக் கற்றுக்கொள்ள முடியும்.

தமிழ் இணையக் கல்விக் கழகம் தமிழ் மொழிக் கற்றலுக்கு மிகச்சிறந்த முன்னோடியாகத் திகழ்வதை மறுக்க இயலாது. அதுபோல, உலகின் பல நாடுகளில் செயல்படும் தமிழ்க் கல்விக் கழகங்கள் அல்லது கல்விக் கழகத் தமிழ் இருக்கைகள் இணையம் மூலமாகத் தமிழ்க் கல்வியை கற்பித்து வருகின்றன.

தமிழ்க் கல்வியை மேம்படுத்தும் நோக்கத்தோடு செயல்படும் வலைத்தளங்களும் வலைப்பதிவுகளும் அதிக அளவில் இருக்கின்றன . இவை கற்றல் கற்பித்தலுக்கு பயன்படக் கூடிய பல்வேறு கற்றல் மூலங்களை (Veries Learning Sources) வழங்குகின்றன. அதுமட்டுமல்லாது கற்றல் கற்பித்தல் தொடர்பான வலைத்தளங்களையும் வலைப்பதிவுகளையும் உருவாக் கி செயல்படுத்திடவும் இவை பெரும் வாய்ப்பினை வழங்கிக் கொண்டிருக்கின்றன. அதிகம் விலை கொடுத்து மிகப்பெரிய அளவிலான கலைக் களஞ்சியங்கள் , , அகரமுதலிகள் ஆகியவற்றை வாங்கிச் சேமித்து வைக்க வேண்டிய வேலையோ அல்லது எங்கு சென்றாலும் அவற்றைச் சுமந்து செல்ல வேண்டிய அவசியமோ இல்லா நிலையில் இன்று இருக்கும் இடத்திலேயே விரல் நுனியில் விவரங்களைப் பெற்றுக் கொள்ளக் கூடிய வாய்ப்பினை இணையம் வழங்கிக்கொண்டிருக்கிறது .

இது தவிர தமிழ் விக்கிபீடியா மற்றும் தமிழ் விக் ஷனரி ஆகியன இணைய வழிக் கற்றலுக்கு மிகவும் பயன்தருவனவாகும். தமிழ் விக்கிபீ டியாவில் இன்று வரையில் 49, 688 (04.12.2012நாளின் படி) கட்டுரைகள் வலையேற்றப்பட்டுள்ளன. அதேபோல் தமிழ் விக்கிஷனரியில் 2, 75, 283 (04.12.2012நாளின் படி) கலைச்சொற்கள் சேர்க்கப்பட்டுள்ளன .

கல்வி வளர்சிக்கு உதவிடும் சில இணைய தளங்கள்

www.tamilvu.org
http://ta.wikipedia.org
http://ta.wiktionary.org/wiki/
www.tamil.net
www.tamil.org
www.southasia.upenn.edu/tamil
www.languageshome.com
www.tamildigest.com
www.tamilheritage.org
http://hongkongtamil.blogspot.com
http://www.thozhilnutpam.com
http://www.tamildict.com
http://tamilelibrary.org
http://www.chennaiilibrary.com

எல்லோருக்கும் இணைய வழிக் கல்வி

ஒரு நாட்டின் முன்னேற்றத்திற்கும், வளர்ச்சிக்கும் மிக முக்கியமானது அந்நாட்டில் வெளிப்படும் புதிய படைப்பாற்றல் .மாணவர்களின் படைப்பாற்றல் திறனை அதிகரிக்கக் கணினி மென்பொருள் வசதியுடன் கூடிய மொழிப் பயிற்று முறை அனைத்து வகுப்பறைகளிலும் கொண்டுவரப்படல் வேண்டும். அன்றைய ஓலைச்சுவடி முதல் இன்றைய இணையம் வரை தகவல் பகிர்தலின் பரிணாம வளர்ச்சியை எடுத்துக் காட்ட வேண்டும் .தமிழில் கற்பிக்க மிகமிக எளிமையான சொற்களையே பயன்படுத்துதல் வேண்டும் .கற்பவருக்குக் ஆர்வத்தைத் தூண்டும் வகையில் பாடத்திட்டங்களும் பாடப்பொருளும் அமையப் பெறல் வேண்டும் .

முடிவுரை

நேற்றைய உலகம் இருண்டு கிடந்தது, இன்றைய உலகம் வெளிச்சமாக இருக்கிறது. நாளைய உலகம் பிரகாசமாக இருக்க வேண்டும். அதற்குத் தேவை அனைவருக்கும் தரமான அடிப்படை கல்வி. அதை இணையம் வழி இனிதே வழங்குவோம்! இன்சுவைக் கனிகள் பலவும் புசித்து மகிழுவோம்.

மாணவர்களைச் சுயவழிக் கற்றலுக்கான தயார்ப்படுத்துதலில் தகவல் தொடர்புத் தொழில்நுட்பப் பயன்பாடு

கோவிந்தசாமி சந்தன்ராஜ்

தமிழ்த்துறை ஒருங்கிணைப்பாளர், பயா லேபா மெத்தடிஸ்ட் பெண்கள் உயர்நிலைப்பள்ளி, 296 லோராங் ஆ கு

Govindasamy_santhanraj@moe.edu.sg

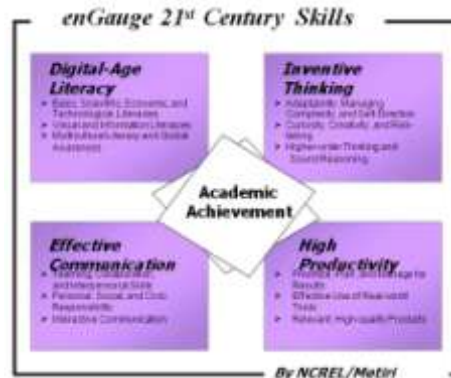
மொழியும் தகவல் தொடர்புத் தொழில் நுட்பமும்

மொழி தகவல் பரிமாற்றத்திற்கான கருவியாக இருந்துவருகிறது . அம்மொழியின் வாயிலாக அறிவினைப் பெருக்கிக்கொள்ளும் மனிதன் புதிய அறிவினைக் கட்டமைக்க புதிய நுட்பங்களை நாடியதின் விளைவே தகவல் தொழில்நுட்ப வளர்ச்சியாகும் . 21- ஆம் நூற்றாண்டில் கற்றலில் ஈடுபடும் யாவரும் புதிய அறிவினைக் கட்டமைக்கத் தகவல் தொடர்புத் தொழில்நுட்பம் பெருந்துணை புரிகிறது . இன்றைய மாணவன் நாளை உலகக் குடிமகனாக, உலகத்தை வழிநடத்துபவனாக மேம்பட, சுயத்தின் கசிவினைச் செம்மைப்படுத்திப் புதிய அறிவினை நாளும் கட்டமைக்க வேண்டியது அவசர உலகத்தின் அவசியமாகிறது .

-21ஆம் நூற்றாண்டுக் கற்றலின் பரப்பும் நோக்கும்

-21ஆம் நூற்றாண்டுக் கற்றலில், மின்னியல் அறிவு (DIGITAL LITERACY), கற்பனையோடு கூடிய படைப்பாக்கச் சிந்தனை (INVENTIVE THINKING), பயன்முனைப்புமிக்க கருத்துப்பரிமாற்றம் (EFFECTIVE COMMUNICATION) அதிக உற்பத்தித் திறன் (HIGH PRODUCTIVITY) எனும் நான்கு பரப்புகள் முக்கியப் பங்கு வகிக்கின்றன . இன்றைய மாணவர்கள் நடைமுறைக்கேற்பக் கற்றுக்கொள்வதும் நடைமுறையோடு பொருத்திப் பார்த்துக் கற்றலில் தாம் அடைந்துள்ள நிலையைப் புரிந்துகொண்டு மேம்பட வழி வகுக்கவும் மற்றும் கற்றவற்றை நடைமுறைக்கேற்ப வெளிப்படுத்திச் செயலாற்றிப் பயன் பெறவும் 21-ஆம் நூற்றாண்டுத் திறன்கள் மிகவு ம் அவசியமாகின்றன. இவ்வகையில் மாணவர்கள் தங்களுடைய கற்றலை எவ்வித அச்சுறுத்தலும் இல்லாத சூழலில் நிகழ்த்திக்கொள்ளவும் மனமகிழ்ச்சியுடன் விருப்பமான முறையில் கற்றுக்கொள்ளவும் தங்களின் கற்பனைகளுக்கு உயிர்கொடுக்கும் வசதி வாய்ப்புள்ள களமாகப் பயன்படவும் சிறு முயற்சி யின் மூலம் பன்மடங்குச் சிறப்பு வாய்ந்த உற்பத்தித்திறன்மிக்க படைப்புகளைப் படைக்கவும் அப்படைப்புகளைப் பன்முக நிலையில் அலசி ஆராய்ந்து மேம்படுத்தவும் தகுதி வாய்ந்தது கணினியும் கணினிவழி ஊடகமான இணையமுமே ஆகும் என்பதில் மாற்றுக்கருத்து இல்லை.

எனவே, சுயவழிக் கற்றலில் திறம்பட ஈடுபட்டு முழுமையான கற்றலை மேற்கொள்ள , மாணவர்கள் பெற்றிருக்க வேண்டிய பல்வேறு திறன்களைப் பெற மிகப் பொருத்தமான கருவியாக இருப்பது தகவல்தொடர்புத் தொழில்நுட்பம் மட்டுமே என்பதில் ஐயமில்லை .



சுயவழிக் கற்றலின் நோக்கும் தேவையும்

மரபுவழிக் கற்றல் சூழலில், ஆசிரியரிடமிருந்து பெறும் அளவுக்கு அதிகமான உதவிகள் துணைகொண்டு கற்பதிலும், புறச்சூழல்களிலிருந்து பெறப்படும் தொடர் தூண்டுதல்கள்மூலம் கற்பதிலும் , கண்டிப்பு, தண்டிப்பின் விளைவு கருதி இறுக்கமான சூழலில் கற்பதிலும் விருப்பமின்றி ஈடுபட்டுவந்த மாணவர்களை உடனடியாகச் சுயவழிக் கற்றலில் ஈடுபடுத்துவது என்பது கண்ணைக் கட்டிக் காட்டில் விட்டதுபோன்றுதான் இருக்கும் .

சுயவழிக் கற்றல் என்பது முற்றிலும் புதுமையான கோட்பாடு அல்ல எனினும் இந்த நூற்றாண்டில் அதைப் பலரும் வெவ்வேறுவிதமாக வரையறை செய்துள்ளனர் . கற்றல் கற்பித்தலுக்குப் பொருத்தமானதாக நான் தேர்ந்தெடுத்தது திரு மால்கம் நோல்ஸ் அவர்களின் வரையறையாகும் .

Malcolm Knowles definition of SDL

“In its broadest meaning, ‘self-directed learning’ describes a process by which individuals take the initiative, with or without the assistance of others, in diagnosing their learning needs, formulating learning goals, identify human and material resources for learning, choosing and implement appropriate learning strategies, and evaluating learning outcomes.” (Knowles, 1975, p. 18)

திரு மால்கமின் வரையறைப்படி , சுயவழிக் கற்றலில் மாணவர்களை ஈடுபடுத்தத் தகுந்த ஊடகம் அவசியமாகின்றது. மேலும், புதிய அறிவினைக் கட்டமைக்கும் நோக்கில் சூழலையும் வசதிகளையும் உருவாக்கித் தருவது இணையம் என்பதில் மாற்றுக்கருத்து இல்லை . அவ்வகையில் சுயவழிக் கற்றல் கோட்பாடு என்பது-21 ஆம் நூற்றாண்டில் தகவல் தொடர்புத் தொழில்நுட்பத்தை முதன்மையாகக் கொண்டு அமைந்துள்ளது என்பதை எவரும் மறுக்கவியலாது .

சுயவழிக் கற்றலில் முக்கிய மூன்று நிலைகள் உள்ளன . அவை பின்வருமாறு அமைந்துள்ளன . 1.கற்றலுக்குத் தாமே பொறுப்பேற்றல் (Ownership of learning) .2 சுய கற்றலை நிர்வகித்தலும் மற்றும் கண்காணித்தலும் (Self-management and self-monitoring) .3வகுப்பறைக்கு அப்பாலும் கற்றல் அல்லது விரிவான கற்றல் (Extension of learning) . இவ்வாறு அமைந்துள்ள சுயவழிக் கற்றலின் நிலைகளைச் செயற்படுத்துவதற்கு மாணவர்கள் பயிற்சி எடுத்துக்கொள்வது சாலச் சிறந்தது . எனவே அந்நோக்கத்தை மனத்திற்கொண்டு 2011 –ஆம் ஆண்டு முதல் பயா லேபா பெண்கள் உயர்நிலைப்பள்ளியில் படிக்கும் மாணவிகளுக்குச் சுயவழிக் கற்றலில் தகவல்தொடர்புத் தொழில்நுட்பப் பயன்பாடு தொடர்பாகப் பல கற்றல் நடவடிக்கைகள் மேற்கொள்ளப்பட்டன.



Source: The ICT Connection [http://ictconnection.edumail.org]

கற்றலுக்குத் தாமே பொறுப்பேற்றல்

மாணவர்கள், பொதுவாக ஆசிரியர் வகுப்பில் கூறுவதைச் செய்து முடித்தலின்வழிக் கற்றுக்கொள்வதை விடுத்துத் தாங்களே தங்களின் கற்றலுக்கான பொறுப்பை ஏற்றுக் கற்க வேண்டும் .அவ்வகையில், என் பள்ளியில் பயிலும் உயர்நிலை இரண்டு தமிழ் மாணவர்கள் சுயவழிக் கற்றல் தொடர்பான கலந்துரையாடலின்போது தாங்களாகவே முன்வந்து படைத்துக் கற்றல் என்ற முறையினைத் தேர்ந்தெடுக்க முடிவு எடுத்தனர் . அதன் அடிப்படையில் உயர்நிலை இரண்டு விரைவுப் பிரிவின்கான பாடநூலில் 'நினைவில் நிற்பவர்கள்' என்ற நான்காம் தொகுதியில் இடம்பெற்ற 'நடிகர் திலகம்' என்ற பாடப்பகுதியைத் தேர்வு செய்தனர்.

சில மாணவர்கள் , சிவாஜியின் வாழ்க்கை வரலாற்றைத் தேர்ந்தெடுத்தனர் . வேறு சிலர் எம் .ஜி.யாரையும் சிவாஜியையும் ஒப்பிட்டுப் படைக்க முடிவுசெய்தனர் . மற்ற சிலர் , சிவாஜி தமிழ்த் திரைப்படத்தில் ஏற்படுத்திய மாற்றங்களைப் பற்றிப் படைக்க நினைத்தனர் . படைப்பு எவ்வாறு அமைய வேண்டும் என்பது பற்றிய வழிகாட்டுதல் குறிப்புகள் ஆசிரியரால் வழங்கப்பட்டன . மாணவர்கள், தாங்களாகவே இணையத்தில் காணப்படும் பொருத்தமான கருத்துகளையும் ஒலிப்பகுதிகளையும் (Audio), காணொளிகளையும் (Video), படங்களையும் தேர்ந்தெடுத்துப் பயன்படுத்திப் படவில்லை (Powerpoint Slide) மூலம் படைத்தனர்.

ஒவ்வொரு குழுவின் படைப்பையும் மற்ற குழுக்கள் மதிப்பீடு செய்து கருத்துகள் கூறும் நோக்கில் மதிப்பீட்டு அளவைகளை மாணவர்களே உருவாக்க வழிகாட்டப்பட்டது . படைப்பின் முடிவில் தங்களின் படைப்புகளைத் தாமே சுயமதிப்பீடு செய்தலின் வழியும் , குழு மதிப்பீட்டின்வழிப் பெறப்பட்ட கருத்துகள்வழியும் , ஆசிரியர்க் கருத்துரைத்தல் வழியும் பெற்ற தகுந்த கருத்துகளைக் கொண்டு சுயமாகக் கற்றலில் மேம்பாடு காண முடியும் என்ற நம்பிக்கையைப் பெற்றார்கள் . மாணவர்கள் தங்களது நேரத்தைச் சரியாகப் பயன்படுத்துவதிலும் தேவைப்படும் தகவல்களைத் தேர்ந்தெடுக்க முடிவெடுப்பதிலும் செய்திகளை முறைப்படுத்திக்கூறுவதிலும் தம் திறமையை வெளிக்காட்டினர்.

வகுப்பறைக்கு அப்பாலும் கற்றல் அல்லது விரிவான கற்றல்

இன்றைய எல்லைகளற்ற உலகமயமாக்கல் கற்பதற்கும் சாலப்பொருந்தும் . உலகமே ஒரு பள்ளிக்கூடம் என்பது உயர்ந்த சிந்தனை. ஆசிரியர் கற்றுக்கொடுப்பதை வகுப்பறையில் கற்பதைத் தவிர , தாங்களாகவே கேள்விகேட்டுக் கற்றுக்கொள்ளும் வாய்ப்பினை ஏற்படுத்தவும் தகவல் தொழில்நுட்பத்தைப் பயன்படுத்தி நாடு எனும் எல்லை கடந்து வகுப்பறைக்கு அப்பாலும் விரிவான கற்றலை நிகழ்த்தவும் ஏற்பாடு செய்யப்பட்டது .

உயர்நிலை 2 உயர்தமிழ் மற்றும் மூன்று விரைவு /வழக்கநிலை மாணவர்களுக்கு , வகுப்பறைக்கு அப்பால் நிகழும் கற்றலின் முக்கியவத்தை விளக்க , மதுரை அருகேயுள்ள மேலூர்க் கல்லூரியில் பணி புரியும் இரு பேராசிரியர்களக் கொண்டு வெவ்வேறு காலக்கட்டத்தில் இரண்டு பயிற்சிப் பட்டறைகள் நடத்தப்பட்டன . அதற்குத் தொலைஒளிவழிக் கருத்துப்பரிமாற்றத்திற்கான செயலி (Skype) பயன்படுத்தப்பட்டது.

.1'இந்தியச் சிற்பங்கள் காட்டும் வரலாறும் பண்பாடும் ' எனும் தலைப்பில் விரிவுரையாளர் மணிவண்ணன் அவர்கள் புகைப்படங்கள்வழியாகவே நிகழ்த்திய கலந்துரையாடல் , மாணவர்களை 'மரபும் பண்பாடும்' என்னும் கருப்பொருளையொட்டிப் பல கேள்விகள் கேட்கத் தூண்டியது . தமிழ்ப்பண்பாடு விளக்கும் தமிழர் வரலாற்றினையும் பண்பாட்டைப் பேணிக் காக்க வேண்டியதன் நோக் கத்தையும் புதிய நோக்கில் அவற்றை ஆராய வேண்டியதன் அவசியத்தையும் அறிந்துகொண்டனர் .

.2'கதை எழுதுதலில் கற்பனையும் படைப்பாக்கச் சிந்தனைகளும் ' என்னும் தலைப்பில் தொலைஒளிவழிப் பயிற்சிப் பட்டறையை நடத்தியவர் பேராசிரியர் இளங்கோவன் . பய லேபா பெண்கள் பள்ளி உயர்நிலை 3-ஆம் வகுப்பு மாணவர்களும் செயின்ட் மார்கரெட்ஸ் பள்ளி மாணவர்களும் இணைந்து கற்றுக்கொண்டனர் .

படைப்பிலக்கியச் சிந்தனைகளுக்குத் தகுந்த உதரணங்களும் மாதிரிக் கதைகளும் காட்டப்பட்டன . மாணவர்கள் புதுமையான அனுபவத்தைப் பெற்றது மட்டுமல்ல கதை எழுதவும் தூண்டப்பட்டனர் . குழுநிலையில் கதைகளை எழுதி உடனடியாக வகுப்பில் படைத்ததோடு மட்டுமல்லாமல் அதனைத் திறனாய்வு செய்யவும் கற்றுக்கொண்டனர். ஆசிரியரின் வழிகாட்டுதல், மாணவர்களைச் சுயமாகக் கற்றுக்கொள்ளத் தூண்டியது .

சுய கற்றலை நிர்வகித்தலும் கண்காணித்தலும்

கற்றல் என்பது தொடர்நிகழ்வாகும் . சுயவழிக் கற்றலின்போது பெறப்படும் அறிவை முறைப்படுத்தித் தேவைக்கேற்ப மாற்றங்களை நிகழ்த்துவதற்கு நிர்வாகத்திறனும் சிந்தனை மாற்றங்களைக் கண்காணிக்கும் திறனும் அவசியம் . அவற்றிற்கு, மின்னியல் கற்றல் கற்பித்தல் முனையங்கள் அல்லது கற்றல்சார் வாயில்களும் (Academic Portal) பல்வேறு பல்லாடக வசதிகளையும் தகவல் பரிமாற்றச் செயலிகளையும் கொண்டுள்ள சமூக ஊடக வலைக்கட்டமைப்புகளும் (Social Media Networks) பெரிதும் துணைபுரிகின்றன .

கற்றலை நிர்வகித்தல் மற்றும் கண்காணிப்புத் தொடர்பாக மாணவர்களைத் தயார்ப்படுத்த , பள்ளியால் ஏற்பாடு செய்யப்பட்டுள்ள கற்றல்சார் வாயில் (Academic Learning Portal) (தனியாரால் உருவாக்கப்பட்ட , பல இருவழிக் கருத்துப்பரிமாற்றக் கருவிகள்கொண்ட வாயில்) என்னும் இயங்குதளத்தில் அமைந்துள்ள வாய்மொழிப் பயிற்சிக்கான செயலி பயன்படுத்தப்பட்டது . அச்செயலி, புகைப்படங்கள் (Photos), காணொளிகள் (Videos), ஒலிப்பனுவல் (Audio Text) போன்றவற்றை ஏற்றம் செய்வதற்குத் தகுந்தவாறு வடிவமைக்கப்பட்டுள்ளது . மேலும் ஏற்றம் செய்யப்பட்டவற்றைப் பற்றி மாணவர்கள் காணொளியாகவோ ஒலிப்பகுதியாகவோ தட்டச்சு வாயிலாகவோ கருத்துரைக்கத் தேவையான வசதிகளும் உள்ளன . ஆதலால் மாணவர்கள் தங்களுடைய கற்றல் அனுபவங்களைச் சேகரித்து வைத்துக்கொள்வதோடு அவற்றை நிர்வகிக்கவும் தம் கற்றலைக் கண்காணித்து மேம்படுத்தவும் முடியும்.

மேற்கூறிய நோக்கத்தைச் செயற்படுத்தும்பொருட்டு உயர்நிலை 2 வழக்கத் தொழில்நுட்ப மாணவர்களின் வாய்மொழித்திறனை மேம்படுத்த படம் ஏற்றம் செய்யப்பட்டு , அதைப் பற்றி அவர்கள் தம் கருத்துகளைப் பதிவுசெய்யவும் மற்றவர் பதிவுசெய்த கருத்துகளைப் பற்றிய தம் மதிப்பீட்டையும் பதிவு செய்யவும் வழிகாட்டப்பட்டது. இவ்வமைப்பில் மாணவர்கள் தங்களது கருத்துகளைப் பேசிப் பதிவு செய்தனர் . மற்றவர் கூறிய கருத்தைப் பற்றிய தம் மதிப்பீட்டினையும் பதிவு செய்தனர். இதே பயிற்சியை இன்று மிக எளிதாக Vocaroo எனும் நிகழ்நிலை ஒலிப்பதிவுச் செயலிமூலம் பேசி அதை முகநூலில் (Facebook) ஏற்றம் செய்யலாம் . ஒருவரின் கருத்துக்கு மற்றவர் தம் மதிப்பீட்டை அம்முறையிலேயே பதிவுசெய்து அதனைத் தொடர்ந்து ஏற்றம் செய்யலாம் . இவ்வாறான தொழில்நுட்ப வசதிகள் மாணவர்கள் தம் கற்றலைத் தாமே நிர்வகித்துக்கொள்வதையும் கண்காணிப்பையும் எளிதாக்குகின்றதோடு உடனடியாகக் கற்றலில் மேம்பட வழிவகுக்கின்றன .

சுயவழிக் கற்றலில் சமூக வலைத்தளங்கள்

-2012ஆம் வருடம் , அக்டோபர் மாதத்தில் நிகழ்நிலைத் தமிழ்மொழிக் கற்றல் வாயில் (Online Tamil Language Portal) முகநூலில் (Facebook) தொடங்கப்பட்டுள்ளது. மாணவர்கள் அனைவரும் தங்களுக்கெனத் தனித்த (சொந்த உபயோகத்திற்கு பயன்படுத்தப்படாத) முகநூல் கணக்கைத் தொடங்கி தமிழ்மொழிக் கற்றல் வாயிலில் உறுப்பினராகத் தம்மை இணைத்துக்கொண்டுள்ளனர் . மாணவர்களைப் படிப்படியாகச் சுயவழிக் கற்றலுக்குத் தயார்படுத்துவதற்குத் தகவல் தொடர்புத் தொழில் நுட்பத்தைப் பயன்படுத்திய எம் முயற்சி மாணவர்களிடம் ஆர்வத்தைத் தூண்டியுள்ளதோடு இணையத்தைப் பொறுப்பாகப் பயன்படுத்த வேண்டியதன் அவசியத்தையும் வளர்த்துள்ளது எனலாம் . மாணவர்கள் உற்சாகமாகக் கற்க ஆரம்பித்துள்ளனர் . தங்களின் கற்றலுக்குத் தாங்களே பொறுப்பு என்பதையும் உணரத்தொடங்கியதுடன் ஆர்வமாகக் கற்கவும் தலைப்படத் தொடங்கியுள்ளனர் .

நான் சிம் பல்கலைக் கழகத்தில் பகுதி நேர விரிவுரையாளராகப் பணிபுரி வதால் நாட்டுப்புறவியல் பாடத்தை வகுப்பறைக்கு அப்பால் கற்பதற்கு வசதியாக முகநூலைப் பயன்படுத்தினேன் . இந்தியாவில் உள்ள

பேராசிரியர்களும் தங்களது கருத்தை மாணவர்களிடம் பகிர்ந்துகொள்ள வேண்டும் நோக்கில் அவர்களும் முகநூல் குழுவுடன் இணைந்தனர். இள வயதினரும் மூத்த குடிமக்களும் இப்பகுதிநேரப் பட்டய வகுப்பின் மாணவர்களாக உள்ளனர். மூத்த குடிமக்கள் தகவல் தொழில்நுட்பப் பயிற்சி பெற்றிராத காரணத்தால் தகவல் தொழில்நுட்பப் பயன்பாட்டின் வழிச் சுயக் கற்றலுக்கு அவர்களை உட்படுத்துவது தொடக்கத்தில் சவாலாக இருந்தது. ஆயினும் அவர்கள் தொழில்நுட்பம் தொடர்பு அறிவினை சக மாணவர்கள் செய்வதைப் பார்த்துக் பெருக்கிக்கொள்ள வாய்ப்பிருந்ததால் அவர்களும் காலவோட்டத்தில் முழுமையாக நிகழ்நிலைக் கற்றலில் (Online Learning) ஈடுபடத் தொடங்கினர். பகுதி நேரப் பட்டப்படிப்பை மேற்கொள்ளும் அம்மாணவர்களில் தகவல் தொழில்நுட்பப் பயன்பாட்டில் திறன்பெற்றவர்களின் கற்றலில் மாபெரும் மாற்றம் நிகழ்ந்துள்ளது மிகவும் மகிழ்ச்சியளிக்கிறது. (2012FolkloreSIM@groups.facebook.com and <http://future-tamil.blogspot.sg/>)

புதிய தொழில்நுட்பக் கருவிகள்/செயலிகள் வழிக் கற்றல் –ஒரு பார்வை

இன்று கை அசைவில் கணினித் திரைகளை லாவகமாக இயக்கும் வசதிகொண்ட புதிய கருவிகள் (Touch screen gadgets) உருப்பெற்றுவிட்டன. மேலும் இணையத்தில் சுயவழிக் கற்றலுக்கு உகந்த ஏராளமான செயலிகள் இலவசமாகவே காணப்படுகின்றன.

இணையப் பலகை அல்லது நுண்ணறிவுப் பலகை எனக் கருதப்படக்கூடிய ஐ-பேடின் (I-pad) வழிக் காணொளி எடுக்கத் தேவையான திறன்களை வளர்ப்பதற்கான பயிலரங்கு உயர்நிலை மூன்றாம் வகுப்பு மாணவர்களுக்காக நடத்தப்பட்டது.

-2013ஆம் ஆண்டு முதல் சுயவழிக்கற்றல் முகநூல் வாயிலாகச் செயல்படும். வொக்காரோ (Vocaroo), கூகுள்டாக் (GoogleTalk), இன்ஸ்டாகிராம் (Instagram), முகநூலிலேயே அமைந்துள்ள கருவிகள் எனப் பல மின்னியல் வசதிகள் பயன்படுத்தப்படுவதன்வழி மாணவர்கள் சுயவழிக் கற்றலைத் திறம்பட மேற்கொள்வர் என்பது திண்ணம்

சுயவழிக் கற்றலுக்கான பயிற்றுவனங்களை உருவாக்கத் தேவையான வழிகாட்டுதல்கள்

சுயவழிக் கற்றலுக்காக மாணவர்களைத் தயார்ப்படுத்துதலில் தொடக்கத்தில் இடவயச் சூழலும் மின்னியல் சூழலும் இணைந்து பயன்படுத்தப்பட வேண்டியது அவசியமாகிறது. சுயவழிக் கற்றலுக்காகத் மாணவர்கள் தங்களுக்குத் தேவையான கற்றல் வளங்களை உருவாக்கிக் கொள்வதற்காக உயர்நிலை ஒன்றாம் வகுப்பு மாணவர்களுக்கு இணையக் காணொளி விளம்பரங்கள் உருவாக்குவதற்கான பயிற்சி வழங்கப்பட்டது.

வாய்மொழித் திறனையும் படைப்பாக்கத் திறனையும் ஒருங்கே வளர்க்கும் நாடகத் திறன் மேம்பாட்டிற்காக 2012-ஆம் வருடம் உயர்நிலை 2 மாணவர்களுக்கு தனி நடிப்புப் பயிலரங்கு நடத்தப்பட்டு மாணவர்கள் தாங்களாகவே தலைப்பைத் தேர்ந்தெடுத்து நடிக்க ஏற்பாடு செய்யப்பட்டது. மேலும் அவர்களின் செயல்பாடு ஒளிப்பதிவு செய்யப்பட்டு அவர்களாலேயே மதிப்பீடும் செய்யப்பட்டது.

-2013ஆம் ஆண்டில் மாணவர்கள் சுயமாகவே விளம்பரங்களையும் சிறு காணொளிப் பகுதிகளையும் தயாரித்துத் தங்களின் கற்றலுக்கான பயிற்றுவனங்களாகப் பயன்படுத்துவர்.

எதிர்காலத் தேவை

தமிழ் இனித் தனித்து நிற்க, தலை நிமிர்ந்து நிற்க, தரணி போற்ற நிற்க இன்றைய மாணவர்கள் எப்போதும் உலாவிக்கொண்டிருக்கும் இணையத்திலும் கைக்குள் அடக்கிக்கொள்ளும் மின்னியல் கருவிகளிலும் தமிழ் கற்றல் கற்பித்தலுக்குத் தகுந்த வளங்களை இருவழிச் செயல்பாட்டு (interactive) முறையில் உருவாக்க வேண்டியது தமிழர்தம் கடமையாகும். அவ்வகையில் செயல்படின் தமிழ் கற்பிக்க ஆசான் தேவையில்லை. ஆளுக்கொரு மின்னியல் கருவி இருந்தால் போதும். கன்னித் தமிழ் கணனித் தமிழாக என்று ம் இளமை குன்றாமல் அவையத்தை ஆட்சி செய்வாள் என்பது திண்ணம்.

எழுதுதல் திறன் மேம்பாட்டில் “கணினிவழிக் குறைதீர் கற்றல் கருவி”

ச. ஜாஸ்லின் பிரிசில்டா

(ஆய்வாளர்), துணைப் பேராசிரியர், மேரிமாதா கல்லூரி, தேனி.

முன்னுரை

ஒரு மொழியைக் கற்கும் பொழுது கற்போர் பல்வேறு காரணங்களால் பிழைசெய்வது இயற்கை. மாணவர்களின் பிழைகளை இனங்கண்டு அதனைக்களையச் செய்து மொழியினை முறையாகப்பயன்படுத்த கற்றுக்கொடுக்கவேண்டும். மாணவர்களின் பிழைகளைச்சுட்டிக் காட்டுவதாலோ அவர்களின் பிழைக்குப் பதில் சரியான மொழியமைப்பில் எழுதச்செய்தாலோ பிழைகள் முற்றிலும் நீக்கப்படுவதில்லை. மாறாகப் பிழைகளைத்தொகுத்து காரணம் கண்டறிந்து துல்லியமாகப் பிழை நீக்க உதவுவதே பிழை ஆய்வு ஆகும்.

மொழிக்கல்வி என்பது பேசுதல் எழுதுதல் வாசித்தல் கேட்டல் ஆகிய நான்கு திறன்சார் கற்றல் ஆகும் . அதில் எழுதுதல் திறனில் சிறப்புப்பெற இலக்கண ப் பிழையற்ற எழுதுதல் இன்றியமையாதது . பள்ளிக்கல்வி முழுமையிலும் இலக்கணக் கூறுகள் கற்றிருந்தும் கல்லூரிக்கல்வியிலும் இலக்கணப் பிழைகள் நேரும்பொழுது எழுதுதல்த் திறனைக் குறைக்கின்றது . ஆகவே, தவறுகள் நிகழும் கூறுகளில் மட்டும் குறைதீர் கற்றல் வழி பயின்றுகொண்டால் எளிதில் இலக்கணப்பிழைகளைக் களையலாம் . அதனையும் கணினிவழிக் கல்வி ஏற்பாட்டில் தாமே கற்றல் அடிப்படையில் கற்பதற்குத் தேவையான குறைதீர் பயிற்சிக் கருவி வடிவமைத்தல் இன்றையத்தேவையாகும் .

தமிழ் மொழிப்பாடப் பிழைகள்

மொழியின் இயல்பான நிலைக்கு மாறாகப் பேசும்போதும் எழுதும்போதும் ஏற்படும் தவறினை பிழை என்று குறிப்பிடலாம். எழுதும் பொழுது ஏற்படும் பிழைகளை எழுத்துப்பிழை , சந்திப்பிழை, மரபுப்பிழை, திணை-பால்-எண் - இடப் பிழை தொடரமைப்புப் பிழை , கருத்துப்பிழை, பொருள்பிழை எனப் பகுக்கலாம் .. இதனை மொழியியல் அடிப்படையில் முறையே ஒலியியல் , உருபனியல், தொடரியல், பொருளியல் ஆகிய நிலைகளில் ஏற்படும் பிழைகள் எனக் கூறலாம் . மரபிலக்கண அடிப்படையில் எழுத்துநிலை , சொல்நிலை, பொருள்நிலைப் பிழைகள் எனப் பகுக்கலாம் . மேலும், கல்வியில் பிழை ஆய்வு அடிப்படையில் , மாணவர்கள் தமிழ் மொழி கற்கும் பொழுது செய்யும் பிழைகளா கப் பின்வரும் ஏழினைக் கூறலாம் : ஒப்புமையாக்கம், விதிவிலக்கு அறியாமை , தவறான கருத்துருவாக்கம், எளிமையாக்கல், தூண்டுதல், பேச்சுவழக்கின் தாக்கம், பயிலாமை.

பிழைகளே செய்யாமல் மாணவர்கள் கற்றுக்கொள்ள இயலாது. மாணவர்கள் ஒவ்வொரு பிழை செய்யும் பொழுதும் கற்றுக்கொள்கிறார்கள். பிழைகளைக் களைந்து சரியாகக் கற்கும் பொழுது கற்றலில் முன்னேறுகிறார்கள். பிழைகள் மாணவர்களின் கற்றலில் நேர்ந்த குறை என்பதால் ஆசிரியர்கள் பிழைகளுக்கான காரணத்தைக்கண்டுபிடித்து களையா விட்டால் மனதில் ஆழப்பதிந்து கல்மேலெழுத்துப் பிழை என மாற வாய்ப்புள்ளது.

பிழைகளை இனம் காணல்

எந்த ஒரு துறையிலும் கோப்புகளைப் பாராமரித்தல் , ஊடகங்களுக்கு எழுதுதல் , படைப்பிலக்கியம், வணிகக்கடிதங்கள் என்று எல்லா மொழிவழி இயக்கத்திற்கும் எழுதுதல்திறன் மிகவும் முதன்மையானது . ஆகவே மாணவர்களின் எழுதுதல்திறனில் இலக்கணத்திறனை மேம்படுத்துதல் வேண்டி அவர்களின் இலக்கணக்

குறைகளைக் களையும் பொருட்டு கணினிவழிக் குறைதீர் பயிற்சிக் கருவி உருவாக்குதலான இவ் வாய்வில் மதுரைக் காமராசர் பல்கலைக்கழக அங்கிகாரம் பெற்ற மேரிமாதா கல்லூரி , நல்லகருப்பன்பட்டி, பெரியகுளம் வட்டம், தேனி மாவட்டத்தில் இளங்கலை முதலாமாண்டு இரண்டாமாண்டு பொதுத் தமிழ் பயிலும் மாணவர்களின் மாதிரித் தேர்வுத்தாளில் காணப்படும் பொதுவான கீழ்க்கண்ட இலக்கணப் பிழைகள் ஆய்வாளரால் தேர்வுத்தாள் வழி ஆய்வு செய்து கண்டறியப்பட்டுள்ளது .

1. ஒரே மாதிரியான உச்சரிப்பு ஒலியினைப் பெற்று வேறு வரிவடிவத்தினை கொண்டுள்ள எழுத்துக்களை எழுதுவதில் ஏற்படும் பிழைகள் .

"இந்த அண்புடைமை மிகவும் மனிதனுக்கு இருக்க வேண்டிய ஒன்றாகும் ."கேள்வி எண் 17, திருமந்திரம்- அண்புடைமை, தேர்வு எண்: B1120256, துறை: IT (2nd Year) .

2. ஒற்று மிகும் மிகா விதி அறியாமையால் ஏற்படும் பிழைகள் .

"....இவர் காமராசர் பல்கலைக்கழகத்தில் தமிழ் துறை தலைவராக விளங்கியவர்"கேள்வி எண்:15, புதுக்கவிதையின் தோற்றமும் வளர்ச்சியும், தேர்வு எண்: 12MMC2525, துறை: BCA (First Year)

3. மரபு அறிவின்மையால் ஏற்படும் பிழைகள் .

4. திணை-பால்-எண் -இட வழக்கலால் உண்டாகும் பிழைகள் .

"சீதை கோபத்துடன் வேண்டாம் நீங்கலே பிடித்து வாருங்கள் என்றான் ."- கேள்வி எண் 13. 'தெய்வம் நீ என்று உணர் ' (நாடகம்), தேர்வு எண்:12MMC3206, துறை: B.comE.com (First Year)

5. தொடரமைப்பில் உண்டாகும் பிழைகள்

6. கருத்துத் தெரிவின்மையால் உண்டாகும் பிழைகள் .

சரியாகக் கற்காமையால் மழைக்குறி வெள்ளம் என்பதை திருக்குற்றாலக் குறவஞ்சியின் மலை வளத்தோடு கலந்து தவறாக எழுதப்பட்ட விடை."மலைக்குறி வெள்ளம் என்ற தலைப்பில் சிற்றிலக்கியங்களில் ஒன்றாக இது திகழ்ந்து வந்துள்ளன."கேள்வி எண்:18, முக்கூடற்பள்ளு, தேர்வு எண்:B1119985, துறை: B.Sc.CS (2ndYear)

7. பொருள் வழக்கலால் உண்டாகும் பிழைகள் .

இதனை மேலும் இரண்டாகப் பகுத்துக் காணலாம் .

1. மரபிலக்கண தேர்ச்சி இன்மையால் ஏற்படும் பிழைகள் .
2. பாடப்பொருள் பயிலாமையால் உண்டாகும் பிழைகள் .

குறைதீர் கற்றல் கருவி

மாணவர்கள் ஏற்படுத்தும் பிழைகள் என்பவை கற்றலில் உள்ள தடைகள் . பிழைகளுக்கு ஏற்ப குறைத்துக் கொண்டு மதிப்பெண் வழங்குதல் என்பது கற்றலை முழுமைப்படுத்த உதவாது. ஆகவே, மாணவர்களிடம் பிழை

ஆய்வு செய்தவுடன் அதனை களைவதற்குத் தேவையான குறைதீர்க் கற்றல் கருவியை உருவாக்கி மாணவர்களைக் கற்கச் செய்து முழுமையாகப் பிழைகள்நீக்கிக் கற்றலில் மேம்பாடு அடையச் செய்ய வேண்டும் . கற்றலில் முழுமைத்தன்மை ஏற்பட மரபு முறை கற்றலைவிட நு ட்பவழி அல்லது கணினிவழிக் கற்றலே மிகுந்த பயனுடையது .

மாணவர்களின் தேர்வுத்தாளில் சந்திப்பிழைகளின் விகிதம் பொதுவாகவே அதிகமாகவும் அனைவர் தாட்களிலும் காணப்பட்ட சூழலில் ஆய்வாளர் கணினிவழிக் குறைதீர் கற்றல் கருவிக்கு சந்திவிதிகள் பகுதியினைப் பாடப்பொருளாக எடுத்துக் கொண்டு தாமே கற்றல் அடிப்படையில் கற்றல் பாட மாதிரியை உருவாக்கி இருக்கிறார். இது html மற்றும் visual basic கணினி நிரல் மொழிகளில் உருவாக்கப்பட்டது . இது இந்த ஆய்வுத் தாளுக்காக மிகக் குறுகிய கால அளவில் தயாரிக்கப்பட்டது . இந்த குறைதீர் கற்றல் கருவியினை மாதிரியாகக் கொண்டு மேற்கொண்டு ஆய்வாளர்கள் பிற கணினி மொழிகளிலும் பல்வேறு உத்திகளைப் பயன் படுத்தியும் தயாரிக்கலாம். ஆய்வாளர் இக் கருவியில் புறவய வினா அமைப்பில் வினாக்கள் அமைத்து விடை சொல்லத் தவறும் நிலையில் பாட விதிகளைப் படித்து பின் மீண்டும் வினாவிற் குப் பதிலளிக்கும் வண்ணமாய் வடிவமைத்து உள்ளார். இதன் பயனாக முழுமையாக ஒரு விதியினைக் கற்றுத்தேராமல் அடுத்த நிலைக்குச் செல்ல இயலாது .

மரபுமுறைக் குறைதீர் கற்றலில் உள்ள தடைகள்:

1. முன்னதாக, வகுப்பறையில் குழுக் கற்பித்தலில் , மாணவர்களிடம் ஏற்பட்ட கற்றல் குறைபாடுகளைக் களைய மீண்டும் கு ழு கற்பித்தல் முறை சிறப்பாக உதவாது . குறைதீர் கற்றலுக்கு தனிப் பயிற்சி என்பதே சிறந்தது .
2. ஒவ்வொரு மாணவனும் தனித்துவமானவன் . வகுப்பில், தனித்தனியே ஒவ்வொருவரையும் சந்தித்துத் தனிப்பயிற்சி அளிக்க ஒரு ஆசிரியரால் இயலாது .
3. வகுப்பில் அத்துணை மாணவர்களுக்கும் ஒரு ஆசிரியர் தனிப்பயிற்சி அளிக்க கால நேரம் போதாது .
4. மேலும் அனைத்து மாணவர்களும் ஒரே மாதிரிப் பிழைகளைச் செய்வதில்லை . ஆகவே, அவரவரின் பிழைகளுக்கு ஏற்ப குறைதீர் கற்றல் தேவைப்படுகிறது .

முன் கூறிய அனைத்துத் தடைகளையும் களைந்து கற்றலை மேம்பாடு அடைய இயலாமையே மரபுமுறைக் கற்றலின் மிகப்பெரும் குறைபாடு அல்லது தடைகள் ஆகும் .

கணினி வழிக் குறைதீர் கற்றலில் நன்மைகள்

1. மாணவர்கள் அனைத்துப் பேருக்குமான குறைதீர் கற்றல் கருவி உருவாக்கினாலும் தனியொரு மாணவரின் தேவைக்கு ஏற்ப கற்றுக்கொள்ள முடியும் .
2. மாணவர் வகுப்பறைச் சூழல் தாண்டி தாமே குறைதீர் கற்றல் நிகழ்த்த கணினி வழிக் குறைதீர் கற்றல் கருவி பயன்படுகிறது .
3. ஒரே நேரத்தில் அனைத்து மாணவர்களும் தத்தமக்குரிய குறைதீர் கற்றலை நிகழ்த்திப் பிழைகளைக் களைய கணினி வழிக் குறைதீர் கற்றல் கருவி மிகவும் உதவுகிறது .
4. தனியொரு மாணவனின் கற்கும் திறனுக்கு ஏற்ப ஒருமுறை யோ பலமுறையா மீண்டும் மீண்டும் தாமே கற்றுத் தேற கணினி வழிக் குறைதீர் கற்றல் கருவி துணை செய்கிறது .
5. முழுமையான நூறு சதம் கற்றறிதல் என்பதனை அடையும் வரை பல்வேறு பயிற்சிகள் மற்றும் மதிப்பீட்டு முறைகளைப் பயன்படுத்தி குறைதீர் கற்றலை நிகழ்த்த இயலும்

கணினி வழிக் குறைதீர் கற்றலில் தடைகள்:

1. போதுமான கணினிவழிக் குறைதீர் கற்றல் பாடப்பொருள் பயிற்சிக் கருவி இன்மை .
2. தமிழகச் சூழலில் ஆசிரியருக்கும் மாணவருக்கும் நுட்பவழிக் கற்றலில் பயிற்சியின்மை .

3. தமிழ் மொழிக்கல்வி குறித்தான உயர்மனப்பான்மை இன்மை . ஆகவே, முப்பத்தி ஐந்து சதவிகிதத் தேர்ச்சி போதுமென்ற மனப்பான்மை.

எதிர்காலத் தேவை

1. கணினிவழிக் கற்றல் பாடப்பொருள் உருவாக்கும் பொழுதே அதனைச் சோதித்துப் பார்த்துக் குறைதீர் கற்றலை நிகழ்த்தத் தேவையான பாடப்பொருள் கருவிகளை உருவாக்குதல் வேண்டும் .
2. இதுவரை கற்றோரின் கற்றல் குறைகளைத் தீர்க்கத் தமிழக மெங்கிலும் தமிழ் மொழிப்பாடப் பிழை ஆய்வினைச் செய்து பிழைகளைக் குறைதீர்க் கற்றல் கருவிகளின் துணை கொண்டு களைய வேண்டும் .
3. தமிழ் மொழிப்பாடம் குறித்தான உயர் மனப்பான்மையைத் தமிழகப் பள்ளி கல்லூரிகளில் ஏற்படுத்த வேண்டும். அதற்கு அரசாங்கம் ஆறாம் வகுப்பு புகுவ தற்கு முன் மாணவர்களுக்கு அடிப்படைத் தமிழ்மொழி இலக்கணக் கட்டமைப்பு அறிதலுக்கான தேர்வினை தமிழக அளவில் கட்டாயமாக்க வேண்டும் . அதில் குறையிருப்பின் அதனை நிவர்த்தி செய்துவிட்டு ஆறாம் வகுப்பில் நுழைவதை வலியுறுத்த வேண்டும்.
4. எட்டாம் வகுப்பு இறுதியில் தமிழ்மொழி வன்மை யைச் சோதித்தலுக்கானத் தேர்வினை கட்டாயமாக்க வேண்டும். சான்றாக, எழுதுதல் திறனில் கடிதம் , கட்டுரை, சுருக்கி வரைதல் இவை போன்ற புலமை சோதிக்கப்பட வேண்டும்.
5. எல்லாவற்றிக்குமாகக் கணினி வழியிலான கற்றல் பாடப் பொருளும் பயிற்சியும் சோதித்தல் கருவியும் தமிழ்மொழிப்பாடத்தில் உருவாக்க வேண்டியது உடனடித் தேவை . இதற்குத் தமிழ் வல்லுனர்கள் கல்வியியல் வல்லுனர்கள் மென்பொருள் தயாரிப்பாளர்கள் அனைவரும் இணைந்து பணியாற்ற வேண்டியது இக்காலத் தேவை.

கட்டுரை ஆக்கத்திற்குப் பயன்பட்ட நூல்கள்

1. சங்கர், வே ., இயல் தமிழ் இலக்கணம், நன்மொழிப் பதிப்பகம், 41, கங்கை வீதி, வசந்த் நகர், புதுவை, முதற்பதிப்பு, ஆகஸ்டு 2004.
2. சங்கர், வே ., மொழித்திறனும் மொழிப்பயிற்சியும் , நன்மொழிப் பதிப்பகம் , 41, கங்கை வீதி , வசந்த் நகர், புதுவை, முதற்பதிப்பு, டிசம்பர், 2004.
3. Rekha Yadav Dr., EDUCATION TECHNOLOGY, University Science Press, 113, Golden House, Daryaganj, New Delhi, First Edition, 2011.
4. Subbaiah pon., Balakumar, M., ASSESSING MASTERY OF LANGUAGE, Central Institute of Indian Languages, Mysore, India, Publication no. 505, First Publication, Feb. 2003.
5. Subbaiah pon. (Editor), Balakumar, M. (Assistance), TEST OF LANGUAGE PROFICIENCY for Secondary (std.X) level, Central Institute of Indian Languages, Mysore, India, Publication no. 490, Second Publication, Jan. 2011.

கணினி வழி தமிழ் கற்றல் கற்பித்தல் குறித்த ஆய்வுகள்

Siva Pillai

Visiting Lecturer, Goldsmiths, University of London. UK

Chief Examiner Edexcel- Tamil Language- London-UK

Chief Examiner/ item writer: Cambridge ASSET Examination Board. UK- Tamil Language

pillaisiva@gmail.com

புலம்பெயர் நாடுகளில் உள்ள பல பிரச்சனைகளுக் குள் முக்கியமாகப் பார்க்கப்படுவது எமது சந்ததியினருக்கு தமிழ் கல்வி அறிவை ஊட்டுவதே ஆகும் , அதற்குப் பெற்றோர்களும் சமூக புத்திஜீவிகளும் பல வழிகளில் முயற்சிக்கிறார்கள். ஆனால் அவர்கள் மாணவர்களின் கருத்தை உள்வாங்குவதில் பின் தங்கி நிற்கிறார்கள் .இந்த சூழ்நிலையில் இளம் தலைமுறையினருக்கு தமிழ் கற்பதில் வெறுப்பு உண்டாகிறது .

இதே சமயம் அவர்கள் தமது முன்பள்ளிகளில் ஐரோப்பிய மொழிகளை விரும்பிப் படிக்கிறார்கள் . அம்மொழி வீட்டில் பேசப்படாது இருந்த போதும் அந்த மொழிகளில் புலமை அடைகிறார்கள் . புலம் பெயர் நாடுகளில் எவ்வாறு தமிழ் மொழியை மாணவர்கள் விரும்பும் வகையில் கற்பிக்கலாம் என்பதே இன்றைய கேள்வி.

சமூக மன்றங்களால் நடத்தப்படும் தமிழ்ப் பள்ளிகளில் உள்ள கற்றல் முறையைப் பற்றி மாணவர்களிடம் கேட்கையில் அவர்கள் அது “போரிங்”, அதாவது ஆர்வம் ஊட்டும் வகையில் தமிழ் கல்வி கற்பிக்கப்படுவதில்லை என்கிற கருத்தினை முன்வைக்கிறார்கள் .

ஐரோப்பிய மொழிகள் முன்பள்ளிகளில் பயிற்றப்பட்ட ஆசிரியர்களால் கற்றுக் கொடுக்கப்படுகின்றன .புலம்பெயர் நாடுகளில் சமூக மன்றங்களால் நடத்தப்படும் , வார இறுதி நாள் களில் மற்றும் வார நாள்களில் முன்பள்ளி முடிந்ததும் நடக்கும் பள்ளிகளில் தமிழ் கற்பிக்கும் ஆசிரியர்களுக்கு, அந்தந்த நாட்டுக்கேற்ப கல்வி கற்பிக்கும் ஆசிரியப் பயிற்சி பெறும் வாய்ப்பு இல்லாததால், அவர்கள் தமக்கு ஏற்ற வகையில் (மாணவர்களுக்கு ஏற்ற வகையில் அல்ல) தமிழைக் கற்பிக்க முயற்சிக்கிறார்கள் . இவ்வாசிரியர்கள் எல்லோரும் தன் னார்வம் கொண்ட, ஊதியமி ன்றிக் கற்பிக்கும் ஆசிரியர்கள் ஆவார்கள். இதில் சிலர் தத்தமது ஊர்களில் பயிற்சி பெற்ற ஆசிரியர்களாக இருக்கிறார்கள் .பழைய முறையிலேயே கற்றுக் கொடுக்க முற்படும் போது மாணவனுக்கு வெறுப்பு உண்டாகிது.

கற்பித்தலில் புதிய வழி முறைகளைக் கையாளா து விடும் போது கற்றலில் மாணவர்களுக்கு ஆர்வம் குறைந்து, அதில் ஈடுபாடு அற்றுப் போகிறது . தமிழ் மொழியை கேட்டல், பேசுதல், வாசித்தல், எழுது தல் என நான்கு வகையாகப் பிரித்துக் கற்பி த்தால் மாணவனுக்குத் தமிழைக் கற்பதில் ஆர்வமுண்டாகும். ஏனெனில், அவன் ஏனைய மொழிகளுடனான தனது கற்கையுடன் தனது தமிழ் கற்கையையும் ஒப்பிட்டுப் பார்க்கிறான் . அது அவனுக்கேற்ற எளியமுறையில் அமையாதபோது அவன் பெற்றோரின் உந்தலுக்காகத் தமிழ் கற்கச் செல்கிறான். இவ்வாறான சூழ்நிலையை நாம் புலம் பெயர் நாடுகளில் காண லாம் .சிலர் ஆர்வம் கொண்டு தமிழைக் கற்கிறார்கள், சிலர் வெறுப்படைந்து வேறுதிசையில் தமது ஆர்வத்தைத் திருப்புகிறார்கள் . கற்றல் கற்பித்தலில் ஆசிரியர் பங்கும் பெற்றோர் பங்கும் முக்கிய இடத்தை வகிக்கிறது .

மாணவர்களுக்குத் தெரிந்ததைக் கொண்டு தெரியாததைக் கற்றுக் கொடுப்பது ஆசிரியரின் முக்கி ய பங்கு ஆகும் . அவர்கள் விரும்பும் கேலிச் சித்திரத்தின் மூ லம் நாம் பேசாமலே யே அவர்களை பேசப் பழக்கலாம் .பல கேலிச் சித்திரப் பகுதிகளை எல்லா மொழிகளுக்கும் ஏற்ற வகையில் பயன்படுத்தலாம்.

முதலில் குழந்தைகக்கு விருப்பமானது என்ன என்பதைக் கண்டறிந்து அதைச் செய்து கொடு த்தால் அவர்களுக்குக் கற்பதில் ஆர்வம் உண்டா கும். ஒரு படத்தைக் காண்பித்து குழந்தைகளை அது பற்றிப் பேசச் சொல்லலாம்.அவர்களிடம் ஏதேனும் ஒரு விஷயத்தை அறிமுகப் படுத்தி அதைப் பற்றி மேலும் விவரங்களை

விவாதிக்கச் சொல்லலாம். இவ்வாறான செயல்கள் குழந்தைகளது சுய சிந்தனைத்திறன் மற்றும் கற்பனைத்திறனை மேம்படுத்துவதில் உதவி புரியும்.

மேலும், ஒரு காட்சிப் பொருளை மாணவனின் முன்வைத்து அவனுக்குத் தெரிந்த வழியில் அதைப் பற்றிப்பேசச் சொல்லலாம். அவனது அன்றாட வேலைகள் என்ன என்பதைப் பற்றி வகுப்பிலுள்ள அனைவரிடமும் எடுத்துரைக்கச் சொல்லலாம். இச் செயல்களின் வாயிலாக ஒரு தன்னம்பிக்கையை மாணவனிடத்தில் உண்டாக்கலாம்.

இன்று கணினியைப் பல வழிவகைகளில் பயன் படுத்த முடியும். இதன் மூலம் மாணவர்களுக்குப் பிடித்த முறையில் மொழியைக் கற்பிக்கலாம்.

முக்கியமாக மாணவர்களிடையே தாம் கற்கும் பாடங்களை கணி வியில் உருவாக்கக் கேட்கலாம். இதன் மூலம் அவர்களையும் கற்பிக்கும் நிலைக்கு கொண்டு வர முடியும். இவ்வழிமுறை மாணவர்களுக்கிடையில் ஒரு தொடர்பாடலை ஏற்படுத்தும்.

மாணவர்களுக்கு நிறைய கதைகளைக் கேட்கச் செய்து மற்றையோருக்குக் கூறச் சொல்வதன் மூலம் அவர்களதுனது ஆக்கத்திறனை மேம்படுத்தலாம் . மேலும் அவர்கள் பேசுவதைப் பதிவு செய்து கேட்கச் செய்யலாம். தமிழ்ச் சொற்களை உச்சரித்துப் பழகுவதற்கும் கணினி மிகவும் பயனுள்ளதாக அமைகிறது. .

இன்றையமின்னணு உபயோகப் பொருட்கள் யாவும் படங்கள் கொண்டனவாகவே உள்ளன, உதரணமாக ஐபோனைப் பார்த்தால் திரையில் ஏதும் எழுத்தில் இல்லை . படத்தில் மட்டுமே தகவல்கள் தரப்பட்டுள்ளன , ஒரு பொத்தானை அழுக்குவதன் மூலம் அந்த மென் பொருள் செயல்படுகிறது . விளக்க எழுத்துரு அதில் இல்லை .



படத்தில் காட்டப்பட்டுள்ளது போல் பிளாஸ்டிக்கில் வெட்டிய தமிழ் எழுத்துகள் இன்று தொட்டுணர்தல் வழி மூலம் கற்கும் வளங்களாக உபயோகிக்கப் படுகிறது .தமிழ் எழுத்துக்களைத் தடவிப் பார்த்து குழந்தைகள் எழுத்துருவை மனப்படம் செய்கிறார்கள்.



மின்வெண்பலகை (IWB)

அது போன்றனவே மின்வெண்பலகைகளும்; அவற்றில் கையால் தொட்டு எழு தலாம். நமது உணர்வுகள் எல்லா அறிவையும் வளர்க்கும் தூண்டுகோள்கள் .இதை ஒவ்வொரு குழந்தையும் த னது வாழ்வில் உணர்ந்து, தன்னிச்சையாகச் செயல்பட முடியும் .தொட்டு உணர்ந்து கற்கும் வளங்கள் இன்று கற்றல் கற்பித்தலுக்கு அதிகம் கிடைக்கின்றன .

நாட்டுக்கு நாடு மாணவர்கள் செய்திகளையும் தமது படைப்புகளையும் பரிமாறக் கூடியதாக உள்ளது . கணப்பொழுதில் நாட்டுக்கு நாடு தொடர்பாடலை நாம் மேற்கொள்ளலாம் . தற்போது மாணவர்களுக்கு முன் பள்ளியிலேயே விடையோ பயிற்சி யும் அளிக்கிறார்கள் .இத்தொழில் நுட்பம் கற்றல் கற்பி ப்பதில் பல மாற்றங்களை ஏற்படுத்திக் கொண்டிருக்கிறது , வருங்காலச் சந்தயினரிடம் தற்காலத்திற்கேற்ற தொழில்நுட்ப வழியில் தமிழ் மொழியைக் கற்கச் செய்வது நமது கடமையாகும்.

உயர்கல்வியில் தொழில்நுட்பத்தைப் பயன்படுத்துவதில் விழித்திறன் குறைந்தவர்களின் பிரச்சினைகள் பற்றிய ஓர் ஆய்வு

முனைவர். ச.அமுதா

உதவிப் பேராசிரியர், கல்வி நுட்பவியல் துறை, பாரதிதாசன் பல்கலைக்கழகம், திருச்சிராப்பள்ளி

ஆய்வுச் சுருக்கம்

அரிது அரிது மானிடராய்ப் பிறத்தலரிது அதனினும்
கூன் குருடு செவிடு பேடு நீங்கி பிறத்தலறிது.

- ஒளவையார்

பண்டைக் காலத்தில் ஒரு குழந்தை குறைபாட்டுடன் பிறந்துவிட்டால் அது அந்த குறைபாட்டுடனே வீட்டிலேயே இருந்து கொண்டு உலகத்தோடு தொடர்பு இல்லாமலே தனக்கென தனி உலகத்தை வகுத்து வாழ்ந்து இறந்துவிடும். பெற்றோர்களுக்கு போதிய விழிப்புணர்வு இன்மையே இதற்கு காரணமாகும். ஆனால் இந்த நிலைமை படிப்படியாக மாற்றம் அடைந்து உடல் குறைபாடுடையவர்களை சமூகத்தில் அங்கீகரிக்கத் தொடங்கிவிட்டனர்.

வளர்ந்துவரும் தொழில்நுட்பம் இப்பூமியை ஒரு உலகளாவிய கிராமமாகச் சித்தரித்துள்ளது. புதுப்புது தொழில் நுட்பங்கள் நாளொருமேனியும், பொழுதொரு வண்ணமுமாய் வந்து கொண்டிருக்கிறது. தகவல் தொழில்நுட்பப் புரட்சியில் முதன்மையானது இணையதளம் ஆகும். கணினி உதவியுடன் இணைய தளத்தைப் பயன்படுத்தி உலகின் எந்த மூலையில் உள்ள தகவல்களையும் உடனுக்குடன் பெற முடியும். உயர்கல்வியில் ஒரு பாடத்திற்கான பாடத்திட்டத்தை உள்ளடக்கிய வளநூல் இல்லை. பாடத்திட்டத்தைக் கற்கப் பல வளநூல்களின் உதவியை நாட வேண்டியுள்ளது. எந்தக் குறைபாடும் இல்லாத சாதாரண மாணவர்களே பாடத்தொடர்புடைய தகவல்களைப் பெற சிரமப்படுகின்றனர். இந்த நிலையில் விழியிழந்தோரின் நிலைமை பற்றி சிந்திக்க வேண்டியுள்ளது. விழியிழந்தோரில் பலர் நமது வாழ்க்கை சூனியமாகிவிட்டது என்று எண்ணி இருந்து விடுகின்றனர். ஆனால், சிலர் சாதாரண மாணவர்களுக்கு இணையாகத் தங்களின் அறிவை வளர்த்துக் கொள்ள வேண்டும், அவர்களுக்கு இணையாக நவீன தொழில் நுட்பத்தைப் பயன்படுத்தித் தகவல் அறிவில் சிறந்தவர்களாகத் திகழ வேண்டும் என்று விரும்புகின்றனர். மேலும், உயர்கல்வியை முடித்த பலரும் ஆசிரியராகவே விரும்புகின்றனர். கற்பித்தல் சரிவர நிகழ வேண்டுமானால் கற்றல் சரிவர நடைபெற வேண்டும். உயர்கல்வி பயிலும் போதே தகவல்களைப் பெறுவதில் சிரமங்கள் ஏற்பட்டால், கற்றலில் தொய்வு ஏற்பட்டு பின்நாளில் கற்பித்தலில் தடை ஏற்படும்.

வெவ்வேறு தகவல்களைத் திரட்ட வேண்டுமானால், இணைய தளம், பல்வகை மின் புத்தகங்கள், மின்ஆய்வுக் கட்டுரைகள், செய்திகள் இவற்றை நாடவேண்டியுள்ளது. ஒவ்வொரு செய்திகளையும் அறிய தனக்கென ஒரு துணை தேவைப்படுகிறது. நடைமுறையில் இது சாத்தியப்படவில்லை. ஆகவே, ஒவ்வொரு மாற்றுத்திறனாளியும், தன்னிச்சையாகச் செயல்பட வேண்டிய கட்டாயச் சூழல் ஏற்படுகிறது. அந்தச் சூழலை எவ்வாறு எதிர்கொள்வது என்று பார்த்தால், தொழில்நுட்பத்தின் தேவை அவசியமாகிறது. கணினித் தொழில்நுட்பத்தினை அறிந்த விழியிழந்தவர்கள் தன்னம்பிக்கையோடு இருக்கிறார்கள் என ஆராய்ச்சிகள் கூறுகின்றன. ஆனால் விழியிழந்த மாற்றுத் திறனாளிகள் நடைமுறையில் தொழில்நுட்பத்தினைப் பயன்படுத்திப் படிப்பதில் என்னென்ன சிரமங்கள் எதிர்கொள்ளப்படுகின்றன, எந்த அளவு அவர்களுக்கு தொழில்நுட்பத்தினைப் பற்றிய விழிப்புணர்வு உள்ளது என்பது பற்றியும் தகுந்த ஆதாரங்களுடன் நேரடியாகக் கண்டுணர்ந்து, அதை வெளிப்படுத்துவதே இவ்வாய்வின் நோக்கமாகும்.

நவீன சிறப்புக் கல்வி முறையின் கீழ் பயன்படுத்தப்படும் உயர் தொழில்நுட்ப சாதனங்கள், கணினி மென் பொருட்கள் தற்போதைய பயன்பாட்டிலிருந்து மறைந்து போன ஒலி நாடாக்கள், இன்று பயன்பாட்டில் உள்ள சக்தி வாய்ந்த நவீன ஐ-பேடுகள் இவை கற்றலில் ஏற்படுத்தக்கூடிய நன்மைகள், இவற்றால் ஏற்படும் சவால்கள் என்ன என்பது குறித்தும் சமுதாயத்தில் எந்த அளவு விழிப்புணர்வு உள்ளது என்பது பற்றியும் இவ்வாய்வின் மூலம் தெரிந்து கொள்ளலாம்.

ஆய்வின் நோக்கம் (Objectives)

கீழ்க்கண்டவை ஆய்வின் நோக்கங்களாக உள்ளன .

- விழியிழந்தவர்களின் தகவலறியும் திறனை அறிய .
- விழியிழந்தவர்களின் உயர்தொழிநுட்பக் கருவிகளின் பயன்பாடு பற்றி அறிய
- விழியிழந்தவர்கள் கல்வி மேம்பாட்டிற்காக அடையும் சிரமங்களைத் தெரிய .
- விழியிழந்தவர்கள் பல்வேறு தொழில்நுட்பக் கருவி களைக் கையாளும் விதத்தினைத் தெரிய .
- தொழில்நுட்பக் கருவிகளின் பயன்பாட்டில் ஏற்படும் பிரச்சினைகளும் , தீர்வுக்கான ஆலோசனை வழங்கவும்.

ஆய்வு வினாக்கள் (Research Questions)

பின்வரும் ஆய்வு வினாக்களுக்கு விடை தேடுவதற்காக ஆய்வு நோக்கத்தை நோக்கி இவ்வாய்வு முன்னெடுத்துச் செல்லப்பட்டுள்ளது.

- விழியிழந்தோர் தகவல்களைப் பெற ஏற்படும் இடர்பாடுகள் யாவை ?
- உயர்கல்வியில் தொழில்நுட்பப் பயன்பாடு விழியிழந்த மாற்றுத்திறனாளிகளை எவ்வளவு சென்றடைந்துள்ளது?
- கணினி பயன்பாடு குறித்த விழிப்புணர்வு விழியிழந்த மாற்றுத்திறனாளிகளிடம் எவ்வளவு உள்ளது ?
- எந்தெந்தக் காரணிகள் விழியிழந்த மாற்றுத்திறனாளிகளின் கற்றல் மேம்பாட்டிற்குத் தடையாக உள்ளது ?
- கற்றலில் உயர்தொழில்நுட்பக் கருவிகளின் பயன்பாடு மாற்றுத்திறனாளிகளிடையே எவ்வாறுள்ளது?

ஆய்வுப் பிரதேசமும், மாதிரி எடுத்தலும்

ஆய்வுக்கான மாதிரிகள் தமிழ்நாட்டின் பல்வேறு பகுதியிலிருந்து வந்து திருச்சிராப்பள்ளியில் கல்லூரி பயிலும் மாணவர்களாவர். இவர்களுள் இரண்டு மாணவர்கள் பல்கலைக் கழகத்திலும் மற்றைய ஏழு மாணவர்கள் அரசு உதவி பெறும் கல்லூரிகளிலும் கற்கின்றனர் .

வினாநிரல்

ஆய்வாளரால் தரவுகளைத் திரட்டுவதற்காகத் தயாரிக்கப்பட்ட வினாநிரலானது 15 வினாக்களைக் கொண்டது . சேகரிக்கப்பட்ட தகவல்கள் யாவும் , ஆய்வாளரால் முன்வைக்கப்பட்ட ஆய்வு வினாக்களை சோதித்தறியும் வகையில் உள்ளது . இதன் பொருட்டு மாதிரிகள் விழியிழந்தோர் என்பதால் அவர்களால் கூறப்பட்ட பதில்கள் ஒலிநாடாவில் பதிவு செய்யப்பட்டது . வினாநிரல் மட்டும் தராமல் நேர்காணல் முறையிலும் அவர்களிடமிருந்து தகவல் பெறப்பட்டது . தரப்பகுப்பாய்வு செய்யப்பட்டு இனங்காணப்பட்ட பிரச்சினைகள் மூலம் பெறப்பட்ட தகவல்கள் (Findings) தீர்வுக்கான ஆலோசனைகள் கீழே தரப்பட்டுள்ளன :

1. தொழில்நுட்பத்தை உயர்கல்வியில் பயன்படுத்தும் மாணவர்கள் மிகவும் குறைவான எண்ணிக்கையில் காணப்படுகின்றனர் .
2. கணினி அறிவு இருந்தால்தான் தொழில் நுட்பங்களைப் பயன்படுத்த முடியும் .
3. அமெரிக்க உச்சரிப்பில் இருப்பதால் எல்லாவற்றையும் புரிந்து கொள்ள முடிவதில்லை . மேலும் மிகுந்த சிரமத்துடன் உற்று கவனித்தால்தான் வார்த்தைகள் புரியும் .
4. பல ஆங்கில வார்த்தைகளுக்கு அர்த்தங்கள் புரிவதில்லை .
5. புத்தகங்களில் உள்ள சொல்தொகுதி (Vocabulary) அதிக கடினத்தன்மையுடன் இருப்பதால் புரிந்து கொள்வது சிரமமாக உள்ளது .
6. இணையத்தின் மூலம் தகவல்களைப் பெறுவதற்கு மற்றொருவர் உதவி தேவைப்படுகிறது .
7. தாய் மொழியான தமிழ் மொழியில் இல்லாததால் தமிழ் பயின்ற மாணவர்கள் தொழில்நுட்பத்தினால் பயன் அடையவில்லை .
8. பிரைலி மூலம் படிக்கும்போது செலவு அதிகமாகிறது .

9. வெவ்வேறு வகையான மாற்றுத்திறனாளி களுக்கு தேவைப்படும் தொழில்நுட்பத்தை பயிற்றுனர் கற்று வைக்கவில்லை.
10. பாட ஆசிரியர்கள் தொழில்நுட்பத்தைப் பயன்படுத்தும் போது அருகில் இல்லாததால் அர்த்தம் புரியாமலே படிக்க நேரிகிறது.
11. பயிற்றுனர் அதிகநேரம் மாணவர்களின் குறைபாடுகளைப் பற்றி பேசி நேரத்தை வீணடிக்கின்றனர்.
12. தொழில்நுட்பம் சரிவர பயிற்றுவிக்கப்படாததால் நேரத்தை சரிவர பயன்படுத்தத் தெரிவதில்லை .

தீர்வுக்கான ஆலோசனைகள்

- மாற்றுத்திறனாளிகளுக்கு தொழில்நுட்பத்தை பயன்படுத்தும் வகையில் கல்லூரியில் ஆற்றுப்படுத்தல் (orientation) பயிற்சி வழங்கப்பட வேண்டும்.
- உயர்கல்வியில் கணினிப் பயிற்சி கட்டாயம் அளிக்கப்பட வேண்டும்.
- வாரத்தில் ஒருநாளாவது கல்லூரி நேரத்திலேயே மாற்றுத் திறனாளிகள் மின்நூலகத்தைப் பயன்படுத்த அனுமதிக்க வேண்டும். அப்பொழுது அந்தத் துறையைச் சார்ந்த ஆசிரியர்கள் உடனிருக்க வேண்டும்.
- எளிமையான, புரிந்து கொள்ளக்கூடிய வார்த்தை களுடன் உள்ள புத்தகங்களை மின் புத்தகங்களாக மாற்றி மாற்றுத்திறனாளிகளின் பயன்பாட்டிற்கு வைக்க வேண்டும்.
- மின் நூலகத்திலுள்ள பயிற்றுநர் அனைத்துவிதமாக மாற்றுத் திறனாளிகளுக்குமான உதவித் தொழில்நுட்பத்தை (Assistive Technology) தெரிந்து வைத்திருக்க வேண்டும்.
- நேர ஆளுமைக்கான பயிற்சியை மாணவர்களுக்குக் கற்றுத்தர வேண்டும்.
- மாற்றுத்திறனாளிகள் பயன்படுத்தக்கூடிய கற்றல் வள மையங்களை உருவாக்க வேண்டும்.

பரிந்துரைகள்

- கல்லூரி நிர்வாகிகள் தங்கள் கல்லூரிகளில் மாற்றுத்திறனாளிகளுக்கான மின் நூலகத்தை உருவாக்கித்தர வேண்டும்.
- பிரைல் தொழில்நுட்பத்தில் பிரதிகளின் விலை மிகவும் அதிகமாக உள்ளது . அரசாங்கங்கள் விலையைக் கட்டுப்படுத்த ஆவன செய்ய வேண்டும்.
- கணினியை எளிமையாகப் பயன்படுத்த கணினித் துறையைச் சார்ந்த வல்லுனர்கள் தகுந்த மென்பொருளை உருவாக்க வேண்டும்.
- ஒவ்வொரு மாணவனுக்கும் மேற்படிப்பிற்காக அரசாங்கம் உதவித்தொகை அளிக்க வேண்டும்.
- தமிழ் மொழியில் கற்க உரிய மென்பொருள் உருவாக்கம் செய்திட வேண்டும்.
- கல்வி நிறுவனங்கள் அல்லாமல் மாற்றுத் திறனாளிகளுக்கும் கணினி இணையம் இவற்றை எந்த நேரத்திலும் பயன்படுத்த இணைய மையம் (Internet Cafe) உருவாக்கிட வேண்டும்.
- Internet Cafe ல் அனைத்து மாற்றுத்திறனாளிகளுக்குமான உதவித் தொழில்நுட்பத்தை கற்றுக்கொடுக்க பயிற்சி பெற்ற ஆசிரியர்களை நியமிக்க வேண்டும்.

முடிவுரை

நவீனத் தகவல் தொழில் நுட்பப் புரட்சி யுகத்தில் தொழில் நுட்பங்கள் பயன்படுத்தாமல் வாழ்வது என்பதும் , அதிலும் உயர்கல்வி பெ றுவது மிகக் கடினமாகும் . பல்வேறுபட்ட உடல்குறைபாடு உடையவர்களையும் சமுதாயத்தில் ஒதுக்கி வைக்கக் கூடாது . அவர்களுக்கு அங்கீகாரம் அளித்து மற்ற மாணவர்களைப் போல் நடத் வேண்டும் என்ற உயர்ந்த எண்ணத்தின் அடிப்படையில்தான் உள்ளடங்கிய கல்வி சர்வ சிஷ் அபியானால் (SSA) அறிமுகம் செய்யப்பட்டது. இந்த ஆய்வின் மூலம் விழியிழந்த மாணவர்கள் தங்களுக்கான பள்ளிகள் எங்கெல்லாம் இருக்கிறது; படிப்பதற்கும், பயன்படுத்துவதற்கும் என்னென்ன கருவிகள் உள்ளன . அதன் பயன்பாடு என்ன என்பது பற்றித் தெளிவாகப் புரிந்து கொள்வர் . இந்த ஆய்வேடு எதிர்காலத்தி ல் விழியிழந்தோர் பலருக்கு மிகவும் பயனுள்ளதாக இருக்கும் என்பது திண்ணம் .

உசாத்துணை நூல்கள் மற்றும் ஆராய்ச்சிகள்

1. சின்னத்தம்பி. க (1999) அறிகைத் தொழிற்பாடும் ஆசிரியரும் , கொழும்பு: ஜெயா ஓவர்சீஸ் பிரின்டர்ஸ் , (2ம் பதிப்பு).
2. நடராஜன் (2005 கல்வியில் ஆராய்ச்சி, சாந்தா பப்ளிஷர்ஸ், சென்னை.
3. விக்ரமசிங்க் (2002) சமூக விஞ்ஞான ஆய்வுகள், கொழும்பு.
4. *Aids & Materials for the Visually impaired*, Wayne: SIFB Products, P. 48.
5. American Foundation for the Blind (1989): *Products for People with Vision Problems*, New York: American Foundation for the Blind, P. 22.
6. *Braille Equipment for the Blind*, 1999 Catalogue, Sweden: Index Braille, P. 16
7. Burns, B.R. (2000) Introduction to Research Methods, Pearson Education Pvt. Ltd., Australia.
8. Daniel Bishop and Daniel J.A. Rhind (2011) Barriers and enablers for visually impaired students at a UK Higher Education Institution, *British Journal of Visual Impairment* 29: 177-195.
9. Mani, M. N. G. (1989): *Education of the Visually Handicapped: Perspective, opportunities, Strategies and Targets*, Paper presented at the National Seminar on Pre-school Intervention, Prevention, and Early Intervention, Dehradun.
10. National Rehabilitation Engineering Institute (1998): *Aids and Appliances for the Disable*: Price Catalogue Ahmedabad: Blind Men's Association.
11. Royal National Institute for the Blind (1977) *Illustrated Catalogue of Apparatus and Games for the Blind*, London.
12. Shukla, S.R. (1987): *A Study of Teacher Rated Behaviour, Scholastic Achievements and Aptitude of Visually Handicapped Children*, Indian Journal of Disability and Rehabilitation, P. 67-73

குழந்தைகளுக்கான தமிழ் இணையப் பயன்பாடும் தரவுகளும்

முனைவர் கல்பனாசேக்கிழார்

உதவிப்பேராசிரியர், தமிழியல்துறை, அண்ணாமலைப் பல்கலைக்கழகம்

குழந்தைகளுக்கு அறிவு கொடுப்பவர்களாகப் பெற்றோர்களும், சுற்றத்தாரும், ஆசிரியர்களும் இருக்கும் நிலையில், இன்றைய நவீன சூழலில் ஊடகங்களும் குழந்தைகளுக்கு அறிவைப் புகட்டுவனவாக உள்ளன. அவற்றுள் தொலைக்காட்சி, குழந்தைகளின் நேரத்தைப் பெருமளவு ஆக்கிரமித்துக்கொள்கிறது . தொலைக்காட்சி விளம்பரங்கள், தொடர்கள், குழந்தைகளுக்கான நிகழ்ச்சிகள் போன்ற அனைத்து நிகழ்ச்சிகளும் குழந்தைகளின் எண்ணங்களில் நல்லெண்ணத்தை விதைப்பதற்கு அடிப்படையாக இருக்கின்றனவா ? என்பதும், உலகமயமாக்கலின் கோரபடியில் மரபார்ந்த அடிப்படை வாழ்க்கைமுறைகள் நசுக்கப்பட்டுக்கொண்டு இருக்கும் இன்றைய சூழலில் , நம்முடைய மரபு சார்ந்த வாழ்வில் சிந்தனைகளை மீட்டெடுத்து குழந்தைகளுக்கு அறிமுகப்படுத்துகின்றனவா? என்பதும் சிந்திதற்குரியது.

பார்த்தல், கேட்டல், படித்தலின் வழியாக குழந்தைகள் தங்களுக்கான உலகத்தை உணரத்தொடங்குகின்றனர் . இவற்றுள் கேட்டல் என்னும் கதை சொல்லும் மரபு என்பது , தொல்காலம் தொட்டு உலகளாவி இருக்க கூடியது என்றாலும், அவரவர்களின் பண்பாட்டு விழுமியங்களுடன் தொடர்புடையவையாக அமைந்திருக்கும் . தமிழில் கதைசொல்லும் மரபும் நீண்ட பாரம்பரியத்தைக் கொண்டது . கதை கேட்டு வளர்ந்த எவரும் இன்று தங்கள் குழந்தைகளுக்கு கதை சொல்லும் நிலையில் இல்லை . அந்த இடத்தை இட்டு நிரப்புவை யாக இணையமும் வீடியோகேம்களும் தொலைக்காட்சிகளும் இருக்கின்றன . நிகழ், எதிர் காலங்களில் இணையம் , கற்றல், கற்றபித்தலிலும், குழந்தைகளுக்கத் தேவையான , அவர்கள் விரும்பு தரவுகளைத் தேடும் வெளியாக அமையும் என்பதில் கிஞ்சித்தும் ஐயமில்லை . அதற்கான செயல்பாடுகள் இணையத்தில் என்னென்ன நடந்துள்ளன என்பது பற்றியும் செய்ய வேண்டியவற்றைப் பற்றியும் ஆய்வதாக இக்கட்டுரை அமைகின்றது .

பொதுவாக, குழந்தைகள் பாடம் தொடர்பான தகவல்களைத் திரட்டுவதற்கும் , புதிது புதிதாக தெரிந்து கொள்ளவேண்டும் என்பதற்கும் இணையத்தை நாடுகின்றனர் . அண்மை காலங்களில் சமூக வலை தளங்களில் கூட குழந்தைகள் தங்களுடைய கருத்துக்களைப் தமிழில் பதிவு செய்வதும் , அது சார்ந்த உரையாடலை நிகழ்த்துவதும் காணமுடிகின்றது . இணையத்தை நோக்கி வரும் குழந்தைகளின் தேவையை பூர்த்தி செய்யக்கூடிய அளவிற்குத் தமிழில் தரவுகள் தொகுக்கப்படவில்லை . தொகுக்கும் முயற்சிகள் நடந்துள்ளன. ஆனால் அவை ஆரம்பநிலைகளிலேயே இருக்கின்றன. 12.10.2012 முதல் பத்து நாட்கள் அண்ணாமலைப் பல்கலைக்கழக மொழியியல் துறையுடன் இந்திய மொழிகளின் மத்திய நிறுவனம் இணைந்து குழந்தைகளுக்கான கதையெழுதுதல் என்னும் பணிப்பட்டறையினை நடத்தினார்கள். உலக அளவில் குழந்தைகளின் உலகத்தைப் புரிந்துகொண்டு, அவர்களின் தேவை என்னென்பதையும் அறிந்து, குழந்தைகளின் உலகத்திற்கான மொழியாடலுடன், அச்சமூக பண்பாட்டுடனையும் இணைத்து கதைகளை உருவாக்கி வருகிறார்கள். தமிழில் அதுபோன்று நடைபெறுவது இல்லை அதற்கான முயற்சியாக இப்பணிப்பட்டறையை ஒருங்கிணைத்திற்பதாகக் கூறினார்கள். இன்று அச்சுவடிவ நூல்களை வாசிப்பதை விட மின்வடிவ நூல்களை வாசிப்பதில் குழந்தைகளின் அதிக விருப்பம் காட்டுகின்றனர். ஆண்ராய்டின் வழியாக அண்மையில் செலபேசியில் குழந்தைகளுக்கான சித்திரக்கதை முழுவதும் படிக்கும்படி படங்களுடன் பதிவேற்றம் செய்யப்பட்டிருந்தது.

குழந்தைகளுக்கான தமிழ் இணைத் தரவுகள் என்று எண்ணும் பொழுது,

குழந்தைக் கதைகள் - வாசித்தல் / கேட்டல் / காட்சி அமைப்போடு பார்த்தல் / சித்திரங்களின்வழி
உருவாக்குதல் / நடிப்புவழி வெளிப்படுத்துதல்
குழந்தைப் பாடல்கள் - எழுத்து வடிவம் / இசையோடு கேட்டல் காட்சி அமைப்போடு பார்த்தல் /
குழந்தைகளுக்கான களஞ்சியம் / அகராதி

என்ற நிலையில் அமைந்துள்ளனவா என்று தேடினேன். கூகுளில் குழந்தைகளுக்கான கதைகள் தமிழில் தேடிய பொழுது இரண்டு இலட்சத்திற்கு மேல் காணப்பட்டன. ஆனால் சில கதைகளாக காணப்படுகின்றதே ஒழிய முழு கதைகளின் தொகுப்பு காணப்படவில்லை. பரமார்த்த கதைகள், பீர்பால், முல்லா, விக்கிரமதித்தன், புராண கதைகள் போன்றவை முழுமையாக காணப்படவில்லை. சில கதைகள் மட்டுமே காணப்படுகின்றன. அம்புலி மாமா, குழந்தை கதைகள் போன்ற இணைய தளங்கள் காணப்பட்டாலும் முழுமையானவை அல்ல. இவை போன்றே காட்சி அமைப்போடு கூடிய குழந்தைகள் கதைகளும் அமைந்துள்ளன. இணையத்தில் எஸ்.இராகிருஷ்ணன் இணையப் பக்கத்தில் குழந்தைகளுக்கான கதைகளின் தேவைகளைப் பற்றி குறிப்பிட்டு, கதை சொல்லிகளின் மெய்பாடுகள் எப்படி இருக்க வேண்டும் என்பதற்குச் சில இணையக் காட்சி பகுதியினைக் குறிப்பிட்டுள்ளார்.

<http://www.youtube.com/watch?v=bBkSEVzn46I>
<http://www.youtube.com/watch?feature=endscreen&v=nS9YHrZQrq8&NR=1>
http://www.youtube.com/watch?v=WP_LTtFYt3A
<http://www.youtube.com/watch?v=gAk5auEICDo>
<http://www.youtube.com/watch?v=izSfwz2Ewh4>
<http://www.youtube.com/watch?v=zYrOii-zHZo>
<http://www.youtube.com/watch?v=ephWySu6lQk&feature=related>
<http://www.youtube.com/watch?v=Vaf0esKLMZg>
<http://www.sramakrishnan.com/?p=503>

இந்த பகுதிகளில் காணப்படுகின்ற காட்சிகளைப் பார்க்கும் பொழுது, உலகின் பல பகுதிகளில் இருக்க கூடிய கதை சொல்லிகளின் கதை சொல்லும் முறைகள் பதிவு செய்யப்பட்டுள்ளன நம்முடைய தமிழ் மரபில் . கிராமப்புரங்களில் சொல்லகூடிய பல கதைகள் இந்த தலைமுறையில் இருக்ககூடிய பெரியவர்களின் வாய்மொழியோடு சென்றுவிடுமோ? என்ற நிலை காணப்படுகிறது முடிந்தவரை இவற்றை எல்லாம் நேரில் . சென்றுஅவர்களின் வாய்மொழி வழியாகவே கதைகளைச் சேகரித்து இணையத்தில் பதிவேற்றலாம் செய்யலாம் . இதன் மூலம் பாரம்பரியமான கதைகள் கதை சொல்லும் முறைகளை அடுத்த தலைமுறைக்குக் கடத்த முடியும் அழிவுறாமல் காக்கமுடியும். நாட்டுப்புற கதைகளைத் தொகுத்துள்ளதும் குறிப்பிடத்தக்கது .ரா.படைப்பாளர் கி .

குழந்தைகளுக்காக தமிழில் பல பாடல்கள் உருவாக்கப்பட்டுள்ளன . அவை அனைத்தும் இணையத்தில் ஏற்றப்படவில்லை. அச்சுவடிவில் இருக்க கூடிய நூல்களை மின் புத்தகங்களாவும், காட்சி அமைப்போடு கூடிய ஒலி ஒளி யுடனும் அமைக்கப்பட வேண்டும்.

மேலும் குழந்தைகளுக்கான களஞ்சியம், அகராதி தமிழில் தேடியபொழுது கிடைக்கவில்லை. இவைகளும் தொகுக்கப்படவேண்டும். இவற்றைத் தொகுக்கும் பொழுது குழந்தைகளின் தேவைகளை மனதில் கொண்டு உருவாக்கப்படவேண்டும். புதியபுதிய நவீன குழந்தைகளுக்கான சிந்தனைகள், கலைச்சொற்கள் அனைத்தும் குழந்தைகளுக்கு உடனுக்குடன் கிடைக்கும் படி தமிழ் மொழியில் மொழிப்பெயர்க்கப்படவேண்டும்.

கணினி வழி தமிழ் கற்றல் மற்றும் கற்பித்தல் குறித்தான ஆய்வுகள்

Louis Isack Kumar

Education Officer, Ministry of Education, Singapore

isack_kumar_louis@moe.edu.sg

சிங்கப்பூர் கல்வி அமைச்சின் பிரதான குறிக்கோள் தாய்மொழியைத் தேர்வுப்பாடமாக மட்டுமின்றி வாழ்க்கைக்குப் பயன்படும் மொழியாகவும் – வாழும் மொழியாகவும் நிலைபெறச் செய்தல் ஆகும் , இதனைச் செயல்படுத்த சிங்கப்பூர் கல்வி அமைச்சு 2010 ஆம் ஆண்டு தாய்மொழி மறுஆய்வுக் குழு ஒன்றை அமைத்தது, அதில் தமிழை வாழும் மொழியாக்குவதற்கான வழிவகைகளை அலசி ஆராய்ந்து சில பரிந்துரைகளை முன் வைத்தது . ஆய்வறிக்கையில் மொழித் திறன்களான கேட்டல் , பேசுதல், படித்தல், எழுதுதல் திறன்களோடு இருவழிக் கருத்துப்பரிமாற்றத் திறன்களும் பரிந்துரைக்கப்பட்டன . பேச்சு, எழுத்து ஆகிய இருவழிகளிலான கருத்துப்பரிமாற்றத் திறன்களை மேம்படுத்துவதற்கும் சுய முனைப்புடன் கற்றல் , உடனிணைந்து கற்றல் ஆகியவற்றிற்கு வாய்ப்பளிக்கவும் தகவல் தொடர்புத் தொழில்நுட்பம் உகந்த கருவியாகும் என முடிவு செய்யப்பட்டன. இந்த ஆய்வின் முடிவையே எனது ஆய்வின் தொடக்கமாகக் கொண்டு பார்ட்லி உயர்நிலைப் பள்ளியில் கணினியையும் இணையத்தையும் தமிழ் கற்றல் மற்றும் கற்பித்தலுக்குப் பாலமாக்கினேன் .



கணினிவழி இணையத்தில் இலவசமாக கிடைக்கும் மென்பொருள்களும்

(Screencast-o-matic, Vocaroo, Camstasia, Camstudio) சமூக ஊடகங்களும் (Facebook, Twitter, YouTube etc...)

இன்றைய 21-ஆம் நூற்றாண்டுக் கற்றல் கற்பித்தல் சூழலுக்குப் பயன்படுகின்றன . மாணவர்கள் கற்றலில் விருப்பத்தோடு ஈடுபட இவற்றில் சிலவற்றை தமிழ்மொழி கற்றல் கற்பித்தலுக்கு எவ்வாறு பயன்விளைவுமிக்கதாய்ப் பயன்படுத்த முடியும் என்பதை கண்டறிந்தேன் . இதற்காக முகப்புத்தகத்தில் பார்ட்லி தமிழ் அரங்கம் என்ற பக்கத்தை உருவாக்கி அதில் எங்கள் பள்ளியின் தமிழ் மாணவர்களை சேர்த்து ஒரு கருத்துதெரிவிப்பு தளத்தை உருவாக்கினேன் .



இணையத்தில் இலவசமாக கிடைக்கும் மென்பொருளான Screencast-o-matic மூலம் பட உரையாடல் , அகன்ற படிப்பு, தலைப்பையொட்டிய உரையாடல் போன்றவற்றை ஒளிப்பதிவாகவோ அல்லது ஒலிப்பதிவாகவோ பதிவு செய்யலாம். இவ்வாறு செய்வதன் மூலம் மாணவர்களை ஈடுபாட்டுடன் கற்க செய்து அவர்களுடைய பேச்சுத்திறனை மேம்படுத்தலாம் . இதற்காக தமிழ்மொழி மாணவர்கள் www.screencast-o-matic.com என்ற இணையப்பக்கத்திற்கு சென்று தங்கள் குரலை பதிவு செய்தனர் , ஒலிப்பதிவு அல்லது ஒளிப்பதிவு செய்யப்பட்ட படஉரையாடல், படித்தல் போன்ற தங்களுடைய படைப்புகளை முகப்புத்தகம் (Facebook) என்னும் சமூக வலைத்தளத்தில் பதிவேற்றம் செய்து ஆசிரியருடனும் சக மாணவருடனும் பரிமாறிக்கொண்டனர் ; அதில் படைப்புப்பற்றி ஆசிரியர், சக மாணவர் ஆகியோரின் கருத்துத் தெரிவிப்பு இடம்பெற்றது. அவ்வாறு வழங்கப்படும் கருத்துகளைக் கொண்டு மாணவர்கள் தத்தம் படைப்பினைச் செம்மைப்படுத்தித் தரமானதாகப் படைத்தல் அல்லது புதியதோர் படைப்பினை உருவாக்குதல் நிகழ்ந்தது .



மேற்கண்டவாறு என்னுடைய மாணவர்கள் செயற்பட்டதால் அவர்கள் பெற்ற பயன்கள் பின்பவருமாறு

- இணைய மென்பொருளையும் சமூக வலைத்தளங்களையும் கற்றலுக்குரிய முறையில் பயன்டுத்தினர் . இதன்வழி சக மாணவருடன் உடனிலைந்து கற்றலிலும் , சுயமாகக் கற்றலிலும் ஈடுபட்டனர் .
- சக மாணவரின் படைப்பிற்குக் கருத்துத்தெரிவிப்பு வழங்கவும் , தங்களது படைப்பிற்கு ஆசிரியர் , சக மாணவர் ஆகியோர் வழங்கும் கருத்துத்தெரிவிப்பினைக் கொண்டு தம் படைப்பினைச் செம்மைப்படுத்தவும் கற்றுக்கொண்டனர் .
- சமூக வலைத்தளங்கள் வாயிலாக வேறுபடுத்தப்பட்ட கற்றல் சூழலில் ஆசிரியருடனும் சக மாணவருடனும் இருவழிக் கருத்துப்பரிமாற்றத்தை மேற்கொண்டனர் .
- ஒருங்கிணைக்கப்பட்ட மொழித்திறன்களும் இருவழிக் கருத்துப்பரிமாற்றத் திறன்களும் மேம்பட்டன .

Theoretical underpinning

Recognising the need to prepare our Tamil Language students for life in the 21st century, Enhancing the Teaching and Testing of Mother Tongue Languages (MTL) to nurture Active Learners and proficient users – MTL Review Committee released its recommendations to nurture students to be proficient users of the language. Great emphasis will be placed on helping students to like learning their MTLs and become proficient in using them in a variety of real-life settings to communicate effectively, and to leverage on students' increasing info-communications technology (ICT) savvies more ICT. This includes a variety of interactive content, resources and activities to suit the different leaning needs of students and MOE's vision is to make MTL a living language for students.

Research has shown that MTL teachers should explore the use of technology in engaging pupils to develop their oral skills. Hence ICT tools have been used to engage the pupils to enhance oral skills. The pupils engage in authentic learning as they collaborate on an interactive media platform to conduct research on global issues.

Strategies & Learning activity

Self – directed learning and Collaborative learning is widely carried out as pupils are free to post comments, suggest and share their work on social media (You tube, Facebook, Twitter etc...) . After a particular oral project work is completed, pupils upload their completed projects on to social media under their assigned groups. Their classmates are able to comment and add on any information which could enrich their learning. This activity helps pupils consolidate what they have learnt as well as discuss interesting experiences. They will also undergo peer assessment. Pupils who are absent are also able to find out what they have missed. This would not have been possible with a written entry. Pupils also take their own initiative to comment on their classmates’ project work. Pupils will consolidate what they have learnt by using an online tool, and share their thoughts on social media (Facebook) . The teacher acts as an administrator and facilitator by providing guiding questions, feedback, suggestions and commendation. The platform provides students with the opportunity to share their oral and reading project work on their own accord. As such, Self – directed learning of Bartley Secondary School is also evident in this activity.

Evidence of impact on student learning

Students in Singapore have been using various ICT tools (screencast online tool, Vocaroo online tool, Web cam recorder & Social medias such as Facebook, YouTube) for Tamil Language to practise and peer - assess their speaking/ conversation skills during Mother Tongue lesson time in Bartley Secondary School, Singapore, they did an ICT oracy project based on self-directed learning and Collaborative learning.

Their oracy project helped them with self - improvement as they listened to their own recordings again and again and re-recorded themselves for better performance before they saved their work and submitted their projects. Peer assessment helps to create a more interesting and fun environment in class especially when they do a screenplay of a short film Team work.

The screencast-o-matic online recording tool and vocaroo online recording tools are very useful for picture conversation and extensive reading and it gives the students an exposure to real-life scenarios in which they are likely to speak Tamil. All students are encouraged to save their oracy projects in their own thumb drives or in a system workspace, thereafter the students’ project work will be uploaded on social media websites such as YouTube and Facebook. Before uploading students’ project students form a group in social media according to their levels under the teacher’s guidance, after the uploaded project, students will do their peer assessment and will post comments about their peers’ project improvement. This helps them to communicate with one another very well. This also provides a collaborative learning platform, so that they could retrieve or do peer assessment at any time. This exercise helps increase an interest in MT reading and speaking, which is usually considered a tough task for students who mostly prefer speaking English.

This simple yet effective usage of ICT emphasises on spoken language. Also, self-initiative is shown by students in using available audio recording tools for learning mother tongue. The more tech-savvy pupils help their classmates whenever they can.

தமிழ் இணையப் பயன்பாட்டில் கல்லூரி மாணவர்கள் எதிர்கொள்ளும் சிக்கல்கள் ஓர் ஆய்வு

தமிழ்ப்பரிதி. மாரி

உதவிப் பேராசிரியர், இதழியல் மற்றும் மக்கள் தொடர்பாடல் துறை, பெரியார் பல்கலைக்கழகம், சேலம்
tparithi@gmail.com

அறிமுகம்

உலகின் அனைத்து நாடுகளிலும் தமிழர்கள் பரவியுள்ள இன்றைய சூழலில் , தமிழ் இணையம் , உலகளாவியத் தமிழர்களை இணைக்கும் ஓர் ஒப்பற்ற இணைப்புப்பாலமாக விளங்கி வருகின்றது . பல நாடுகளும் , பல அரசுகளும், பற்பல தனி நபர்களும் தமிழ் இணையத்திற்குத் தொண்டாற்றி , பல அரிய கண்டுபிடிப்புகளை நிகழ்த்தி செயலாக்கி வ ருகின்றனர். இந்த கண்டுபிடிப்புகளுக்கிடையே ஓர் ஒருங்கிணைப்பும் , ஒத்திசைவும், ஆய்வும் தேவைப்படுகின்றது. உலகெலாம் பரவி வரும் மின் ஆளுகைக் கோட்பாடு இணையத்தமிழின் மீதும் தன் அழுத்தத்தினை தானாகவே கொடுக்கத் தொடங்கிவிட்டது . இச்சூழலில் உலகளாவிய தமிழ் இணையப் பயனீட்டாளர்களின் எண்ணிக்கைப் பெருகி , அரசு சார்ந்த இணையங்களிலும் , தமிழ் அமைப்பு சார்ந்த இணையங்களிலும், தனி நபர் செயற்படுத்தும் இணையங்களிலும் , அலைபேசி சார்ந்த தமிழ் இணையப் பயன்பாடுகளிலும் மற்றும் தனி நபர் வலைப்பூக்களிலும் தமிழ் இணையத்தின் தேவை நாளுக்குநாள் அதிகரித் து வருகின்றது. கணினிப்பயன்பாட்டை கல்லூரி மாணவர்களிடையே அதிகரிக்கும் வகையில் தமிழ்நாடு அரசு , கல்லூரி மாணவர்களுக்கு விலையில்லா மடிக்கணினி வழங்கியுள்ள இன்றையச்சூழலில் , தமிழ் இணையப்பயன்பாட்டில் கல்லூரி மாணவர்கள் எதிர்கொள்ளும் சிக்கல்களை ஆராய்வது பொருத்தமுடைய ஒன்றாகும். தமிழ் இணையப்பயன்பாட்டில் கல்லூரி மாணவர்கள் எதிர்கொள்ளும் சிக்கல்களை ஆராய்வதே இந்த ஆய்வுக்கட்டுரையின் நோக்கமாகும் . தமிழ் இணையப்பயன்பாட்டிற்கு அடிப்படையாக இருக்கும் இயங்குதளங்கள், இணைய உலாவிகள் , தமிழ் ஒருங்குகுறி எழுத்துருக்கள் , தமிழ் ஒருங்குகுறி எழுதிகள், தமிழ் ஒருங்குகுறி எழுத்துரு மாற்றிகள் , தமிழ் விசைப்பலகை , தமிழ்த்தேடுபொறிகள் ஆகியன தொடர்பாக எதிர்கொள்ளும் பல்வேறு சிக்கல்களையும் அதற்கான தீர்வுகளையும் ஆராய்வதே இவ்வாய்வுக்கட்டுரையின் நோக்கமாகும்.

இணையத்தின் வரலாறும் இணையத்தமிழும்

1960 இல் தொடங்கிய இணையத்தின் வரலாற்றுப்பாதையில் தமிழ் ஒருங்குகுறியின் வரவு ஒரு புதிய வெளிச்சமாகும். தமிழ் ஒருங்குகுறிக்கான ஒத்திசைவான இயங்குதளங்கள் , இணைய உலாவிகள் , தமிழ் ஒருங்குகுறி எழுதிகள் , தமிழ் ஒருங்குகுறி மாற்றிகள் , தமிழ் ஒருங்குகுறிக்கு ஒத்திசைவான சமூக வ லைதளங்கள், மென்பொருட்கள் எனத்தொடரும் தமிழ்க்கணினியின் பயணத்தில் உள்ள இடைவெளிகளைக்கண்டறிவதும் குறைகளைக்களைவதுமே தமிழ் மின்னாட்சி எளியோரை அடையவும் தமிழர் ஏற்றம் பெறவும் வழிவகை செய்யும் .

ஆய்வுமுறையும் ஆய்வெல்லை

ஒரு கட்டற்ற வினாத்தொகுப்பு உருவாக்கப்பட்டு அதனடி ப்படையில் சேலம் பெரியார் பல்கலைக்கழக , கல்லூரி மாணவர்கள் 100 பேருடன் குழு விவாதம் நிகழ்த்தப்பெற்று , தமிழ் இணையம் தொடர்புடைய தமிழ் எழுத்துரு , இணைய உலாவிகள் , இயங்குதளம், திறந்த நிலை மென்பொருட்கள் , அகரமுதலிகள், தமிழ்த்தேடுபொறிகள் ஆகியவற்றின் பயன்பாட்டுச் சிக்கல் குறித்து இணையங்களைப் பயன்படுத்தும் கல்லூரி மாணவர்களின் கருத்துகள் ஆய்விற்குட்படுத்தப் பெற்றுள்ளது . ஆய்வில் பங்குபெற்றோர் 100 பேர், இவர்களில் 50 பேர் ஆண்கள், 50 பேர் பெண்கள், ஆய்வில் பங்குபெற்ற 100 பேரில் 40 பேர் இளங்கலையும் 60 பேர் முதுகலையும் பயிலு வோர் ஆவர். ஆய்வில் பங்குபெற்ற 100 பேரில் 47 பேர் அறிவியல் புலத்தினர் , 53 பேர் கலைப்புலத்தினர் ஆவர் . இந்த ஆய்வு சேலம் பெரியார் பல்கலைக்கழக, கல்லூரி மாணவர்களுடன் மட்டுமே நிகழ்த்தப்பெற்றுள்ளது .

ஆய்வு முடிவுகள்

விசைப்பலகையின் பயன்பாடு

கல்லூரி மாணவர்களில் 15 விழுக்காட்டினர் பாமினி விசைப்பலகையிலும் 8 விழுக்காட்டினர் தமிழ் 99 விசைப்பலகையிலும் 17 விழுக்காட்டினர் ஆங்கில ஒலிபெயர்ப்பு முறையிலும் 60 விழுக்காட்டினர் எந்தவகை விசைப்பலகையிலும் தமிழ்த்தட்டச்சு செய்யத் தெரியாதவர்களாக உள்ளனர் .

தமிழ் ஒருங்குகுறியின் பயன்பாடு

கல்லூரி மாணவர்களில் 12 விழுக்காட்டினர் மட்டுமே தமிழ் ஒருங்குகுறி பற்றி அறிந்தவர்களாய் உள்ளனர் . தமிழ் ஒருங்குகுறியில் தட்டச்சு அதன் மென்பொருள் , தொடர்புடைய இணையதளங்கள் பற்றி நன்மைகள் பற்றி மாணவர்களிடம் போதிய விழிப்புணர்வு இல்லை . சமூக வலைதளங்களான ஃ பேஸ்புக்கில் 68 விழுக்காட்டினர் தமிழ் ஒருங்குகுறியைப் பயன்படுத்துகின்றனர். 32 விழுக்காட்டினர் சமூக வலைதளங்களைப் பயன்படுத்துவதில்லை.

தமிழ் மின்னஞ்சலின் பயன்பாடு

கல்லூரி மாணவர்களில் 12 விழுக்காட்டினர் மட்டுமே தமிழ் ஒருங்குகுறியைப் பயன்படுத்தி மின்னஞ்சல் விடுக்கின்றனர். 88 விழுக்காட்டினர் ஆங்கிலத்தில் மின்னஞ்சல் விடுக்கின்றனர் . தமிழ் ஒருங்குகுறியைப் பயன்படுத்தி மின்னஞ்சல் விடுப்பதில் மாணவர்களிடையே தயக்கம் காணப்படுகின்றது . தமிழ் எழுத்துருவை இயக்கத்தெரியாமை, தமிழ்த்தட்டச்சில் முறையான பயிற்சி இல்லாமை போன்றவை இ தற்கால இடைவெளிக்கூறுகளாக உள்ளன .

இணைய உலாவிகளில் தமிழ் ஒருங்குகுறியின் ஒத்திசைவு

கூகிள் குரோம் , ஃபயர்ஃபாக்ஸ், இண்டர்னெட் எக்ஸ்புளோரர் , அவண்ட், ஒபேரா என்னும் வரிசையில் இணைய உலாவிகளைப் பயன்படுத்துகின்றனர் . தமிழ் ஒருங்குகுறியின் ஒத்திசைவு கூகிள் குரோமில் 60 விழுக்காடு இருப்பதாகவும், ஃபயர்ஃபாக்சில் 30 விழுக்காடு இருப்பதாகவும் , இண்டர்னெட் எக்ஸ்புளோரர் , அவண்ட் ஆகிவ்வற்றில் முறையே 5 விழுக்காடு இருப்பதாகவும் , ஒபேராவில் தமிழ் ஒருங்குகுறியின் ஒத்திசைவின்மை 90 விழுக்காடு இருப்பதாகவும் கூறுகின்றனர் .

அலைபேசி உலாவிகளில் தமிழ் ஒருங்குகுறியின் ஒத்திசைவு

ஆண்ட்ராய்டில் குரோம் உலாவியைப் பயன்படுத்துவோர் 13 விழுக்காடு, ஒபேரா மைபைல் உலாவியைப் பயன்படுத்துவோர் 2 விழுக்காடு, மினி ஒபேரா உலாவியைப் பயன்படுத்துவோர் 5 விழுக்காடு, யுசி உலாவியைப் பயன்படுத்துவோர் 20 விழுக்காடு என்பதாய் உள்ளனர்.

தமிழ்த்தேடு பொறிகளின் பயன்பாடு

90 விழுக்காட்டினர் கூகிள் தேடு பொறியையும் , 5 விழுக்காட்டினர் யாகூ தேடு பொறியையும் , 5 விழுக்காட்டினர் பிங்க் தேடு பொறியையும் பயன்படுத்துகின்றனர் . மாணவர்களின் தேடுபொறிப்பயன்பாடு 88 விழுக்காடு ஆங்கிலத்திலும் 12 விழுக்காடு ஒருங்குகுறித் தமிழிலும் உள்ளது .

அதியன் இணைப்பின் பயன்பாடு

அதியன் இணைப்பு வாயிலாக கணினித்திரையில் சரியாக படிக்கவியலா தமிழ் எழுத்துருக்களை 1 விழுக்காட்டினர் மட்டுமே ஒருங்குகுறித் தமிழில் எழுத்துரு மாற்றும் செய்து பயன்படுத்துகின்றனர் .

தமிழ் ஒருங்குறி எழுதி, மாற்றிகளின் பயன்பாடு

சுரதா, அழகி, என்எச்எம் எழுதி , முரசு ஆகிய தமிழ் ஒருங்குறி எழுதிகளில் , என்எச்எம் எழுதியை 7 விழுக்காட்டினரும், தமிழா மென்பொருளை 4 விழுக்காட்டினரும், அழகி மென்பொருளை 1 விழுக்காட்டினரும் பயன்படுத்தி வருகின்றனர் . கூகிள் எழுத்துப்பெயர்ப்பி வாயிலாக 5 விழுக்காட்டினர் தமிழ் ஒருங்குறியைப் எழுதியைப்பயன்படுத்தி வருகின்றனர் . என்எச்எம் தமிழ் ஒருங்குறி மாற்றியை 7 விழுக்காட்டினரும், சுரதா தமிழ் ஒருங்குறி மாற்றியை 5 விழுக்காட்டினரும் பயன்படுத்தி வருகின்றனர்

திறந்தநிலை மென்பொருட்கள், இயங்குதளங்களின் பயன்பாடு மற்றும் ஒத்திசைவு

தமிழ் லினக்ஸ் பயன்பாடு 1 விழுக்காடாகவும், பாஸ் இயங்குதளப்பயன்பாடு 4 விழுக்காடாகவும், ஜிம்ப் மென்பொருளின் பயன்பாடு 2 விழுக்காடாகவும் உள்ளது . தமிழ் ஒருங்குறிப்பயன்பாட்டில் விண்டோஸ் இயங்குதளத்தைப் பயன்படுத்துவோர் 93 விழுக்காடாகவும், லினக்ஸ், மேக் ஆகிய இயங்குதளங்களைப் பயன்படுத்துவோர் 7 விழுக்காடாகவும் உள்ளனர் .

அகரமுதலிகளில் தமிழ்த்தேடல்

அகரமுதலிகளின் பயன்பாட்டில் தமிழ் விக்சனரியின் பயன்பாடு 30 விழுக்காடாகவும், தமிழ் இணையக் கல்விக்கழகத்தின் அகரமுதலி பயன்பாடு தமிழ்வழித்தேடல் 17 விழுக்காடாகவும், அகரமுதலிகளைப் பயன்படுத்துவோரின் விழுக்காடு 53 ஆகவும் உள்ளது .

பரிந்துரைகளும் ஆய்வுப்பரவலும்

தமிழ்கணினியின் வளர்ச்சியையும் தமிழ் மின்னாட்சியையும் கருத்தில் கொண்டு தமிழக அரசு பின்வரும் செயற்பாடுகளை மேற்கொள்ள வேண்டும் . (1) தமிழகப் பள்ளி, கல்லூரி மாணவர்கள் மற்றும் தமிழக அரசுப்பணியாளர்கள் அனைவருக்கும் தமிழ்க்கணினிப்பயிற்சி அளிக்கப்பெறுவதோடு , தமிழ்க்கணிமைச் செயலாக்கத்திற்கான மென்பொருட்கள் , நூல்கள், கையேடுகளை அரசு வழங்க வேண்டும் , (2) விண்டோஸ், லினக்ஸ், மேக் உள்ளிட்ட அனைத்து இயங்குதளங்களி லும் தமிழ் ஒருங்குறி எழுதியையும் , தமிழ் ஒருங்குறி மாற்றியையும் அடிப்படை இணைப்பாக இணைக்க வேண்டும் , (3) தமிழ் ஒருங்குறி உள்ளீடு , வெளியீடுகளுக்கு ஒத்திசைவுள்ள கணினி, அலைபேசி, கணினி சார்ந்த கருவிகள், மென்பொருட்களுக்கு மட்டுமே விற்பனை ஒப்புதல் வழங்க வேண் டும். (3) தமிழக அரசின் இணையதளங்கள் கண்டிப்பாக தமிழில் இருக்க வேண்டும் , (4) தமிழக அரசின் இணையதளங்களின் முதன்மைப்பக்கம் கண்டிப்பாக தமிழில் இருக்க முதல் முன்னுரிமை அளிக்கப்பெற வேண்டும்; தமிழக அரசு விரும்பினால் இரண்டாம் முன்னுரிமையை ஆங்கிலத்திற்கு வழங்கலாம் , (5) தமிழ்க்கணிமை மற்றும் தமிழ் மின்னாட்சிக்கென தனித்துவ ஆராய்ச்சி நிறுவனத்தை தமிழக அரசு அமைப்பதோடு தற்போது செயல்பட்டுவரும் தமிழ்க்கணிமை கல்வி , ஆராய்ச்சி நிறுவனங்களோடு ஒருங்கிணைப்ப ஏற்படுத்தப்பெற வேண்டும்.

சான்றாவணங்கள்

1. tamil.sg பார்வையிடப்பெற்ற நாள்: 10.12.2012
2. onbile.com/info/the-most-popular-mobile-browsers/ பார்வையிடப்பெற்ற நாள்: 11.12.2012
3. software.nhm.in பார்வையிடப்பெற்ற நாள்: 06.12.2012
4. android.com பார்வையிடப்பெற்ற நாள்: 05.12.2012
5. unicode.org பார்வையிடப்பெற்ற நாள்: 14.12.2012
6. thamizhagam.net பார்வையிடப்பெற்ற நாள்: 11.12.2012
7. microsoft.com/downloads/Search.aspx?displaylang=ta பார்வையிடப்பெற்ற நாள்: 13.12.2012

எதிர்கால இணைய வகுப்பறைகள்

முனைவர் ப.சு. மூவேந்தன்

உதவிப்பேராசிரியர், தமிழியல்துறை, அண்ணாமலைப் பல்கலைக்கழகம்,
psmnthn757@gmail.com

முன்னுரை

இலக்கியக்கல்வி கற்றல்- கற்பித்தல் கருத்தாடலில் இணையவழிக் கல்வி (Virtual Classrooms), இணையவழிக் கலந்துரையாடல் (Web Conferencing System) ஆகியவற்றின் தேவை, பயன்பாடு, அவற்றின் செயற்பாடுகள் குறித்து இக்கட்டுரை விவாதிக்கின்றது.

ஆய்வுச்சிக்கல்

இணைய வகுப்பறைகளின் எதிர்காலச் செல்வாக்கு.

ஆய்வின் நோக்கம்

- இணைய வகுப்பறைகளின் தேவை, பயன்பாடு.
- எதிர்கால வகுப்பறைகள்- இணைய வகுப்பறைகளில் இலக்கியம் கற்றல் – கற்பித்தல் குறித்து இக்கட்டுரை எடுத்துரைக்கின்றது.

கல்விமுறை

பழங்காலத்தில் பெற்றுவந்த கல்விமுறைக்கும், தற்காலத்தில் பெறும் கல்விமுறைக்கும் மிகுந்த மாற்றம் உள்ளது. குருகுலக்கல்வி முதல் இணையக்கல்வி வரையில் ஏராளமான கற்பித்தல் முறைகள் பின்பற்றப்பட்டு வந்துள்ளன. கல்வி பெறுவதில் தொடக்கக்காலத்தில் இருந்த சிக்கல்கள், சிரமங்கள் தற்காலத்தில் குறைவு. மனிதர்களிடையேயான நேரடிக்கருத்துப் பரி மாற்றத்தின் வாயிலாகப் பெற்றுவந்த கல்விமுறை தற்காலத்தில் தகவல் தொழில்நுட்பத்தின் வாயிலாகச் சாத்தியமாகியுள்ளது.

‘வாணை அளப்போம். கடல்நீரை அளப்போம்’ என்று உலகைத் தன் கவிதைக்குள் கட்டிவைத்த தமிழ்ப்புலவன் பாரதியின் எண்ணப்போக்கு மெல்லமெல்ல நிறைவேறி வருகின்றது. தற்காலத்தில் கல்வியறிவினைப் பெறும் வாயில்கள் பல உள். எனினும், தரமான கல்வியின் தேவை இன்று உணரப்பட்டுள்ளது. அதன் தேவையை நிறைவுசெய்யும் விதமாக இணைய வகுப்பறைகள் செயல்பட்டு வருகின்றன.

கடந்த இருபது ஆண்டுகளாக இணையவழிக் கல்வி, பெருமளவில் வளர்ந்து தற்போது பரவலாக்கம் பெற்றுவருகின்றது. இணையவழியாகக் கல்வி தருதல்- பெறுதலில் வேகமும் மாற்றங்களும் ஏற்பட்டு வருகின்றன. என்றாலும் இந்தியாவில் கல்விமுறையில் மாற்றங்கள் இன்னும் பரவலாக்கம் பெறவில்லை.

சூழல், இடம், காலம்

எந்தவொரு செயல்பாட்டுக்கும் இடம், காலம் என்பன முக்கியக் கூறுகளாக அமைந்துள்ளன. அதைவிடச் சூழல் முக்கியப் பங்காற்றுகின்றது. இணைய வகுப்பறைகள் காலம், இடம், சூழல் போன்றவற்றை நிறைவு செய்கின்றன. அவை கல்வி பெறுதலின் தரம் உணர்ந்து நடத்தப்படுகின்றன. மேலும் அது மாணவனின் தேவையை நிறைவு செய்வதாய் உள்ளது. எதிர்கால இணைய வகுப்பறைகளில் இலக்கியக் கல்வி எவ்வாறு அமையும் என்பதனைக் காணலாம்.

தற்காலத்தில் குறிப்பாக, இலக்கியக் கல்வி கற்றல்- கற்பித்தல் என்பது, பாடல் விளக்கப் பகுதிகளுக்குப் பொருள் விளக்குவதாகவும், தெளிவுரை, பொழிப்புரைகூறித் தெளிவுறுத்துவதாகவும் அமைந்துள்ளது. பாடலின்

புறவடிவங்களான யாப் பு, அடி, தளை, சீர், தொடை, அணி, உவமை முதலானவற்றைக் கூறுவதும் , அகவடிவக் கூறுகளான கருத்து , சிந்தனை, சமுதாயத்தாக்கம் முதலானவற்றை எடுத்துரைப்பதும் மரபான செயல்பாடுகளாகவும், கல்வி புகட்டும் முறையாகவும் இருந்து வருகின்றது .

இதனால் மாணவர் பெறும் பயன் யாது ? வினாவுக்கு விடை எதுவுமில்லை . ஒவ்வொரு காலகட்டத்திலும் ஒவ்வொரு நாட்டிலும், ஒவ்வொரு சூழலிலும் எழுந்த இலக்கியங்கள் ஒவ்வொரு கருத்தினைப் பதிவு செய்துள்ளன . பதிவு செய்யவேண்டிய நிர்ப்பந்தத்திற்குத் தள்ளப்பட்டே , இலக்கிய உருவாக்கம் நிகழ்ந்துள்ளது . இதனைப் பற்றிய தெளிவான அறித ல் தேவை . இலக்கியத்தின் வாயிலாக நாம் பெறும் செய்தி யாது என்பதை அறியவேண்டும் . மேலும் பாடற்புனைவின் தேவை , சமூகக்களம் குறித்த பின்புல அறிவு பற்றிய புரிதல் தேவை . இன்றைய கல்வி புகட்டும் முறையில் இவை பெரிதும் தவிர்க்கப்படுகின்றது .

இந்நிலையில் மாற்றம் வேண்டும் . அதற்கு இணைய வகுப்பறைகள் பயன்தரும் . இணைய வகுப்பறைகளின் தனிச்சிறப்புகளைப் பின்வருமாறு வகைப்படுத்தலாம் .

இணைய வகுப்பறைகள் (VC)

இணைய வகுப்பறைகள் (Virtual Classrooms) இணையவழி இயங்குகின்றன. இவை பல்வேறு தொழில்நுட்பச் சாதனங்கள் குறிப்பாக, பாடம். ஒலிப்பேழை. பல்லாடகத் தொழில்நுட்பக் கருவிகள் , காணொளி. படங்கள் போன்ற (text, audio and video chat, whiteboards, streaming audio and video, file sharing, etc.,) பல்லாடக மென்பொருட்களின் பங்களிப்பு வழி நடைபெறுகின்றன.

இணைய வகுப்பறை பெற்றுள்ள சாதகங்கள்:

- அறிவுசார் தொழில்நுட்பத்துடன் கூடிய , மரபான கல்வி பெறுதல் (Presentations, discussions and case studies by traditional method)
- ஆசிரியர்-மாணவர்களுடனான நேரிடைத் தொடர்பு, கருத்தாடல்கள், இலக்கியப் புரிதல்கள்.
- பாடம் புகட்டுவதில் ஆசிரியர் மற்றும் மாணவரின் வேட்கை, ஆர்வம், புரிதல் மற்றும் இலக்கியப் பங்களிப்பு.
- ஆசிரியரின் முழுத்திறன் வெளிப்படும் வாய்ப்பு.
- ஆசிரியர்-மாணவர் விருப்ப அடிப்படையிலான நேரம். வினாக்கள் உடனடியான மதிப்பீடுகள்.
- உலகின் எந்தப் பகுதியில் இருக்கும் மாணவரும் தரமான கல்வி பெறும் வழிவகை
- ஆசிரியர்-மாணவர் இருசாராரும் பயன்படுத்தும்படியான பொதுவான அறிபலகை.
- ஒருங்கிணைந்த நேரிடைச் செயல்பாடுகள்
- பாடஞ்சார்ந்த கருத்துப் பரிமாறுதல்கள், உலகளாவிய சிந்தனைப் போக்குகளும் புரிதல்களும் . சிந்தனை ஊடாட்டம்.
- மாணவர் விருப்ப அடிப்படையிலான வினாநிரல்
- மாணவர் கற்கும் திறன், ஆசிரியரின் கற்பிக்கும் திறன் பற்றிய உடனடியான மதிப்பீடு (student ability, teacher skill)
- ஆசிரியரின் கற்பிக்கும் திறன் குறித்த உடனடி மதிப்பீடு (feedback)
- கல்வி பெறுதலில் பின்தங்கியுள்ள மாணவர்கள் இணைய வகுப்பறைகளில் நடத்தப்பட்ட பாடங்களைப் பதிவுசெய்து மீள்பார்வை செய்வதன் மூலம் தன் கல்வித்திறனை மேம்படுத்திக் கொள்ளமுடியும் .
- மாணவர்களுக்கான பாடத்திட்ட நூல்களை மின்னூல்களாகத் தருவதன் மூலம் எல்லோர்க்கும் பாடநூல் கிடைக்க வகை ஏற்படும்.

இவை எல்லாமே தற்போது சாத்தியமாகியுள்ளன. இவற்றின் வாயிலாகத் தரமான இலக்கியக்கல்வியைப் பெறலாம்-புகட்டலாம்.

இணையவழிக் கலந்துரையாடல் (WCS) (கல்விமுறை)

தற்போது வணிகச் செயல்பாடுகளுக்கும், தொழில் நிறுவனங்களின் கலந்துரையாடல்களுக்கும் பயன்படுத்தப்பட்டு வரும் இணையவழிக் கலந்துரையாடல் (WCS-Web Conferencing System) மூலமாகக் கல்வி புகட்டும் முயற்சிகள் மேற்கொள்ளப்பட்டு வருகின்றன. இம்முறையினைக் கல்வி புகட்டுவதற்குப் பயன்படுத்துவதன் வாயிலாகப் இணைய வகுப்பறைகளை மேலும் வளமானதாக ஆக்கமுடியும் .

இணையவழிக் கலந்துரையாடல் மூலம் பெறும் பரிமாற்றச் சாதகங்கள்

- குழுவினரின் தகவல்கள், விவாதங்கள், கலந்துரையாடல் மற்றும் கருத்துப் பரிமாற்றங்கள் (file sharing)
- White boards
- Text chat – to deliver live presentations.
- உலகின் எங்கோ ஒரு பகுதியிலிருக்கும் ஆசிரியருடன் பல்வேறு பகுதிகளைச் சார்ந்த மாணவர்களும் பங்கேற்றுப் பயிற்சிபெறும் வசதி. (Direct interaction between the learners and the teachers) உலகளாவிய சிந்தனையோட்டம்.
- துறை சார்ந்த சிறந்த ஆசிரியர்கள் - பல்வேறு மாணவர்கள் ஒன்றிணைந்து கற்றல் – கற்பித்தல்
- பங்கேற்பாளர்கள் தங்கள் கருத்துக்களை வடித்தெடுக்க முன்முயற்சிகள்
- பங்கேற்றல் - தகவல் பரிமாறல்
- பல்லாடக வழித் தகவல் தொடர்பு
- தொலைதூரக் கல்வியின் சாத்தியப்பாடு.
- மாணவர்களுடனான நேரிடைத் தொடர்பு மற்றும் புரிதல்கள் (Direct interaction)
- வகுப்புகளைப் பதிவு செய்து மீண்டும் பார்த்தல் (Recording and playback)
- அளவீடு மற்றும் மதிப்பீடு செய்தல் (Surveys and feedback)

உலகத்தின் பல்வேறு இடங்களில் உள்ள மாணவர்களை ஒன்று திரட்டி ஒரேநேரத்தில் பாடங்கள் நடத்துவதனால் இலக்கியத்தின் தேவையினையும் , பொதுநோக்கினையும், உலகம் தழுவியுள்ள தற்கால இலக்கியத்தின் போக்கு , உலகளாவிய அறிவுச்சிந்தனைகளையும் , பொதுச் சிந்தனைகளையும் , உலகு தழுவிய சிக்கல்களையும் அறிய முடியும்.

இணையவழிக் கலந்துரையாடலின் மூலமாக இலக்கியத்தில் சிறந்துவிளங்கும் விற்பன்னர்களைக்கொண்டு இலக்கியச் செய்முறைப் பயிற்சிகள் தரலாம் . சிறந்த இலக்கியம் பெற்றிருக்க வேண்டிய கூறுகளை மாணவர்களுக்குக் கற்றுத் தரலாம் .

ஆர்வமுள்ள மாணவர்களை இலக்கியம் படைக்கத் தூண்டி அவரது இலக்கியப் படைப்பை மதிப்பீடு செய்துதரலாம். மேலும் தரப்படுத்தலாம் . இதனால் இலக்கியக்கல்வி வளரும் . நல்ல தரமான இலக்கியங்கள் உருவாகும்.

இலக்கியக் கல்வி பெறுவதில் வயது , கல்வி முதலானவை பெரும் இடர்ப்பாடாக உள்ளன . இலக்கியப் படைப்பாளனுக்கும், இலக்கிய வாசிப்பாளனுக்கும், இலக்கியக் கல்வியாளனுக்கும் வயது ஒரு தடையாக இருத்தல் கூடாது. உலகமெங்கும் இலக்கியப் படைப்பாளர்கள் பல்துறையைச் சார்ந்தவர்களாக உள்ளனர் . ஆதலால் பல்வேறு அறிவுசார் தளங்களைச் சார்ந்தவர்களும் இலக்கியக்கல்வி பெறுதல்- புரிதலில் பங்கேற்றல் வேண்டும்.

எதிர்கால இலக்கியக் கல்வி முறை ஆற்றல் , அறிவுத்துறையைச் சார்ந்தே அமைந்திருக்கும் . அதன் வழியே கல்வி பெறும் மாணவர்கள் நல்ல தரமான உயர்கல்வியைப் பெறுவார்கள் . அதற்கு இணைய வகுப்பறைகள் பெருந்துணை புரியும். இலக்கியக்கல்வி சமூகப் பயன்பாட்டுக்கு உகந்ததாக இருக்கும் .

இணைய வகுப்பறைகள் மற்றும் இணையவழிக் கலந்துரையாடல்களின் மூலம் கல்வி கற்பிக்கும் வகுப்பறைகளே எதிர்காலக் கல்வித் திட்டத்தில் பெரும்பங்காற்றப் போகின்றன. இதன் முன்முயற்சிகள் உலகெங்கும் பரவிவிட்ட நிலையில் தமிழிலும் இதுபோன்ற முயற்சிகள் பெரும்பயனைத் தரும், மாணவர்கள் தரமான உயர்கல்வி பெறும் வழிவகை ஏற்படும்.

பரிந்துரைகள்

இணைய வகுப்பறைகளை நமது கல்வித்திட்டத்தில் கொண்டுவரப்பட வேண்டும் . நாட்டின் எங்கோ ஒரு மூலையில் இருந்துகொண்டு , சரியான கல்விகற்க வாய்ப்பில்லாமல் தவிக்கும் ஏழை மாணவனுக்குத் தரமான கல்வி இதன்மூலம் கிடைக்கும்.

தனிநபர் தேவையின் அடிப்படையில் பயன்பாட்டில் உள்ள இணைய வகுப்பறைகளை ஒன்றிணைத்து , அதனை எல்லோரும் எளிமையாகப் பெறும்படியான வசதிவாய்ப்புகள் ஏற்படுத்தப்பட வேண்டும் .

இணைய வகுப்பறைகளை உயர்கல்வி பெறுவதற்கு மட்டும் பயன்படுத்தப்படும் நிலை படிப்படியாகக் குறைந்து , பள்ளி, கல்லூரிகளிலும் பரவலாக்கம் செய்யப்பட வேண்டும்.

முடிவுரை

எதிர்கால இணைய வகுப்பறைகள் ஆசிரியர்- மாணவர் பணிச்சுமையைக் குறைக்கும் . இலக்கியம் கற்றலை எளிமையாக்கும். எல்லோர்க்கும் தரமான இலக்கியக்கல்வி கிட்டும் .

இணையத்தின் மூலமாகக் கல்வி (Virtual Classroom) பெறுவதோடு, மட்டுமன்றி இணையவழிக் கலந்துரையாடலின் (Web Conferencing) மூலமாகக் கல்வி பெறுவதால் கல்விமுறை வளர்ச்சி பெறும் . தகவல் பெறும் முறை எளிமையாகும். தரமான கருத்துக்கள் உலகெங்கும் சென்று சேரும்.

இணையதளத்தில் தமிழ் வழிக்கற்றலில் மேன்மை

முனைவர் சு.கணேசன்

முத்து உயர்திறன் கல்வி ஆசிரியர் பயிற்சி நிறுவனம், வேலம்மாள் வித்யாஸ்ரம் கல்வி வளாகம், சென்னை

அறிவு வளர்ச்சியே மனித வளர்ச்சி . இத்தகைய அறிவு வளர்ச்சி மகிழ்ச்சியாக அமைய இணையதளத்தில் தமிழின் மேன்மையை உணரும் வண்ணம் தமிழ் கணினி யின் பங்கினை ஈண்டு காண்போம் . புதிய தொழில்நுட்ப சாதனைகளை வகுப்பறையில் உருவாக்கினால் மாணவனின் எதிர்காலமும் நாட்டின் எதிர்காலமும் சிறப்பாக அமைந்திட தமிழ் இணையதளம் அமைதல் வேண்டும் . நல்வழி அர்ப்பணிப்பு உறுதிப்பாடு ஒழுக்கம் வரையறை என்ற நிலையில் வளர்ந்து வரும் பல்லாடகத் தகவல்தொழில்நுட்பத் துறையில் தமிழ் கணினி உலகம் முழுவதும் வலம்வர செய்யுள் உரைநடை இலக்கணம் துணைப்பாடம் போன்றவை பவர்பாயிண்ட் முறையில் அமைத்துப் ப்ராஜெக்ட் பேஸ்ட் லேர்னிங் முறையில் ஆக்கம் சிந்தனை உரிமை நயம் ஆர்வம் என்ற முறையில் மாணவனின் உள இயலுக்குத் தகுந்த வண்ணம் செய்திகள் பாடக்கருத்துகள் புகைப்படம் போன்றவற்றை மென்பொருளில் சேமித்து சிறிய வட்டுகளில் பதிவு செய்ய வேண்டும் . கூடுதல் தகவல்களை இணைய வலை மூலம் பார்த்தும் , ஆய்வரங்கம் க்ளோனிங் மூலம் செல்போன் போன்ற நவீனகாலத் தொழில்நுட்பத்துடன் தகவல் தொழில்நுட்பத் திறனும் மதிப்பீடு செய்வதில் தமிழ்மொழி இலக்கியங்களும் செவ்வியல் மொழியாக இணையதளத்தில் அமைந்திட வேண்டும் .

திகட்டாத உரைநடை , வலிமிகு இடங்கள் , இலக்கணம் தமிழ் எண்கள் அமைத்தல் நலம் . நழுவங்கள் துண்டுப்படச்சுருள் ஒலிப்பதிவு சலனப்படங்கள் பவர்பாயிண்ட் வழிக் கற்பித்தல் என்னும் அணுகுமுறையில் தமிழ் இணையதளத்தில் உருவாகினால் புதிய நவீன கணினி உலகம் ஒவ்வொரு மாணவன் கையிலும் தமிழ் கணினி உலா வரும் என்பதில் ஐயமில்லை . பாடப்பொருளுடன் தொடர்புடைய புள்ளிவிபரங்களை அட்டவணையுடன் இணையதளத்தில் அறியலாம் . கற்றல் கற்பித்தலில் தமிழ் தொழில்நுட்பக் கருவிகள் மாணவர்களின் ஆர்வத்தினை வளர்க்கும் என்பதில் ஐயமில்லை . குரல்மூலம் கணினி ஈஸ்பீக்கிங் என்ற மென் பொருளை தரவிறக்கி அதை கணினியில் சேமித்துச் சில கமாண்ட்களை அமைத்துக் குரல் கொடுக்க வேண்டும் .

சிறந்த இணையதளம் அட்வான்ட்பெர்சிங்டன்ஸ் . சைபர்குற்றவாளி . சைபர் வாரியர்கள் . வர்த்தக கயவர்கள் இல்லா தமிழ் இணையதளம் உருவாகவேண்டும் . ஜாவா புரோகிராமிங் அரசியாகிய ஜவாமூலம் ஒவ்வொரு குறியையும் இழைகளையும் கட்டுப்படுத்தலாம் ஒன்றுக்கு மேற்பட்ட தர்ட்டுகளை கையாளும் முறை ஒரே நேரத்தில் ஆயிரக்கணக்கான இணைய நுகர்வோர்களை அணுக அனுமதிக்கும் திறனை இணையதளத்தில் பயன்படுத்தலாம் . மேப்புகள் உருவாக்குதல் : 3ஜி இன்டர்நெட் இணைப்பைப் பெற்றிருக்கும் ஒரு செல்பேசியில் நாம் கூகுள் மேப் என்கிற ஒரு சிறிய மென் ஒருங்கை நிறுவிக்கொண்டால் அது சர்வரில் வடிவமைக்கப்பட்டுள்ள மேப்புடன் ஒப்பிட்டு நாம் இருக்கும் இடத்தை மேப்பில் காட்டும் . மென்பொருளை தமிழ் இணையத்தில் பயன்படுத்தலாம் . மைக்ரோசாப்ட் நிறுவனம் அறிமுகப்படுத்திய ஸ்கைட்ரையில் சேமிக்கப் பயன்தரும் . தன்னார்வ பங்குகள் தகவல்கள் ஆகியவற்றைத் தமிழ் விக்கி மீடியாவில் பயன்படுத்திக்கொள்ளலாம் . விக்கி நூல்கள் மேற்கோள் இனங்கள் செய்திகள் பயிற்சிப்பட்டறைகள் மிகவும் பயன்தரும் .

ஆண்ட்ராய்டு அப்ளிகேஷன்ஸ் : 1.மை பிட்னஸ்பால் 2.ஒர்க்அவுட் ட்ரெய்னர் 3.வெப் எம் டி 4.மேப் மை பிட்நெஸ் 5.ஐ டி ரை ஏஜ் ஆகியவை கலோரி உடற்பயிற்சி பாதுகாப்பு உடல் எடை மருத்துவம் ஆகியவை செல்பேசி மூலம் பயன்தரும் .

இணையதளத்தில் கணினி மடிக்கணினி செல்பேசி ஆகியவற்றைத் தொடுதிரையின்மூலம் இயக்கிப் பயன்படுத்த முடியும். கடவுச்சொல் விதிமுறைகளைச் சரியான முறையில் உருவாக்கிக் கொண்டால் காகிதங்களுக்கு விடை கொடுத்துத் தமிழ் இணையதளத்தில்காப்பாற்றலாம் . உபுண்டு பென்ட்ரைவ் மூலம் தமிழ் கணினியை வளப்படுத்தலாம்.பேஸ்புக்கிங் ரிவர்ஸ் லாஜிஸ்டிக்; மென்பொருளைக் கொண்டு வர்த்தக அறிவை உயர்த்தலாம் .

பட்ஸலா இணையதளம் தமிழ் கணினிக்கு அரசனாகத் திகழ்கிறது . ஓராண்டுப் பலனுக்கு நெல்லை நடு நூற்றாண்டுப் பலனுக்கு நல்ல தமிழ்க்கல்வியை இணையதளத்தில் நாடு . கணினி உலகில் தமிழ் அறிவியலை வளர்க்க தமிழ் இணையதளத்தின் மூலம் உலகையே வலம் வரலாம் .தமிழ் கணினியில் குறுந்தகடுகள் இக்கால உலகில் மாணவ சமுதாயத்திற்கு எளிமையாக திறன் வளர்க்கும் . கிங்ஸ் ஸாஃப்ட் 2012ரைட்டர் பிரஸென்டே ~ன் இணையதள வலைப்புகளுக்கான வடிவங்கள் தமிழ் மொழிக்கு வளம் சேர்க்கும்.கண் இமைக்கும் நேரத்தில் வீடியோ படங்களைத் தேடி எடுத்துப் பார்க்க கூகுளில் பைப்பர் இணையதளம் தமிழ்கணினி வளர வகை செய்யும் . இதன் வேகம் ஒரு நொடிக்கு 1.ஜிகா பைட் என்று கணக்கிடப்பட்டுள்ளது .

பைதான் என்ற மென்பொருள் தமிழ் இணையதளத்தில் தகவல்கள் மென்பொருள்கள் உதவி க்குறிப்புகள் விரிவாக்க மொழியாகப் பயன்படும் . சிர்ப் என்ற புதிய தொழில்நுட்பத்தால் செல்போன் கேமராவில் படம் பிடிக்கப்படும் படங்கள் ஒளியாக மாற்றி மற்ற செல்போன்களுக்கு ஸ்மார்ட்போன் மூலம் பயன்பெறலாம் . காம்போஃபோன்களில் தனியாக ஒரு அடாப்டர் அமைத்து ஸ்கைலைட் நிறுவனம் இமேஜ் கோப்புகளை போர்டு கோப்புகளாகத் தமிழ் இணையதளத்தில் பயன்படுத்தலாம் .

விரைவு அறிவு சார்ந்த தகவல்கள் சனவரி 15ல் விக்கிப்பீடியா தொடங்கப்பட்டது .இதில் வரையறை இல்லாமல் அரசியல் திரைப்படம் சட்டம் மருத்துவம் இலக்கியம் விளையாட்டு போன்றவைத் தமிழ் மொழியில் அமைத்து தமிழ் கணினியில் பயன்படுத்தலாம் . விளையாட்டு முறை விளக்கமுறை தமிழ் கணித முறை பொருத்துக் தேர்க இலக்கணக்குறிப்பு உறுப்பிலக்கணம் அகம் புறம் கவிதை போன்றவற்றைக் குறுந்தகட்டில் அமைத்து பயன்படுத்தி அதன் கோப்புகளை உலகளாவிய செயல்திட்டமாக மாற்றலாம் . செயல்திறன் மனக்கிள ர்ச்சி ஊக்கம் கவனம் விருப்பம் சிந்தனை தோல்விகளில் துவளாமை போன்றவற்றை இணையவழித் தமிழ்க்கல்வியில் தமிழ் இணையப் பல்கலைக்கழகத்தின் மூலம் அடிப்படைக்கல்வியை அயல்நாடுகளில் வாழ்கின்ற தமிழர்களுக்கும் பிறருக்கும் பயன்படுத்தலாம். மரபு வழிக் கல்வியைக் காட்டிலும் தமிழி ல் இணையவழிக் கல்வியை கரும்பலகை இல்லாமல் பேசுதல் திறன் எழுதுதல் திறனை வளர்க்கலாம் . கரும்பலகையிலே உள்ள செய்தியைப் பார்த்துத் தாளிலே எழுதும் பயிற்சியை நீக்கி நவீன உலகில் தமிழ் இணையதளம் வழி எழுத்துப்பயிற்சி மிகவும் உயர்வைத் தரும் . 18 தேசிய மொழிகள் 1652 மொழிகள் உலாவும் இவ்வுலகில் தமிழ் கணினியால் சாதனை படைப்போம்

தமிழ் வழியில் பிற நு ிலில் உள்ள கருத்துகளை படக்காட்சி ஒலிநாடாப்பதிவு கையடக்கத்தட்டு செல்கள் வலைதளங்கள் படக்காட்சியுடன் கூடிய தொலைக்காட்சி இலக்கமுறைத் தொலைக்காட்சிகளை ஆற்றல் வாய்ந்த தொலைபேசி ஒளி இடை கம்பிகள் ஈமெயில் அச்சிடும் ஊடகம் அச்சிடப்படா ஊடகம் ஆகியவற்றைப் புதிய உத்திகள் கொண்டு தமிழ் இணையதளத்தில் உருவாக்குவோம் அறியாமை இருளை அகற்றும் தமிழ்ப் பாடப்பகுதியை உருவாக்கி ஆற்றலுடன் உள்ள தமிழ் கணினியைப் பயன்படுத்தும் மாணவர்களுக்கு இனிய இணையதளம் அமைக்க உலகில் ஈடில் லாத் தமிழில் உண்மையான மென்பொருளை உருவாக்கி ஊக்கமுடன் தமிழ்ப்பாடப் பகுதிகளை எளிமையாக்குவோம் .என்றும் எளிமையாகத் தமிழ் இணையதளத்தை உருவாக்குவோம்.ஏணிப்படியாகத் தமிழ் கணினி உலகம் உயரும் ஐயமில்லாமல் செம்மொழியின் சிறப்பினை ஒன்றுபட்ட பாரதத்தில் ஓங்காரமாகி ஒளிமயமா ன் மாணவசமுதாயத்திற்கு புதிய புதிய உத்திகளை உருவாக்கி மென்பொருளை ஒளவை உண்ட நெல்லிக்கனியின் சாற்றாக அமைப்போம் .

குறில், நெடில், வினாஎழுத்து. சுட்டெழுத்து முதல் சார்பு சொற்களின் வகைகள். வழு. வழாநிலை வழக்கு போலி பதம் புணர்ச்சி பொருள்கோள் பா பாவகைகள் பொது வேற்றுமை உவமை உருவகம் அணிவகைகள் ஆகிய இலக்கணத்திறன்களுக்கு சிந்தனைவரைபடம் மூலமும் குறுந்தகடுகள் மூலமும்பதிவுசெய்து பிறகு இணையதளத்தின்மூலம்தமிழின் திறன்களை மாணவச்சமுதாயத்திற்கு பயிற்சியளிக்கலாம் ;.

உரைநடைப்பகுதியில் பழமொழிகள் மேற்கோள்நூல்கள் பத்தியமைத்தல் நிறுத்தற்குறிகள் எழுத்துநடை சொல்நடை பொருள்நடை வட்டாரமொழிநடை , மரபுநடை, உச்சரிக்கும் முறைகள் , ஆகியவற்றை தமிழ்இணையதளத்தின் மூலம் தமிழ்வளங் காணச்செய்யலாம். சொற்களின் பொருள் தொடர்பை அறியும் பயிற்சிகளில் எழுதுதல்.குழந்தைகளின் படவிளக்க அகராதியைப்பயன்படுத்தும் திறன் எழுத்துகளின் உரியவடிவம் ஒலிவடிவம் வரிவடிவம் மயங்கொலிப்பிழைகள் . வருணனையும். விவரிப்புகளையும் சொல்விளையாட்டுப்பதிர்களையும். நேரிணை, எதிரிணை, செறியிணை போன்ற தமிழ்மொழி வளர்த்திறன்களை தமிழ்இணையத்தின்மூலம் தெளிவாக உணரலாம் . விழிப்புணர்வுப்பாடல்கள் கலந்துரையாடல்கள். நாடகங்கள். எளியஎண்விவரங்கள். தனிநடிப்பு. பாடுதல். ஆடுதல் விவாதம் கருத்தரங்கம் வில்லுப்பாட்டு. பட்டிமன்றம். ஜெயித்துக்காட்டுவோம் போன்றவற்றை தமிழ் இணையத்தளத்தின் மூலம் வழிக்காட்டுதல், செய்தால் செம்மொழியாம் நம் தமிழ்மொழி சிறக்கும் என்பதில் ஐ யமில்லை எனலாம்.

சங்ககால இலக்கியங்கள் எட்டுத்தொகைநூல்கள் பத்துப்பாட்டு மேல்கணக்கு நூல்கள் கீழ்க்கணக்கு நூல்கள் ஐம்பெரும்இலக்கியங்கள் ஐஞ்சிறுங்காப்பியங்கள் தற்கால இலக்கியங்கள் பக்தி இலக்கியங்கள் சமயஇலக்கியங்கள் தேவாரம் திருவாசகம் தற்காலக்கவிதைகள் புதுக்கவிதைகள் புதினம் இசைப்பாடல்கள் கிராமியப்பாடல்கள் வேதங்கள் புராணங்கள் சித்தர்பாடல்கள் மற்றும் புத்தர் காந்தி மகாவீரர் இயேசு வள்ளலார் வேதநாயகம்பிள்ளை நாமக்கல்லார் கவிமணிபோன்றவர்களின் தமிழ்ப்படைப்புகளை தமிழ்மொழியில் தெளிவாக உணர்த்திட தமிழ் இணையத்தளத்தில்வெளியிட தன்னார்வத்தோடு தமிழ்கணினியாளர்கள் முன்வரவேண்டும் கேட்டல் பேசுதல் படித்தல் எழுதுதல் படைப்பாற்றல் நடை முறை இலக்கணம் தானே கற்றல் சொல்லாட்சித்திறன் போன்ற அரிய பெரிய திறன்கள் தமிழ் இணையத்தின் மூலம் செம்மொழியாம் நம் தமிழ்மொழி உலகளாவிய தமிழ்மொழியாக வளர்ந்து உலகத் தமிழர்களை உருவாக்கும் என்பதில் ஐயமில்லை . உலகப்பொதுமறையாம் திருக்குறள்நல்லநல்ல நீதிநூல்களாம் நாலடியார்பழமொழி நீதிநெறிவிளக்கம் சீறாப்புராணம் பெரியபுராணம் சைவசித் தாந்தம் பிள்ளைத்தமிழ் போன்றவற்றை இணையதளத்தின் வாயிலாக கற்பிப்பதினால் தமிழ்மொழிசிறக்கும் என்பதில் ஐயமில்லை எனலாம் தொல்லியல் தொல்காப்பியம் நன்னூல் நிகண்டுகள் தண்டியலங்காரம் யாப்பருங்கலக்காரிகை யாப்பருங்கலவிருத்தி வீரசோழியம் போன்ற நூல்களில் உள்ள செய்திகளை தமிழ் இணையதளத்தின் வாயிலாக வெளியிட்டால் தமிழின் பெருமையெலாம் உலகறிய செய்யமுடியும்.

ஸ்கேம்பர் உத்தியில் குறிப்பிட்டுள்ள ஏழு உத்திகளில் பதிலிடுசெய்தல் இணைத்தல் சேர்த்தல் பொருந்தச்செய்தல் வேறுவடிவம்தருதல் வேறுபயன்பாடுகளைக்கண்டுபிடித்தல் நீக்குதல் எதிரும் புதிருமாக அமைத்தல் போன்றவற்றை தமிழ்இணையத்தின்மூலம் கற்பிக்கலாம் .பத்து வினாக்கள் உத்திகள் பனுவல்துண்டுகள் வட்டமேஜை வரிசை வட்டக் கருத்துப்பரிமாற்றம் அறிவு வட்டம் கற்றல்மூலைகள் விழுமங்கள் வரிசை பிரச்சனையை அனுப்புதல் மூன்றுபடி நிலைப்பேட்டி சிந்தி -இணை- பகிர் மற்றும் கதைத்துண்டுகள் ஆகியவற்றை தமிழ் இணையத்தின் மூலமாக மாணவர்களுக்கு சிறந்த மிகுந்த பயன் தரும்வகையில் உயர்ந்த முறையில் பயிற்சியளிக்கலாம்.

நான்தெரிந்து வைத்திருப்பது நான் தெரிந்து கொள்ள விரும்புவது நான் கற்றுக்கொண்டது சொற்சிதறல் கருத்து தீர்ப்பு கருத்தெழுதுதல் கருத்துப்பெட்டகம் செய்முறை எழுத்துத்திறன் பிரச்சனைக்குத்தீர்வு வளர்ச்சி பேசும் தகடுகள் மொன்டாஜ்போன்ற தமிழ்கற்பித்தல் உத்தியில் உள்ள செயல்பாடுகளை தக்க வழியில் தமிழ் இணையத்தளத்தில் பயன்படுத்தினால் தமிழ்வளங்கானும் அகில முழுவதும் என்பதில் ஐயமில்லை . எனவே மாணவர்களின் உள இயல் திறனுக்கு ஏற்ப தமிழ் பாடப் பகுதிகளை எளிமையாக்கி பள்ளித்தளம் கல்லூரி

பல்கலைக்கழக உலக அரங்கில் நம் உயிரின் மேலான முத்தமிழின் மேன்மைகளை பாமரரும் அறியும் வண்ணம் தமிழ் இணையதளத்தில் அமைப்பேர்த் நாம்அனைவரும் ஒன்றுபட்டால் நமது திறன்கள் எதிர்கால இந்தியாவை உருவாக்கும் மாணவனின் கையில் உள்ளங்கையில் உள்ள அழியாத உயர் நெல்லிக்கனியாக தமிழ் கணினி தமிழ் இணையதளத்தில் உலகளாவி உயர்ந்திடும் இரு சுடரன்றோ !

இணையம் வழி தமிழ் கற்றல் -கற்பித்தல் – ஓர் ஆய்வு

பி.ஆர்.இலட்சுமி

தமிழ்த்துறை வல்லுநர், வேலம்மாள் வித்யாஸ்ரம், PEARL ACADEMIC LEADERSHIP INSTITUTE

”எட்டுத்திக்கும் சென்றிடுவீர்! கலைச்செல்வங்கள்
இங்கு கொண்டுவந்து சேர்ப்பீர்“. – பாரதியார்.

இன்றைய உலகம் அறிவியல் உலகம் . அத்தகைய உலகில் தமிழ்மொழி சற்று பின்னோக்கிய நிலையில் சென்று கொண்டிருக்கிறது. கற்றல் திறன் இணையம்வழி மிகவும் குறைவான நிலையில் நடைபெற்றுக் கொண்டிருக்கிறது . கற்றல் திறன் இணையம்வழி நடைபெற மொழித்திறன் குறைபாடு , மொழியில் ஆர்வமின்மை , புதுமைக் கருத்துகளைப் படைக்கும் திறன் இல்லாமை போன்றவை காரணமாக அமைகின்றன . அவ்வகையில் இணையம்வழி கற்றல் – கற்பித்தல் பற்றிய செய்திகளை எடுத்துக்கூறுவது இவ்வாய்வின் நோக்கமாகிறது .

இன்றைய இளைஞர்கள் படிப்பு வேறு , சமூகத்தில் வாழும் வாழ்க்கைச் சூழல் வேறு என நினைத்து வாழ்ந்து வருகின்றனர். அன்றே மலர்ந்து அன்றே வாடும் கவர்ச்சிப் பூக்களாகிய சில வலைத்தளங்களைப் பார்த்து தாம் வாழும் வாழ்க்கைப் பாதையை மாற்றி அமைத்துக் கொள்கின்றனர் . இதனால் இணையத்தளம் என்றாலே சற்று விலகி நிற்கும் நிலை உருவாகியுள்ளது. வேலைவாய்ப்பிற்காகவும், ஆய்வுக்காகவும், தகவல் தொடர்பிற்காகவும் இணையத்தினை நாடிச் செல்லும் மாணவர்கள் தமிழ்க்கல்வி கற்பதற்காகவும் இணையத்தைப் பயன்படுத்த வேண்டும். தமிழ் கற்றலுக்கெனப் பல்வேறு வலைத்தளங்களும், மென்பொருட்களும் உருவாக்கப்படவேண்டும் .

'கல்வி கரையில்' என்பது நாலடி. அது போல்இணையம் வழி கற்றல் - கற்பித்தல் உத்திகளுக்கு இது தான் எல்லை என்று வரையறுக்க இயலாது. கற்பித்தலில் புதுமைகள் நிறைந்து <http://ccat.sas.upenn.edu/plc/tamilweb> சிறப்பாக இயங்கிக் கொண்டு வருகிறது . தமிழ் எழுதும் முறை (மென்பொருள் பயன்பாடு), ஃபிளாஷ்அட்டை, இதழ்கள், மின்புத்தகங்கள், வலைப்பூக்கள் தொடர்பு போன்ற நிலைகளில் - இணையம்வழியாகத் தமிழ் கற்பித்தல் நடைபெறுகின்றன. இச்சூழ்நிலையில் தமிழாசிரியருக்கு நழுவல் தயாரிப்புகள், அசைவூட்டம், ஒளி, ஒலி வடிவத் தயாரிப்புகளைத் தரவேற்றம் செய்யக் கற்றுக் கொள்ளுதல், இணையம் பற்றிய முழுமையான தெளிவுநிலை போன்றவை அவசியமாகிறது.

இணையம் வாயிலாகத் தமிழ் கற்பித்தலின்போது

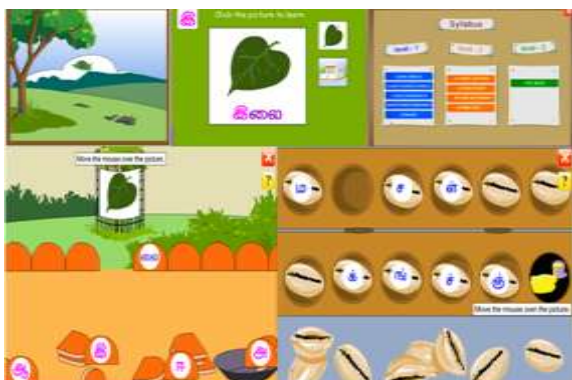
கணினியில் ஏற்படும் பிரச்சினைகள்
மென்பொருள் பற்றிய அறியாமை
எழுத்துருப் பிரச்சினைகள்
அகராதிகளைத் தெளிவுறப் பார்க்கும் நேரமின்மை

போன்ற காரணங்களினால் இணையம் வழி- தமிழ் கற்பித்தல் நடைபெற சிக்கல்கள் ஏற்படுகின்றன. ஆசிரியர்கள் தட்டச்சு பயிற்சி பெற Kiran's Typing Tutor போன்ற மென்பொருளைப் பயன்படுத்தலாம். NHM Writer (ஒருங்குறியீடு) மென்பொருளின் மூலம் தட்டச்சு செய்வதினால் பலவிதமான எழுத்துருப் பிரச்சினைகளைத் தவிர்க்க இயலுகிறது. கணினி பற்றிய போதிய தெளிவுபெற www.eegarai.net, velanblogspot.com தினமலர் - கம்ப்யூட்டர் மலர் (இணைய இதழ்) போன்றவற்றைப் பயன்படுத்தலாம். தினமலர் - கம்ப்யூட்டர் மலர் (இணையஇதழ்) தமிழ் கணினி - தொடர்பான பல செய்திகளைத் தருகிறது. கற்றல் திறன் சிறப்பாக நடைபெற கேட்டல், பேசுதல், படித்தல், எழுதுதல் போன்ற நான்கு படிநிலைகளில் நடைபெறுகின்றன. இந்நான்கு படிநிலைகளிலும் திறன் பெற்றவரால் மட்டுமே படைப்பாக்கத் திறனை உருவாக்க இயலும் .

	கேட்டல்	பேசுதல்	படித்தல்	எழுதுதல்
1	செய்தி - கேட்டல்	எழுத்துகள் - ஒலிப்பு	எழுத்துகள் - ஒலிப்பு	வரிவடிவம் - எழுதுதல்
2	உரையாடல் - கருத்துகள்	சொற்கள் - ஒலிப்பு ர, ற, ல, ள, ழ - ஒலி வேறுபாடு - ஒலிப்பு	சொற்கள் - படித்தல்	கொம்பு, துணைக்கால், வளைவு - எழுதுதல்
3	கதை - உட்கருத்து	கதை - கூறுதல்	பொருள் விளங்கும் வண்ணம் நிறுத்திப் படித்தல்	இடம் விட்டு அழகாக எழுதுதல்
4	நாடகம் - உணர்ச்சிகள்	எளிய செய்தி, சொற்களஞ்சியம் பெருக்குதல்	நிறுத்தற் குறி - படித்தல்	நிறுத்தற்குறியிட்டு எழுதுதல்
5	ஒலி - பாடல் கேட்டல்	படித்தவை, நன்றியுரை - செய்யுள் - படித்தல்	கதைகள், செய்தித்தாள் படித்தல்	தான் கேட்ட , படித்த பொருள் குறித்து - எழுதுதல், (இலக்கணம்)
6	சமூகம் - குறித்துக் கேட்டல்	எளிய தலைப்பு - பேசுதல்	பொதுக் கட்டுரைகள் , கடிதங்கள் -	விண்ணப்பம், கடிதம், இதழ் - எழுதுதல் (படைப்பாற்றல்)

தெரிந்ததிலிருந்து தெரியாதவற்றிற்குச் செல்லுதல் , தெளிவானவற்றிலிருந்து சிக்கலானவற்றிற்குச் செல்லுதல் , காட்சிப் பொருளிலிருந்து கருத்துப் பொருளுக்குச் செல்லுதல் போன்றவை கல்வியின் அடிப்படைக் கொள்கையாகும்.

www.tamilunlimited.com - வலைத்தளம் 2011 ஆம் ஆண்டுமுதல் புதிய வடிவமைப்புடன் தமிழ்க்கல்வியினைச் சிறப்பாக ஆங்கிலமும், தமிழும் கலந்து அளித்து வருகிறது.



இணையம்வழி – மழலைக்கல்வி கற்பித்தலில் ஒளி வடிவக்காட்சி (youtube - கதை, பாடல், எழுத்துகள் எழுதும் முறை) சிறப்பாக அமைந்துள்ளது .

ஒரு மொழியினைக் கற்பிக்க

திட்டமிடுதல்
ஆர்வமூட்டல்
ஆலோசனை வழங்குதல்
மதிப்பிடுதல்
ஊக்குவித்தல்
ஒழுங்கினை ஏற்படுத்துதல்
வினாவுதல்

போன்ற உத்திகள் கையாளப்படுகின்றன . நவீனத் தொழில் நுட்ப வசதிகளுக்கேற்ப இத்தகைய உத்திகளைக்கொண்டு கல்வி முறையில் தரப்படுத்துதலும் , புதுமையும் கொண்டு www.tamilvirtualacademy.org விளங்குகிறது .



சிங்கப்பூர், மலேசியா, இந்தியா, ஸ்ரீலங்கா போன்ற நாடுகளில் தமிழ் கற்பிப்பது சற்று எளிதான செயலாக இருப்பினும், எழுத்துகளைக் கற்பித்தல் சற்று அரிதான செயலாக உள்ளது . இன்றைய சூழலுக்கு ஏற்ற வகையில் எழுத்துகளை ஆசிரியர் கற்றுத் தருவது போல் www.kidsone.com வலைத்தளம் அளிக்கிறது . எழுதும்முறை, கதைகள், பாடல்கள், பயிற்சித்தாள்கள் என மூன்று மொழிகளில் அமைக்கப்பட்டுள்ளது சிறப்பாக உள்ளது .

பயிற்சித்தாள் இல்லா வகுப்பறையும் , வினையாட்டுமுறைக் கல்வியற்ற வகுப்பறையும் பயனற்றது . இதனைக் கருத்தில் கொண்டு

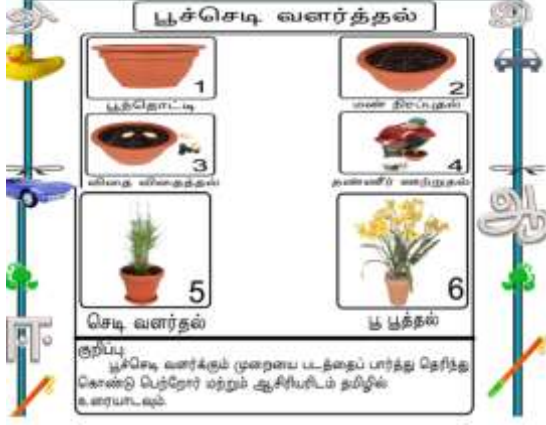
www.kidspark.com
www.noolagam.com
<http://www.balmitra.com/languagebook/Tamil/tamilflashcards>

போன்றவை இயங்கி வருகின்றன .

இன்றைய சமுதாயத்திற்குப் பயிற்சித்தாளுடன் கூடிய தமிழ்க்கல்விமுறை மாணவர்களின் படைப்பாற்றல் திறனை வளர்க்கும் . தமிழ்மொழிக் கல்வியினை இணையம் வழி வெளிக்கொணர்தல் தமிழரின் கடமையாகக் கருதலாம் . பாடத்தைக் கற்கும் பொழுது அகராதி, பழமையான தமிழ் செய்திகள் சுவடி பற்றிய தெளிவு இவற்றை

www.tamilheritage.org
www.tamilvirtualacademy.in
www.tamilmanam.com
www.virtualvinodha.com
www.cict.in

போன்றவற்றின் வழி அறிய இயலுகிறது.



இதற்கு இன்னமும் பல தொழில்நுட்ப வசதிகள் பயன்படுத்தி உலகம் முழுவதும் இணையவழி தமிழ் கற்றலை எளிதாக்கலாம் என்பது தெளிவாக்கப்பட்டுள்ளது.

ஆய்விற்குப் பயன்பட்ட இணையத்தளங்கள்

www.tamilvirtualuniversity.org
www.tamilunltd.com

எடுபுண்டு (edubuntu) வழி கல்வி

மேகவர்ணன் ஜெகதீசன்

மாணவ ஆசிரியர், மலேசிய ஆசிரியர் கல்விக் கழகம், இராஜா மெலேவார் வளாகம், சிரம்பான்.

முன்னுரை

எழுத்தாணியில் ஆரம்பித்து இன்று விரல் நுனியில் விளையாடும் வரை முன்னேறி உள்ளது நமது தமிழ் எழுத்துத் துறை. அது போல் மரத்தடியில் அமர்ந்து மணலைப் பயன்படுத்தியே ஆசிரியர்கள் கல்வியை முற்காலத்தில் போதித்தனர். இன்று கணினியைப் பயன்படுத்திக் கல்வி பல பள்ளிகளில் போதிக்கப்படுகிறது. ஆனால், அப்படி பயன்படுத்தப்படும் கணினி மென்பொருள்களில் தமிழ் மொழியின் ஆதிக்கம் மிக மிகக் குறைவு. எடு-புண்டு எனப்படும் கணினி இயங்குதளம் முழுக்க முழுக்க தமிழ் மொழியையும் தனது காட்சி மொழியாக பயன்படுத்துகிறது.

லினக்ஸ் எனப்படும் இயங்குதளத்தை அடிப்படையாக வைத்து உருவாக்கப்பட்டது உபுண்டு எனப்படும் கணினி இயங்குதளம். எடுபுண்டு எனப்படுவதை உபுண்டு கல்வித் தொகுதி எனவும் கூறலாம். கோவையில் நடந்த தமிழ் இணைய மாநாட்டில் திரு.இரவிசங்கர் மற்றும் தமிழ் விகிபீடியா குழுமமும் அங்கு வந்த மக்களுக்குக் கட்டற்ற மென்பொருளின் பயன்களை விளக்கி அங்கு வந்த மக்களுக்கும் உபுண்டு 10.04 விநியோகித்தனர்.

பயன்பாடு

நம் மலேசிய நாட்டில் பெரும்பாலும் மக்கள் மைக்ரோசப்ட் இயங்குதளத்தையே பயன்படுத்துகின்றனர். பள்ளிகளிலும் மைக்ரோசப்ட் இயங்குதளமே பயன்படுத்தப்படுகிறது. மைக்ரோசப்ட் இயங்குதளத்தைப் பயன்படுத்தி நாம் மாணவர்களுக்கு கற்றல் கற்பித்தல் நடவடிக்கை செய்ய வேண்டுமானால் நாம் கூடுதலாக வேறொரு மென்பொருளை வாங்கி, அதனைப் பயன்படுத்த வேண்டும். ஆனால் எடுபுண்டு இயங்குதளத்தைப் பயன்படுத்துவதன் வழி நாம் கூடுதல் மென்பொருளைப் பயன்படுத்த வேண்டியதில்லை. ஒரு DVD-ஐ பயன்படுத்தி அதன் வழியே நாம் கணினியை முழுக்க முழுக்க இயக்கலாம். இதனால் ஆசிரியர் கூடுதல் மென்பொருளை வாங்க வேண்டியதில்லை. மாணவர்களுக்காகப் பல மென்பொருள்கள் விளையாட்டு அடிப்படையில் இந்த இயங்குதளத்தில் பொருத்தப்பட்டிருக்கின்றன.

மாணவர்கள் தங்களது ஓவியக் கலையை வளர்க்க இதில் உள்ள தக்ஸ் பெயிண்ட் எனப்படும் மென்பொருள் உதவுகிறது. இது மாணவர்களின் கற்பனை வளத்தைப் பெருக்கவும் அவர்களிடத்தில் உள்ள கலைத் திறமையை வெளிக்கொணரவும் பேருதவியாக இருக்கிறது. இந்த மென்பொருளும் தமிழ் மொழியில் இயங்க வல்லது. மாணவர்கள் தங்களின் கைவண்ணத்தை அச்சிட்டு அதனைக் கோப்பு முறையில் சேமித்து வைத்துக்கொள்ளலாம்.

மாணவர் மற்றும் இன்றி ஆசிரியர்களும் தங்களின் அலுவலக பணிகளை இந்த இயங்குதளத்தின் வழி செய்யலாம். லைபர் ஆபிஸ் எனப்படும் மென்பொருளின் வழி ஆசிரியர்கள் மைக்ரோசப்ட் ஆபிஸைப் பயன்படுத்தி செய்யும் அலுவல் பணிகளைச் செய்யலாம். மைக்ரோசப்ட் ஆபிஸில் செய்த அலுவல்களையும் நாம் லைபர் ஆபிஸில் திறக்கலாம், தொடரலாம் மற்றும் திருத்தலாம்.

விவொயெம் எனும் மென்பொருளின் வழி ஆசிரியர் மனவோட்டவரை செய்து மாணவர்களுக்குக் கற்றல் நடவடிக்கையைப் போதிக்கலாம். பெரும்பாலும் கட்டுரைகளைப் போதிக்கும் போது ஆசிரியர் இதனைப் பயன்படுத்தலாம். கேஸ்டார் எனும் மென்பொருளின் வழி ஆசிரியர் மாணவர்களுக்கு விண்வெளி பற்றிய அறிவியல் அறிவுகளைப் புகுத்தலாம். விண்வெளியில் வான்கோள் வழி பார்ப்பது போல் இந்த க்ஸ்டார் வழி மாணவர்கள் நமது பால் வீதியை பார்க்கலாம்.

எடுபுண்டுவில் உள்ள தக்ஸ் மேத் எனப்படும் மென்பொருள் கணித பாடத்திற்காக தயாரிக்கப்பட்டது. ஆரம்பப் பள்ளி மாணவர்கள் முதல் இடைநிலைப்பள்ளி மாணவர்கள் வரை கணிதப் பாடம் கற்கும் அனைவருக்கும் ஆசிரியர்கள் இதனைப் பயன்படுத்தலாம். நான் தொடக்கத் தமிழ்ப்பள்ளியில் செய்த ஓர் ஆய்வின் வழி, இந்த தக்ஸ் மேத் மென்பொருளின் பயன்பாடு ஆசிரியர்களிடமும் மாணவர்களிடமும் நல்ல வரவேற்பினை பெற்றுள்ளது.

இந்த நவீன தகவல் தொழில்நுட்ப உலகத்தில் எடுபுண்டுவின் பயன்பாடுகள் இன்றியமையாதவை. ஆசிரியர்கள் கணினி ஆய்வுக்கூடத்தில் எடுபுண்டுவைப் பயன்படுத்தித் தாம் இருந்த இடத்திலிருந்தே மாணவர்களுக்குப் பயிற்சிகளை வழங்கலாம். வழங்கிய பயிற்சிகளையும் உடனுக்குடன் திருத்தலாம். இது மாணவர்களிடத்தில் கல்வியின் பால் உள்ள ஆர்வத்தைத் தூண்டி அவர்களைச் சிறப்பு தேர்ச்சி பெற செய்கிறது. பரிகார போதனையில் பயில்கின்ற மாணவர்களுக்கும் பாலர் பள்ளியில் பயில்கின்ற மாணவர்களுக்கும் ஜி-காம்பரீஸ் கல்வி தொகுதியில் உள்ள மென்பொருளைப் பயன்படுத்தி ஆயத்த பயிற்சிகளை வழங்கலாம்.

அதுமட்டுமல்லாமல், மாணவர்களுக்குக் கணினியின் ஆரம்பக்கட்ட பயன்பாட்டையும் இதில் உள்ள மென்பொருள்களின் வழி போதிக்கலாம். உதாரணத்திற்கு, விசைப் பலகையைப் பயன்படுத்தும் முறையை கேதட்ச் எனும் மென்பொருளின் வழி ஆசிரியர் அல்லது பெற்றோர்கள் மாணவர்களுக்குப் போதிக்கலாம். இந்தப் போதனை அனைத்தும் மாணவர்களுக்கு விளையாட்டு முறையிலேயே போதிக்கப்படுகிறது. மாணவர்களுக்கு இது மிகவும் சவால் மிகுந்ததாகவும் ஈர்க்கும் வகையிலும் அமைந்து அவர்களின் முயற்சியைத் தூண்டும் தூண்டுகோலாகவும் இருக்கும்.

அறிவற்றம் காக்கும் கருவி செறுவார்க்கும்
உள்ளழிக்கல் ஆகா அரண்

எனும் வள்ளுவரின் 421-வது குறளுக்கேற்ப அறிவே அழிவு வராமல் காப்பாற்றும் ஆயுதம். பகைவராலும் அழிக்க முடியாத பாதுகாப்பும் ஆகும். அப்படிப்பட்ட ஒரு ஆயுதம் ஒருவனுக்குத் தான் தன் குருவிடம் கற்கும் கல்வியின் வழி கிடைக்கப்பெறுகிறது. அந்தக் கல்வியை ஓர் ஆசிரியர் திறன்பட மாணவர்களுக்குப் போதிக்க வேண்டும். முற்கால கற்றல் கற்பித்தல் போல் கரும்பலகையும் வெண்கட்டியும் இல்லாமல் தகவல் தொழில்நுட்பத்தின் வழி மாணவர்களுக்குக் கல்வியறிவைப் புகுத்த இந்த எடுபுண்டு பெரும் பங்காற்றுகிறது.

இந்த இயங்குதளத்தைப் பாடத்துணைபொருளாகவும் மதிப்பீட்டுக் கருவியாகவும் ஆசிரியர்கள் பள்ளியில் பயன்படுத்தலாம். இதில் மாணவர்கள் விளையாடிய விளையாட்டின் புள்ளிகளைக் கொண்டு ஆசிரியர்கள் மாணவர்களின் அடைவு நிலையை எளிதாக மதிப்பீடு செய்யலாம். அது மட்டுமின்றி மாணவர்கள் எந்தப் பகுதியில் சிக்கல்களை எதிர் நோக்குகின்றனர் எனவும் ஆசிரியர் அறிய இந்த இயங்குதளம் உதவுகிறது.

உற்று நோக்கல்

பள்ளிசார் பட்டறிவு போதும், பயிற்றுப் பணியின் போதும் மாணவர்களிடையே இந்த இயங்குதளம் ஆர்வத்தை ஏற்படுத்தியது. மாணவர்கள் போட்டியிட்டு இதில் வரும் கேள்விகளுக்கு விடை கொடுக்க முனைந்தனர். மாணவர்கள் வெண்பலகையைப் பயன்படுத்தி கற்பதை விட எடுபுண்டுவைப் பயன்படுத்தி கல்விக் கற்பதை அதிகம் விரும்புகின்றனர். கற்றவற்றை மாணவர்கள் நீண்ட காலத்திற்கு நினைவில் வைத்திருக்கவும் முடிகிறது. எடுபுண்டுவைப் பயன்படுத்தி கல்வியைப் போதிப்பதின் வழி பள்ளிக்கு மட்டம் போடும் மாணவர்களும் தினமும் பள்ளிக்கு வந்து ஆர்வமாக பயில்கின்றனர். இது மாணவர்களுக்கு ஒரு புதைநிலை உந்து சக்தியாகவும் செயல்படுகிறது.

நேர்காணல்

நகர்புறப் பள்ளிகளிலும் புறநகர் பள்ளிகளிலும் நடத்திய ஆய்வின் வழி, ஆசிரியர் எடுபுண்டுவைப் பயன்படுத்தி பாடத்தைப் போதிப்பதின் வழி அன்றைய பாட நோக்கம் எளிதில் அடைகின்றது. ஆசிரியர்களும் தங்கள் அலுவல்களையும் எளிதில் செய்ய முடிகிறது என பள்ளி ஆசிரியர்கள் தெரிவித்தனர். இதனைப் பயன்படுத்தி பாடத்தைப் போதிக்கும் போது, மாணவர்கள் வகுப்பில் மிக ஆர்வமாகவும் முழு கவனத்தையும் செலுத்துகின்றனர் என ஆசிரியர்கள் கருத்துரைத்தனர்.

குறைகள்

இந்த இயங்குதளம் மலாய் மொழியில் இன்னும் வெளியிடப்படவில்லை. ஏனெனில் மலேசிய திருநாட்டில் இதனை முழுமையாக யாரும் பொறுப்பேற்று செய்யவில்லை. ஆனால் இதில் இந்தோனேசிய மொழி இணைக்கப்பட்டுள்ளது. இதில் உள்ள தகவல் மேத் மற்றும் சில மென்பொருள் ஆங்கில மொழியை மட்டுமே காட்சிமொழியாக வைத்துள்ளன. அனைத்து மென்பொருள்களும் மொழியாக்கம் செய்யப்பட்டால், அந்தந்த பள்ளிகளின் மொழிக்கேற்ப எளிதாக பயன்படுத்திக் கொள்ளலாம். இந்த இயங்குதளத்தில் நாம் லினக்ஸ் இயங்குதளத்திற்கான மென்பொருள்களை மட்டுமே கூடுதலாக நிறுவ முடியும். அதாவது .tar மற்றும் .taz வகையான கோப்புகளை நிறுவ முடியும். மைக்ரோசாப்ட் இயங்குதளத்தில் இயங்கும் சில மென்பொருள்களை நாம் இந்த இயங்குதளத்தில் நிறுவ முடியாது. அனைத்து மென்பொருள்களும் இந்த இயங்குதளத்தில் நிறுவ முடிந்தால், இன்னும் சிறப்பாக இருக்கும்.

முடிவுரை

பன்னாட்டு ஆசிரியர்களும் கணினி வல்லுநர்களும் ஒன்றிணைந்து தயாரித்த இந்த இயங்குதளம் முழுக்க முழுக்க இலவசமே. சமசீர் கல்வி போல் இந்த இயங்குதளமும் அதில் உள்ள மென்பொருள்களும் அனைவரும் பயன்படுத்தும் வகையில் அமைக்கப்பட்டிருக்கிறது. இந்த இயங்குதளத்தை இலவசமாகவே இணையம் வழி பதிவிறக்கம் செய்யும் வகையில் வழிசெய்யப்பட்டிருக்கிறது. இதனை நாம் நமது கணினி வன்டட்டில் எந்த மாற்றங்களும் செய்யாமல் 'லைப்' முறையில் பயன்படுத்தலாம். இந்த இயங்குதளத்தை www.edubuntu.org எனும் அகப்பக்கத்திற்கு சென்று நாம் நமது பெயரைப் பதியாமலே இந்த இயங்குதளத்தைப் பதிவிறக்கம் செய்யலாம். எடுபுண்டு ஆசிரியர்களுக்கும் மாணவர்களுக்கும் ஒரு நல்ல கல்வித்துணையாகும்.

மேற்கோள் பட்டியல்

1. Benjamin Mako Hill (2011): *The Official Ubuntu Book (6th Edition)* ; Boston, Pearson Education.
2. <https://wiki.edubuntu.org/Education/UbuntuSchools>
3. <http://ubuntu-tam.org/vaasal/>
4. https://wiki.ubuntu.com/Tamil_Team
5. <http://www.edubuntu.org/>
6. <http://en.wikipedia.org/wiki/Edubuntu>

கணினி வழி தமிழ் கற்றல் மற்றும் கற்பித்தல் குறித்தான ஆய்வுகள்

முனைவர் சு. சாந்தி

செல்லம்மாள் மகளிர் கல்லூரி, சென்னை

முன்னுரை

மொழி, பேசும் மக்களுக்கு அடையாளமாகத் திகழ்கிறது . மொழியே மக்களை இணைக்கிறது . கருத்துகளைப் பரிமாறிக் கொள்ள உதவுகிறது . கலை, இலக்கியம், பண்பாட்டின் ஊற்றுக்கண்ணாகத் திகழ்வது மொழியே , மொழியைக் காப்பதும் வளர்ப்பதும் அம்மொழியைப் பேசுவோரின் கடமையாகும் . உடலுக்கு உயிர் மூச்சே உயிர்ப்பு, எண்ணக் கருத்து வெளிப்பாட்டுக்கு உயிர் மூச்சு மொழி . காலங்காலமாகத் தமிழ்மொழி வாய்மொழியாகவும், ஓலைச் சுவடிகளாகவும் , தாள்களிலும் வெளிப்பட்டு வந்தது . அதன் இன்றைய நிலை வெளிப்பாடு கணினித் தமிழ், கணினியிலும் மின்-பதிப்புகளாக்கிச் சேர்ப்பது மொழிக்காப்போரின் கடமை .

பொருள்

கற்றல் - கற்பித்தல் பலமுறைகளைப் பின்பற்றி நடைபெறும் நிகழ்வாகும் . கணினி வழிக் கற்றல் - கற்பித்தலில் சிறந்தமுறையாகக் கருதப்படுவது நிரல்வழிக் (Programmed Teaching) கற்பித்தலாகும். இம்முறையில் ஒன்றைக் கற்றுத் தெளிந்தபிறகே அடுத்தபடிக்குச் செல்லமுடியும் . கணினியில் பல்லாடகக் கற்பித்தல் (Multimedia), மதிப்பீடு (Evaluation), அசைவூட்டுப் படங்கள் (Motion Picture), இசை (Music), விளக்கங்கள் (Description), படங்கள் (Pictures), வாய்மொழி உரைகள் (E-Lectures), எழுத்துரைகள் (Text) எனப் பல வசதிகளும் வாய்ப்புகளும் உள்ளன . ஆதலால், கற்றல் எளிதானதாகவும் நிலையானதாகவும் மேம்பாடுடையதாகவும் திகழ்கிறது , கற்போர்க்கும் ஆர்வமும் , விருப்பமும், நாட்டமும் ஏற்படுகிறது . இணைய வழிக்கல்வி (Online Education), தொலை தூரக்கல்வி (Distance Education) வாயிலாகச் சான்றிதழ்கள் , பட்டயங்கள் (Diploma), பட்டங்கள் (Degree), உயர்கல்வி (Higher Education) முதலியன பெறுமளவிற்குக் கணினிவழிக் கல்வி சிறப்புற்றிருக்கிறது .

கற்றலும் பயிலுதலும்

மொழியியல் திறன்களான கேட்டல் (Listening), பேசுதல் (Speaking), படித்தல் (Reading), எழுதுதல் (Writing) முதலிய திறன்களையும் மொழியியல் பகுப்பாய்வு, பண்பாடு, தொழில், சான்றாண்மை போன்ற நிலைகளையும் மையமாகக் கொண்டு கணினி வழி கற்றலும் பயிலுதலும் நடைபெறுகிறது .

ஒலியியல் (Phonetics), ஒலியனியல் (Phonology), உருபனியல் (Morphology), தொடரியல் (Syntax), பொருண்மையியல் (Semantics) என்ற நிலையில் மொழியியல் பகுப்பாய்வு உள்ளது .

ஒலியியல் (Phonetics)

பேச்சு வழக்கில் இடம் பெறும் சொற்களின் உச்சரிப்பில் மொழியியறுதியில் வரும் -l (ல்), -r (ர்), - ! (ழ்), - n (ன்) போன்றவை அப்படியே உச்சரிக்கப்படுவதில்லை . சில இடங்களில் உகரம் இணைக்கப்பட்டும் சில இடங்களில் இவை உச்சரிக்கப்படாமலும் இருக்கின்றன .

எ.கா.

எழுத்து வழக்கு	பேச்சு வழக்கு
கால்	காலு
ஆள்	ஆளு
வந்தாள்	வந்தா

பேச்சு வழக்கில் ல் , ள், ழ் வேறுபாடுகளும் ன் , ண் வேறுபாடுகளும் ற் , ற் வேறுபாடுகளும் சரிவர உச்சரிக்கப்படுவதில்லை. இதனால் எழுத்து வழக்கில் எழுதும்போது எழுத்துப் பிழைகள் மலிந்து காணப்படுகின்றன. ஆகவே இவ்வொலிகளுக்கிடையே காணப்படும் உச்சரிப்பு வேறுபாடுகளைத் தெளிவாக மாணவருக்குக் கற்பிப்பதோடு இவ்வொலிகள் ஒவ்வொன்றிற்குமுள்ள உச்சரிப்போடு எழுத்து வடிவங்களைப் பொருத்திக் காட்டி மாணவர் அவற்றின் வேறுபாடுகளை நின்கு உணருமாறு செய்ய தற்போது கணினி கற்றல் - கற்பித்தல் முறையில் எளிமைப்படுத்தப்பட்டுள்ளது.

தமிழில் இரு உயிரெழுத்துகளுக்கிடையே எழுதப்படும் . வல்லெழுத்துகளை க , ச, ட, த, ப, ற போன்றவற்றை உச்சரிக்கக் கற்றுத் தரும்போது ஒலி விதிகளை எழுத்து வடிவத்தை முன்னிறுத்தி அவற்றோடு ஒலிகளைத் தொடர்புபடுத்திக் கற்றுத் தரவேண்டும்.

எ.கா.

க- (k, g, x)
ச- (c, j, s)
ட- (t, d)
த- (t, d)
ப- (p, b)

இது போன்று ஒலியின் சேர்க்கைகள் மொழி முதலில் வர என்பது கணினி வழி கற்றல் - கற்பித்தல் முறையில் எளிமைப்படும்.

ஒலியியல் ஒலியனியல் ஆய்வு

ஒரு குறிப்பிட்ட மொழியில் மொழிக்கு முதலில் மற்றும் இறுதியில் வரக்கூடிய ஒலிகள் யாவை ? எந்தெந்த ஒலிச்சேர்க்கைகள் அனுமதிக்கப்படுகின்றன மற்றும் தவிர்க்கப்படுகின்றன என்பன போன்ற ஆய்வுகள் தக்க தொடரணைகள் (Algorithms) எழுதப்படுவதன் மூலம் மாணவனுக்கு எளிதில் கணினி வழி கற்பிக்க முடியும்.

உருபனியல் நிலை (Morphology)

உருபனியல் நிலையில் கணிப் பொறியைக் கொண்டு பெருமளவிலான ஆய்வுகளை மேற்கொள்ள முடியும் புணர்ச்சி மாற்றங்கள் யாவும் விதிகளின் அடிப்படையில் (rules) எழுதப்பட்டு, அவற்றைத் தக்க கட்டளை நிரல் (Programs) களின் மூலமாக கணினி மயமாக்கிவிடலாம்.

நிலை மொழியின் இறுதி ஒலியும் வருமொழியின் முதல் ஒலியும் புணர்ச்சியின் போது அடையும் மாற்றங்கள் மொழியியல் (ஒலியியல்) முறையில் விதிகளின் படிவத்தில் எழுதப்பட்டு அவற்றிற்கேற்ற கட்டளை நிரல்கள் தொகுக்கப்பட்டு மாணவர்களுக்கு கணினி வழி கற்பிக்க முடியும்.

எ.கா.

மிகுதல் (தோன்றல் விதிகள்)
பூ + கொத்து - பூங்கொத்து
அப்பா +ஐ- அப்பாவை

தொடர் உருவாக்கம் (Text generation)

உருபன் மற்றும் புணர்ச்சி நிலைகளில் ஒரு மொழியில் கணினி வழியிலான ஆய்வுகள் மேற்கொள்ளப்பட்ட பின்னால் அவற்றின் உதவி கொண்டு தொடர்களை உருவாக்கும் முயற்சி (Text generation) கணினியின் வாயிலாக மேற்கொள்ளப்படுகிறது.

உரைக்கோவை ஆய்வு

இலக்கிய உரைக்கோவை மற்றும் அறிவியல் உரைக்கோவை ஆகியனவற்றுக்கிடையே உள்ள வேறுபாடுகள் மிகவும் சிறப்பான விதத்தில் வரையறுக்கப்பட்டுள்ளன . இவற்றின் அடிப்படையில் நிரல் எழுதப்படும்போது . கணினி வழி உரைக் கோவையை பிரித்தரிய முடியும் . கணினி வழி மொழி ஆய்வில் தன்னிச்சையான இயந்திர மொழிபெயர்ப்பு எளிமைப்படும்.

கற்பித்தல் செயலை அலகுகளாகப் பிரித்து செப்பமாக வரைவதற்கான விரிவான உட்கூறுகள் , கல்வியியல் அறிஞர்களால் முன் மொழியப்பட்டுள்ளன . எந்தப் பாடத்திட்டத்திற்காகக் கணினியில் செயல்திட்ட வரைவு மேற்கொள்ளப்பட்ட வேண்டுமென்று முடிவு செய்தபிறகு , யாருக்காக (Target) என்பதையும், அவர்களுடைய கொள்திறன் (Input Resource) என்ன என்பதையும் முடிவு செய்தபிறகு இந்த நிரல்வழித் திட்டக் கற்பித்தலுக்கான செயல்திட்டம் வரையப்பட வேண்டும் .

ஒலியியல், ஒலியனியல், உருபனியல், தொடரில் உரைக்கோவை ஆய்வு போன்ற ஆய்வுகள் கணினி வழி நடைபெறுவதால் கணினி வழி மொழி கற்றல் / கற்பித்தல் (Computer Assisted Language Learning / Teaching (All/Calt) மேலும் வளர நின்கு வாய்ப்பு உள்ளது .

மொழியியல் பகுப்பாய்வை கொண்டு கற்பித்தலுக்கென வடிவமைக்கப்பட்ட மென்பொருள்களை -ன்று பிரிவுகளாக பகுக்கலாம். (1) செயல்பாட்டுச் செயல்திட்ட வரைவுத் தொகுதி (2) ஆக்கச் செயல்திட்ட வரைவுகள் (3) ஆக்க மொழிகள்.

இன்றைய நிலை

ஐந்தாம் தலைமுறைக் கணினிப் பயன்பாட்டில் மிக முக்கியமான செயலாக்கக் கூறு இயற்கை மொழியை ஆய்வு மற்றும் பயன்படுத்தலும் ஆகும் . இயற்கை மொழி வழியான தொடர்பாக்கம் மற்றும் இயற்கை மொழிப் பயிற்சிக்கு இந்தக் கணினி ஊடகத்தைப் பயன்படுத்திக் கொள்வது இன்றைய காலக் கட்டத்தின் சிறப்பான செயற்பாடாகும். இன்றைய மொழிப் பயன்பாட்டுச் சூழலில் , மொழிபெயர்த்தல், தகவல் தொடர்பு , இயற்கை சூழலில், மொழிபெயர்த்தல், தகவல் தொடர்பு இயற்கை மொழியை நிவீனப்படுத்துதல், தகவலைக் கோரிப் பெறுதல், அறிவை ஆழ்ந்தறிதல், கணினி வழிக் கற்பித்தல் ஆகியன நடைமுறைச் செயலியக்கங்களாகும் .

மாணவனுக்கு மொழிப் பற்றிய அறிவாக்கம் தரும் தளத்தில் -ன்று நிலைகளில் கணினியின் பயன்பாடு கருத்த தக்கதாகும்.

மொழி கற்பித்தல்

ஆசிரியர் நோக்கில் மாணவனுக்கு மொழி கற்பித்தலுக்கான பயன்பாடு நெறிகளை உருவாக்குவதாகும் .

பெயர்ச் சொல் ஆய்வு

தமிழ் மொழியில் பெயர்ச் சொற்களுடன் வேற்றுமை உருபுகளைச் சேர்ப்பதால் ஏற்படும் மாற்றங்கள் , பண்மை உருபு போன்ற பிறவற்றைச் சேர்ப்பதால் ஏற்படும் மாற்றங்கள் ஆகியனவற் றை விதிகளின் வடிவத்தில் நிரல்கள் எழுதுவதன் -லமாக மாணவர்களுக்கு கணினி வழி எளிமையாக கற்பிக்க முடியும் .

எ.கா.

மரம் + ஐ- மரத்தை

நாம் + ஐ- நம்மை

குடம் +கள்- குடங்கள்

வினைச் சொற்கள்

தமிழ் மொழியைப் பொறுத்த அளவில் கணிப் பொறி கொண்டு வினைச் சொற்கள் ஆய்வு அதிக அளவில் நடைபெறுகிறது. காலங்காட்டும் உருபுகளின் அடிப்படையில் வினைச் சொற்கள் பிரிக்கப்பட்டு , நிரல் எழுதப்பட்டு மாணவர்களுக்கு கணினி வழி கற்பிக்க முடியும் .

மொழி பயில்தல்

மாணவரின் நோக்கில் ஆசிரியரிடமிருந்து மொழியைக் கற்றுக் கொள்வதற்கான செயற்பாட்டமைவைக் குறிப்பதற்கும்.

தமிழ் எழுத்து வழக்கில் சொல் முதலில் ல , ட ஒலியன்கள் வருவது இல்லை. ஆனால் பேச்சு வழக்கில் அவை வரும் .

(எ.கா) லட்சு, லாடம், டப்பா, டமாரம்

எழுத்து வழக்கில் அமைகின்ற மூன்று மெய்க்கூட்டுகள் (-ர்க்க - ; -ழக்க-; -ய்க்க - போன்றவை) பேச்சு வழக்கில் இரு மெய்க்கூட்டுகளாக அமைகின்றன.

எ.வ பே. வ

பார்க்கிறேன் பாக்கிறேன்
வாய்க்கால் வாக்கால்

தமிழ் ஆங்கிலம்

நான் புத்தகத்தைப் படித்தேன் I read the book
அவள் காய்களை நறுக்கினாள் She cut the vegetables

தமிழ்: எழுவாய் + செயப்படு பொருள் + பயனிலை என்ற அமைப்பினைப் பெற்றுள்ளது .

ஆங்கிலம்: எழுவாய் + பயனிலை + செயப்படு பொருள் என்று அமையும் . மாணவர்கள் ஆசிரியரிடமிருந்து மொழி பயிலும் போது மொழியின் நிலையை எளிமையாக கணினி வழி கற்க முடியும் .

மொழி சார்தல்

குழந்தை தாய்மொழியைச் சூழல் வழியாக இயல்பாகக் கற்றுகொள்வதைக் குறிப்பதற்கும் .

தமிழில் இடம்பெறும் உறவுமுறைச் சொற்கள் , விளிச்சொற்கள், தன்மை பன்மை அமைப்பு , வியர்கோள் அமைப்பு , ஏவல் அமைப்பு போன்றவற்றைக் கற்றுக் தரும்போது தமிழில் இல்லாத இரு கூறுகள் ஆங்கிலத்தி ல் உள்ளன இதை கணினி வழி எளிய முறையில் வேறுபடுத்தி காட்டமுடியும் .

எ.கா.

தமிழ்	ஆங்கிலம்
அண்ணன் - தம்பி	(Elder, Younger) brother
அக்கா - தங்கை	(Elder, Yonger) Sister
நாங்கள், நாம்	we
நீ, நீங்கள்	you
நண்பர்கள்	O ! friends

இந்த மூன்று அமைவுகளிலும் கணினியின் பயன்பாடு குறிப்பிடத் தக்க விளைவுகளை ஏற்படுத்துகிறது .

தமிழ் மொழியைக் கற்பிக்கும்போது அடிப்படைத்திறன்களான (Basic Competencies) கேட்டல், பேசுதல், படித்தல், எழுதுதல் திறன்களை இரு வழிகளில் கற்பிக்க முடியும் . இதனால் திருத்தமான பேச்சு , வெற்றொலி, மயங்கொலிப் பிழைகள், குறில், நெடில் வேறுபாடறிதல் முதலியவற்றை மாணவர்களுக்கு எளிதாகக் கற்பிக்க முடியும் .

முடிவுரை

கேள்வியோடு காட்சியும் இணையும் போது கற்றல் மேம்படுகிறது . இவ்வாறான காட்சிப்படுத்தலை ‘உருப்போலி’ (Simulation) என்பர். ஒலி - ஒளி இரண்டும் இணையும்போது ஈர்ப்புடைய துணைக்கருவியாகவும் , தானே ஆர்வத்துடன் தொடர்ந்து சலிப்பில்லாமற் கற்கவும் கற்பிக்கவும் , துணை செய்கிறது . கணினி வழியாகக் கற்பிக்கும்போது விரைவு , நுட்பம், தெளிவு, மாணவர் ஈர்ப்பு , அறிவுத்திறனை நிலைநிறுத்துதல் , தன்னார்வம் முதலிய பல விளைதிறன்கள் வலுப்படுத்துகின்றன .

சிங்கப்பூரில் கணினிவழித் தமிழ்மொழிக் கற்றல் கற்பித்தல்

த. பொன்னி

முனைவர் பட்ட ஆய்வாளர், அறிவியல் தமிழ் மற்றும் தமிழ் வளர்ச்சித்துறை, தமிழ்ப் பல்கலைக்கழகம், தஞ்சாவூர்
ponniaditya@gmail.com

முன்னுரை

தமிழ்மொழி தகவல் தொழில்நுட்பத்தின் வளர்ச்சிக்கு ஈடுகொடுத்து வருகின்றது . அறிவியல், பொறியியல் பாடங்கள் மட்டுமின்றித் தமிழ்ப் பாடங்களையும் கணினிமூலம் கற்பிக்க முடியும் என்பதைச் சிங்கப்பூர் உண்மையாக்கி நடைமுறைப்படுத்தி வருகின்றது . புலம்பெயர்ந்த நாடுகளில் தமிழ்மொழியைக் கணினி மூலம் கற்பிப்பதில் சிங்கப்பூர் முன்னோடியாக விளங்குகின்றது . இதன்மூலம் தமிழ்மொழியை அறிவியல் மயமாக்கி அறிவியல் தமிழை உருவாக்குகிறது .

சிங்கப்பூரில் தமிழ்க் கற்பித்தல் முறைகள்

சிங்கப்பூர் பள்ளிகளில் மொழியின் அடிப்படைத் திறன்களான கேட்டல் , பேசுதல், படித்தல், எழுதுதல் ஆகியவற்றை வளர்த்து மாணவர்களிடையே கருத்துப் பரிமாற்றத் திறனை மேம்படுத்தும் கற்பித்தல் முறைகள் பின்பற்றப்படுகின்றன.

மேலும் வாழ்க்கையைத் திறம்பட நடத்திச் செல்ல உதவும் சிந்தனைத்திறன் வளர்ச்சிக்கு முக்கியத்துவம் அளிக்கப்படுகிறது. இத்திறன்களை வளர்க்கும் நோக்கில் புதிய கற்பித்தல் முறையான கருப்பொருள் சார்ந்து அமையும் ஒருங்கிணைந்த பாடமுறை பள்ளிகளில் பின்பற்றப்படுகிறது .

“தமிழ் வழங்கும் வேறு எந்த நாட்டையும் விடச் சிறந்த துணைக் கருவிகளையும் , நவீன சாதனங்களையும் பயன்படுத்தித் தமிழ் கற்பிக்கக்கூடிய வாய்ப்பு வசதி கள் சிங்கப்பூரில் மிகுதியாக உள்ளன ” (கா. இராமையா, ப. 101)

சிங்கப்பூர்ச் சூழலுக்கும் வளர்ந்துவரும் தொழில்நுட்பத்திற்கும் ஏற்ப மாணவர்கள் சிறந்து விளங்கும் வகையில் கணினியும் பயிற்றுக் கருவியாக அமைந்துள்ளது . இதனால் ஒவ்வொரு வகுப்பறையிலும் கணினி இணைக்கப்பட்டு அதன்வழி மாணவர்கள் கற்கவும், கற்பிக்கவும் முடிகிறது.

தமிழ்க் கற்பிக்கப் பயன்படுத்தும் மென்பொருள்களில் சில-

மென்பொருள் என்பதை இரண்டு வகையாகப் பிரிக்கலாம் . அவை,

1. சந்தையில் தயார் நிலையில் விற்கப்படும் மென்பொருள்
2. கல்வி அமைச்சகம் வெளியிட்டுள்ள மென்பொருள், கணினி வட்டுகள்

என்பனவாகும்.

எளிமையான மென்பொருள் MS-Word, Power Point

உயர்நிலை மென்பொருள் – Director, Authorware ஆகியவற்றைக் கொண்டும் தமிழ்மொழி கற்பிக்கப்படுகிறது .

மாணவர்களின் தேவைக்கேற்ப மரபுத் தொடர்களைக் கற்பிக்கவும் , சொற்களின் பொருள்களைக் கற்பிப்பதற்கும் இம்மென்பொருளைப் பயன்படுத்திப் பயிற்சிகள் தயாரிக்கின்றனர் . இப்பயிற்சிகள் இருவழித் தொடர்பிலான பயிற்சிகளாக (Interactive Exercise) அமைக்கப்பட்டுள்ளது. ஆசிரியர் ஓர் இடத்திலும் மாணவர் ஓர் இடத்திலும் இருந்துகொண்டு தமிழ்ப் பயிற்சிகள் மிக எளிதான முறையில் செய்வதற்கு இம்மென்பொருள் பயன்படுத்தப்பட்டு வருகிறது. இம்மென்பொருளின் வழிக் கீழ்க்கண்ட பயிற்சிகள் செய்யப்படுகிறது .

1. தெரிவு விடை வினா பயிற்சிகள் (Multiple Choice Question)
2. மாறிக் கிடக்கும் சொற்களைச் சரியான தொடராக்கும் பயிற்சிகள் (Jumbled Words)
3. இடதுபக்கம் சொற்கள் வலது பக்கம் மாறிக் கிடக்கும் பொருள்கள் ஆகியவற்றை இணைக்கும் பயிற்சிகள் (Match it)

இதனால் மாணவர்களது பிழைகளை மாணவர்களே திருத்திக் கொள்ள முடியும் .

Dream Viewer

கட்டுரை எழுதுதல், கதை சொல்லுதல் ஆகியவற்றிற்கு இந்த மென்பொருளைப் பயன்படுத்துகின்றனர் .

Video Editing

ஒரு பாடத்தின் அறிமுகமாகச் சிலவற்றைக் காண்பிக்கவும் பாடத்தின் அல்லது பாடலின் பொருளை விளக்கவும் திரைப்படக் காட்சிகள் பயன்படுத்தப்படுகின்றன .

பாடம் தொடர்புடைய திரைப்படப் பகுதிகளைத் தனியாக வெட்டியெடுத்து அவற்றில் ஆசிரியருக்குத் தேவையான விதத்தில் ஒலி , ஒளிக் கூறுகளை மாற்றி அமைத்துக் கொண்டு பாடம் நடத்தப்படுகிறது . இம்முறையால் மாணவர்கள் தமிழ்ப் பாடத்தை ஆர்வமுடன் கற்கின்றனர் .

எம்.எஸ். ஆபிஸ் பேக்கேஜ் (MS-Office Package)

இணைத்து என்ற மென்பொருள் உள்ளது . இதனால் சொல் அமைப்பு மென்பொருளில் எந்த மொழியில் உருவாக்கியதையும் மாற்றம் செய்யாமல் இந்த மென்பொருளுக்குள் கொண்டுவர முடிகிறது .

பவர் பாயின்ட் மென்பொருள் மிகச் சுலபமான ஒரு பல்லுடகச் சாதனமாகும் . மீஇணைப்புச் (Hyper links) செய்வதற்கும் இதில் வழியுண்டு . சொற்கள், படங்கள், அதன் ஒலிகள் இவற்றைப் பல வழிவகையில் பலவிதத்தில் திரையில் வந்துபோகச் செயற்படுத்தலாம் .

இணையதளங்கள்வழித் தமிழ்க் கற்பித்தல்

பாடக் கருப்பொருளையொட்டி இணையத் தளங்களில் இடம்பெற்றுள்ள பக்கங்களை முதலில் தேர்ந்தெடுத்துக் கொண்டு அவற்றில் உள்ள செய்திகளை ஒட்டி நடவடிக்கைகளைத் திட்டமிட்டுப் பாடத்திட்டத்தை உருவாக்கிக் கொள்கின்றனர் .

சிங்கப்பூரில் சில பள்ளிகளில் ஒவ்வொரு பருவத்திலும் ஒரு வாரம் மாணவர்கள் வீட்டில் இருந்து இணையதளத்தின்வழி மாணவர்கள் அவர்களுக்குக் கருத்துப் பரிமாற்றம் செய்யவும் ஆசிரியரோடு கருத்துப் பரிமாற்றம் செய்யவும், மாணவர்களுக்கு வீட்டுப்பாடங்களும் இணையம் மூலம் கொடுக்கப்படுகிறது . மாணவர்கள் வீட்டுப்பாடம் செய்து முடித்தவுடன் கணினியே திருத்தி மதிப்பெண் கொடுத்துவிடும் . ஆசிரியர் அதைக் குறித்துக் கொள்வார். இம்முறை மூலம் இணையவழிக் கற்பதில் மாணவர்கள் அதிக ஆர்வமாக ஈடுபடுகிறார்கள் .

இணைய வினா விளக்கம் (Web Quest)

இணைய வினா விளக்க முறையில் மாணவர்கள் கற்றுக் கொள்பவர்களாகவும் , கற்பிப்பவர்களாகவும் செயல்படுகின்றனர் . மாணவர்களிடம் குழு உணர்வை வளர்க்கவும் , குழுவாக இருந்து அறிவைப் பெறவும் இம்முறை பெரிதும் பயன்படுகிறது . ஒரு பணியைத் தொடங்கும்முன் எவற்றை எல்லாம் கவனத்தில் கொள்ள வேண்டும். யார் யார் என்ன என்ன வேலைகளைச் செய்ய வேண்டும் என்றும் குறிப்பிட வேண்டும் . பின்பு அதைப் பகுத்து, தொகுத்து ஒன்றுசேர்க்க வேண்டும் . கருத்துக்களைப் பெறுவதற்கு எவ்வெவ் இணையத் தளத்திற்குச் செல்ல வேண்டும் , யார் யாருக்கு நன் றி கூறவேண்டும் என்னும் பணிகளை உள்ளடக்கியது 'இணைய வினா விளக்கம்'.

மின்னஞ்சல் வழித் தமிழ்க் கற்பித்தல்

மாணவர்கள் ஒவ்வொருவருக்கும் 'மின்னஞ்சல் முகவரி' இருக்க வேண்டியது அவசியமாகும் . ஆசிரியர்கள் தங்கள் மாணவர்களின் நிலை அறிந்து அவர்களுக்கு ஏற்ற வகையில் மின்னஞ்சல் பாடங்களை உருவாக்குகின்றனர் . அப்பாடங்களைச் செய்திக் கடிதங்களாக மாணவர்களுக்கு அனுப்புகின்றனர் . மாணவர்களின் கருத்துக்களை மின்னஞ்சல் மூலமாகவோ நேரிலோ கேட்டறிகின்றனர் .

செய்திக் கடிதத்தின் கூறுகளில் சில

1. மாணவர்களின் கட்டுரை
2. சொற்களஞ்சியம்
3. புதிர் வினாக்கள்
4. விடுகதைகள்

மின்னஞ்சல்வழிக் கற்றலின் பயன்கள்

1. ஆசிரியர் மாணவர் நெருக்கம் அதிகமாகிறது
2. உறவு மேம்படுகிறது
3. மாணவர்களின் நிலையை உணர்ந்து அவர்களின் தேவைக்கேற்பப் பாடங்களை உருவாக்க முடிகிறது.
4. ஆசிரியர்களுக்கு நெகிழ்வுத் தன்மையும் அதிக சுதந்திரமும் கிடைக்கிறது .

கணினிவழிக் கற்பித்தலினால் உண்டாகும் பயன்கள்

- ஒரே மென்பொருளைப் பல பாடநோக்கங்களுக்குப் பயன்படுத்தலாம் .
- பாடங்களைக் கணினியில் தயாரிக்கவும் தயாரித்த பாடங்களை விரைவில் மாற்றவும் திருத்தவும் முடியும் .
- தனிநபர் திறனுக்கேற்பப் பயன்படுத்தலாம் .
- ஆசிரியர் 'பவர்பாயண்ட்' மென்பொருளில் தயாரித்தால் முக்கியமான எந்தக் கருத்தையும் மறக்க முடியாது .
- மாணவர்களின் சரியான விடைக்குப் பாராட்டும் , தவறான விடைக்கு வழிகாட்டுதலும் கணினி கொடுக்கும். இவ்வாறு உடனுக்குடன் பதில் கிடைக்கும் .
- சிங்கப்பூரில் 'சார்ஸ் நோய்' வந்தபோது பள்ளிகள் மூடப்பட்டன . அப்போது மாணவர்கள் வீட்டிலிருந்து கல்வி கற்பதற்கு கணினியே பெரிதும் பயன்பட்டது .
- மாணவர்கள் வீட்டிலிருந்தபடியே தமிழ்ப்பாடம் கற்கமுடியும் .
- வழக்கமான வகுப்பறைச் சூழலில் கிடைக்கும் பயனைவிடப் பன்மடங்கு கூடுதலான பயன் கிடைக்கிறது .
- மாணவர்கள் தமிழ்மொழி கற்பதில் உற்சாகமாகவும், ஆர்வமுடனும் செயல்படுகின்றனர் .

மாணவர்களின் முழுக்கவனத்தையும் தமிழ்ப் பாடத்தின்மேல் ஈர்க்க முடியும் .

சிங்கப்பூரில் கணினியில் தமிழ்ப்பாடம் கற்பிக்கப்பட்ட வரலாறு

- 1987ஆம் ஆண்டு சிங்கப்பூரில் தமிழாசிரியர்களுக்குக் கணினி கொடுக்கப்பட்டது.
- பாரதி, துணைவன் சொற்செயலிகளைக் கொண்டு தட்டச்சு வேலைகளுக்கு மட்டும் கணினி பயன்படுத்தப்பட்டது.
- MS-Word, Power Point போன்ற மென்பொருள்களை ஆசிரியர்கள் அதிகமாகப் பயன்படுத்தினர் .
- பிறகு கல்வி சார்ந்த மென்பொருள்களைக் கற்றல் கற்பித்தலுக்குப் பயன்படுத்தப்பட்டது .
- மேசைக் கணினி யிலிருந்து மடிக் கணினிக்கும் குறிப்புக் கணினிக்கும் கையடக்கக் கணினிக்கும் மாறுவதில் ஆசிரியர்களுக்கு பெரிய சிக்கலோ பிரச்சனையோ ஏற்படவில்லை .
- தற்போது குறுந்தகடு , ஒலிவட்டு, ஒளிவட்டு, ஒளிஒலி நாடாக்கள் என்னும் நிலையில் தமிழ் கற்பிக்கப்படுகிறது.
- 'நம் நாடி' என்ற இணையதளம் சிங்கப்பூர் கல்வி அமைச்சகத்தால் நிர்வகிக்கப்படுகிறது .

- தமிழாசிரியர்கள் ஒன்றுகூடி , தங்கள் பாடங்களை நடத்துவதற்குத் தேவையான பயிற்சிகளை உருவாக்கி <http://singtisc.org/edu/pri/PRIMARY.HTM> என்னும் ஓர் இணையதளத்தில் வைத்துள்ளனர் . இங்கிருந்து ஆசிரியர்கள் எந்தப் பாடத்தை வேண்டுமானாலும் எடுத்துப் பயன்படுத்திக் கொள்கின்றனர் .

முடிவுரை

சிங்கப்பூரில் தமிழ் கற்றல் கற்பித்தலை இனிமையாக்குவதற்குக் கணினி பயன்பாட்டை அதிகரித்துக் கொண்டே வருகின்றனர். இரண்டாம் மொழியாகத் தமிழைக் கற்கும் சிங்கப்பூர் மாணவர்கள் எளிமையானவும் இனிமையாகவும் கணினிவழிக் கற்கின்றனர் . இவ்வாறு அறிவியல் தொழில்நுட்பங்களைத் தமிழ்மொழி கற்பிக்கப் பயன்படுத்தும்போது தமிழ் வருங்காலத்திலும் வாழும் மொழியாக விளங்கும் .

துணைநூல்கள்

1. ஐந்தாவது உலகத் தமிழாசிரியர் மாநாட்டு மலர் , சிங்கப்பூர், 2001
2. ஆ.ரா. சிவகுமாரன், சிங்கப்பூர்த் தமிழ்ப் பாடநூல்களிலும் , தமிழ் கற்பித்தலிலும் அறிவியல் கூறுகள் ஒரு பார்வை, கட்டுரை.
3. வி.ஆர்.பி.மாணிக்கம், அயலகத் தமிழ்க்கலை இலக்கியம்- சமகாலச் செல்நெறிகள் , உலகத் தமிழாராய்ச்சி நிறுவனம், சென்னை, 2001
4. கலைமணி இரத்தினச்சாமி , சிங்கப்பூரில் தமிழ் கற்றல் கற்பித்தலை இனிமையும் எளிமையும் மிக்கதாக்குவோம், தமிழிக் கலை அச்சகம், 2001

தமிழ்மொழி கற்றல் கற்பித்தலில் தகவல் தொழில்நுட்பத்தின் பங்களிப்பு

முனைவர் சாம் கிதியோன்

உதவி பேராசிரியர், முதுகலைத் தமிழ்த்துறை, பிஷப் ஹீபர் கல்லூரி திருச்சிராப்பள்ளி

முனைவர் சீதா லட்சுமி

இணைப்பேராசிரியர், ஆசிய மொழிகள் மற்றும் பண்பாடுகள், தேசியக்கல்விக்கழகம், சிங்கப்பூர்

இக்கட்டுரை இரண்டு பகுதிகளாக அமைந்துள்ளது . முதல் பகுதி தமிழகத்தின் திருச்சியில் உள்ள பிஷப் ஹீபர் கல்லூரியில் தமிழ் மொழி கற்றல் கற்பித்தலில் இடம்பெறும் தகவல் தொழில்நுட்பம் பற்றி விவரிக்கின்றது . குறிப்பாக, பாடப்புத்தகம் வாயிலாக மாணவர்களுக்குச் செய்யுளைக் கற்பித்தலில் உள்ள சவாலை திறம்பட மேற்கொள்ள நவீன கணிப்பொறி தொழில்நுட்பம் மிக இன்றியமையாததாக உள்ளது . வகுப்பறைகளிலே இணையதளம் மற்றும் பவர்பாயிண்ட் மூலம் நேரடியாக காட்சிப்படுத்தல் கற்றல் கற்பித்தலை எளிமையாக்குகிறது. அதைப்பற்றி இக்கட்டுரை விரிவாகப் பேச இருக்கிறது .

சிங்கப்பூரில் தகவல் தொழில்நுட்ப முதன்மைத்திட்டம் என்பது மூன்று நிலைகளில் அறிமுகமாகித் தன் வெற்றியைப்பரப்பிவருகிறது. சிங்கப்பூரில் தமிழாசிரியர்க்குத் தமிழ் கற்பிப்பதில் குறிப்பாக ஆசிரியவியலைக் கற்பிப்பதில் தகவல் தொழில்நுட்பம் எவ்வாறு இடம்பெறுகிறது என்று பார்ப்போம் . விம்பா, வெப்குவெஸ்ட், முகப்புத்தகம், வாய்ஸ்த்திரெட், போஸ்ட்டரஸ், குளோக்ஸ்டர், ஃபோட்டோஸ்டோரி எனப்படவும் தமிழ் மொழியைக் கற்பித்தலில் இடம்பெற்றுள்ளன . அவை பற்றி ய செய்திகளைக் கட்டுரையின் இரண்டாம் பகுதி பகிர்ந்துகொள்ளும்

கட்டுரை

ஈராயிரமாண்டிற்கு மேலான தமிழ்க் கல்வி மரபில் கற்றல்- கற்பித்தல் தொழில் சிறப்பான இடத்தை நவீன காலத்தில் எட்டியுள்ளது . சந்ததிசந்ததியாக, எண்ணங்களை வெறும் மணலில் விரல் வைத்து எழுதிப் பழகிக் கற்ற கற்பித்த நாட்கள் மாறி ஓலைச்சுவடி , அச்சத்தாள் என்று வளர்ந்து , தற்போது கணிப்பொறித் தொழில்நுட்பத்தில் தொடுதிரை (Touch Screen) வசதி வரை நீட்சிபெற்றுள்ளது . கற்றல்-கற்பித்தல் தொழில்நுட்பத்தில் ஏற்பட்ட சிறப்பான வசதிப் பெருக்கங்கள் தமிழ்ச் செய்யுள் கற்றல்- கற்பித்தல் செயல்பாட்டில் எத்தகைய மாற்றங்களைக் கொண்டு வந்துள்ளன என்பது தனியே ஆய்வுக்குரியது . பிற துறையினர் கற்றல்- கற்பித்தல் மற்றும் ஆய்வு வசதிக்காகக் கணிப்பொறித் தொழில்நுட்பத்தை வகுப்பறைகளிலோ கருத்தரங்க நிகழ்வுகளிலோ பயன்படுத்தும் அளவிற்கு தமிழ்த்துறை மக்கள் பயன்படுத்துவதில்லை.. அதற்கு அந்தத் தொழில்நுட்ப அறிவைக் கற்று அன்றாடப் பயன்பாட்டிற்குக் கொண்டுவர முயற்சிக்காததே காரணம் எனலாம் . மேலும், கணிப்பொறி உபகரணங்கள் , இணையதள வசதிகளின் கிடைப்பில் காணப்படும் அசௌகரியங்கள் மின்தடை போன்ற புறக்காரணிகள் சிலவும் எமது சூழலில் கணிப் பொறித் தொழில் நுட்பவழி கற்றல் கற்பித்தலை நடைமுறைப்படுத்துவதைத் தாமதமாக்குகின்றன. அதையும் மீறிய சிலமுயற்சிகள் வரவேற்புக்குரியன .

தற்போது செவ்வியல் இலக்கியங்கள் முழுமையாகவும் பக்தி இலக்கியங்களில் சிலவும் நவீன இலக்கியங்களில் குறிப்பிடத்தகுந்த சில பகுதிகளும் இணையதளத்தில் விளக்க உரைக்குறிப்புகளாக காணக் கிடைக்கின்றன . இவைகளைப் பதிவிறக்கம் செய்துப் பயன்படுத்திகொள்ளலாம் . தமிழ்ச் செய்யுளைக் கற்பித்தலில் கேள்விஞானம் என்பது பழைய காலங்களில் முக்கியத்துவம் பெற்றிருந்தது . சான்றாக, வெண்பாச் செய்யுளாக இருந்தால் ஒருமுறை சொல்வதைக் கேட்டு அதனைத் திரும்ப சொல்லக்கூடிய அளவில் அமைய வேண்டும் , அதற்கேற்ப செய்யுளில் ஓசைநயங்களும் செய்யுள் உறுப்புகளும் இருக்கும் அதனைப் புரிந்து கொண்டால் கற்றல் -கற்பித்தல் எளிமையாக நிகழும் . தற்போது பிரதிகள் (Text) அனைத்தையும் கணினித் திரையில் பார்ப்பதற்கு ஏற்ப நாம் உருவாக்கிக் கொள்ள முடியும் . கரும்பலகைப் பயன்பாட்டில் இல்லாத சில சாத்தியங்கள் கணினித் திரைப்பயன்பாட்டில் உண்டு . ஏற்கனவே சில பயிற்சிகள் செய்து நடத்த வேண்டிய பகுதிகளைப் பதிவேற்றம் செய்து கொள்ள வேண்டும் . அதனை மாணவர்கள் வாசித்துப் புரிந்துகொள்வதற்கேற்ப சந்த நயம் பாதிக்காமல் சீர் அசைகளைப் பிரித்தும் பிரிக்காமலும் என்று இருவகைகளாகப் பிரதிகளை வடிவமைத்துக்கொள்ள வேண்டும் .

பிறகு இதில் இடம் பெறும் சொற்களுக்குரிய அருஞ்சொற்பொருளையும் அருகில் அமைக்க வேண்டும் . ஒன்றுக்கு மேற்பட்ட அர்த்தங்கள் இருப்பின் அதனையும் அங்கு கொடுக்க வேண்டும் . பிறகு செய்யுள் தொடர்களை தனித்தனி அலகுகளாகப் பிரித்துக் காட்டல் அவசியம் . இது மொழியியல் அடிப்படையில் செய்யுள் வாசிப்பை ஊக்கப்படுத்தும். செய்யுளின் எந்தெந்தக் கூறுகள் எதன் எதனோடு இணைந்து எவ்வித அர்த்தங்களை எல்லாம் உருவாக்கலாம் என்பதைத் தெரிந்து கொள்ளலாம் . இந்நடைமுறை கற்றல்- கற்பித்தலை ஒரு விளையாட்டுச் செயல்முறை போன்று விருப்பமுடையதாக்கும் . அதன் மூலம் கற்றல்- கற்பித்தல் புதிய விளக்கங்களை அடைதலும் சாத்தியம்.

சான்றாக, அதியமான் அளித்த நெல்லிக்கனியைப் பெற்றுக்கொண்டு பாடியதாக அமைந்த இப்பாடலை நான் கு அலகுகளாக பகுக்கலாம்.

புறநானூறு 91

வலம் படு வாய்வாள் ஏந்தி, ஒன்னார்
களம் படக் கடந்த கழல் தொடித் தடக் கை,
ஆர் கலி நறவின், அதியர் கோமான்!
போர் அடு திருவின் பொலந் தார் அஞ்சி! 5
பால் புரை பிறை நுதல் பொலிந்த சென்னி
நீலமணி மிடற்று ஒருவன் போல
மன்னுக பெரும! நீயே, தொல் நிலைப்
பெரு மலை விடரகத்து அரு மிசைக் கொண்ட
சிறியிலை நெல்லித் தீம் கனி குறியாது, 10
ஆதல் நின் அகத்து அடக்கி,
சாதல் நீங்க, எமக்கு ஈத்தனையே!

திணை பாடாண் திணை; துறை வாழ்த்தியல்.
அவனை அவர் நெல்லிப் பழம் பெற்றுப் பாடியது.

ஆங்கிலத்தில் வாழ்த்திற்கு பதில் வாழ்த்தாக “Wish You the Same” என்று சொல்லும் தொடர் பிரபலமானது . இப்பாடலும், வாழ்த்திற்கு பதில் வாழ்த்து என்ற வகையில் சிறப்புற அமைந்துள்ளது .

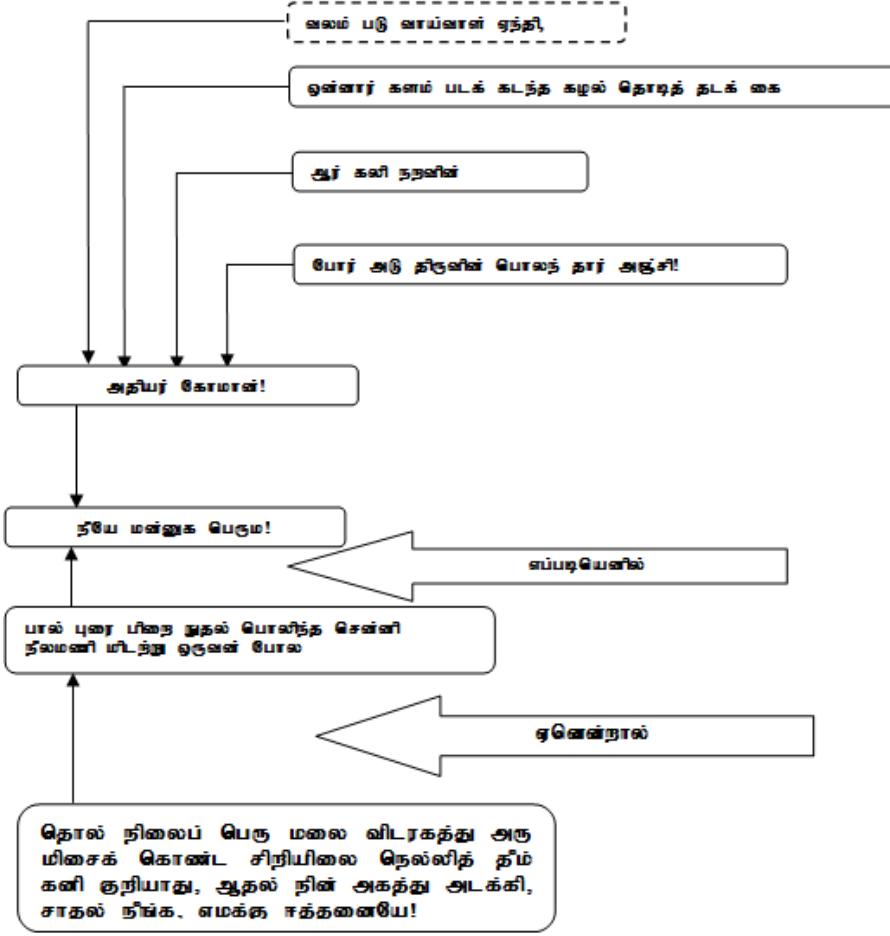
மன்னுகப் பெரும நீயே என்பது தான், இப்பாடலின் அடிப்படை அலகு. அதனை நீயே மன்னுகப் பெரும! (Wish You the Same) என்று எழுதிக்கொள்ளலாம் . நீயே என்பது , இங்கு அதியமானைக் குறிக்கிறது . அந்த அதியமான் யார் எப்படிப்பட்டவன் என்பதை பாடலின் முதல் நான்கு தொடர்கள் ஓர் அலகாக நின்று விவரிக்கின்றன .

மூன்றாவது அலகாக , அதியமான் யார் மாதிரி , எப்படி வாழ வேண்டும் என்பதை “நீலமணி மிடற்று ஒருவன் போல” அழிவில்லாது வாழும் தன்மையைப் பெற்று வாழவேண்டும் என்று விளக்குகிறது .

நான்காவது அலகு, ஏன் அவ்விதம் வாழ வேண்டும் என்பதை கடைசி ஐந்து அடிகள் விவரிக்கின்றன . இதனை, ‘மர வரைபட’ முறையில் கீழ்க்கண்டவாறு குறித்து காட்டலாம் .

இவ்விதமாக, பாடலுக்குரிய அமைப்பை , மொழியியல் அலகுகளாகப் பகுத்து அண்மையுறுப்புப் பகுப்பாய்வு நோக்கில் மாணவர்களுக்கு இலக்கிய ரசனையையும் படைப்பாற்றலையும் மேம்படுத்தி வளர்க்க கணினி வழி கற்றல்-கற்பித்தல் முறை பயனுடையதாய் அமைகின்றது.

கற்றல்-கற்பித்தலுக்கு வசதியாக செய்யுளுக் குரிய இசையையோ கருத்து விளக்கத்திற்கான நிகழ்வுகளையோ விவரிக்கும் ஆடியோ வீடியோக்களை பதிவேற்றம் செய்தோ இணையம் வழி பதிவிறக்கம் செய்தோ ஐம்புலன் ஒருங்கிணைப்பை அதிகப்படுத்தலாம் .



சிங்கப்பூர் சிறிய நாடு . மூன்றாம் உலக நாடாக இருந்த இந்நாடு தனது தொலைநோக்கு மனிதவளத்தைச் சிறப்பாகப் பயன்படுத்துதல் . சிறப்பு நிர்வாகம் , கல்விக்கு முதலிடம் போன்ற சிறப்பு அம்சங்களால் முதல் உலக நாடாக இன்று வளர்ந்து திகழ்கிறது . இங்கு அதன் ஒரே வளமான மனித வளத்தைக் கல்வி வாயிலாக அது தொடர்ந்து மேம்படுத்தி வருகிறது . அவ்வகையில் ஆசிரியர் பயிற்சி என்பது 21ஆம் நூற்றாண்டு உலகக் குடிமக்களை உருவாக்கி மேம்படுத்துவதில் முக்கியப் பங்கு வகிக்கிறது . இரண்டாம் மொழி நிலையில் கற்பிக்கப்படும் தாய்மொழிப் பாடங்களில் ஒன்றான தமிழ் மொழியைக் கற்பிக்கும் ஆசிரியர்கள் தகவல் தொழில்நுட்பத்தைப் பயன்படுத்தி மாணவர்களின் மொழித்திறன் களையும் சிந்தனைத்திறனையும் மேம்படுத்தி வருகின்றனர். அவ்வகையில் சில வெற்றிகரமான முயற்சிகளைப் பார்ப்போம் .

கணினியைப் பயன்படுத்துதல் என்பது இன்றைய மாணவர்களுக்குக் கைகவரப் பெற்றுள்ள அற்புதத் திறன் அத்தகு திறனைப் பயன்படுத்தித் தட்டச்சில் ஈடுபடல் , பவர் பாயிண்ட் போன்றப் பல படைப்புகளைச் செய்தல் , இணையத்தை வலம் வந்து தகவல் திரட்டல் , திரட்டிய தகவல்களை பகுப்பாய்வு செய்து படைத்தல் என்ற நிலையில் மாணவர்களுக்கு உதவும் வகையில் ஆசிரியர் பயிற்சி அமைகிறது . இணையம் என்பது இன்று உலகைத் தன்வசப்படுத்தி வருகிறது . இத்தகு இணையத்தைப் பயன்படுத்தி ஒரு சிக்கலுக்கு விடை காணுதல் , ஒரு புலனாய்வை மேற்கொண்டு கல்வி தொடர்பான விசாரணையை நடத்துதல் என்ற முறையில் இணைய வினா விளக்கம் என்ற அணுகுமுறை ஆசிரியர் பயிற்சியில் அறிமுகப்படுத்தப்பட்டுள்ளது .

முகப்புத்தகம், முகநூல் என்ற சமூக வலைக் கட்டமைப்பு தமிழாசிரியர் பயிற்சியின் தனக்கென ஓரிடம் பெற்றுத் திகழ்கிறது. தமிழ்ப் பண்பாடு , பாரம்பரியம் வாழ்வியல் போன்றவை தொடர்பான அரிய செய்திகளை மாணவர்களுக்கு எடுத்துச் சொல்லும் வகையில் 'ஃபேஸ் புக்' என்ற முகநூல் பயன்படுத்தப்பட்டது . இது மாணவ

ஆசிரியர்களிடையே, தகவல்களைத் திரட்டிப் பகிர்தல் , வாரந்தோறும் எதிரொளிப்பைப் படைத்தல் , தம் சக உறுப்பினர்களுக்குப் பொருள் பொதிந்த வினாக்களை விடுத்துப் புதிய அறிவைக் கட்டமைக்க வழி செய்தல் எனப் பயிற்சி அமைந்தது இது தமிழ் மொழி ஆற்றல் கற்பித்தலில் புதிய பரிமாணத்தை உருவாக்க வழி செய்தது .

‘போஸ்ட்ரஸ்’, ‘வாய்ஸ்த்ரெட்’, ஆரிய இரு மென்பொருள்கள் இணையத்தின் இலவசமாகக் கிடைப்பதில்லை . இவற்றைப் பயன்படுத்திப் பேச்சுத் தமிழை தொடக்க மற்றும் உயர்நிலைப் பள்ளி மாணவர்களிடையே அதிகமாகப் பயன்படுத்த ஊக்குவிக்கும் வகையில் பயிற்சி இடம்பெற்றுள்ளது . இதன் வாயிலாகப் பள்ளி மாணவர்கள் திரையில் இடம் பெறும் படம் , படக்காட்சி ஆகியவற்றைப் பார்த்து அவை தொடர்பான கருத்துக்களை ஒலிவழியே பகிர்ந்து கொள்ள ‘வாய்ஸ்த்ரெட்’, வழிவகுக்கிறது. அதே நேரம் பதினம் வயது மாணவர்கள் வானொலி நிகழ்ச்சித் தொகுப்பாளராகத் தம் அறிவிப்பு , பலரிடம் நேர்முகம் மேற்கொள்ளல், கிடைத்த கருத்துக்களை முறையாக ஒருங்கிணைத்து நிகழ்ச்சியைப் படைத்தல் என ‘போஸ்ட்ரஸ்’ மென்பொருள் மாணவர்கள் பேச்சுத்தமிழில் பேசுவழி செய்கிறது .

புகைப்படக்கதை என்பது கணினியைப் பயன்படுத்தித் தெரிவு செய்யப்பட்ட புகைப்படங்களைக் கொண்டு ஒரு ஒலி ஒளிக் கதையை உருவாக்கிப் படைப்பதாகும் . இதில் படத்திரட்டல் , கருத்துத் திரட்டல் , தேவையான படங்களைத் தெரிவு செய்தல் , படங்களுக்கான வாசகங்களை எழுதிக் குரல் தருதல் , இணையப் பொருத்தமாக இணைத்தல் எனப் புகைப்படக்கதை அமைகிறது .

குளோக்ஸ்டர் என்பது மற்றொரு மென்பொருள் இணையத்தில் இலவசமாகக் கிடைக்கும் இந்த மென்பொருள் , கணினித் திரையில் ஒரே நேரத்தில் உரைநடை , ஒலி, ஒளிக்கூறுகள், அசையும் படங்கள் , உட்படப் பலவற்றைக் காட்டும் ஒரு தொகுப்புக்கருவி பாடத்திட்டத்தில் உள்ள பழமொழிகள் , செய்யுள்கள் ஆகியவற்றை ஒவ்வொன்றாக நடத்திக் காட்டும் வகையில் இக்கருவியைப் பயன்படுத்தியுள்ளனர் .

விம்பா லைவ் என்பது மாணவ ஆசிரியர்கள் தங்களுக்குள் தாங்கள் இருந்த இடத்திலிருந்தபடி தொடர்பை நிகழ்த்த உதவும் மென்பொருள் . இது குழுக் கலந்துரையாடல் , வாய்மொழிப் படைப்பு ஆகியவற்றுக்கு உதவும் மென்பொருள் ஆகும் .

இதுவரை இங்கு குறிப்பிடப்பட்டவை தமிழாசிரியர் பயிற்சியில் இடம்பெற்றுள்ள தகவல் தொழில் நுட்பக் கல்வியின் சில பகுதிகளை கடக்கும் கூறுகள் இவற்றை மாணவர் ஆசிரியர் பயன்படுத்துவதில் படிநிலை என்ற ஒன்று முக்கிய இடம் பெறுகிறது .

எந்த ஒரு அணுகுமுறையும் தகவல் தொழில்நுட்ப மென்பொருளும் முடிவில் ஒரு தயாரிப்பை உருவாக்கும் தன்மையதாக இருக்கவேண்டும் . இங்கு Product என்னும் தயாரிப்பைவிட அதை எவ்வாறு உருவாக்குவது என்ற படிநிலை Process இன்றியமையாதது. அவ்வகையில் ஒவ்வொரு மென்பொருளும் அறிமுகப்படுத்திப் பயிற்சி தரும்போதும், இந்தப் படிநிலை , மதிப்பிட்டுக்கான விளக்க அட்டவணை (Rubrics) ஆகியன மாணவர்களுக்கு விளக்கப்படும் .

இத்தகு மென்பொருள்கள் , தகவல் தொழில்நுட்ப அணுகுமுறைகளை அறிமுகப்படுத்தும்போது தேசியக் கல்விக் கழகத்தில் மின்கற்றலுக்கான மையத்தில் (Centre for E-Learning) உள்ள கற்றல் வடிவமைப்பாளர் (Instructional Designers) தொழில்நுட்பம் உலர்ந்த உதவியை வழங்குவர் . அவர்களது ஒத்துழைப்புடன் ஆசிரியர்கள் சார்ந்த பயிற்சியை விரிவுரையாளர்கள் வழங்கிவருகின்றனர் . சிங்கப்பூரில் கல்வி அமைச்சின் தகவல் தொழில்நுட்ப முதன்மை திட்டம் 1-3 வரையிலான முனைப்புகளுக்கு (IT Master Plans 1-3) மேலும் வளமும் உரமும் சேர்க்கும் வகையில் பள்ளி மாணவர்களின் தகவல் தொழில் நுட்பத் திறன்களை மேம்படுத்தும் வகையில் ஆசிரியர் பயிற்சி அமைந்து வருகிறது .

கணினி, இணையதளம் வழியிலான கற்றல்- கற்பித்தல் செயல்பாடு மாணவர்களின் படைப்பற்றல் திறனை மேம்படுத்துவதோடு ஆசிரியர் மாணவர் உறவையும் உயிரோட்டமுடையதாக இக்காலச் சூழலுக்கு ஏற்ப திறம்பட மாற்றும் .

மேற்கோள் நூல்கள்

1. Seetha Lakshmi (2011) . Naa Go and Tamil Pedagogy. IN Naidu, P., Krishnan P., & Kannabiran Rama (EDS.) *naa govindasamy enum padaippaali (Naa Govindasamy - a Creator)*, 1, 137-149.
2. Seetha Lakshmi (2010A.) Teaching and Resource Building in Teacher Education. IN Vasu Renganathan (ED.) *Tamil Internet 2010: Conference Papers*, 1, 19-26.
3. Seetha Lakshmi, Majal Pratima (2010B) . Introduction of Video Conferencing in Tamil Classroom: A Study. *International Journal of Educational Research, Development and Extension IJERDE*, 1, (1), 33-41.
4. Seetha Lakshmi and Jarina Peer (2009). Use of Tamil Language and IT in Tamil Language Education. IN - (ED.) *CRPP International Conference, E-Conference Proceedings*. (website: http://conference.nie.edu.sg/2009/papers_pdf/PAP575.pdf), 1, 1-14.

குறிப்புகள்