உத்தமம்
INFITT
உலகத் தமிழ்த் தகவல் தொழில்நுட்ப மன்றம்
International Forum for Information Technology in Tamil

Institute of Indology
& Tamil Studies

தமிழ் இணையம்
2 0 0 9

TAMIL INTERNET
2 0 0 9

மாநாட்டுக் கட்டுரைகள்
CONFERENCE PAPERS

Conference papers

Tamil Internet

2009

Cologne, Germany – Oct 23-25

CONTENTS

Topic A. Computer Assisted Learning and Teaching of Tamil

Topic B. Tamil Diaspora: Teaching Tamil as a Second Language and Impact of Technology:

Topic C. Tamil Portal: Blogger, Aggregator and Wikipedia

18. எதிர்கால இணையத்தில் தமிழ் யூனிகோடு: வளர்தமிழ் ஆய்வில் தரவுத்தளங்களும், திரட்டிகளும்
முனைவர் நா. கணேசன், ஹ்யூஸ்டன், அமெரிக்கா

19. Automatic E-Content Generation
E.Iniya Nehru, India

20. ICANN and IDN Internet Corporation for Assigned Names and Numbers International Domain Names. S.Maniam, Singapore.

21. Development of Social Networking Website and Tamil E-learning Software Using Unicode / ISCII standards. Dr.A.Muthukumar, India.

22. தமிழ் விக்கிபீடியாவும் துணைத்திட்டங்களும் : அறிமுகம், தமிழ் இணையத்தில் அவற்றின் வகிபாகம், எதிர்காலம், செயன்முறை விளக்கம்
முரளிதரன் மயூரன், Sri Lanka.

Topic D. Tamil Localization, Tamil Keyboard and Open Source Software

23. தமிழ் திறவூற்று மென்பொருள்கள் - 'தமிழா' தோற்றமும் தொடர்ச்சியும் சி.ம.இளந்தமிழ், Malaysia.

24. Indic-Keyboards – A multilingual Indic keyboard interface
A.G. Ramakrishnan, Akshay Rao, Arun S., Abhinava Shivakumar, India.

25. Tamil Localisation Process – A case study
Kengatharaiyer Sarveswaran, Sri Lanka.

26. தமிழ் மென்பொருள்களும் மக்கள் பாவனையும்
சிவா அனுராஜ் – வட அமெரிக்கா

27. Inside Tamil Unicode
Sinnathurai Srivas, Australia

28. Building Tamil Unicode Fonts for Mac OS X.
Muthu Nedumaran, Malaysia

29. Tamil Encoding, Keyboard Layout and Collation Sequence Standard for ICT Sri Lanka
Balachandran G. Sri Lanka

Topic E. Natural Language Processing

30. An Intelligent System for Picture Based Tamil Sentence Generation
Dr.T.Mala, Dr.T.V.Geetha, India

31. A HMM Based Online Tamil Word Recognizer
Rituraj Kunwar, Shashi Kiran Suresh Sundaram and A G Ramakrishnan, India.

32. Problems related to Eng-Tam Translation
M.B.A.Salai Aaviyamma and Dr.K.Kathiravan

33. Tamil-English Cross Lingual Information Retrieval System for Agriculture Society
D. Thenmozhi and C. Aravindan

Topic F. Tamil in Handheld and Mobile Phones

34. அறிவியல் மற்றும் தற்காலத் தேவைகளுக்கேற்ற தமிழெழுத்துச் சீர்திருத்தமும் அவற்றைக் கணினி மற்றும் கைபேசி செயல்பாட்டிற்குப் பயன்படுத்தும் முறையும்
த. ஞான பாரதி, India.

35. Tamil on Mobile Devices
Muthu Nedumaran, Malaysia.

36. E=mt2 E-Learning = M-Learning makes Tamil learning Squared
S. Swarnalatha, India

Topic G. E-Texts, Corpora and Search Engines

37. Tamil Corpus Generation and Text Analysis
M. Ganesan, India

38. OMNIS/2 Integrating Libraries with Digital Multimedia Database
C.Radha, India

39. மின்பதிப்பு - பல்லூடக தகவல் தரவுக் களஞ்சியம், ஓலைச்சுவடிகளின் பேரட்டவணை உருவாக்கல்
சுபாஷினி டிரெம்மல், ஜெர்மனி

40. Tamil Inscriptions and on line search: Database Compilation, Grammatical analysis, Lexicon and Translation.
Appasamy Murugaiyan, France.

41. பல்லூடகவழிப் பழந்தமிழ் இலக்கியக் கருத்தாடல் இலக்கியம் கற்றல் கற்பித்தல் அடிப்படையில் (Sangam Literature Teaching and Learning through Multimedia)
முனைவர் வா. மு. சே. முத்துராமலிங்க ஆண்டவர்

Topic H. Computational Linguistics

42. தமிழ் வினை வடிவங்கள்: கணினிப் பகுப்பாய்வு
(Computer Analysis of Tamil Verb Forms)
முனைவர் ப. டேவிட் பிரபாகர்

43. Dravidian Wordnet
(திராவிட சொல்வலையாக்கம்)
Dr. S.Rajendran, India.

44. Unsupervised Approach to Tamil Morpheme Segmentation
K.Rajan, Dr. V.Ramalingam, Dr. M.Ganesan, India.

45. தொல்காப்பியத்தின் எழுத்ததிகாரத்துக்கான இடம் சாரா இலக்கணம்
இல. பாலசுந்தரராமன், ஈசுவர் சிரீதரன்

Topic I. Natural Langauge Processing: Morphological Tagger

46. AMRITA Morph Analyzer and Generator for Tamil: A Rule Based Approach
Dr. A.G. Menon, Amrita and Leiden (Netherland), S. Saravanan, R. Loganathan and Dr. K. Soman, India.

47. A Novel Approach to Morphological Analysis for Tamil Language
(தமிழ் உருபனியல் ஆய்வில் ஒரு புதிய அணுகுமுறை)
Anand kumar M, Dhanalakshmi V, Rajendran S, Soman K

48. POS Tagger and Chunker for Tamil Language
(தமிழ் சொல்வகை அடையாளப்படுத்தி மற்றும் தொடர் பகுப்பான்)
Dhanalakshmi V, Anand kumar M, Rajendran S, Soman K

Topic J. Electronic Tamil Dictionaries and Glossaries

49. தகவல் தொழில்நுட்ப கலைச்சொற்களை வளப்படுத்துங்கள்
மா. ஆண்டோ பீட்டர், இந்தியா.

50. Electronic Dictionary for Sangam Literature
(சங்க இலக்கியத்திற்கான மின் அகராதி)
Dr. K.Umaraj, India

51. கணிணியியலில் நேர்பெயர்ப்புச் சொற்களும் ஒலிபெயர்ப்புச் சொற்களும்
இலக்குவனார் திருவள்ளுவன், இந்தியா

52. Critical Editions of Tamil Works: Exploratory Survey and Future Perspectives
Jean-Luc Chevillard, France

53. The English Dictionary of the Tamil Verb: What can it tell us about the structure of Tamil?
Harold F. Schiffman, U.S.A.

டாக்டர் பூங்கோதை ஆலடி அருணா,
தகவல் தொழில்நுட்பத் துறை அமைச்சர்

தலைமைச் செயலகம்
சென்னை - 600 009

நாள் 18.09.2009

வாழ்த்து செய்தி

முத்தமிழறிஞர் டாக்டர். கலைஞர் தலைமையில் நடைபெற்ற சென்னை இரண்டாம் தமிழ் இணைய மாநாட்டின் மூலமாக கணினி தமிழுக்கு தீர்வு ஏற்பட்டது. உலகெங்கும் வாழும் தமிழ் மக்களுக்கு கணினி தமிழுக்கு ஒருங்கிணைக்கப்பட்ட எழுத்துரு மற்றும் விசைப்பலகையை தமிழக அரசு 1999 ஆம் ஆண்டு அரசாணை மூலமாக வெளியிட்டது. தமிழக முதல்வரின் இப்படிப்பட்ட தகவல் தொழில் நுட்ப நலப்பணிகள் தொடர்பான சாதனைகளை நாம் கூறிக் கொண்டே இருக்கலாம்.

அதேபோல 23–10–2009 முதல் 25–10–2009 ஆகிய நாட்களில் ஜெர்மனி, கோலோன் பல்கலைக்கழகத்தில் நடைபெறும் தமிழ் இணைய மாநாட்டின் மூலம் உலகெங்கும் வாழும் தமிழர்கள் மட்டுமல்லாது பிற மொழியினரும் தமிழை எளிய முறையில் கணிப்பொறி மூலம் பயன்படுத்துகின்ற வகையில் நல்ல பல திட்டங்களை, கொள்கைகளை வகுத்து தமிழை மேன்மேலும் ஏற்றமுறச் செய்திட ஜெர்மனியில் நடைபெறும் மாநாட்டையும் அதில் கலந்து கொள்ள செல்லும் அறிஞர் குழுவையும் மனமார வாழ்த்துகிறேன்.

பூங்கோதை

டாக்டர்.பூங்கோதை ஆலடி அருணா
தகவல் தொழில்நுட்பவியல் துறை அமைச்சர்

மலேசிய மூலத்தொழில் அமைச்சர்
டத்தோ டாக்டர் எஸ்.சுப்ரமணியம் அவர்களின் வாழ்த்துச்செய்தி

தமிழ் இணைய மாநாடு இவ்வாண்டு ஜெர்மனியில் நடைப்பெறுவதை எண்ணி மகிழ்ச்சியடைகிறேன். தமிழ் கணினி வளர்ச்சிக்கும் தமிழ் மொழியின் வளர்ச்சிக்கும் இணைய மாநாடுகள் முக்கிய பங்காற்றி வருவது அறிந்து மகிழ்கிறேன். தமிழ் மொழி வளமான மொழி என்பது அனைவருக்கும் தெரியும், வளமான இம்மொழியை அடுத்த தலைமுறையினருக்கு கொண்டு செல்லும் கருவியும் தொழில்நுட்பமுமாக கணினியும் இணையமும் திகழ்கின்றது. இத்தொழில்நுட்பத்தைக் கொண்டு தமிழுக்காக சீரிய பணிகள் ஆற்றி வரும் தமிழ் இணையம் 2009 ஏற்பாட்டுக் குழுவினரை மனமுவந்து வாழ்த்துகிறேன். தமிழ் தகவல் தொழில்நுட்பத்திற்கான அனைத்துலக மன்றமான உத்தமமும், ஜெர்மனியில் அமைந்துள்ள கோலன் பல்கலைக்கழகத்தின் தமிழ் ஆய்வு மையமும் இணைந்து நடத்தும் இம்மாநாடு எல்லா வகையிலும் சிறப்புடன் நடைப்பெற என்னுடைய வாழ்த்துகள்.

மலேசிய, கூட்டரசு பிரதேச துணை அமைச்சர்
டத்தோ சரவணன் அவர்களின் வாழ்த்துச்செய்தி

தமிழ்க் கணினி வளர்ச்சி என்பது இன்று தமிழ் மொழி வளர்ச்சியில் முக்கிய கூறாக விளங்கி வருகிறது. எதிர்வரும் அக்டோபர் 23 தொடங்கி 25 வரை ஜெர்மனியில் நடைப்பெற உள்ள தமிழ் இணைய மாநாட்டிற்கு வாழ்த்து தெரிவிப்பதில் மட்டற்ற மகிழ்ச்சி மகிழ்கிறேன். தமிழ் தகவல் தொழில்நுட்பத்திற்கான அனைத்துலக மன்றமான உத்தமமும், ஜெர்மனியில் அமைந்துள்ள கோலன் பல்கலைக்கழகத்தின் தமிழ் ஆய்வு மையமும் இணைந்து நடத்தும் இம்மாநாடு எல்லா வகையிலும் வெற்றி பெறும் என்பது திண்ணம். இதுவரை சென்னை (1997, 1999, 2003), சிங்கை (2000, 2004), மலேசியா (2001), அமெரிக்கா (2002) நடைப்பெற்றுள்ள இம்மாநாடு, இவ்வாண்டு புலம் பெயர்ந்த தமிழர்கள் அதிகம் வாழும் ஐரோப்பாவில் நடைப்பெறுவது தமிழ் மொழியின் வளர்ச்சிக்கு விரைவாக வித்திடும். 2001-ல் ம.இ.காவின் உதவியுடன் கோலாலம்பூரில் நடைப்பெற்ற தமிழ் இணைய மாநாட்டினை பெருமிதத்துடன் திரும்பி பார்கின்றேன் 'கணினி வழி காண்போம் தமிழ்' எனும் கருப்பொருளில் நடைப்பெறவிருக்கும் இம்மாநாடு தமிழ் தகவல் தொழில்நுட்ப உலகில் ஏற்பட்டுவரும் சவால்களையும் மாற்றங்களையும் ஒருங்கே ஆராய்ந்து தீர்வினை கொண்டு வரும் என பெரிதும் எதிர்பார்கின்றேன்

தமிழ் இணையம் 2009 சிறப்பாக நடைப்பெற சீரிய பணிகள் ஆற்றி வரும் ஏற்பாட்டுக் குழுவினரை மனமுவந்து வாழ்த்துகிறேன்.

Prof. M. Anandakrishnan

Chairman, IIT-Kanpur & Science City, Chennai,

Tamilnadu Chair, INFITT 2000-2003

Cologne INFITT Conference Reflects Optimism with Realism

Starting from late seventies and early eighties, the developments in personal computers provoked a number of Tamil professionals to undertake significant efforts to achieve incorporation of Tamil language in computers and the fledgling internet. Most of them like Srinivasan, from Quebec, Canada and Kalyanasundaram from Lausanne, Swizerland and many others like them were not formally trained computer professionals. Some like Muthu Nedumaran from Kuala Lumpur, Malaysia were computer experts with accomplishments in language computing. There were many others like them in France, Germany, USA, Canada, Australia, Singapore, Malaysia, Sri Lanka and India.

Their initial contributions were mostly boot-strapping part-time efforts out of genuine enthusiasm but had very little scope for exchange of ideas. By late eighties, the degree of incompatibility among these efforts rendered the value of their contributions mainly local and limited. They did not subscribe to the prevailing international standards either because of their not being aware of them or due to serious disagreement with them.

The first initiative to identify the dimensions of disorder in Tamil computing was taken in 1997 in a conference in Singapore convened by Na. Govindaswami. The meeting consisted of some forty concerned persons mostly from Singapore, Malaysia, Sri Lanka and India. The issues raised at this meeting were further articulated in the conference in Chennai in Feb.1999.

Since the Chennai conference was attended by senior government leaders from Malaysia, Mauritius, Sri Lanka, Singapore and India along with a number of experts and officials from these countries as well as the experts in language computing from several other countries such as USA, France, UK, Germany, Switzerland, Australia and so on, there was some hope that solutions can be found for some of the issues raised in Singapore meeting. No doubt there were some agreements like Tamilnet-99 keyboard and encoding for TAB and TAM Tamil Characters. At the same time there was a sense of non-fulfillment of many other crucial tasks. This conference had the enthusiastic support of Kalaigner Karunanithi, the Chief Minister of Tamil Nadu and active propulsion by (late). Thiru Murasoli Maran, then Member of Parliament. As fallout of this conference, Tamil Nadu got a Tamil Virtual University, the Tamil Software Development Fund and more substantially the attention of policy makers, government officials, academic institutions and business community on the value of promoting Tamil Computing in multiple fronts.

At this stage there was a keen recognition to give an institutional foundation for continued attention to the emerging developments in Tamil Computing and Internet applications. Mr. Thondaman, a Minister in Sri Lanka Cabinet, extended an invitation to a few of the representative participants at Chennai conference to visit Sri Lanka and interact with their experts. The delegation, consisting of persons from India, Singapore, and Malaysia and visited Sri Lanka. It was during this visit that the organizational framework called INFITT was evolved. Even as the rules and regulations for INFITT

were being evolved, thanks to the dynamic initiatives of Arun Mahiznan and A. Narayanan with the able support of Prof. Tan Tin Wee, a home base in Singapore was established with generous support from Singapore Government. The existence of this base was in a large measure responsible for propelling more INFITT Conferences in different countries, as well as creation and maintenance of the INFITT website and a home for sharing information, and archiving the documents and proceedings.

It is axiomatic that no major initiative relating to Tamil Community will be free from serious controversies - some political, some philosophical and occasionally personality differences. It is a matter of satisfaction that INFITT has survived all these occurrences. The Tamil Community is suddenly waking up to all the missed opportunities after realizing how far Tamil language is falling behind many other languages. It is evident from the degree of enthusiasm and cooperation seen in putting together the Cologne Conference, thanks to the proactive initiatives of the present INFITT managers, particularly Kalyan and Kaviarasan. The agenda and the papers reflect the current international realities in terms of technologies and applications as also a high degree of optimism. I hope that the recommendations of Cologne conference will be further elaborated in the World Tamil Conference in Tamil Nadu among the World Tamil Leaders and experts. Given the nature of preparations and responses I am highly confident of the successful outcome of the cologne INFITT deliberations.

The Cologne Conference, I hope, will be a landmark event setting the challenging agenda for the next decade enlisting the world wide co-operation among the Tamil community and the Computer professionals. Already there are many breakthroughs in harnessing the power of technologies for Tamil applications such as News Portals, Voice Recognition and several areas of human computer interaction. The tone set by this conference and the spirit fostered at this time will go a long way in realizing many dreams.

Mr. Muthu Nedumaran

Murasu Systems,

Kuala Lumpur, Malaysia

Chair, INFITT 2003-2007,

Vice-Chair INFITT 2000-2003

Technology evolves very rapidly and we do not want Tamil to be in catch-up mode all the time. We need to think ahead, understand the ground rules and move forward so that Tamil is always ready when ground-breaking technology reaches the user community across the globe.

In this light, it is heartening to see INFITT continuing to organize the Tamil Internet Conferences (TICs). This is certainly reflective of the mission with which we started INFITT and of the spirit in which we organized the previous conferences.

I am also pleased to see that the 8th conference will be the first TIC in Europe. This gives us an opportunity to study and explore usage of Tamil IT and specific needs there may be in this geography. The Internet has made the world so small. Co-existence with other languages and cultures on a single application is no longer a novelty but a necessity. Let us take advantage of the Cologne TIC to make Tamil seamlessly co-exist with the rest of the world scripts and be readily available in all new and emerging platforms.

I take this opportunity to congratulate and thank Kalyan, the chair of INFITT, and his team to make this conference possible.

Mr. Arun Mahizhnan

Institute of Policy Studies,

Singapore

Executive Director, INFITT 2000-2007

தமிழ் இணைய மாநாடு ஐந்தாண்டுகளுக்குப்பின் மீண்டும் நடைபெற இருப்பது எனக்குப் பெருமகிழ்வைத் தருகிறது. கல்யாண், வாசு தலைமையில் கூடியிருக்கும் இந்த மாநாடு இணையத் தமிழ் எதிர்நோக்கியிருக்கும் தொழில்நுட்பச் சவால்களுக்குத் தீர்வு காண்பதோடு உத்தமம் இனிச் செல்ல வேண்டிய நோக்கையும் போக்கையும் தெளிவாக்கும் என்று நம்புகிறேன். ஒன்பது ஆண்டுகளுக்கு முன் நிறுவப்பட்ட உத்தமம் அடுத்த ஆண்டு தனது பத்தாவது வயதை எட்டவிருக்கிறது. அது ஒரு முக்கியமான காலக் கட்டம். அந்த மகிழ்ச்சியான நேரத்தில் நாமெல்லோரும் பெருமை கொள்ளத் தக்க வகையில் ஒரு சில திட்டங்களின் சாதனைகளை நாம் மக்கள் முன் வைக்க வேண்டும். அந்தச் சாதனைகளுக்கு அடிக்கல் இந்த மாநாட்டில் நிறுவப்பட வேண்டும் என்பது என்னுடைய தாழ்மையான வேண்டுகோள்.

ஜெர்மானியப் பேராசிரியை உல்ரிக்கா நிக்லாஸ் உதவியோடு நடைபெறும் இந்த மாநாடு பல நாட்டுத் தமிழர்களையும் தமிழ் இணைய ஆர்வலர்களையும் ஒன்று கூட்டியுள்ளது. இந்த வேளையில், இந்த மாநாடு சிறப்பாக நடந்தேறச் சிங்கப்பூர்த் தமிழர்களின் சார்பாக அவர்தம் வாழ்த்துக்களைத் தெரிவித்துக் கொள்கிறேன்.

அன்புடன்,

அருண் மகிழ்நன்

I'm truly delighted to see the revival of INFITT's annual conference since the last one in Singapore in 2004. The INFITT conferences have always functioned as a gathering place for Tamil Internet experts and enthusiasts and as an intellectual forum to discuss common issues and share success stories. I'm sure the meeting in Koeln would once again offer the collegiate and constructive forum that we need to engage the issues that have accumulated in the last five years. I'm very encouraged by the large number of papers that have been submitted to the conference and I'm confident that under the leadership of Kalyan and Vasu the conference would yield commendable results.

It would also be a great occasion to renew old friendships and forge new ones. I want to pay a special tribute to Prof Ulrike Niklas and her Institute of Indology and Tamil Studies (IITS) of the University of Koeln for making this conference possible. It reminds me of how another non-Tamil, Prof Tan Tin Wee of the National University of Singapore, together with fellow Singaporean Naa Govindasamy seeded the very first Tamil Internet conference in this series, held in Singapore in 1997. The Tamil Internet community is fortunate to count among its ardent supporters people like Ulrike and Tin Wee.

On behalf of my fellow Singaporeans who helped in the birth of INFITT in Singapore in 2000 and who supported the INFITT Secretariat till 2004, I want to offer our congratulations and best wishes to Kalyan and his able team for organising the 8th Tamil Internet Conference.

Arun Mahizhnan

A Global Movement for an Ancient Language in the Internet age

Tan Tin Wee

All the world's our neighbourhood, all peoples our kinsmen

In the pre-recorded history of mankind, the invention of language signified a major milestone but the written form of language provided the basis for the recording of history and culture for the benefit of those to come. Geographically bound by word of mouth and limitations of travel, the spread of local language to faraway lands was first by the vehicle of migration, then by the written word, the printed word, and more recently by the advent of the Internet and the World Wide Web. How can an ancient language such as Tamil be carried by this new wave of change?

As a person of Chinese origin, with English as my first language and Mandarin Chinese as my second language in school, and Hokkien a Chinese dialect as the mother tongue, I could by 1991, immediately adapt to the technologies of email, and then gopher (1992), and finally the Web (1993), first in English, and then in the written form of Chinese (1994), both in traditional and simplified versions. Chinese traditions, culture and language could survive in this new age of the Internet because we were one of the first to develop tools to convert the written electronic form of Chinese into online Web images of Chinese characters and display it on the Web in 1995. But what of those without this facility?

As a Singaporean, grilled as a child in our daily national pledge, "regardless of race, language or religion to build a democratic society based on justice and equality", it was natural for us to do the same for the other languages of our fellow countrymen, Malay and Indian languages, starting with Tamil. Malay was anglicized and easy, but Tamil lacked a standard universally accepted keyboard and a universal encoding. Email sent in one encoding required the same keyboard software to visualize and respond to, or else everything looked computer gibberish. The Tamil Unicode block was not acceptably usable for Tamil readers and writers.

How are we to declare ourselves proudly Singaporeans on the Internet if we did not plant a multilingual flag on the Internet, as a multilingual and a multiracial society? Our National InfoMap (www.sg) in 1994 was just in plain English. As the then manager of the first Internet provider for Singapore, Technet, the responsibility fell on my shoulders to drive this initiative forward. It was destiny that brought me to Naa Govindasamy, a Tamil poet and teacher at our Singapore National Institute of Education, and in his spare time, an inventor and technopreneur.

The first and the most important thing Naa Govindasamy gave me was the solution to put Tamil on the Internet. His language-optimized Kanian keyboard, fonts and software provided the inspiration to put the Tamil language on the Web. Working together in partnership with my trusty staff, Mr Leong Kok Yong, we put up PoemWeb containing Asian poems and most significantly, our National Pledge, in all four official languages of Singapore on a single webpage, English, Chinese, Malay and Tamil. Launched by none other than the then President of the Republic of Singapore, Mr Ong Teng Cheong, we had planted our multilingual flag on the Internet using Unicode encoding and automated text-to-image conversion by 1995. Next was a Java input engine, JIME, in 1996, followed by the invention of multilingual Internationalized Domain Names (IDNs) in 1998 in with domain names would be rendered operational in Tamil.

Secondly, it was Naa Govindasamy's vision that led us to convene the first Tamil Internet conference at the National University of Singapore in 1997. One fine day, he accosted me with this idea that he could pull off an international conference for Tamil Internet if I could provide him with the logistics. He had this vision of unifying Tamil language keyboard, input system, font display, email and web applications and making them all fully interoperable on the Internet.

To achieve it, he had to gather all key players in the field, from Professor Harold Schiffman to Professor M Anandakrishnan, from Sujatha, to Maalan and, most generous of all, his own competitors in Tamil software. He had the sponsorship and the support of many, including S Maniam. All we had to do was to execute it. 1997 was a big success. That success led us to think of strengthening the local organization for Tamil Internet activities. We consulted some key leaders, including Mr S R Nathan (who shortly thereafter became President of Singapore) and that process brought in Mr R Ravindran, a Member of Parliament as an Advisor and, at my invitation, Arun Mahizhnan to lead the organization together with Naa Govindasamy and Maniam. Unfortunately, Naa Govindasamy passed away most unexpectedly.

However, I was heartened to see that Naa Govindasamy's vision and legacy were enshrined in the subsequent Tamil Internet Steering Committee set up by the Singapore government and co-chaired by Mr Ravindran, MP and Arun Mahizhnan, with S Maniam as Secretary and myself as an Advisor. I was even more gratified when TISC was able to join hands with Tamil Internet enthusiasts all over the world to set up the International Forum for IT in Tamil (INFITT) in Singapore with Professor Anandakrishnan as Chair, Arun Mahizhnan as the Executive Director and Nara Andeappan as Administration Executive. I had the privilege of being appointed as an Advisor.

For me it was the culmination of a long effort started serendipitously with a stranger but concluding with many new and deep friendships and the satisfaction of having played a small role in enabling an ancient language enter the internet age. This year is the 10th Anniversary of his passing, and it continues to be my privilege and blessing to be part of the unfolding of his visionary ideas. He may be no more, but his spirit of universal endeavour and brotherhood lives on within our hearts.

For it is all of you, my friends and my brothers and sisters who have made Tamil Internet Chennai 1999, Singapore 2000, Kuala Lumpur 2001, Foster City 2002, Chennai 2003, Singapore 2004, and now Koeln 2009, a continuing legacy of success. It is my privilege to have worked closely in the past in addition to all the names listed above, with super-enthusiastic colleagues such as Mrs Devi Balasundaram, Muthu Nedumaran, Dr K Kalyanasundaram, who was later to go on to win the Sundara Ramasamy Award for Tamil Information Technology in 2008 for contributions in Tamil fonts, font encoding standards such as TSCII, Project Madurai and Chairmanship of INFITT. It is all of you who have moved forward with the formation and the growth of INFITT, to realize the vision of a strong and vibrant ancient language in a modern medium, uniting people in its wake.

I had said in my speech in Foster City TI2002, and I repeat it at Koeln 2009 because it is all the more appropriate: Since we have been enlightened by the vision of wise, we would not be amazed by the great in their glory, and we would even less despise the meek. Today, INFITT needs your support to realise its dream.

Everyone – both meek and great – needs to get involved, from those like me who don't speak Tamil at all or very well, to those who are experts. As we break down artificial barriers, and together, advance the cause of an ancient Language dear to us all, it is my privilege to recall Professor Hart's bilingual

rendering of the great poet Kaniyan Punkundranar, As my best friend S Subbiah's mother gave me a hug before I came into this hall in Foster City, I could not but feel: *Your world is my world, and mine yours, you are my family.(Yaathum oore, Yavarum Kelir)*

Let's all work together in INFITT as one family, and build a better tomorrow for Tamil Internet. Have a wonderful 2009 conference in Cologne!

Nandri, Vannakkam

Dr. K. Kalyanasundaram

Chair, INFITT 2007-Lausanne, Switzerland

Chairman, INFITT

Amongst Indian languages Tamil stands unique due to its large Diaspora population and a distinct status as a state-recognized language in many countries of the world. Tamil language has a long history dating back to two thousand years. In addition to the evolution of the script of the language over this long period, there has been an evolution in the methods by which the rich cultural, literary heritage and traditions are passed on from one generation to the next. So we have Tamil history preserved and propagated in caves, copper plate inscriptions to palm leave manuscripts to printed books published during the last century.

At the present transition to 21st century, we are heading towards another technology based preservation and propagation of our language and its cultural and literary heritage. Computers and internet have become parts of our life and as principal means of communication and information interchange. In this context, non-profit global body, "International Forum for Information Technology for Tamil" (INFITT) has an important role to play in helping development of required hardware, software and related standards. Through several technical working groups and international conferences, INFITT is trying to bring together main technology developers from academia, IT industries and free lancers. We are pleased to organize "Tamil Internet 2009" in Cologne, Germany as the 8th International Conference of the series started in 1997. The slogan for this conference "கணி வழி காண்போம் தமிழ்" (kaNi vazi kANpOm tamiz) reflects our commitment to take Tamil language, its literary and cultural heritage to its next level through the use of computers and Internet. We are also pleased that the conference is being hosted by a premier European Institute for Teaching and Scholarly Research in Tamil, viz, the Institute of Indology and Tamil Studies of the University of Cologne, Germany

We are extremely pleased that, in spite of severe global recessions, response to our call-for-papers has been tremendous. With paper presenters coming in large numbers from India, Sri Lanka, Malaysia, Singapore, countries of Europe and North America, our initial plans to hold a mini-Tamil Internet Conference for two days have to be expanded to having two parallel tracks. We look forward to more than 50 technical papers presented and discussed in nearly all key areas of Tamil Computing and Tamil Internet as we see it evolving today. On behalf of the entire body of INFITT, I take this opportunity to welcome all the conference participants to Cologne and wish them a very enjoyable and productive conference.

Universität zu Köln
Institut für Indologie und Tamilistik
Geschäftsführende Direktorin: Prof. Dr. Ulrike Niklas

Pohligstr. 1, D- 50969 Köln,

Telefon (0221) 470-5346 / 5345, Fax (0221) 470-5385, E-Mail: u.niklas@uni-koeln.de

http://www.indologie.uni-koeln.de

In the year 2000, while I was an Assistant Professor at the South Asian Studies Programme at NUS in Singapore, I first came into contact with INFITT during the conference held at Singapore that year. Ever since, I kept contacts and friendships with people I had met during that event, and I ardently followed on the WWW the development of Tamil computing, made possibile by this organization.

I feel proud and thankful that INFITT gives me the opportunity to host this year's Conference at Cologne University. My special thanks go to Kalyan who first convinced me that we would be able to do it, and who afterwards shouldered most of the more complicated organisational work – leaving to me only those things that have to be done here in Cologne itself.

Hosting this conference in Cologne coincides well with the new "India-Initiative" of both, the City of Cologne and its University. During the last few years, several big IT companies from India have chosen Cologne as their European "home", and the University is at present developing its ties to India through MOUs and through offices set up in Delhi, Bangalore and Pondicherry.

On behalf of the City of Cologne and its University, I extend a warm welcome to the delegates from all over the world. Let us do our best to make this conference a success!

Vasu Renganathan
Department of South Asia Studies
University of Pennsylvania
Philadelphia, PA 19104 U.S.A.
Conference Program Committee Chairman

It is heartening to note that the organization of the International Forum for Information Technology for Tamil (INFITT) formed a dedicated and powerful community around the world that thrives for the betterment of the field of Information Technology in Tamil. When we conducted a similar INFITT conference in 2000 in Singapore this field was in its offspring stage and there were only very limited number of papers read in almost all of the areas namely E-texts, computational linguistics, natural language processing, computer assisted Tamil teaching and so on. In comparison, this conference proceedings, which carries around fifty papers in ten different subfields, demonstrates the state of the art innovations in Tamil information Technology. The papers in this book are divided into ten different topics namely a) Computer Assisted learning and Teaching of Tamil, b) Tamil Diaspora: Teaching Tamil as a second language and impact of technology, c) Tamil Portal, Aggregator and Wikipedia, d) Tamil localization, keyboards and open source software, e) Natural Language Processing: OCR, Information Retrieval and Artificial Intelligence , f) Tamil in handheld and mobile phones, g) Preparation E-Texts, Corpora, Search Engines, h) Computational Linguistics, i) Development of Morphological Tagger, and j) Database of Tamil technical terms, glossaries and dictionaries.

It has now become a fact that Learning Tamil as a second language implies the use of electronic resources, and simply using books and audio tapes, as we always had in the past, does not become a fruitful methodology for language pedagogy any more. Around twenty papers included in this book on this topic demonstrate various techniques and methodologies for how to use computers efficiently and productively in second language curriculum. The papers included in the topics on Tamil Portal, Tamil localization and keyboard standards demonstrate as to how internet became an important resource for information dissemination and how the tools that are available now for Tamil are exceptional in nature.

With the standard set for Tamil fonts via Tamil Unicode a completely new world of viewing electronic Tamil texts and making web portals emerged and it paved the ways very simple even for the lay computer users. This is obvious from the extensive use of Tamil in email communications, bloggers, websites, online electronic magazines and so on. One of the major advantages of conducting research in Tamil using electronic resources is looking at the data from various dimensions, which is not otherwise possible using the hard copies such as books. The papers that are included under the section on Tamil Web portals describe in detail how this tradition grew over the years. A quick example would be searching for the word மொழி in the entire canon of Tamil texts from Sangam to modern Tamil, one would immediately notice how this word went through a transformation of meanings such as ‘to utter’, ‘speech’, ‘sayings’ and so on and finally to mean ‘language’. With electronic approach, facts like this are comprehensible instantaneously.

The papers on the topics of E-texts, Corpora, electronic dictionaries, glossaries and search engines describe many revolutionary methods of using Tamil electronic materials. These papers reveal how

any Tamil research conducted without using electronic resources can easily become handicapped in furnishing appropriate facts in the respective field. This is true with respect to not only for studying the canon of texts from Sangam, medieval and modern Tamil, but also for studying Tamil inscriptions and palm-leaf manuscripts, which preserve a depth of knowledge about our precious Tamil tradition from time immemorial. Study of these texts has become immensely simple, and the papers included in this topic demonstrate it very meticulously. The researchers who devote their entire professional time on digitizing and researching Tamil literatures, inscriptions, modern Tamil texts etc., describe their findings in the papers that are included in this section. The papers on electronic dictionaries, glossaries and online databases demonstrate how the kind of facts that are derived from the nature of Tamil words, sentences, texts etc., were possible only with the advent of Tamil electronic resources, and how such facts could not have been discovered without this medium of research.

Not to mention the fact that studying Tamil language through the eyes of electronic bytes laid a foundation for the emergence of new fields of Natural Language Processing, Artificial Intelligence and Computational Linguistics. A complete array of papers included under these topics illustrate inspiring works related to Tamil on Optical Character Recognition, computer analyzing of Tamil words and sentences, Tamil text to speech, English to Tamil translation, sentence generation and so on so forth. Use of Tamil in handhelds and mobile phones has been popular ever since graphics enabled mobile phones were introduced. The section on 'Tamil in mobile phones and handhelds" illustrate the technology extensively with suitable demos. This made the medium of "Tamil texting" as simple as that of voice communication. A decade ago, typing Tamil in computers was felt to be a very difficult task, but it has now become the simplest task ever.

It may not be an exaggeration to say that all of the accomplishments as outlined in the research papers of this conference book were entirely possible due to hard work and dedication of INFITT members all around the world. Importantly, all of the researchers who have presented their works in this volume have benefited from each other in one way another by sharing the knowledge and resources that they evolved in their respective fields. In this sense it is an obvious fact that this "International community comprising of researchers on Tamil Information Technology" was shaped entirely by INFITT.

We hope this conference will be a turning point for every Tamil researcher in terms of getting to know more about collaborative research and information dissemination in the areas of Information Technology in Tamil. We believe the future research in these areas, as covered in all of the papers that are included in this conference book, will take a new dimension due to this conference, as it always had been due to the past conferences of INFITT.

எங்கும் தமிழ்! எதிலும் தமிழ்!
உலகெங்கும் தமிழ் பரப்பும் உத்தமம் புகழ்
எங்கனமும் எக்கணமும் ஓங்கட்டும்!

I. Executive Director's Report from January 2008 to October 2009

Introduction

The formation of International Forum for Information Technology in Tamil (INFITT) was the culmination of a year-long effort that began soon after the TamilNet 99 conference held in Chennai, India. INFITT was the first global organization set up to represent Governments, Corporations, interest groups and individuals who are concerned with the development and use of Tamil computing and Tamil Internet. It remains the only such organization in existence till now.

INFITT, since its inauguration on 23 July 2000, has focused its efforts on establishing the infrastructure to function as a global organization. Its annual international conference held annually till 2004 have gained international acceptance as the major platform to discuss issues relating to Tamil computing and Tamil internet.

INFITT, in close collaboration with its regional chapters, Government, Tamil Internet/Computing Steering committees and other organizations interested in Tamil computing, strive to archive the following through various initiatives.

- To coordinate the efforts of institutions and individuals interested in Tamil IT and to facilitate dialogue and cooperation among them;
- To identify key application areas of the development of Tamil IT, to define broad guidelines for their implementation and to provide technical assistance wherever and whenever possible;
- To develop norms and standards for Tamil IT;
- To promote knowledge and use of Tamil IT;
- To organize Tamil Internet Conferences (TIC) regularly;
- To act as a representative of the Tamil IT community in international, regional and national IT organisation, and to function as a liaison body and voice for Tamil Information Technology (IT)

After its inauguration in 2000, INFITT was officially registered as non-profit organisation in the United States on 20th May 2002. Though INFITT has had a membership scheme from its inception, the official membership of registered NGO was launched from 9th August 2002.

January 2004 and 2008 elections were conducted with a suspended GC, due to the membership not reaching the critical level.

New EC and office bearers elected

Thanks to INFITT advisor Professor M. Ananda Krishnan for the smooth conduct of EC election 2008. The election process was completed as planned new EC and office bearers were elected for a two year term ending December 31,2009. In addition to the new EC Thillai Kumaran replaced Kumar Kumarappan as treasurer of INFITT, since Kumar could not continue due to his prior commitments.

Treasurer is attached to secretariat and is responsible for keeping the books and be compliant per federal and state laws. INFITT sincerely thanks Kumar Kumarappan's help in preparing the tax statements and filing with Federal and State Governments. INFITT also thanks Thillai for agreeing to volunteer for this critical post to help us to be compliant legally according to the federal and state laws of US and State of California.

உத்தமம் 2008-2009 தேர்தல் முடிவுகள்.

உத்தமம் 2008-2009 க்கான ஆட்சிக் குழு :

தலைவர் : முனைவர் கு. கல்யாணசுந்தரம், சுவிட்சர்லாந்து.
துணைத்தலைவர்: திரு தி.ந.ச. வெங்கடரங்கன், சென்னை.
செயலர் - இயக்குநர் : திரு வா.மு.சே. கவிஅரசன், அமெரிக்கா.

உத்தமம் 2008-2009 க்கான செயற் குழு **:**

இந்தியா
திரு மாலன், சென்னை.
முனைவர் இலக்குவனார் மறைமலை, சென்னை.
திரு இராம.கிருட்டிணன் (இராம்.கி), சென்னை.
திரு அ.இளங்கோவன், சென்னை.
திரு தி.ந.ச. வெங்கடரங்கன், சென்னை.
முனைவர் பத்ரி சேஷாத்ரி, சென்னை.
திரு சதீஷ் குமார் நி., பெங்களூரு.

சிங்கப்பூர்
திரு மணியம் சிங்கை.

வட அமெரிக்கா
முனைவர் வாசு. அரங்கநாதன், நியூ ஜெர்சி.
திரு வா.மு.சே.கவிஅரசன், ஒகயோ.

ஐரோப்பா
முனைவர் கு. கல்யாணசுந்தரம், சுவிட்சர்லாந்து.

இலங்கை
திரு த.தவரூபன், யாழ்ப்பாணம்.

மலேசியா
திரு இரவீந்தரன் பால், கோலாலம்பூர்.

ஆஸ்திரேலேசியா
திரு ஜெயதீபன், சிட்னி, ஆஸ்திரேலியா.

பன்னாட்டு உறுப்பினர்
திரு கலைமணி, சிங்கை
திரு முனைவர் நா. கண்ணன், தென் கொரியா

பொதுக்குழு தேர்ந்தெடுக்க போதிய உறுப்பினர் இன்மையால், அனைத்து உறுப்பினர்களும் கொண்ட உறுப்பினர் குழுவே, உத்தமம் 2008-2009 க்கான பொதுக் குழுவாகவும் செயல்பட்டது**.**

உத்தமம் குழுமங்களில் தமிழ்

உத்தமத்தின் குழுமங்களில் தமிழில் உரையாடுவது மேலும் வலுப் பெற்றது. தமிழ்க் கணினி தொடர்பான சொற்றொடர்களுடன் உரையாடுவது எளிதான காரியமன்று. இருப்பினும் உத்தமத்தின் உறுப்பினர் பெரும்பாலும் தமிழில் உரையாட விழைவது மன நிறைவான நிகழ்வு. கணித் தமிழ் சொற்றொடர்களுடன் தமிழில் மடலாடும் உத்தமம் உறுப்பினர்களின் எண்ணிக்கை 2008-2009ல் பல மடங்கு அதிகரித்துள்ளது குறிப்பிடத்தக்க செய்தி.

Membership

Efforts are on to get more institutional memberships. Region wise current membership as of October 2nd 2009: India-31, Ameicas-21, Singapore-10, Sri Lanka-7, Malaysia-7, Australia-3, Asia-Pacific-1, Adding to a total of 88 members for 2009. It is very interesting to note that some members have renewed their 2010 membership as early as March 2009.

About 20 % of the existing members have already renewed their membership for 2010. A welcome and noticeable change to see more and more members extending their faith on continuing long term relationship with INFITT. We also continue to loose membership every year. We need to come up ways to retain the members. Some of the questions from non-renewing membership include what benefit INFITT membership is bringing me. Efforts are on way to continue the communications with the members who have not renewed their members and win back them to INFITT fold.

Institutional membership seems to be a challenge for INFITT. Efforts are on to with the state and national governments and with MNC to impress them to become institutional members and to involve INFTTT in their policy decisions relating to Tamil IT.

Workshops and Conferences

As you are aware INFITT did not do any major conference after TIC 2004. One of the major tasks of the new EC was to revive Tamil Internet Conferences. INFITT sincerely thanks all representatives in Singapore, India and Americas who have spent many hours and days to help us organise an event. After continuous deliberations, since a conference was never held in Europe in the past, it was felt a conference at Europe will be more appropriate and hence TIC2009 was born with due approval of EC for conducting a TIC 2009 at Germany. Thanks to Ulrike, CO-Chair TIC2009 and University of Cologne authorities for helping us to organise the flagship event of INFITT TIC 2009 at Germany.

INFITT North American Chapter in co-ordination with FeTNA has held INFITT Workshop in 2008 at Orlando, Florida and in 2009 at Atlanta, Georgia. About 50 FeTNA participants attended both the workshops. INFITT is planning continue its tie-up with FeTNA, for 2010 Workshop at Connecticut and future FeTNA events. Singapore EC members conducted teacher training programs. Glad to note INFITT EC members are leading initiatives in their regions.

Publications

Revival of INFITT also reflected on the publication front. Dr. N. Kannan was appointed as editor of Minmanjari. Thanks to the efforts of the editor Dr. N. Kannan, Minmanjari 2008 Deepavali Malar was released last year. Minmanjari, is also a bringing a special souvenir to commemorate TIC 2009. Efforts are on to start a new Technical journal on Tamil IT. EC approved an expense to get ISBN numbers for the conference book and ISSN number for Minmanjari.

Regional Chapters

INFITT realizes the need for a strong regional chapter. Since INFITT is registered in USA, our focus was to create a regional body, which will help us to continue to grow and create interest in US residents. USA GB members has accepted the proposal on NA leadership. USA chapter is now operational with Dr. Na. Ganesan is now the president and Thiruthangal Vetri Pandian is the secretary of INFITT NA chapter. Hopefully in the years to come INFITT NA-chapter will help us maintain the legal requirements of INFITT, as INFITT is registered in USA.

Efforts are on to create a strong regional chapter in Malaysia and to strengthen the regional chapters in Singapore. In India and Sri Lanka, we have not made much head way on forming the regional chapters. We are still working closely with the regional EC and GB members to form active chapters in every region.

Canada and Middle East with large Tamil population, is a huge potential for us to expand our membership bases. We are in the process of identifying the right persons for leading the charge in these two regions.

Working Groups

Many General Body members showed interest in becoming members of Working groups. We are in the process of reviving some of the working groups, including changing of Chairman, adding more active members etc. Dr. Na. Ganesan took over as Chair of Working Group 2 – Unicode. Other inactive working groups are being reviewed, the working groups will either be closed, if it has completed it's purpose or will be revived with new active persons as the case may be.

Domain names

We are happy to inform that infitt.com domain is also now owned by us. All references for infitt.com are now redirected to infitt.org

Finance

Unfortunately INFITT's main source of income continues to be from the membership dues from the members. Although per constitution INFITT can accept donations from MNC's and Private and Public institutes, not much progress is made to build a strong viable financial base for INFITT. Efforts are on to identify ways that INFITT can generate revenue, including Conference fees, Technical journal revenues, making Minmanjari as a revenue generating resource etc.

Discussion threads in EC

Discussions in EC rose to a new level. Till date around 610 messages were posted to ECINFITT group and it is all set to surpass the previous record of 654 messages in 2004. Kudos to EC for active participation. It also needs to be noted that some EC members posted as few as 5 messages. When we accept the fact that the numbers of messages are in no way measure of our performance, we do need to acknowledge the time spent by each EC member in reading, digesting and participating in the threads. Thanks to all members of EC who has participated in the threads.

Apart from transacting normal business, some of the interesting discussions include Constitutional amendments, composition of EC, Role and powers of ED, Chair and Vice-Chair role changing on yearly basis, Relationship with Government dignitaries, etc.

Discussion threads in GC / GB

From INFITT inception GC was never elected from GB, as there was not enough region wise membership. We believe we now have enough membership to constitute a GC from GB. Hence a GC will be elected for the year 20010 and 2011. Interesting discussions and topics in GB included posting of Text to Speech Converter by Prof A.G. Ramakrishnan, received comments and applauded by members of General body. Another interesting email was the recommendation on the naming of the sessions by Tamil IT pioneers who are not with us anymore. A gentle and nice way to remember the Tamil IT pioneers during TIC 2009.

பொதுக்குழுவின் பரிந்துரைகளை ஏற்று, இணைய மாநாடு 2009ல் சிங்கைத் தமிழர் நா. கோவிந்தசாமி அரங்கு , சென்னைத் தமிழர் சுஜாதா அரங்கு, ஈழத் தமிழர் சன்முகலிங்கம் இராமலிங்கம் அரங்கு, அரபுத் தமிழர் உமர் அரங்கு என நான்கு அரங்குகள் பெயரிடப்பட்டு, அவர்களின் பணியை மணியம், வெங்கட், கவி, கணேசன் நினைவு கூறவும் உள்ளனர்.

2010 and ahead

Election Schedule will be announced during TIC2009 GB meet and new GC, EC and the Chair, VC, ED (Trio) team will start functioning from 1st January 2010.

Proposed INFITT 2010 Election Schedule (Elects GC, EC and Trio per constitution – Annexure A)

Suggested Returning officer	: Dr. M.Ananda Krishnan
Call of GC Nominations	: November 18 to November 24 2009
Election of GC, if needed	: November 25 to December 1 2009
Call for EC nomination	: December 2 to December 8 2009
Election of EC, if needed	: December 9 to December 16 2009
Call for nomination for Chair, VC and ED (Trio)	: December 17 to December 24 2009
Election of Trio (Chair, VC, ED) if needed	: December 24 to December 31, 2009
New Trio, EC and GC assume office	: 0 1/01/2010.

Conclusion

Over the nine years, from inception some initiatives of INFITT have gained recognition, the most significant being the annual Tamil Internet conferences. These have also become popular platforms for launch of Tamil computing programs by interested parties to reach out to a wider audience. In recognition of the efforts of INFITT, governments have also extended their support to promote Tamil Internet and Tamil computing. With this institutional support, INFITT hopes to strive for bigger things in the future and seeks strong support and commitment from its members to make its global presence. Hopefully a new era is now started with the onset of Tamil Internet conference again in 2009 after a gap of five years.

II. ED's Report from February 2004 to December 2007

Changes in the Office bearers (Chair, Vice Chair, ED)

Upon completion of their terms, Mr. Muthu Nedumaran and Mr. Arun Mahizhnan resigned as Chair and Ex. Director of INFITT. They indicated their desire to leave office way back in Dec. 2004, but continued on request of EC members. INFITT EC promptly elected Dr. K. Kalyanasundaram of Switzerland (up till now as Vice- Chair) and Mr. T.N.C Venkatarangan of Tamilnadu, India as the Vice-Chair of INFITT. EC could not find a suitable Executive Director for INFITT and was filled only on January 2008.

Changes in the Executive Committee (replacements for Singapore & Malaysia)

To enable induction of new members into EC, Mr. Muthu Nedumaran and Mr. Arun Mahizhnan decided to leave INFITT EC. Mr. S. Maniam and Mr. R. Kalaimani of Singapore have been co-opted to the EC to represent this key region of Tamil Diaspora. Similarly Mr. Ravindran K. Paul of Kuala Lumpur has been co-opted to represent Malaysia. We thank these professionals (who have been participating in INFITT activities such as Tamil Internet Conferences and Working Groups) for their willingness to join the EC and help build INFITT.

INFITT Secretariat

When INFITT was launched in 2000, Singapore Govt. has been kind enough to provide direct support (through their IDA) for the running of INFITT Secretariat in Singapore. This support enabled having the services of Mr. Nara Andiappan as Administrative Manager. Initial promised support for 2 years was extended by another year. Since then Nara has been kind enough to provide voluntary support. In view of this situation, early this year, Mr. Arun Mahizhnan initiated discussions with several key Tamil IT personalities of Chennai to see if the INFITT Secretariat can be relocated in Tamilnadu. On a suggestion from our INFITT Advisor Prof.Anandakrishnan, Mr.Arun contacted Kani Thamizh Sangam (KTS) of Chennai and they have expressed interest to support INFITT on this. Suggestions have been made that the next Ex. Director (ED) of INFITT also be from Chennai to help facilitate this migration of the Secretariat from Singapore to Chennai. While there has been broad support for the proposal to move the Secretariat to Chennai and run it with the help of KTS, we must acknowledge that there has not been any consensus on the linking of this with the election of the next ED for INFITT. ED is the Chief Adm. Officer of INFITT directly responsible for the running of the Secretariat. EC is still working on the detail. ED was formally elected in January 2008.

Individual and Associate membership /new definition, scope

As part of the reactivation process, EC will be launching a major grass roots membership drive. Since the term of office for the elected GC and EC has expired, there is an urgent need to revamp the INFITT electoral body "GB" made up of registered members. INFITT is an international organization committed to promote Tamil IT across the globe. Bodies such as INFITT cannot sustain and grow without the support and help of main stream Tamil Diaspora. INFITT need to attract in particular younger generation Tamils interested in Tamil IT. Hence the EC decided to offer "Associate membership" to this community without payment of annual dues. In view of absence of major activities during the past 2 years, the EC also decided to offer 50% discount of the annual dues, for a limited period of 3 months to those who sign up as "individual members". Online registration with option to pay dues with credit card will open soon in our website. To promote interactions between members of a given region, INFITT has regional chapters. We have already such regional Chapters for Europe and North America. Discussions were still underway to start INFITT Chapter for Singapore and Malaysia.

Changes in the INFITT website (new look, main and mirror sites)

Thanks to the support of Dr. Badri Seshadri of Chennai, INFITT website has been running for nearly two years from one of his web-servers. When INFITT website was launched in 2000, it was designed with the help of IT professionals of Singapore IDA. Good part of the site was based on specific scripts. Now that IDA support is no longer available, we could not make changes in the code-base. We need to find alternative ways of running our website. After examining various options, EC decided to use one of the popular and well supported open source CMS package. INFITT website now has a new face-look and is based on the new CMS. We are still in the process of moving the contents of the old site, while restructuring the site itself. Since paid web-hosting prices have come down considerably during the past few years, EC also decided to go for paid web-hosting to run its main website "infitt.org" (as was the case during 2000-2003) and use the web server of Dr. Badri Seshadri to host a mirror site (mirror.infitt.org). New site is already running from a paid webhost based in USA.

Plans for a new Discussion list on Tamil in FOSS

While exploring various CMS options for the INFITT website, EC also examined the possibility of running INFITT website truly as a bilingual site. A bilingual site where the user can choose the view the contents of the site in Tamil or in English. This requires "Tamil locale" enabling in the open source

CMS. "Tamil Localisation" is much broader in scope, applicable to operating systems to nearly all major software. There have been lots of efforts scattered across the globe on "Tamil enabling" in Linux OS, Open source Office and with Mozilla Browsers. We in the EC feel that there is an urgent need to bring together all these scattered efforts on "Tamil locale enabling" in all of "Free and Open Source Software FOSS". EC has decided to launch a discussion list exclusively on this topic. We will be opening up a "Forum" section in the website where we will start this discussion list soon. We urge all those working on this area and all those interested in this topic to join this DL so that collectively we can solve required problems and even share the technical manpower available.

Tamil Internet Conference from 2005 to 2008

Tamil Internet Conferences (TIC) has been the flagship events of INFITT, reaching out to the Tamil Diaspora living in key regions of the world. In spite of repeated attempts TIC 2009 could not organized after 2004 until 2009.

Workshops

INFITT Sri Lankan Chapter organised a one day seminar on 06th June 2004 and INFITT Chairman at that time Mr.Muthu Nedumaran was the main resource person.

Singapore EC member Mr.R.Kalaimani organized a Podcasting workshop in Singapore in Oct 2007. It was very well received by the Singapore school teachers.

In summer 2007 Dr.K.Kalyanasundaram, Mr.T.N.C.Venkatarangan and Mr.Badri Seshadri gave talks on Tamil Computing/IT to computer science students of few private Engineering colleges in and around Chennai, all organized jointly with the Tamilnadu chapter of the Computer Society of India Meenakshi College of Engineering for Women, RMD Engineering College, and Jeppiyar Engineering College. Dr.Kalyanasundaram also gave a talk on INFITT and its activities to the members of Chennai chapter of Computer Society of India at Hotel Kanchi, Chennai in July 2007.

Institutional Meetings

In summer 2007 Dr.Kalyanasundaram and Mr.Venkatarangan visited Microsoft and Yahoo India Regional R& D development centers at Bangalore and had fruitful discussions with the senior managers there on possible collaboration between INFITT and these lead IT MNCs.

III. INFITT Member news

Dr. K. Kalyanasundaram

In Summer 2008, Dr.Kalyanasundaram received the Sundara Ramasamy Award for Tamil Computing given by the Tamil Literary Garden Group of Canada at a function held at the University of Toronto. During February 2009, he also received the University of California Berkeley Tamil Chair Award at their annual Tamil Conference held at the University of California, Berkeley, CA campus.

Dr. Naga. Ganesan

'மாதவிப்பந்தல்', முனைவர் நாக கணேசனுக்கு பட்டாம்பூச்சி விருது வழங்கி கௌரவப்படுத்தினர்

திரு.மாலன் (நாராயணன்)

மூத்த பத்திரிக்கையாளர் திரு.மாலன் (நாராயணன்), இளைஞர்களுக்கான "புதிய தலைமுறை" என்ற வார இதழைத் தொடங்கியுள்ளார் .http://www.puthiyathalaimurai.com/

முனைவர் திரு.மறைமலை

தமிழ்க்கவிஞர்கள் பலருக்கு வலைப்பூ அமைத்து அவர்தம் கவிதையின் சிறப்பை உலகறியச்செய்யும் முயற்சியில் ஈடுபட்டுள்ளார் .கவிதைகளின் ஆங்கிலமொழியாக்கமும் கவிஞர்களின் வாழ்க்கைக் குறிப்பும் வெளிநாட்டார்க்கு -குறிப்பாக ஆய்வாளர்க்கு நன்கு பயன்படுகின்றன .இப்போது தமிழிலும் வலைப்பூக்கள் உருவாக்கிவருகின்றார் .இத்துணைக் கவிஞர்களுக்கு யாரும் இதுவரை இப்படி வலைப்பூ அமைத்ததில்லை என்பது இதன் தனிச்சிறப்பு.

Mr. Siva Pillai

Mr. Siva Pillai has the honor of being a Principal Examiner for Cambridge ASSET Examination (Tamil Language). He also has the honor of being a Chief Examiner for London Edexcel Examination (Tamil Language). He is the winner of European Languages 2007. He is the winner of OURLANUAGES Project 2008/09. He is an Honorable member of UKFCS- United Kingdom Federation of Chinese School.

Annexure A.
Refer the constitution document and amendment–I to the constitution available online at http://bit.ly/infittc (redirects to the actual pages in Infitt.org website)

COMPUTER ASSISTED LEARNING

AND

TEACHING OF TAMIL

E-Learning for Enhancing Language Proficiency

by

Dr.A.Devaki
Senior Lecturer in Education,
Govt. College of Education,
Komarapalayam-638183,
Namakkal District, Tamil Nadu, India.
E-Mail: devi_mathi2006@yahoo.co.in
Mobile: +919865354869.

Prof.D.Mathialagan
Head, Department of English,
Institute of Road and Transport Technology,
Erode-638316,
Tamil Nadu, India.
E-Mail: mathi_d2001@yahoo.co.in
Mobile: +919842753370.

E-learning has been in vogue for more than a decade and includes all technology enhanced learning. It is akin to distance learning with few more advantages for the learner. Today, particularly in the third world countries, where it is difficult to provide on-campus learning for all the learners, it is imperative that e-learning is taken up and encouraged in a big way to make it accessible and affordable for all learners. Students of e-learning rarely or never meet face-to-face, nor access on campus educational facilities. E-learning guides the students through information or helps them perform in specific tasks.

E-learning is capturing a large portion of learning activities both in academics and industry. The use of self-placed e-learning is gaining currency all over the world. Many higher educational institutions are offering on-line classes.

While creating content for e-learning one has to be flexible in one's approach. An educator has to effectively create educational materials while providing the most engaging educational experiences for the student at the same time. E-learning system not only provides learning objectives, but also evaluates the progress of the student and credit can be earned toward higher learning institutions. This reuse is an excellent example of knowledge retention and the cyclical process of knowledge transfer and use of data and records.

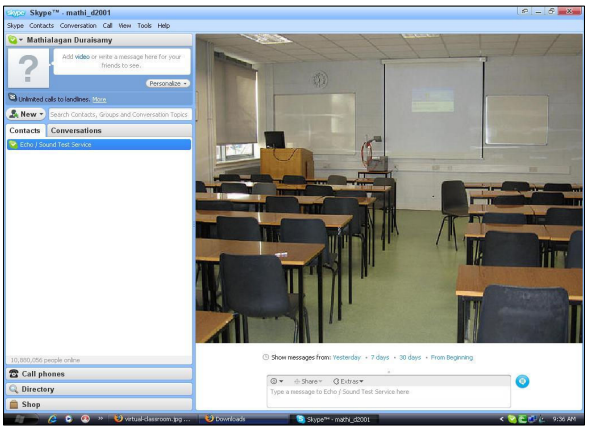
Today many technologies are used in e-learning, from blogs to collaborative software, e-portfolios and virtual classrooms. Most e-learning situations use combinations of these techniques. E-learning, however, also has implications beyond just the technology and refers to the actual learning that takes place using these systems. E-learning is naturally suited to distance learning and flexible learning, but can also be used in conjunction with face-to-face teaching, in which case the term Blended learning is commonly used. E-learning pioneer Bernand Luskin says that the 'e' should be interpreted to mean exciting, energetic, enthusiastic, emotional, extended, excellent and educational in addition to electronic. Information based e-learning content communicates information to the student. In information based content, there is no specific skill to be learned. In the performance based content, the lessons build of a procedural skill in which the student is expected to increase proficiency.

The major benefits of e-learning are that it is eco-friendly because it takes place in a virtual environment and thus avoids travel and reduces the usage of paper. An internet connection, a computer and a projector would allow an entire classroom in a third world university to benefit from knowledge sharing by experts. E-learning is self-paced and can be done at anytime of the day. Students generally appear to be at least as satisfied with their online classes as they are with traditional ones. Properly trained staff must also be hired to work with students online. These staff members need to understand the content area, and also be highly trained in the use of computer and internet. The recent trend in the e-learning sector is screen casting. The web based screen casting tools allow the users to create screen casts directly from their browser and make the video available online so that the users can stream the video directly. From the learners point of view this provides the

ability to pause and rewind and gives the learner the advantage to move at their own pace, something a classroom cannot always offer.

The challenge before the Tamil Diaspora is to make use of technologies such as blogs, wikis and discussion boards to promote the teaching of Tamil language skills. The presenter has experience in creating online content for e-learning course modules of the Tamil Nadu Virtual University. The teaching of Tamil to native and non-native speakers using technology through the internet has a bright future and the potential has to be tapped. The available technologies have to be put to right use. A catch-all phrase that included any form of technology assisted learning, e-learning is poised to revolutionize the process of education. The sectors which are entering the field of e-learning serve as a testimony to the growth of e-learning. Telecom, banking, finance and government are rapidly moving towards e-learning. The primary driver is not just to decrease cost but also to increase reach. Universities are also looking at e-learning modules to supplement their regular curriculum courses.

In this context, it becomes necessary to understand how effective e-learning courses are. More simulation-based training based on games are being incorporated in e-learning. And a high level of acumen is required to develop such e-learning modules. And for an e-learning programme to work, it is important to first understand whether something is suitable for e-learning or not. There are two layers to a successful e-learning programme-the technology component and the learning component.



Virtual Classroom environment in Skype

Screen shot of Hi Class Software used for testing Language Proficiency

In India, e-learning courses could be made more popular through availability of broad-band connections at competitive rates, regional language-based content for technical subjects, two-way interaction for doubts and performance feedback with students. A shift in mindset is required to adopt e-learning. It is the same barrier that exists with any adoption to technology. But once that is overcome, e-learning would prove beneficial. As knowledge is socially constructed, learning has to take place through conversations. One of the best ways to learn something is to teach it to others. Teachers of Tamil will have to venture out of the classrooms and move beyond the textbooks to create a conducive environment for the language learners using technology. Students can be encouraged to use Skype, Facebook and Second life, which have become providers of Virtual Classroom environments.

The paper is an attempt to project teaching of Tamil using technology and keep pace with the changing times.

References

1. How to Design Effective Blended Learning, by Julie Marsh and Paul Drexler, November 2001, brandon-hall.com.
2. Ravet S. & Layte M., Technology- based training – a comprehensive guide to choosing, implementing, managing, and developing new technologies in training, Gulf Publishing company, 1998, Houston, Texas.
3. www.elearningmag.com – an online magazine about e-learning.
4. Revenaugh, Mickey. (2005) **Virtual schooling: legislative update** techLearning. Retrieved April 10, 2006, from http://techlearning.com/showArticle.jhtml;jsessionid=4YMRLLWIX5VQMQSNDBGCKHSCJUMEKJVN?articleID=160400812&_requestid=234790.
5. **Resources, statistics, and distance learning resources.** United States Distance Learning Association. http://www.usdla.org/html/aboutUs/researchInfo.htm
6. Siemens, George. (2002, September 30). **Instructional design in e-learning.** elearnspace Retrieved May 2006 from http://www.elearnspace.org/Articles/InstructionalDesign.htm
7. Koskela M, Kiltti P, Vilpola I and Tervonen J, (2005) **Suitability of a virtual learning environment for higher education.** The Electronic Journal of e-Learning Volume 3 Issue 1, pp 21-30, available online at http://www.ejel.org

இணைய வகுப்பறை

தொழில் நுட்பமும் கற்போர் உளவியலும்

Virtual Class room: Technology and Learner's Psychology

முனைவர் வெ. கலைச்செல்வி
தமிழ்த்துறைத் தலைவர், அரசு கல்வியியல் கல்லூரி
குமாரபாளையம் - 638 183. தமிழ்நாடு. இந்தியா
vkalai30@rediffmail.com

Abstract: This paper introduces the concept of virtual class room based techniques and learners psychology. Teachers and students are refrain from each other by place and time. Teaching– learning by virtual mode is the recent modern environment which has influenced the Tamil teaching climate. H.T.M.L., Power point, E-learning, Video conferencing which facilitates changes in the teaching learning process. For e-learning websites like Black Board Web CT, Moodle, Joomla LMS are available. Software like hot potatoes helps the learners to modify the virtual class room as they wish.

Based upon Internet, Learning Management System and Educational Management System are developed. We have to approach teaching learning components not only on the basis of technology but also on the basis of learner's psychology. Learning is a continuous process. Reinforcement is needed through out the teaching process. Teaching should be from simple to difficult and known to unknown. Learning based on Cognitive, Affective and Psychomotor domain. Understanding application, Synthesis, Analysis, Receiving, Responding, Evaluation, Naturalization are some of the important functions that we expect from a learner. So we have to develop virtual learning based on learner's psychology.

We can expect the realization of the set goals only when web based technology and learners psychology come hand in hand. So the text, graphics, colors and action everything should be created on the bases of technology as well as learners psychology. In the democratic class room situation either the teacher or the students should not dominate. Hence, this article describes the approach of language programme pioneered by Tamil virtual University in imparting Tamil scripts, phonemes and words to non Tamils.

ஆசிரியர்களும் மாணவர்களும் இடத்தாலும் காலத்தாலும் விலகியுள்ள நிலையில், இணையத் தொழில்நுட்பம் வழி நிகழும் கற்றல் கற்பித்தல் செயல்பாடு தமிழ்ச் சூழலில் புதிய வரவாகும்.'இணைய வகுப்பறை' எனும் கருத்தாக்கத்தையும் இது ஏற்படுத்தியுள்ளது.

எச்.டி.எம்.எல்., பவர் பாயின்ட்,பி.டி.எப். போன்ற பனுவல் ஆவணங்கள், மீ உரை, மீ இணைப்பு வழி பெறப்படும் பாடங்கள்,வீடியோ மாநாட்டு முறையில் முகமுகமாக அமையும் உரையாடல்கள் இன்றைய கற்றல் கற்பித்தல் சூழலில் புதிய மாற்றங்களை ஏற்படுத்தி வருகின்றன. மின் கற்றலுக்கு 'Black board', 'Web CT',' Moodle', 'Joomla LMS' போன்ற தளங்கள் உள்ளன. தாம் விரும்பும் வண்ணம் இணையப்

பாடங்களை அமைத்துக் கொள்வதற்கான 'Hot Potatoes' போன்ற மென்பொருள்கள் தமிழில் பயன் கொள்ளப்பட்டிருக்கின்றன. இணையத்தையொட்டி, கற்றல் மேலாண்மை அமைப்பு, கல்வி மேலாண்மை அமைப்பு போன்றவையும் உருவாகியுள்ளன. கற்றலுக்கான பாடக் கூறுகளையும் கற்பித்தல் கூறுகளையும் தொழில்நுட்ப ரீதியாக மட்டுமின்றி, கற்போர் உளவியல் சார்ந்தும் அணுகவேண்டியுள்ளது. கற்றல் என்பது தொடர் நிகழ்வு ஆகும். கற்றலைத் தொடர்ந்து ஊக்கப்படுத்தும் காரணிகளை நாம் படிப்படியாக வழங்க வேண்டியுள்ளது. எளிமையிலிருந்து கடினத்திற்கும், தெரிந்ததிலிருந்து தெரியாததற்கும் கற்பித்தலை மேற்கொள்ள வேண்டியுள்ளது.

கற்றல் என்பது அறிவு சார் களம், உணர்வு சார் களம், உள இயக்க சார் களம் ஆகிய புலங்களைக் கொண்டது. புரிந்து கொள்ளுதல், பயன்படுத்துதல்,பகுத்தல், இணைப்பாக்கம் முதலிய செயல்பாடுகளும், ஏற்றல், பதிலளித்தல், மதிப்பீடு, ஒழுங்குபடுத்துதல், பண்பாக்கிக் கொள்ளுதல் போன்ற செயல்பாடுகளும் கற்போரைச் சார்ந்து நிகழ வேண்டியுள்ளன. இவற்றையும் கருத்தில் கொண்டே இணைய வழி கற்றலுக்கான நெறிகளை நாம் உருவாக்க வேண்டியுள்ளது.

கணினி சார்ந்த தொழில்நுட்பமும் கற்றல் சார்ந்த கற்போர் உளவியலும் ஒன்றோடொன்று சரியாக இணையும் நிலையில், நாம் எதிர்பார்க்கும் விளைவுகளை ஏற்படுத்த முடியும். சரியான தூண்டல்களே பொருத்தமான துலங்கல்களை ஏற்படுத்தும்.

எனவே, திரையில் தெரியும் பனுவல், வரைபடம், வண்ணம், இயக்கச் செயல் போன்றவை தொழில்நுட்ப ரீதியாக மட்டுமின்றி, கற்போர் உளவியல் சார்ந்தும் அமைதல் வேண்டும். இக்கட்டுரை இணைய வழி அமைந்த தமிழ் கற்றலுக்கான தளங்கள் சிலவற்றைக் கருத்தில் கொண்டு மேற்சுட்டிய காரணிகளின் அடிப்படையில் ஆராய்கிறது. இக்கட்டுரை தமிழ் இணையப் பல்கலைக் கழகத்தின் தமிழ்க்கல்வித் திட்டத்தின் இணைய வகுப்பறையில் கையாளப்படும் தமிழ் கற்பித்தல் தொழில்நுட்பம் மற்றும் கற்போர் உளவியலைப் பற்றிய மதிப்பீடாக அமைகிறது.

இணைய வகுப்பறை

மாணவா;கள் விரும்பினாலொழிய கற்க முடியாது என்கிறார் ஸ்ரீ அரவிந்தா;. மாணவா;கள் கற்பதற்கு உயிரோட்டமுள்ள வகுப்புச் சூழல் தேவை என்று கல்வியாளா;களும் உளவியல் அறிஞா;களும் குறிப்பிடுகின்றனா; மாணவா;கள் உள்ளங்கொள்ளும் வகையில் அவா;தம் விருப்பத்திற்கேற்ப தரமான, செறிவான கருத்துகளை இணைய வகுப்பறையில் ஆசிரியா;கள் வழங்குகின்றனா; இணைய வகுப்பறையில் ஆசிரியா; ஒருவா; மாணவா;கள் எதிரில் இருப்பதாகக் கற்பனை செய்து கொண்டு மாணவா;களுக்குத் தோன்றும் ஐயங்கள், வினாக்கள் ஆகியவற்றிக்கு ஏற்ப தாமாகவே முன்திட்டமிட்டுக் கற்பிக்கிறார்;. மாணவா;களுக்கு உற்சாகம் குன்றாதவாறும் சலிப்பு ஏற்படாமலும் கற்பிக்க வேண்டியது இன்றியமையாதது. மாணவா;களின் வயது, கற்றல் திறன், நுண்ணறிவு, தேவை, சூழல் ஆகியவற்றைப் பின்புலமாகக் கொண்டு இணைய வகுப்பறைக்கான பாடங்கள் உருவாக்கப்படுகின்றன.

பியாஜே என்னும் உளவியலாளா; கற்றலில் திறன்களை 'ஸ்கிமேடா' என்று சுட்டுகிறார். இவை குழந்தைகளின் மனதில் ஏற்கெனவே பதிந்துள்ளன. கற்றலுக்குப் புலன்களும் புலக்காட்சியும் அடிப்படையாக அமைகின்றன. இணையக் கல்வியில் கிடைக்கும் தகவல்களும் வண்ணத் திரையில் கவனத்தை ஈர்க்கும் அசையும் மற்றும் அசையா உருவங்களும், கேட்கத் தூண்டும் இனிமையான குரல் ஏற்றத்தாழ்வுகளும், இசையும் மாணவா;களுக்குப் புலக்காட்சி அனுபவத்தை அளிக்கின்றன. எனவே இணைய வகுப்பறையில் தன்வயப்படுதலின் மூலம் மாணவா;களின் கற்றல் வலுப்பெறுகிறது.

தமிழ் இணையப் பல்கலைக்கழகத்தின் கல்வித் திட்டம்

இத்திட்டத்தில் தொடக்கக் கல்வி, உயர்கல்வி, பிற இணைய வழிக்கல்வி என்னும் மூன்று பிhpவுகள் உள்ளன. தொடக்கக் கல்வியில் மழலைக்கல்வி, சான்றிதழ்க் கல்வி, மேற்சான்றிதழ் கல்வி என்னும் நிலைகளில் தமிழ்மொழித் திறன்கள் கற்பிக்கப்படுகின்றன. சான்றிதழ்க் கல்வியில் 1 முதல் 6-ஆம் வகுப்பு வரையிலான மொழிப்பாடத் திறன்கள் அடிப்படை நிலை, இடைநிலை, மேல்நிலை என்னும் மூன்று

நிலைகளில் அளிக்கப்படுகின்றன. இதற்கான பாடங்கள் இணைய தளத்தில் இடப்பட்டுள்ளன. சான்றிதழ் கல்வியில் அடிப்படைநிலை இணைய வகுப்பறையின் தொழில் நுட்பமும் கற்போர் உளவியல் ஆகியன இங்கு மதிப்பீடு செய்யப்படுகின்றன.

பாட அமைப்பு முறை

இதில் அடிப்படை நிலையில் மொத்தம் 39 பாடங்கள் வடிவமைக்கப்பட்டுள்ளன. பாடம் 1 முன்னுரையாக அமைந்துள்ளது. பாடம் 2 மற்றும் 3 வாய்மொழிப் பயிற்சியாகவும், பாடம் 4 எழுத்துகளை அறிமுகப்படுத்தும் எழுத்துப் பயிற்சியாகவும், பாடம் 5 முதல் 39 வரை உயிர், மெய், உயிர்மெய் எழுத்துகளை அறிமுகப்படுத்தும் பாடங்களாகவும் அமைந்துள்ளன.

கற்பித்தல் தொழில் நுட்பம்

இணைய சான்றிதழ்கல்வி நிலையில் அடிப்படைக்கல்வி "தமிழ்க் கற்போம்" என்று பின்னணி இசையுடன் வண்ணத்திரையில் காட்டப்படுகின்றது. 39 பாடங்களும் சற்றேறக்குறைய 20 மணி நேரம் கற்பிக்கப்படுகின்றன. எழுத்துப்பயிற்சி மற்றும் வாய்மொழிப் பயிற்சியை முனைவர் நன்னன் அவர்கள்; தமிழ் பயிற்றுமொழி மூலமும் பேராசிரியர் சித்தலிங்கையா ஆங்கிலப் பயிற்று மொழி மூலமும் கற்பிக்கின்றனர். பாடம்-1 தமிழ்க் கற்பித்தல் பற்றிய முன்னுரையாக அமைந்துள்ளது. சர்க்கரைப் பொங்கலைத் திரையில் காண்பித்து, தமிழ்க் கற்பித்தலை அதனோடு தொடர்புபடுத்துகிறார் ஆசிரியா;. கற்றல் கற்பித்தல் நிகழ்வில் மாணவர்களை ஊக்குவித்தல் அல்லது ஆயத்தப்படுத்துதல் என்பது இன்றியமையாதது. ஆனால் 30 நிமிட நீண்ட முன்னுரை கற்போருக்குச் சலிப்பூட்டக்கூடும்.

ஒரு பாடத்தைப் படித்தவுடன் அடுத்த பாடத்திற்குச் செல்ல மீண்டும் முதன்மை மெனுவிற்குச் செல்ல வேண்டியிருக்கிறது. வாய்மொழிப்பயிற்சியில் சில சொற்கள் மாணவர்களுக்குக் கூறப்பட்டு, உச்சாறப்புப் பயிற்சி அளிக்கப்படுகிறது. நீங்கள் சொல்லின் பொருளைப் பற்றிக் கவலைப்பட வேண்டாம் என்று ஆசிரியர் கூறுகிறார். அதற்கு மாற்றாக மாணவர்கள் அறிந்த சொற்களைக் கூறி வாய்மொழிப் பயிற்சி அளிப்பது கற்றலை நீண்டநாளுக்கு நினைவில் நிறுத்தும். எழுத்துப் பயிற்சியில் ப, ட, ம ஆகிய எழுத்துகள் கற்றுத் தரப்படுகின்றன. வாறவடிவம் எழுதிக்காட்டப்படுகிறது. இங்கு எளிமையிலிருந்து கடினத்திற்குச் செல்லுதல் என்னும் நுட்பம் கையாளப்பட்டுள்ளது.

6வது பாடம் முதல் 25 பாடம் வரை மெய்யெழுத்துகளும், 26 முதல் 39 வரை உயிர்மெய் எழுத்துகளும் அறிமுகம் செய்யப்பட்டுப் பயிற்சி அளிக்கப்படுகின்றன. ஒவ்வொரு எழுத்தையும் 'ஈ' காரம் முதல் ஔகார உயிர்மெய் என்று பட்டியலிட்டு விட்டு அகரத்தையும் ஆகாரத்தையும் மெய் எழுத்துகளுடன் சோ;த்துக் கற்பிக்கிறார் ஆசிரியா; எழுத்துகளை வகைப்படுத்தி இருப்பது ஒரே மாதிரியாக இல்லை.

புதிய சொற்கள் அறிமுகப்படுத்தப்படும்பொழுது அவற்றை வாக்கியங்களில் இடம் பெறச்செய்து ஒலித்துக்காட்டப்படுகின்றன. பின்னா; எல்லாப் பாடங்களிலும் சொற்களுக்கேற்ற ஒருசில கருப்பு வெள;ளைப் படங்களே காட்டப்படுகின்றன. ஒவ்வொரு சொல்லுக்கும் பொருத்தமான வண்ணப் படங்களை இணைத்துக் கற்பிப்பது புலக்காட்சி அனுபவத்தை மேம்படுத்தும். தமிழைத் தாய் மொழியாகக் கொள்ளாதோருக்குப் படங்கள் இல்லாமல் தமிழில் புதிய சொற்களைக் கற்பிப்பது கடினமானதாக அமையும். முழுமையிலிருந்து பகுதிக்குச் செல்லுதல் என்னும் கற்பித்தல் தொழில் நுட்பம் கையாளப்படுகிறது.

பயிற்சியில் விடுபட்ட எழுத்தை எழுதுதல், எழுத்தை மாற்றி எழுதுதல் ஆகிய பயிற்சிகள் அளிக்கப்பட்டுள்ளன. பாடம் 39 இல் ஔகார உயிர்மெய் எழுத்துகளைக் கற்பிக்கும்பொழுது ஆய்த எழுத்தும் கற்பிக்கப்படுகிறது.

எழுத்துகள் அறிமுகத்தைத் தொடர்ந்து ஒவ்வொரு பாடத்திலும் எழுத்துகள் இடம் பெறும் சொற்களை எழுத்துக் கூட்டிப்படிக்கக் கற்றுத் தரப்படுகிறது. இடம் பெறும் சொற்களுக்கேற்ற அனைத்துப் படங்களும் காட்டப்படாதது குறையாக உள்ளது. கற்பித்தலின் பொழுது ஆசிரியாறன் குரலும்

வாயசைப்பும் சில சமயங்களில் பொருந்தவில்லை. பொது முன்னுரையில் 39 பாடங்களைப் பற்றிய சுருக்கமும் அளிக்கப்பட்டிருக்கலாம். எனினும் எழுத்துகள் அறிமுகமும் எழுத்துக்களை மனதுக்குள்ளும் வாய்விட்டும் படித்தல் என்பதும் அடிப்படை நிலையில் தமிழ் கற்றலுக்குப் பொறதும் துணை செய்கின்றன. தெளிவான பெரிய வண்ணப்படங்கள் பொருத்தமான அசைவுகளுடன் மேலும் மெருகூட்டப்பட்டால் இப்பயிற்சி சிறப்பானதாக அமையும்.

இணைய வகுப்பறையும் கற்போர் உளவியலும்

கற்றலில் கற்கும் பொருள், கற்போரது மனப்பான்மை, கற்போரது ஆh;வம், நாட்டம், விளைவைப் பற்றி அறிந்திருத்தல், கற்கும் பொருளின் அளவு, சிக்கல், நினைவு வீச்சு, இடைவிட்டுக் கற்றல் ஆகியன முக்கியத்துவம் பெறுகின்றன. தாண்டைக் என்னும் உளவியல் அறிஞர; ஆயத்த விதி, பயிற்சி விதி, விளைவு விதி ஆகியவற்றை ஆசிரியா;கள் நன்கு அறிந்து கற்பிக்க வேண்டும் என்கிறார். மாணவனின் உடல், உள்ளம் இரண்டின் முதிர்ச்சி அடிப்படையில் பாடங்கள் அமைய வேண்டும். தூண்டல் துலங்கல் கற்றலை வலுப்படுத்துகின்றன. இணைய வழிப்பாடங்கள் மாணவா;களின் தனியாள் வேற்றுமைக் கேற்ப வடிவமைக்கப்பட வேண்டும். மீத்திறமிக்க மாணவா;கள் வேகமாகக் கற்றுவிடுவா; சராசரி மாணவா;கள் பின்தங்கியிருப்பா; இவா;களுக்கு பயிற்சி, மீள் பயிற்சி தேவைப்படும்.

ஸ்கின்ரின் கற்றல் விதிப்படி, இணைய வகுப்புக் கற்றலில் பயிற்சி, மீள் பயிற்சி, வெகுமதி அல்லது வலுவூட்டும் சொற்கள் ஆகியன கற்பித்தல் கூறுகளாக உள்ளன. கற்பித்தல் என்பது ஆசிரியருக்கும் மாணவருக்கும் இடையே ஏற்படும் இருவழித் தொடர;பாகும். பிளாண்டரினுடைய வகுப்பறை ஊடாட்டப் பகுப்பாய்வினை இணையக் கற்றலில் (Flander's Interaction Analysis) முழுமையாகப் பயன்படுத்த இயலுவதில்லை. இங்கு இணைய வழிக்கற்றலில் ஒருவழித் தொடர்பே உள்ளது.

இணைய வகுப்பில் ஆசிரியா; மட்டுமே பேசுகிறார். மாணவா;களின் உணர;ச்சிகளை ஏற்றுக்கொள்ளுதல், அவர;தம் கருத்துகளைத் தெளிவுபடுத்துதல், மாணவா; பேசுதல், மாணவா;கள் பேச்சைத் துவக்குதல், அமைதி அல்லது குழப்பம் ஆகிய கூறுகள் இணைய வழிக்கற்றலில் இடம் பெறுவதில்லை. ஆனால் மாணவரது செயல்களையோ நடத்தையையோ புகழ்தல், வலுவூட்டம் செய்தல், வினா எழுப்புதல், மாணவா; ஏற்கும் வகையில் அதைச் செய், இதைச் செய் என்று கட்டளை இடுதல் ஆகிய நிகழ்வுகள் கற்றல் கற்பித்தலைப் பொருளுடையதாக ஆக்கி இணைய வகுப்பறையை ஓரளவு இயல்பான வகுப்பறைச் சூழலுக்கு இட்டுச் செல்கின்றன.

வகுப்பறைச் சூழலை ஆசிரியா; மற்றும் மாணவா; இடைவினைகள், மனப்பான்மைகள் ஆகியவற்றை அளவிடுவதன் மூலமே கண்டறிய இயலும். இணைய வகுப்பறைச் சூழல் ஆசிரியரின் ஆதிக்கம் நிறைந்த சூழலாகவே உள்ளது. ஆசிரியா; மாணவா; இணைந்து செயல்படும் சூழலுக்கு அதிகம் வாய்ப்பளிப்பதாக இல்லை.

கற்போரின் தேவைகளைப் புரிந்து கொண்டு இணைய வகுப்பறைக் கற்றலுக்கான பாடத்திட்டங்களைக் கற்றல் கொள்கைகள் அடிப்படையில் அமைப்பதும், உளவியலறிஞரின் துணைக்கொண்டு செயற்படுத்துவதும் இன்றியமையாதது. இணைய வகுப்பறை மூலம் தமிழைக் கற்றுக்கொடுக்கும் பொழுது தனியாள் வேற்றுமைகளுக்கு முக்கியத்துவம் அளிக்க வேண்டும். எனவே, மீத்திறமிக்க மாணவா;கள், மெதுவாகக் கற்போர், பின்தங்கியோர் ஆகிய மூன்று நிலை மாணவா;களையும் கவனத்தில் கொண்டு கற்பித்தலை மேற்கொள்ள வேண்டும். ஆசிரியாறன் கற்பித்தலுக்குத் துணை செய்யும் வண்ணம் கற்றல் - கற்பித்தல் துணைக்கருவிகளைப் பயன்படுத்துவது கற்பித்தலைத் திறன் உடையதாக ஆக்கும்.

Preparing pedagogy for E-learning courses

A pilot plan for Tamil Nadu

Dr. R. Natarajan
(Visiting Professor, IIPM, Chennai)
16/2 Jagadambal Street, T. Nagar, Chennai 600 017, India
Tel: + 9144 - 2815 1160 Mobile: 9841036446
E-mail: hindunatarajan@hotmail.com

Nowadays one hears such expressions as 'Education industry,' 'Education business.' Abhorring, but we have to grin and bear it. No other go. When education has become a big business proposition parents who cough up hefty fees want substantial return on investment. The current crop of students, familiar with computers even at primary level, can easily take to the state of the art teaching aids. However, merely installing computers in schools and colleges is not enough. The whole education system will have to go the e-way in the upcoming decades. There will be e-learning everywhere by the time this century draws its curtains.

At this juncture, E-learning presupposes E-teaching; hence it is incumbent on the academia to prepare E-teachers before launching E-learning in schools and colleges. Earlier supplementary attempts like occasional film shows, radio broadcast, UGC's TV telecast of lessons were attempted; but they were little efficacious. A centralized education telecast system was not effective for various reasons; the main reason being the tradition bound classroom togetherness of the teacher and the taught was not there.

However, the advent of computers replaced glass-slides as teaching aids for science subjects. Seminars and conferences have switched to power point presentations. The presentations have entered classrooms of management institutions. But all colleges offering MBA do not have teachers who use Power Point Presentation in classrooms. There is a clear rural-urban divide in the academia in using electronic teaching aids.

When video tapes came up, some academics wondered whether all education material could be packed into the new mode. Alas, the contemplation suffered infant mortality. Though video tapes had their role in the 1980s as entertainment sources, they did not find place in academia for teaching purposes.

With laptops booming, school and college fees soaring, teaching aids could as well be electronic now. E-learning is possible from primary to university courses. Possible, but can we take up right now? Why did the centralized teaching by broadcast, and UGC telecast fail? This question should open our eyes. That way of teaching was rigid by timing frigid by content. Gathering students at a place at a particular time to receive the centrally injected education was very difficult. It is so even now. But, with E-learning all students across the country could gain. E-tools can be livelier and personalized; hence students will welcome them.

Thus the computer era has accorded us enough scope for variety and flexibility to handle E-tools. Power Point Presentation of texts and visuals is quite handy, for the teacher and the taught. Teachers of the past who relied on chalk and talk, used to keep in mind all that they had to lecture in the classroom. Some teachers considered it beneath their dignity to carry cue cards. They loaded everything onto their mind. Some had hints on hand. Repetitive exercises made the teachers turn out like biped tape-recorders, except the creative lot among them who continued to enrich their knowledge by wider reading and fresh output in the classroom. They were very few.

The Power Point Presentation, I should say, helps the teacher first before it reaches the student, provided the teacher takes it right earnest, with all sense of creativity. Here is a rider. Before he clicks the slides and start explaining, the teacher should have done homework. Along side the PPP, Power Point Presentation, he should not be the fourth P – Parrot, just repeating what is on the slides.

If he has wide and sustaining reading habit, his presentation would be rich and different from others. A monotonous power point presentation will not enrich students in any way. PP has its limitations. Enrichment should come from the studious teacher. If the teacher, slack in avocation, just modifies the hard copy to a soft one, without applying his mind, the ideal pedagogy would be put to shame. In such a sorrowing situation neither the teacher nor the student gains anything. What could be an ideal situation of E-teaching vis-à-vis E-learning?

E-Text Book Societies

There is a Textbook Society in most states to help the government publish school textbooks. That is a governmental body. These committees should be reconstituted with a judicious mix of E-savvy young teachers and much experienced old timers. The committees should have experts for all subjects. The reconstituted E-textbook committees should have as many sittings as required to draw the course content and a basic power point program for all subjects, besides the requisite reference material.

Then teachers should be given orientation programs. They should be trained in the new methodology. The participants should be advised to follow the core-presentation model. But they can take creative deviation and help increase the up-take capacity of students. Here the individual's creative role also matters much. The world is not going to be same anymore and the academia will have far-reading changes very soon.

Stage I

The classrooms, in the initial stage, should be equipped with a screen and a projector. A white board can double up for this purpose. The newly trained teachers must use this facility. At this stage one cannot expect all students to use lap-tops. So, stage I is restricted to the E-savvy teacher. Students can take down the presentations and additional information provided by the teacher beyond Power Point Presentations. Stage I conceives E-tool as one of the factors and not as the absolute teaching aid. It matters little whether any student brings to the classroom – Laptops or not.

Stage II

Stage II envisages the classrooms being equipped with PCs. I would advocate a model that I saw recently at Hannan University, Osaka. The desk of the students has three PCs installed. The bench accommodates only two students. While the PC in the middle carries what the teacher projects on the screen, the other two are for the students. They see what is there in the PC in the middle and copy the same on to their PCs. Possibly they copy in the pen-drive also and do homework in their own systems.

They need not carry laptops to the classroom, enough if they carry a pen–drive; if needed, they can carry a paper file and needed books, just one or two.

Stage III

This is the total E-learning / E-teaching phase. This stage envisages all students carry laptops; each desk is provided with cable consoles for instant copying of what is projected. The students are obliged to listen to the teachers absorbingly and learn their lessons. Back home they click their computers and revise their course content. Those who browse find some individuals and groups offering E-learning packages. It is only at the nascent stage and the prompters will consolidate themselves by trial error methods in content development and market share.

However, we have to accord welcome to the initial enthusiasm. It is laudable. How far these packages will be useful, or would they turn just money-spinners will have to be ascertained later. There is no statutory body now to rate and regulate these e-learning service providers. I consider it is the duty of the Government and the NGOs to regulate such private offers to create joint an E-repository.

Stage IV

Examinations could also be conducted the e-way. Question papers could be flashed on the screen. Students can key in answers in their systems for mailing to the central system where the teacher can evaluate answers and award marks.

As a teacher, who defected to other walks of life and then returned to teaching after three decades, I wish to insist on infusing practical bearings on pedagogy by training the teachers first. The current psychological quotients in teacher education courses should stay on; but the new teacher education courses should inculcate all aspects of E-teaching. Brilliant persons should be drafted to teaching profession at the e-turn. They should be motivated to innovate and should be engaged to keep on updating.

When introduced extensively E-teaching can eliminate private tuitions. E-tools offer education at home. Once we introduce as a pilot project in specific locations, E-learning could be extended to almost everywhere. Here again, I wish to state that E-teaching should be accorded priority.

E-teaching, above all, will revolutionize the tradition bound paper-based, postal delivery linked distance education realm. The old system keeps really both the lessons and the students at distance, besides the teacher. The new correspondence courses, fully relying on E-tools will be a boon to distance education students; no hassles in getting by snail mail text book stuff as a torn bunch of papers, after inordinate delay.

E-learning and E-teaching will re-write the teaching/learning methodology in schools, colleges and distance education provided the money of the Education ministry and mettle of the academics joins hands in molding the future generation that is familiar with computers even from school days.

With the support of the government, the NGOs and computer companies, it is easy to replace the black board, the chalk and talk. What is needed is the will of the rulers to revolutionize the education system. Let them give color TVs, before the elections. But let them give after the elections computers, not free but at subsidized price in the interest of the rising generation's extensive and effective education by the e-way.

இணையம் வழி மொழிக் கற்றல்-கற்பித்தலில் புதிய அணுகுமுறைகள்

முனைவர் சு. குழந்தைவேல் பன்னீர்செல்வம்
உதவிப்பேராசிரியர், கல்வியியல் துறை,
பாரதிதாசன் பல்கலைக்கழகம்இ திருச்சிராப்பள்ளி, தமிழ்நாடு.
e-mail : skpdiet@yahoo.com

முன்னுரை

மாறிவரும் உலகில் நாள்தோறும் ஏற்படும் மாற்றங்களில் அனைத்தும் அண்மையாய் உணர வைக்கிறது. வளர்;ந்து வரும் தொழில் நுட்பங்கள் தகவல் நுட்பங்கள் தகவல் தொழில் நுட்பத் துறையை மாற்றியமைத்து வருகின்றன. இந்த மாற்றங்களை தமிழ்மொழிக் கற்பித்லுக்கு எவ்வாறு உட்படுத்திக்கொள்ளலாம் என்பதைப் பற்றிய எனது கருத்துக்களை இக்கட்டுரை மூலம் விளக்க விரும்புகிறேன். உலகத் தமிழர்களை தமிழ்மொழியில் ஒருங்கிணைக்க இணையம் முதன்மையானதாக உள்ளது. இதனை மேலும் எளிமைப்படுத்தினால் பயன்பாடுகள் விரைந்து அதிகரிக்கும். உலகந்தழுவி வாழும் தமிழ் மக்கள் தமிழ்மொழியைக் கற்பதற்கும், தமிழர்களின் வரலாறு, கலை, பண்பாடு உள்ளிட்ட வாழ்வியல் கூறுகளைப் பற்றி அறிந்து கொள்வதற்கு தேவையான மாற்றங்களை அல்லது அணுகுமுறைகளை பின்பற்றிட வேண்டும்.

இணையத்தின் வழி தமிழ் கற்றல்

மொழிப்பாடம் கற்பது எளிதன்று. ஆனால் அதனை எளிதாகக் கற்கும் வகையில்;, ஆர்வத்தைத் தூண்டி பாடப்பொருள்களைஇ பாடக்கருத்துக்களை நினைவில் கொள்ளும் வகையில் இணையத்தில் உருவாக்கிட முடியும். தமிழ்க் கல்வியை – தமிழ் மொழிக் கல்வியை அரிச்சுவடி முதல் ஆராய்ச்சிப் படிப்புவரை வழங்கிட பல்லூடகத்தின் (multimedia) வழியே அசைவுப் படங்களுடன் அமைக்கப்படுவதால், கேட்டல் (listening), பேசுதல் (talking), படித்தல் (reading), எழுதுதல் (writing) - LSRW என்ற அடிப்படை மொழித்திறனை எளிதாகப் பெறமுடியும். காரணம், கற்றல்-கற்பித்தலுக்கான பாடங்கள் பல்லூடக வசதிகளால் அச்சுவடிவம், ஒலிவடிவம், ஒளிவடிவம், மூலமாக ஆர்வத்துடன் எளிய, இனிய முறையில்; கற்றுக்கொள்வர். இவைபோன்றே உயர் கல்விப் பாடத்திட்டத்தின் மூலம் தாய்மொழி, பண்பாடு, வரலாறு, கலைகள் உள்ளிட்ட வாழ்வியல் கூறுகளையும், கோட்பாடுகளையும் அறிமுக நிலையிலிருந்து ஆராய்ச்சிநிலை வரையில் பல்வேறு பிரிவுகளில் பாடத்தினை வடிவமைத்திடலாம். இப்பாடங்களில் ஏற்படும் ஐயங்களைப் போக்கிக்கொள்ள பாட ஆசிரியர் களுக்குரிய திட்டங்களும் இணையத்தில் கொண்டுவரப்பட வேண்டும்.

தமிழ் மொழிக் கற்பித்தலில் எளிமையை உள்ளடக்கிய மாற்றங்களை உலகத் தமிழாசிரியர்கள் அறியும் வகைகளில் ‘கற்றல் - கற்பித்தலில் இணையம்’ என்ற தனியொரு இணையத்தை உருவாக்கிடலாம். இது அயல் நாடுகளில் வாழும் மொழியியல் வல்லுநர்களுக்கும் மொழியாசிரியர்களுக்கும் உதவியாக அமையும். வளர்ந்து வருகின்ற கணினியுலக நுட்பங்களை தமிழ்மொழியின் கற்றல் - கற்பித்தலுக்கு பயன்படுத்திக் கொள்ள வேண்டும். மிக வேகமாக வளர்ந்துவரும் e-learning துறையை இன்றையச் சூழலில் தமிழ்மொழிக் கற்பித்தலுக்குரிய நுட்பங்களை பயிற்சிகள் மூலம் ஆசிரியர்களுக்கு தரப்பட்டு, தரப்படுத்திய கற்பித்தலுக்கு (Standardized Teaching) தயார் செய்ய வேண்டும். இதன்மூலம் இணையத்தின் அனைத்து வசதிகளையும் தமிழுக்குப் பயன்படுத்திக் கொள்ளும் வாய்ப்பைப் பெறலாம்.

தமிழ்க் கற்றல்- கற்பித்தலைப் பொறுத்த வரையில் உலகளாவிய ஒரு முயற்சியாக அல்லது திட்டமாக இன்னும் வரவுமில்லைஇ வளரவுமில்லை. புலம் பெயர்ந்த தமிழ்களை இணைக்கும் வகையிலும் தமிழ்மொழியை உலகளாவிய முறையில் கற்கவும் - கற்பிக்கவும் (Universal Teaching-Learning Process) எந்தெந்த உத்திகளைப் பயன்படுத்தலாம் என்பது பற்றியும் தமிழறிஞர்களும் மொழியியல்

வல்லுநர்களும் ஆசிரியர்களும் கலந்தாய்ந்து ஒரு அமைப்பினை உருவாக்கிஇ எந்த நிலையில் கற்க விரும்புகிறார்கள் அதற்கான பாடங்கள் எப்படி அமைய வேண்டும் என்பதை கலைத்திட்ட (Curriculam) வடிவமைப்பின்போது கவனத்தில் கொள்ளவேண்டும்.

தமிழ் மொழியை தமிழ்நாட்டைத் தவிர்த்து இரண்டாம் மொழியாகக் கற்கும் நிலையிலுள்ள மாணவர்களுக்கு மொழித்திறன்களை வளர்த்துக்கொள்ளும் வாய்ப்புகளை

1. http://www.southasia.upenn.edu/tamil
2. http://www.tamil.net/projectmadurai
3. http://wwwtamil-heritage.org
4. http://www.tamil.net
5. http://www.tamil.org

போன்ற இணையத் தளங்கள் வழங்குகின்றன.

கல்வித் திட்டத்தில் இணையவழிக் கல்வியை மாணவர்களுக்கு அளிப்பதால் அவர்களது சிந்தனையாற்றல் வளர்ச்சியடைந்து மாறிவரும் தொழில்நுட்பத்திற்கு தங்களை வலுப்படுத்திக் கொள்ள முடிகிறது. இதற்கு Artificial Intelligence (AI) என்று சொல்லப்படும் தொழில்நுட்பத்தை கற்றல்-கற்பித்தலில் நடைமுறைப்படுத்திட வேண்டியது காலத்தின் தேவை. கல்வியியல் வல்லுநர்களைக் கொண்ட அமைப்பால் மேம்படுத்தி வரையறுக்கப்பட்ட கற்றல் - கற்பித்தலில் ஏஐ இணைக்கப்பட வேண்டும். இந்த முயற்சிகளை கல்வி நிறுவனங்களும் பாடத்திட்ட மேம்பாட்டு மையங்களும் ஆய்வு செய்து பள்ளிகளில் நடைமுறைப்படுத்திடலாம். இணையம் வழியே கற்றல்-கற்பித்தலில் வளர்ந்த நாடுகளைப் போல வளரும் நாடுகளிலும் பரீட்சார்ந்த முறையில் நடைபெற்று வருகிறதென்றாலும் இந்த (AI) தொழில்நுட்பக் கூறுகளை எந்த நிலையில் பாடத்திட்டத்தோடு இணைக்கப்படலாம் என்று ஆய்வுகள் செய்யப்பட்டு வருகின்றன. கணினி தொழில்நுட்பம் சார்ந்த கற்றல் மாணவர்களின் அறிவு வளர்ச்சியை (Logistic Development) வளர்ப்பதில் முக்கிய பங்காற்றுகிறது. மாணவனின் அடிப்படை அறிவு, திறமை இவற்றின் அடிப்படையில் மாறுபட்ட பயிற்சிகளை ஏஐ நுட்பங்கள் மூலம் கற்றல் - கற்பித்தல் நடைபெறுவதால் மாணவன் புதிய கருத்துக்களை செய்திகளைப் பெறுகிறான்.

புதிய முறைகள்

1. கற்றல்-கற்பித்தலில் ஒரு மாணவனின் கற்றல் போக்கினையும் அவனது சிந்தனைகளையும் ஒரு ஆசிரியர்; பலவிதமான கோணங்களிலிருந்து ஆய்ந்து கற்றல் முறைகளை உருவாக்குவதற்கு Modularity என்று பெயர். Modularity என்பது ஒரு செயலை செய்வதற்கு மூளையின் ஒவ்வொரு சிறு சிறு தனிப்பகுதியும் ஒன்றோடொன்று இணைந்து செயல்படுவதாகும். ஒரு மாணவன் ஒரு பயிற்சியை செய்யும்போது அம்மாணவன் எளிய வழியில் எவ்வாறு மாறுபட்ட நுட்பங்களைக்கொண்டு பெறமுடியும் என்பதை அடிப்படையாகக் கொண்டதே Modularity தத்துவம்.

2. மாணவர்களின் கற்றல் சிதைவதற்கு கவனச் சிதைவும் ஒரு காரணமாகும். கவனம் (Attention) குறையும்போது அல்லது சிதறும்போது பெறப்படும் தகவல்களை மூளையில் ஆழமாக (Long term memory) பதிய முடியாது. இதனால் கவனத்தை சீர் செய்யவும் நிலை நிறுத்தவும், டீழை கநநன டியஉம என்ற முறையை சில நாடுகள் செய்து வருகின்றன. இதற்கு குறிப்பிட்ட வன்பொருட்கள் தேவை. கணினியோடு இணைக்கப்பட்ட ஒலிவாங்கியை தலையில் அணிந்து கொண்டு கற்றலில் ஈடுபடுவர். மூளையின் அலைவரிசையில் ஒலிவாங்கி பொருத்தப்பட்டுள்ள கணினிக்குத் தெரியப்படுத்தவும். மாணவனின் கவனம் மாறும்போது கணினியில் அந்தத் தகவல்தெரியும். கவனம் சிதறும்போது மாணவனே தன்னை சோதித்து கவனம் மாறாது கற்றலில் Lgl Bio-feed Back முறை துணை செய்கிறது. சிந்தனைத் தூண்டலை வளர்க்கும் பாடங்களைப் போதிக்கும் மென்பொருட்களை உருவாக்கி அதனை

மாணவர்களுக்கு கற்றலின்போது அளித்திட வகைசெய்யும் கட்டகத்தை (Module) உருவாக்கிட வேண்டும். இதற்கு உலகின் கல்வித்துறையைச் சார்ந்த வல்லுநர்கள் ஒருங்கிணைந்து இந்தப் பணியைச் செய்ய கேட்டுக் கொள்ளலாம்.

3. இணையத்தில் கட்டுரை, மொழிப்பயிற்சி, கருத்தறிதல், இலக்கியம் போன்ற பாடங்களை தேர்வு செய்யப்பட்டு வகுப்பு வாரியாக கணினியில் TSC எழுத்துருக்களால் பாடங்கள் அச்சிடப்பட்டு படங்களும் வண்ணங்களும் சேர்க்கப்பட்டு பல்லூடகப் பாடங்களாக வடிவமைக்கப்படும்.

இவ்வாறு தயாரிக்கப்பட்ட இணையப் பாடங்களை ஒரு கணினி நிறுவனம் இயங்கு எழுத்துருக்களாக (dynamic fonts) மாற்றியமைக்கும். இவ்வாறு அமைப்பதால், மாணவர்கள் தமிழ் எழுத்துருக்களை காண்பதில் சிக்கலிருக்காது. இதன்பின் இணையப் பக்கத்தை மாணவர்கள் தங்கள் வீட்டிலிருந்தவாறு காணமுடியும். பாடங்களைக் கற்பித்தல், குறிப்புகள் வழங்குதல், பயிற்சிக்கான வினாக்கள் அனைத்தும் கணினியின் இணைப்பக்கத்திலேயே அமைந்திருக்கும். அவற்றை மாணவர்கள் தமக்குகந்த நேரத்தில் கற்க முடியும். இதில் ஏற்படும் ஐயங்களைப் போக்கிக் கொள்ள மாணவர்கள் ஆசிரியருடன் மின்னஞ்சல் மூலமாகத் தமிழிலேயே தொடர்பு கொள்ளலாம்.

4. அறிவியல் என்பது எவ்வாறு அனைவருக்கும் பொதுவானதோ அதுபோல கணினித் தமிழும் ஒன்றுபோலப் பயன்படுத்தப்பட வேண்டும். ஆங்கில மென்பொருட்களில் உலகளவில் பயன்படுத்தப்படும் File, Edit, View, Format, Print, Save, Exit, Copy, Paste, Cut, Cancel, OK போன்ற சொற்களுக்கு தாங்கள் சரியென்று கருதும் தமிழ் சொற்களைப் பயன்படுத்துகின்றனர். உலகளாவிய அளவில் பயன்படுத்தப்படும் மென்பொருள்களை தமிழில் அனைவராலும் ஏற்றுக்கொள்ளத் தக்க வகையில் தரப்படுத்தப்பட்ட சொற்களைப் பயன்படுத்த வேண்டும். கணினித் தொழில்நுட்ப வசதியை வளமான கல்விச் சூழலாக்க தேவையான வழிகளை மேற்கொள்ள வேண்டும். இதனைப் பற்றி ஆசிரியர்கள், தொழில்நுட்பங்ளை திட்டமிடும் வல்லுநர்கள், ஆசிரியர்கல்வி நிறுவனங்கள், கல்வியாளர்கள் ஆகியோர் ஒன்றிணைந்து ஒரு பொது செயல்திட்டத்தின் மூலம் (Common Minimum Programme) தரஅளவுகளை அமைக்க வேண்டும். இவ்வாறு அமைக்கப்படும் தொழில்நுட்ப கல்விச்சூழலால் கற்றல்-கற்பித்தல் செயலி;ல் முழுமை பெற முடியும்.

5. இன்றையச் சூழலில் மாணவர்களுக்குத் தேவையான திறன்களையும் சிக்கல்களுக்கான தீர்வுகாணும் விதிமுறைகளையும் பெறுகின்ற கல்விச் சூழலை மாணவர்களுக்கு வழங்க வேண்டும். இதற்கு கணினிக் கல்வியின் தர அமைப்புகளை கீழ்க்கண்டவாறு பிரிக்கலாம்.

- கணினி தொழில்நுட்பங்களையும் செயல்முறைகளையும் மாணவர்கள் புரிந்து கொள்ளுதல்
- தொழில்நுட்பத்துடன் தொடர்புள்ள சமூக பண்பாட்டு பிரச்சினைகளையும் மாணவர்கள் புரிந்து கொள்ளுதல்
- கற்றல் மற்றும் படைப்பாற்றல் (Creative Skill) திறனை வளர்த்துக் கொள்ளல்
- கணினித் தொழில் நுட்பத்தைப் பயன்படுத்தி எதிர்கால ஆய்வுகளை செய்தல்
- தேவைக்கேற்ப தொழில்நுட்பப் புதுமைகளை மேற்கொள்ளுதல்
- நடைமுறை வாழ்க்கையில் ஏற்படும் சிக்கல்களுக்கு விடை காணுதல்.

மேற்கூறியவற்றை உள்ளடக்கிய தொழில்நுட்பச் சூழலுடன் பாடத்திட்டத்தை வடிவமைக்கப்பட்டால் பல்வகைப்பட்ட மாணவர்களின் தேவைகளையும் நிறைவு செய்வதாக அமையும்.

6. பாடத்திட்டத்தைப் பின்பற்றிய பாடநூல்களும் அவற்றைக் கற்பிக்கும் ஆசிரியர்களுக்குரிய பயிற்றுமுறைகளையும் வழிகாட்டி கட்டகங்களையும் வெளியிட வேண்டும். இவை மூன்றும் முக்கியத் தேவைகளாகும்.

7. கணினியில் புலமை பெற்ற ஆசிரியர்களால் மட்டும் இணையத்தில் பாடங்களைத் தயாரித்துக் கல்வித்

தளங்களைச் செயல்படுத்திட்ட நிலைமை மாறி தற்போது Hot Potatoes என்ற மென்பொருளின் மூலம் கணினியில் அடிப்படைத்திறன் பெற்ற யாராலும் சிறப்பாக ஊடாட்டு (Templates) பயிற்சிகளைத் தமிழில் தயாரிக்க முடியும். Half Backed Software என்ற நிறுவனத்தினரின் இந்த மென்பொருள் மூலம் இணையம் சார்ந்த மொழிப் பயிற்சிகளைத் தமிழிலேயே உருவாக்க முடியும். இயங்கும் மென்மொழிகளாக Hot Potatoes இருந்தாலும் படிம அச்சுக்களைக் (Templates) கொண்டு மொழிப் பயிற்சிகளை எளிதில் தயாரித்து இணையத்தில் கொண்டு வரலாம். தமிழின் எதிர்காலத்தை தமிழரின் எதிர்காலத்தை தீர்மானிக்கக்கூடிய ஒரு பெரிய சக்தியாக இன்றைக்கு இணையம் வளர்ந்துள்ளது. ஆங்கிலத்திற்கு அடுத்த தமிழில்தான் அதிகமான இணைப்பக்கங்கள் உள்ளன என்பது தமிழுக்கு பெருமை.

முடிவுரை

மொத்தம் 7.5 கோடி தமிழர்களில் 20 சதவகிதம் பேர் அதாவது 1.5 கோடி பேர் தமிழக எல்லைக்கு வெளியே வாழ்கிறார்கள். இந்தியாவில் தமிழ் ஒரு மாநில ஆட்சிமொழி மட்டுமே. ஆனால் இலங்கை, சிங்கப்பூர் ஆகிய நாடுகளில் தமிழ் தேசிய ஆட்சி மொழி. மலேசியா, மொரீஷியஸ், ஃபிஜி, தென்அமெரிக்கா போன்ற பல நாடுகளில் அங்கீகரிக்கப்பட்ட மொழி. தமிழின் இந்த உலகத் தகுதியை நாம் காப்பாற்ற வேண்டும். அவ்வாறு காப்பாற்ற வேண்டுமானால் புலம் பெயர்ந்த அயலகத் தமிழர்கள் தமது தாய்மொழியோடு தொடர்புள்ளவர்களாக இருக்க வேண்டும்.

உலகளவில் பயன்பாட்டிலுள்ள மொழி என்ற பெருமையைப் பாதுகாக்கவும் வலிமைப்படுத்தவும் உலகத் தமிழ்களனைவரும் தமிழைப் பயன்பாட்டிற்கு தொடர்ந்து கொண்டு வரவேண்டும். இதற்கு தமிழ் கற்பது எளிதாக்கப்பட வேண்டும். நமது ஆசைகள் பெரியதாக இருக்க வேண்டும், நமது கனவுகள் பெரியதாக இருக்க வேண்டும். நாம் உலகுதழுவி வாழும் மொழிக்குடும்பம் என்ற பெருமைகொள்ள வேண்டும். அதற்கேற்ப நமது அணுகுமுறை, கல்வித் திட்டங்கள் செயல்படுத்தும்முறை மாற வேண்டும். விரிந்தப் பார்வையில் இந்த வையகத்தை ஆள வேண்டும்.

மேற்கோள் நூல்கள்

1. *Hoven D. (1999). A. Model for listening and viewing comprehension in multimedia environments,language learning&Technology,3(1),88-103. http//iit.msu,edu/vo/13num/ hoven.*
2. *Jones, L. & Plass, J. (2002) suporting listening comprehensin and Vocabulary acquisiton in french with multimedia annotations. The modern language journal, 86 (4), 546-561.*
3. *Learning theories : Ar. Educational perspective, Date H.Schunk, macmillan publishing company, 1991*
4. *Krishnaswamy (1992).upanayan : A Programme for the Developmental Traning of Children with mental Retardation. Action Aid disability news, 3 (2) 42-43.*
5. *Farrel (1999) the development of virutal education : A Global perspective, Vancouver, Canada : the commonwealth of learning*
6. *Tamil Salmi (2000) Higher education : facing the challenges of the 21st century, Technologia, Jan / Feb2000*
7. *The child and Reality, problem of Genetic psychology, Jean piaget penguine books, 1976.*http://www.techlearning.com/db_area/archives/TL/2002/11/topten5.html.

Effectiveness of Multimedia Package in Learning Vocabulary in Tamil

Dr.G.Singaravelu
Reader,UGC-Academic Staff College,
Bharathiar University,Coimbatore-641 046.Tamilnadu
email: singaravelu.bu@gmail.com

Introduction

Learning vocabulary is essential to develop communicative skill of any language and it is a backbone of the language. To develop Tamil language, young learners should acquire thousand vocabularies. Present methods of teaching vocabulary in Tamil are not fruitful to the young learners to improve their competencies in vocabulary of Tamil. Special innovative method can be supported to the young learners acquiring more vocabularies for suitable communication transactions in Tamil. The researcher endeavored to prepare a package for acquiring more vocabularies in Tamil for the young learners at standard V. The study enlightens the effectiveness of Multimedia Package in Learning Vocabulary in Tamil at standard V.

Objectives of the study

1. To find out the problems of conventional methods in learning vocabulary inTamil.
2. To find out the significant difference in achievement mean score between the pre test of control group and the post test of control group.
3. To find out the significant difference in achievement mean score between the pre test of Experimental group and the post test of Experimental group.
4. To find out the impact of Multimedia package in Learning Vocabulary in Tamil at standard V.

Hypotheses of the study

1. Learners of standard V have problems in learning vocabulary in Tamil.
2. There is no significant difference in achievement mean score between the pre test of control group and the post test of control group.
3. There is no significant difference in achievement mean score between the pre test of Experimental group and the post test of Experimental group.
4. Multimedia package is more effective than conventional methods in Learning Tamil Vocabulary at standard V.

Variables

The independent variables namely Multimedia package and the dependent variable namely achievement score were used in this study.

Delimitations of the Study

The responsibility of the researcher is to see that the study is conducted with maximum care in order to be reliable. However, the following delimitations could not be avoided in the present study.

1. The study is confined to 60 students of standard V studying in primary school, Pulluvapatti, Coimbatore.
2. The study is confined to learning Tamil vocabulary of the state board text book

Methodology

Parallel group Experimental method was adopted in the study.

Sample: Sixty pupils of studying in standard V from Panchayat Union Primary school, Polluvapatti, Coimbatore were selected as sample for the study. Thirty students were considered as Controlled group and another thirty were considered as Experimental group.

Tool: Researcher's self-made achievement test was used as a tool for the study. An achievement test consisted of fifty questions

Construction of tools

The investigator's self made Achievement test was used for the pretests and post tests of both control groups and experimental groups. The same question was used for both pre and post tests to evaluate the pupils' skills of vocabulary in Tamil through objective types of question which carried one mark for each question and contained 50 marks.

Pilot study

In order to ascertain the feasibility of the proposed research and also the adequacy of the proposed tools for the study a pilot study had been undertaken. During the pilot study, the problem under study had been finely tuned. Sufficient number of model question papers were prepared and distributed to 10 students of standard V in Panchayat Union Primary school, Polluvapatti,Coimbatore for the pilot study. This exercise was repeated twice over two sets of 10 students each. The clarification raised by the students was cleared then and there and the filled answer scripts were collected by the researcher. These students were selected in such a way that they were not part of either the control group or experimental group.

Reliability of the tool

A test is reliable if it can be repeated with a similar data set and yields a similar outcome. The expectation of a good research is that it would be reliable. It refers to the trustworthiness or consistency of measurement of a tool whatever it measures. Under this study the reliability had been computed using test-retest method and the calculated value comes to 0.84. The value is quite significant and implies that the tools adopted were reliable. Hence the reliability was established for the study.

Validity of the tool

The concept of validity is fundamental to a research result. A result is internally valid if an appropriate methodology has been followed in order to yield that result. A test is said to be valid if it measures what it intends to measure. The expert opinion of the co staff was obtained before freezing the design of the tools. Subject experts and experienced teachers were requested to analyse the tool. Their opinions indicated that the tool had content validity.

Procedure of the study

1. Identification of the problem by administering pre-test to the both groups.
2. Planning.
3. Preparation of package.
4. Execution of activities through using the package.
5. Administering post-test.

Data collection

The researcher administered pretest to the pupils with the help of the teachers. The question paper and response sheets were given to the individual learners and collected and evaluated learning obstacles of the learners were identified by the pretest. The causes of low achievement by unsuitable methods were found out. Multimedia package used in the classroom for learning vocabulary for one week. The post test was administered and the effectiveness of the Multimedia package was found.

Data analysis

Statistical technique **t** test was applied for the study.

Hypothesis Testing

Hypothesis 1

Students of standard V have problems in learning Vocabulary in Tamil at Panchayat Union Primary school, Polluvapatti, Coimbatore. In the pre-test, students score 32% marks in learning Tamil vocabulary through conventional method and the Experimental group students score 68% marks. It shows that Students of standard V have problems in learning Vocabulary in Tamil at Panchayat Union Primary school, Polluvapatti, Coimbatore.

Hypothesis 2

There is no significant difference between the pret test of control group and post test of control group in achievement mean scores of the pupils in learning Vocabulary in Tamil at standard V in Panchayat Union Primary school, Polluvapatti, Coimbatore.

Table -1 *(achievement mean scores between pre test of control group and posttest of Control group)*

Stages	N	Mean	S.D.	df	t- value	Level of significance
Pretest control group	30	45.60	4.454	58	1.73	P<0.05
Post test control group	30	47.60	4.492			

The calculated 't' value is (1.73) greater than table value (2.00). Hence null hypothesis is accepted at 0.05 levels. Hence there is no significant difference between the pre test of control group and post test of control group in achievement mean scores of the learners in learning vocabulary in Tamil.

Hypothesis 3

There is no significant difference between the pre test of Experimental group and post test of Experimental group in achievement mean scores of the pupils in learning vocabulary in Tamil.

Table-2 *(achievement mean scores between pretest of Experimental group and posttest of Experimental group)*

Stages	N	Mean	S.D.	df	t- value	Level of significance
Pretest Experimental group	30	50.30	5.04	58	22.71	P>0.05
Post test Experimental group	30	85.63	6.61			

The calculated 't' value is (22.71) greater than table value (2.00). Hence null hypothesis is rejected at 0.05 levels. Hence there is significant difference between the pre test of Experimental group and post test experimental group in achievement mean scores of the learners of Tamil in vocabulary.

Hypothesis 4

Learning vocabulary by using Multimedia Package is more effective than existing methods.

Achievement mean scores of the learners in post-test of control group is 47.60 and the achievement mean scores of the learners post test of Experimental group is 85.63.Score of the post test of Experimental group (85.63) is greater than Pre test of Experimental group(50.30) It shows that learning vocabulary by using Multimedia Package is more effective than conventional methods.

Findings

1. In the pre-test, students score 32% marks in learning Tamil vocabulary through conventional method and the Experimental group students score 68% marks.It shows that Students of standard V Panchayat Union Primary school, Polluvapatti,,Coimbatore have problems in learning Tamil vocabulary through conventional method.
2. There is no significant difference between the pre test of control group and post test control group in achievement mean scores of the pupil of standard V in learning Tamil vocabulary through Multimedia Package at Panchayat Union Primary school, Polluvapatti, Coimbatore.
3. There is significant difference between the pre test of Experimental group and post test of Experimental group in achievement mean scores of the pupils in learning Tamil vocabulary.
4. Learning vocabulary in Tamil by using Multimedia Package gave significant improvement.

Educational Implications

1. Using Multimedia Package learning different subjects can be extended to primary level, secondary level and higher secondary level.
2. It can be encouraged to implement to use in adult education
3. It may be implemented in teachers education
4. It may be implemented in alternative school
5. Slow learners can improve by using it
6. It may be more supportive to promote Sarva Siksha abiyan in grass root level.

Conclusion

The study reveals that Students of standard V in Panchayat Union Primary school, Polluvapatti, Coimbatore have problems in learning Tamil vocabulary through conventional method. Learning vocabulary in Tamil through Multimedia package is more effective than conventional methods. Hence it will be more supportive to enrich vocabulary in Tamil at primary education.

References

1. **Geetha.T.V and Rajan parthasarathy**, Multimedia Chemistry in Tamil for X standard Students.
2. **James.E Shuman and Thamson wadsworth**(1988) Multimedia Action.
3. **Sampath.K,Paneerselvam.A and Santhanam.S(1998**), Introduction to Educational Technology,Sterling publishers Pvt Ltd.

Enhancing Learning of Tamil Language in a One-to-One Computing Environment

Sivagouri Kaliamoorthy
Beacon Primary School
sivagouri_kaliamoorthy@moe.edu.sg

Abstract: In recent years, there seems to be an upward trend of Indian pupils entering primary one who take Tamil as their Mother Tongue but come from non-Tamil speaking home environments. Pupils are found to be unable to effectively communicate their ideas and opinions in the language. Some even express fear and anxiety when asked to communicate their ideas in Tamil. This paper presents how technology can be leveraged in a one-to-one computing environment to enhance learning of Tamil language. In this environment, there is an eclectic blend of mastery driven approaches as well as constructivist pedagogies. In a ubiquitous computing environment, the teacher is able to tailor lessons and support pupils of varying abilities; thus scaffolding their learning to build their esteem and to eventually help them to gain confidence to communicate their ideas. This paper will show the strategies used in a technology-rich environment and the challenges faced by the Primary one and two classes to achieve the objectives.

Keywords: Integration of Technology, One-to-one computing

Introduction & Purpose

Recent statistics shows that there is a shift of Tamil language usage at home (Ministry of Education – Singapore, 2005). The survey data findings conducted in our school with Tamil pupils during the Primary 1 orientation in 2008 and 2009 also reflected similar trend with close to more then 40% of Tamil pupils coming from non Tamil speaking background. This implied a lack of authentic context of the usage of the Mother Tongue languages at home. As a result pupils faced communication problems both in written and oral presentation of ideas, constructing a grammatically correct sentence and using the language in a particular situation or context. With a greater emphasis in Standard Spoken Tamil pupils are challenged further in the appropriate contextual usage of Tamil language.

Background of One to One computing environment

The school in this research study is Beacon Primary School, one of the future schools under the FutureSchools@SG project jointly initiated by the local Ministry of Education (MOE) and the Infocomm Development Authority (IDA). Its primary purpose is to explore the possibilities of using and leveraging on information communication technologies (ICT) in the educational realm, especially in the area of Mother Tongue languages acquisition among young learners, aged 7 to 8. With this context in mind, series of lessons were designed and implemented emphasis on language building authentic activities with elements of play leveraging on information communication technologies (ICT). All Tamil pupils were given a laptop and are equipped with basic handling of the equipment. All Tamil pupils are taught how to use Microsoft PowerPoint, Microsoft Word and Photostory3 for Windows. The Tamil classroom is equipped with Promethean Interactive Whiteboard.

Studies have shown that ICT could be used to better engage learners (Fontana, Dede, White, & Cates,

1993; Herrington & Oliver, 1998; Jonassen, Peck, & Wilson, 1999; Sarapuu & Adojaan, 1999; Oliver & Hannafin, 2000; Jonassen, 2000; Jonassen & Carr, 2000; Hollingworth & McLoughlin, 2001; Kearney &

Treagust, 2001; Neo & Neo, 2001). Jonassen and Carr (2000) propose the approach of learning with technology where learners are actively involved in the construction of their own knowledge with the help of ICT tools. They propose that technologies could be used as mind tools for the construction of their knowledge and engaging learners in evaluating, analysing, connecting, elaborating, synthesising, imagining, designing, problem-solving, and decision-making.

ICT tools allowed learners to express their thought processes through multimedia presentations, that is, a consolidation of images, text, animation, and sound. Van Scoter (2004) advocates that digital images support language development. When young learners use ICT tools to tell stories they create with a combination of words and pictures, these stories present a wonderful opportunity for students to create an image with meaning for them. Haugland (1992) advocates that children using computers could gain intelligence, structural knowledge, long-term memory, manual dexterity, verbal skills, problem solving, abstraction and conceptual skills over those who did not use computers. The main idea is not to use the computer for itself but to include supporting activities that will allow for meaningful learning.

Rationale, Approach and Design

Rationale

Learning in complex and ill-structured knowledge domains requires accommodation of multiple perspectives embedded in authentic activities and the reconciliation of those perspectives with personal beliefs resulting in conceptual change. We reason that instead of merely flooding the pupils with vocabulary from anywhere, we are constructing knowledge and context through authentic activities. The authentic activities also included elements of play as a pedagogical tool.

Approach

A case study approach was used in this study to look into how authentic activities with elements of play and leveraging on the use of ICT to better engage pupils in learning of Tamil language. A case study approach is being used to better understand the impact and potentials of the strategies used in this study. Case study research is not sampling research and it is also not the primary intent of this study to understand other cases. According to Stake (1995), it may be useful to try to select cases that are typical or representative of other cases, but a sample of one or a sample of just a few is unlikely to be a strong representation of others. The most important criterion of using case study as a research method is to maximise what we can learn from this instance.

Design

The lessons are designed based on the three concepts of authenticity, learning with technology, and play as discussed above. Students were engaged in an authentic setting by playing. Using the experience and resources built during play (e.g., digital images and vocabulary), they created digital stories using ICT tools. A diagram depicting the basic lesson design flow and its stages is presented in Figure 1.

Figure 1 – Lesson Design and stages of implementation
Stage1 – Introduction to topic and vocabulary
Stage 2 – Authentic Learning Experiences with Elements of Play
Stage 3 – Creation of Digital Story (Multimedia)
Stage 4 – Presentation & Assessment
Stage 5 - Editing

Stage1 – Introduction to topic and vocabulary

The pupils were provided platform to enrich vocabulary by tying literacy with context using ICT tools. Examples of this strategy include the using of digital images, e-books, and online resources to build and understand the set of vocabulary used in the theme. Students were prompted to discuss about the topic. Figure 2 and 3 depict the usage of Big Books and Interactive White Board to engage pupils in the initial stage of this lesson design.

Stage 2 – Authentic Learning Experiences with Elements of Play

Pupils will go through an authentic experience or learning journey. These learning experiences help them internalise the information they gather and serve as a platform to verbalise their meaning making. Peer collaboration and interaction is a means for the pupils to articulate their thought processes.

Stage 3 – Creation of Digital Story (Multimedia)

Using the resources accumulated during the authentic activity, pupils to create digital stories. These stories are the outputs of their authentic learning experience.

Stage 4 – Presentation & Assessment

Pupils can present their creations in the following ways:

a) Pupils save their creations in the computer network shared folder for their peers to assess based on a checklist (see Annex 1 for details). Peers to write their feedback by ticking or crossing appropriate boxes with the criteria listed and provide feedback to their classmates.

b) Pupils save their wok in the computer network shared folder for teacher to assess the digital story based on a set of rubrics. Teachers to provide feedback for improvements.

c) Pupils present the digital story to the class. Teacher to ask questions to elicit response from the pupils to explain reasons behind the text, images or audio recorded. Teacher and peers to give feedback for further improvements to the story based on a checklist provided (see Annex 1 for details).

Stage 5 - Editing

Pupils take ownership in learning by editing after feedback was given by peers or teacher. Depending on the time frame, students may edit as many times as they want.

Research Methods

Pupils' Performance

A diagnostic test was conducted at the start of academic term to assess the reading, listening and speaking levels of the pupils. Pupils' performance was also examined using alternative assessment and their end-of-year oral assessment. The components assessed in alternative assessment included oral communication. Pupils' artefacts like the digital stories also provided a good platform to gauge the progress of their speaking skill. It was a good way for teachers to assess the use of vocabulary and the ability to synthesise images and ideas appropriately. It was observed that more than half of Tamil pupils were not confident to speak or were not fluent in the language. About 25% of the Tamil language pupils were not able to read fluently.

Teachers' Reflection Notes and Observations

Teachers' observations were recorded in their journals. The entries included anecdotes and reflections. Observation includes noting pupils' engagement level in the lessons and activities. The indicators for engagement were:

1) 85-100% active participation in group discussions hands-on activities;
2) the number of times students edit or re –record their digital stories;
3) the number of times students contributes an idea;
4) the number of times students ask each other or teacher to clarify their doubts;
5) the participation by students who were less responsive (quiet and shy pupils)

Pupils' and Parents' Surveys and Interview

Teachers conducted a survey to find the language spoken at home. This survey facilitated in understanding the home background and the comfort level of usage of Tamil at home. About more then 60% of the Tamil Language students spoke in Tamil respectively, yet they could not articulate fluently at the first diagnostic test. A pupil survey was also carried out to better understand pupils' interest and motivation of the lessons and activities. Pupils wrote their feedback on the activities they enjoyed best throughout the year. Pupils were also interviewed and parents given a survey on the impact of these activities on the pupils' oral skills.

Discussions of Findings

Authentic Activities

A series of authentic activities with real-world relevance, requiring pupils to examine them from a variety of perspectives, and with opportunities for collaboration were carried out. Pupils were brought to the Jacob Ballas Children's Garden (Singapore Botanical Gardens) to make comparison between their neighbourhood playgrounds with the garden which instil a care for nature. They used PhotoStory 3 for Windows to create their own digital stories. Pupils had an hands on experience making murukku, learning the Malay martial arts, Silat. The projects also required them to collaborate and work together. Although the end products may be done individually, but the accumulation of resources (e.g., digital images, vocabulary, peer editing) were done as a group.

Pupils' Engagement and Behaviour

The engagement level of pupils was notably high during the lesson activities was observed. Pupils were also observed to be more persistent as they recorded their readings many times trying to perfect

their end products. The peer evaluation process also provided the avenue for them to think through more deeply with their productions. Pupils were actively explored different ways to present their digital stories with technologies (e.g., the Tablet PC, presentation software, sound recording software). Pupils interacted in their Mother Tongue languages more frequently during their Mother Tongue classes. The self construction of the digital artefacts encouraged pupils to take more ownership of their learning. In addition, the number of tasks completed within the time given also increased. This was possibly due to the pervasiveness use of ICT tools to augment the learning of the languages. The skills acquired from one digital story to another also taught them to use one tool and adapt it into another context. The programme also realised that students learnt to work together. It was observed that they are more engaged when they work in groups. It was noted that the checklist and observations of each pupil gave them opportunity to value students' little progress. Shy students came out of their shell before the year end.

Learning Abilities

In order to bridge the different language abilities and needs, some groups were given additional time to complete the tasks and additional scaffolding. The tasks were tailored to meet average and lower ability students.

Feedback from Parents

According to the parents' survey, the frequencies of the two Mother Tongue languages being used at home increased. A parent reflected the following, "… We are using Tamil more often at home now as compared to before." Some parents reflected that they had been corrected by their children when they did not use Tamil correctly. A parent also reflected that her child had corrected the way she should pronounce the words in Tamil. The drama, show and tell, and storytelling sessions motivated the pupils to practice their lines at home with the family members. Some parents shared that these practice sessions helped them bond with their children. Pupils' survey showed that the students enjoyed the MT lessons. All the students requested for activities which involved use of more computer based activities in future.

Learning with Technology – Creation of Digital Stories

The process of the creation of digital stories allowed pupils to record their own voices when narrating their own scripts. The creation of digital stories places the technology in the hands of the learner and allowing the pupil to control its use within objectives that were constructed by the teacher. Hence, the creation of digital stories was a possible strategy that supported presentation and writing using ICT. Presentation and writing require skills like deciding goals, sequencing of ideas, composition of message and editing. Simple applications such as Microsoft PowerPoint and Photostory 3 were used for the creation. These software titles were easily available and widely used in the school. Digital story creation as an ICT-mediated strategy could enrich the classroom learning environment, the curriculum, and student learning experiences by providing an open-ended, creative and motivating productive tool in the classroom (Sadik, 2008). Pupils were observed to be motivated and excited in the use the ICT tools to develop their stories which they can relate to.

The element of play also provided an excellent vehicle for learning. Weininger (1978) emphasizes an inner reality (intellectual and emotional life) and an outer reality (world experiences) and the use of play to accommodate and connect these realities. This was evident in the digital stories created by the pupils. (Please elaborate on this point – very interesting if you can elaborate on this) Pupil leveraged

on ICT as an output platform to present each of their learning experience. The digital stories documented the rich experience they had during the play and revisited them to enhance on their projects. Assessment of the project facilitated the teachers in checking on the language literacy and provided the teachers with the pupils' progress.

Issues and Challenges

As with many strategies to learning the usage of ICT has its limitations and challenges.

Pupil ICT readiness

The initial phase of introduction to both hardware and software was challenging and time consuming. Thus, getting pupils on task using the computers was challenging. At Primary One, many were not familiar with the computer notebooks, let alone the other software titles and programs. However, the pervasiveness of ICT mediated lessons soon paid off when pupils become more skilful with each lesson. At times, the pupils may deviate from the task at hand and focus more on the less important features of the presentation. For instance, Microsoft PowerPoint is an easy and powerful to use for language learning. However, the choice given may be a disadvantage when students start to use too many fonts on one slide or spend more time on the graphics and transition motions than the language objective.

School infrastructure and support

ICT-mediated activities could consume many hours when it was an introduction to a new tool and when technical glitches disrupted the smooth running of the lessons. At times, dealing with network problems due to heavy traffic usage was overwhelming.

Conclusion and Recommendations

This study, though descriptive in nature, had shown that the Tamil pupils have been actively engaged in constructing their own knowledge of the Tamil language with the help of ICT tools *(Jonassen and Carr (2000).* Pupils have acquired basic competency in speaking, constructing simple sentences and communicating their ideas in Tamil language. Pupils who come from predominantly English speaking background shows promises of using the language at home. The authentic tasks enabled bonding between parents and child in completing the tasks effectively. As a future direction more authentic activities be introduced in school and laying the context for pupils to leverage on ICT tools to communicate the ideas.

References:

1. Allwright, D. & Bailey, K. M. (1991). *Focus on the language classroom: An introduction to classroom research for language teachers.* New York: Cambridge University Press.
2. Jonassen, D., Howland, J., Marra, R.M. and Crismond, D., (2008). Meaningful Learning with Technology, Pearson Prentice Hall, New Jersey, USA.
3. Lim, C.P. (2002). A theoretical framework for the study of ICT in schools: A proposal. *British Journal of Educational Technology, 33*(4), 415-426.
4. Lim, C.P. & Chai, C.S. (2004). An activity-theoretical approach to research of ICT integration in Singapore schools: Orienting activities and learner autonomy. *Computers and Education,43*(3), 215-236.
5. Yin, R.K. (1994). *Case Study Research: Design and Methods (2nd Edition)*. Thousand Oaks (CA): SAGE Publications.

பேச்சுத்தமிழைக் கற்பிப்பதில் கணினியின் பயன்பாடு

டாக்டர் ஆ ரா சிவகுமாரன்
இணைப் பேராசிரியர்
தலைவர் - தமிழ்மொழி பண்பாட்டுப் பிரிவு
ஆசியான் மொழிகள் மற்றும் பண்பாட்டுத்துறை
தேசியக் கல்விக்கழகம் -நன்யாங் தொழில்நுட்பப் பல்கலைக்கழகம்
சிங்கப்பூர் 637616

உலகின் பல நாடுகளிலும் தமிழர் பரந்துபட்டு வாழ்கின்றனர். அங்கெல்லாம் தமிழ்மொழி வழக்கில் உள்ளது. தமிழ்மொழி இரட்டைவழக்குச்சூழல் (Diglossic situation) கொண்டது.பேச்சுத்தமிழ், எழுத்துத்தமிழ் என இரண்டும் தமிழர்களின் மொழிச்செயல்பாடுகளுக்குப் பயன்படுகிறன. சிங்கப்பூர்வாழ் தமிழ் மாணவர்கள், இப்போது அன்றாட உரையாடல்களுக்கு ஆங்கிலத்தையும் எழுத்துத் தமிழையும் மிகுதியாகப் பயன்படுத்தும் வழக்கம் அதிகரித்துவருகிறது. அவர்களைத் தமிழில் - பேச்சுத்தமிழில் - பேசவைக்கும் முயற்சியில் பள்ளி ஆசிரியர்கள் மிகவும் ஈடுபட்டு வருகின்றனர்.

தமிழகத்தில் தமிழ்க்குழந்தைகள் பேச்சுத்தமிழைக் குழந்தைப்பருவத்தில் கற்றுக்கொண்ட (அல்லது 'பெற்றுக்கொண்ட') பிறகே பள்ளிக்குச் செல்கின்றனர். பள்ளியில் எழுத்துத்தமிழ் அவர்களுக்குக் கற்றுக்கொடுக்கப்படுகிறது. ஆனால் சிங்கப்பூரில் இதற்கு எதிர்மாறான சூழல் நிலவுகிறது. சுமார் 58 விழுக்காட்டுக் குடும்பங்களில் தமிழ்மொழி வீட்டில் பேசப்படாததால் தமிழ்க்குழந்தைகள் பேச்சுத்தமிழை அறியாமல் பள்ளிக்கு வருகின்றனர். சிங்கப்பூர்ப் பள்ளிகளில் சுமார் மூன்று ஆண்டுகளுக்கு முன்புவரை எழுத்துத்தமிழே மிகுதியாகக் கற்பிக்கப்பட்டது. பாடங்கள் எழுத்துத்தமிழை அடிப்படையாகக் கொண்டே இன்றுவரை அமைந்துள்ளன. எழுத்துத்தமிழைக் கற்றுக்கொள்கின்ற குழந்தைகள், பேச்சுத்தமிழைப் பயன்படுத்த சில முயற்சிகளை மேற்கொள்கின்றனர். அப்பொழுது அவர்களுக்கு எழுத்துத்தமிழின் செல்வாக்கு பேச்சுத்தமிழில் ஏற்படுகிறது .இதை எவ்வாறு தவிர்ப்பது? இதில் உள்ள சிக்கல்களைத் தீர்ப்பதில் கணினியின் பயன்பாடு என்ன என்பது பற்றியே இக் கட்டுரை விவரிக்கிறது. (இக்கட்டுரையில் கூறப்பட்டிருக்கும் செய்திகளைக் கணினியின் துணையோடு படிக்கும்போதே இக்கட்டுரையில் கூறப்பட்டிருக்கும் செய்திகள் நன்கு விளங்கும்)

எழுத்துத்தமிழுக்கும் பேச்சுத்தமிழுக்கும் இடையே உள்ள வேறுபாடுகள் சில குறிப்பிட்ட விதிகளுக்கு உட்பட்டே அமைகின்றன. இதுபற்றிய ஆய்வினை மேற்கொண்டவர்கள் ஏறத்தாழ 29 விதிகளை எடுத்துக்காட்டியுள்ளனர். பொதுவாக இவ்விதிகளுக்கு உட்பட்டு இயங்கும் வேறுபாடுகளை மாணவர்கள் அறிந்து கொள்வார்களானால் பேச்சுத்தமிழை அவர்கள் எளிதாகக் கற்றுக்கொள்ள இயலும். இவ்விதிகளுக்கு உட்பட்ட எடுத்துக்காட்டுகளை மாணவர்கள் (கேட்டல்-பேசுதல் முறையில் - Audio-lingual method) பலமுறை பயிற்சி செய்வதன்வழி எழுத்துத்தமிழிலிருக்கும் சொற்கள் பேச்சுத்தமிழுக்கு எவ்வாறு மாறுகின்றன என்பதை அறிவர்.

பொதுவாகப் பேச்சுத்தமிழுக்கும் எழுத்துத்தமிழுக்கும் இடையே ஏற்படும் வேறுபாடுகள் ஒலியனியலிலும் (Phonology) உருபனியலிலுமே (Morphology) பெரும்பாலும் அமைகின்றன.

விதி 1

"இடம்" என்ற எழுத்துத்தமிழ் எட(ம்) என்று பேச்சுத்தமிழில் மாறுகிறது. இச்சொல்லில் இ என்ற ஒலி எ என்று மாறுவதைக் கீழ்க்கண்ட விதி விளக்குகிறது.

இகரத்தில் தொடங்கும் ஒரு சொல்லில் இகரத்தை அடுத்து அகரம் அல்லது ஐகாரம் ஏறிய மெய்யொலிகள் வந்தால் இகரம் எகரமாக மாறுகிறது. இகரத்தை அடுத்து இரண்டு மெய்யொலிகள் வந்தாலோ அல்லது வேறு உயிர் எழுத்துகள் ஏறிய மெய்யொலிகள் வந்தாலோ இகரம் எகரமாக ஆகாது.

இல்லை என்பது எல்லை என்றாகாது. இரும்பு என்பது எரும்பு என்று ஆகாது. (இங்கு இ -யை அடுத்துவரும் மெய்யொலிமீது உகர உயிர் ஏறி வருகிறது).

எடுத்துக்காட்டுக்கு இதழ்-எதழ்; இலவம்பஞ்சு-எலவம்பஞ்சு; இமை-எமை; இமையமலை- எமையமலை; இலை- எலை; இடையூறு- எடையூறு; இடைவெளி-எடைவெளி; இணை-எணை; இதழ்- எதழ்; இறப்பு-எறப்பு; இரந்து-எரந்து; இரக்கம்-எரக்கம்; இறங்கு- எறங்கு; இலவசம்-எலவசம்; இழப்பு-எழப்பு; இழை-எழை; இளைஞன்-எளைஞன். இச்சொற்களை எழுத்துத்தமிழிலும் பேச்சுத்தமிழிலும் உச்சரித்துக் கணினியில் பதிவுசெய்து மாணவர்களிடம் வழங்கவேண்டும். அவ்வாறு வழங்கப்பட்ட சொற்களை மாணவர்கள் மீண்டும் மீண்டும் கேட்கச் செய்கின்றபொழுது மாணவர்கள் மனத்தில் எழுத்துத்தமிழுக்கு நிகரான பேச்சுத்தமிழ் பதியும்.

விதி 2

உகரத்தில் தொடங்கும் சொல்லுக்கு அடுத்து அகரம் அல்லது ஐகாரம் ஏறிய உயிர்மெய் எழுத்துகள் வந்தால் உகரம் ஒகரமாக மாறும். மேற்குறிப்பிட்ட விதி ஒன்று போலவே இவ்விதியும் அமையும். உடல்-ஒடல்; உடம்பு-ஒடம்பு; உடன்-ஒடன்; உடை-ஒடை; உடையார்-ஒடையார்; உனக்கு-ஒனக்கு; உணர்ந்து-ஒணர்ந்து; உணவு-ஒணவு; உதடு-ஒதடு; உதயம்-ஒதயம்; உதறு-ஒதறு; உதை-ஒதை; உபகரணம்-ஒபகரணம்; உபயம்-ஒபயம்; உபசரி-ஒபசரி; உரசு-ஒரசு; உரைப்பு-ஒரப்பு; உரை-ஒர; உலகம்-ஒலகம்; உழவு-ஒழவு; உழை-ஒழ; இவ்வாறு இவ்விதிகளுக்கு உட்பட்ட சொற்களைத் தொகுக்க வேண்டும். பின்னர் அவற்றை உகர ஒலியுள்ள சொற்கள் எவ்வாறு ஒகர ஒலிபெற்று ஒலிக்கின்றன என்பதைக் கணினி வாயிலாகக் கேட்கச்செய்ய வேண்டும். இவ்வாறு திரும்பத் திரும்பக் கேட்கச்செய்யும்போது மாணவர்களுக்கு எழுத்துத் தமிழுக்கும் பேச்சுத்தமிழுக்கும் உள்ள வேறுபாடு புரியும். அதனால் தயங்கம் நீங்கிப் பேச முயல்வர். அவற்றின் முழுப்பொருளையும் அறிவர்.

உகரத்தில் தொடங்கும் சொற்களை அடுத்து இரண்டு மெய்யெழுத்துகளோ அல்லது உகரம் ஏறிய மெய்யெழுத்துகளோ வருமேயானால் உகரம் ஒகரமாக மாறுவதில்லை. உண்மை, உருண்டை, உருளைக்கிழங்கு, உல்லாசம்.

விதி 3

ஒரு சொல்லின் இறுதியாக வரும் “ஐகாரம்” பேச்சுவழக்கில் சிதைந்து அகரத்திற்கும் எகரத்திற்கும் இடைப்பட்டு ஒலிக்கின்றது. கதை-கத; சதை-சத; பாதை-பாத; சிறை-சிற; இலை- இல; பிள்ளை- பிள்ள; இங்குச் சொற்களின் இறுதியில் வரும் “த” என்னும் எழுத்து அகரத்தில் முடிந்தாலும் இச்சொல் ஒலிக்கின்றபொழுது எகரத்திற்கும் அகரத்திற்கும் இடைப்பட்ட நிலையில் ஒலிக்கும்.

சொல்லின் இடையில் வரும் ஐகாரம் தன் உச்சரிப்பிலிருந்து மாறி அகரமாகப் பேச்சுத்தமிழில் ஒலிக்கின்றது. எ.கா. மடையன்-மடயன் இடையன்-இடயன்

விதி 4

இறந்தகால இடைநிலையாகிய “த்த்” என்னும் இடைநிலை இகரத்தை அடுத்து வரும்பொழுது ச்ச் என்றும், ந்த் என்பது ஞ்ச் என்றும் மாறுபடுகின்றன. படித்தான் – படிச்சா(ன்) அடித்தான்- அடிச்சான் கடித்தான்-கடிச்சான் பிடித்தான்-பிடிச்சான்; மடித்தான்-மடிச்சான்; வெடித்தான்-வெடிச்சான்; குடித்தான்-குடிச்சான்; இடித்தான்-இடிச்சான்; ஒடித்தான்-ஒடிச்சான்; கிழித்தான்-கிழிச்சான்; ஒழித்தான் – ஒழிச்சான்; கழித்தான் – கழிச்சான்; ஒழிந்தான்- ஒழிஞ்சான்

விதி 5

தனிக்குறிலை அடுத்துவரும் ணகர, னகர, ளகர, லகர மெய்களோடு முடிவடையும் சொற்களில் மெய்யெழுத்து இரட்டித்து உகரம் சேர்ந்து முடிவடைகின்றது. இந்த உகரத்தை ஒலித்துணை உகரம் (enunciative U) என்று மொழியியலார் கூறுவர். மண்-மண்ணு; கண்-கண்ணு; பண்-பண்ணு; உண்-

உண்ணு; கல்-கல்லு; செல்-செல்லு; பல்-பல்லு; முள்- முள்ளு; பொன்-பொன்னு

நெடிலைத்தொடர்ந்து வரும் லகர ளகர மெய்யோடு முடிவடையும் சொற்களில் லகரம் உகரத்தோடு சேர்ந்து ஒலிக்கப்படுகிறது.

கால்-காலு, பால்-பாலு, மேல்-மேலு, சால்-சாலு, ரால்-ராலு, வால்-வாலு, தாள்-தாளு, வாள்-வாளு. பொதுவாக இவ்வுகரம் குற்றியலுகரமாவே ஒலிக்கின்றது. இதனால் பேச்சுத்தமிழில் இடம்பெறும் குற்றியலுகரம் வல்லினம் மட்டும் அல்லாமல் எல்லா மெய்யெழுத்துக்களின் மேலும் ஏறி ஒலிக்கின்றது என்று கூறலாம்.

விதி 6

தனிக்குறிலைத் தவிர்த்து னகர மெய்யோடும் மகர மெய்யோடும் முடிவடையும் சொற்களில் னகர, மகர மெல்லின ஒலிகள் கெட்டு, அதன் மூக்கொலித்தன்மை முந்தைய உயிரொலியில் ஏறுகின்றன. ஆனால் அந்த மூக்கொலித் தன்மை பெற்ற உயிர் ஒலிகளை எழுத்து வடிவில் காட்ட இப்போது இயலாது. (அரசன்-அரச~; வந்தான்-வந்தா~; செய்தான் - செய்தா~; அவன்-அவ~; மரம்-மர~; தருமம்- தரும~; இலக்கியம்- இலக்கிய~; படம் – பட~

விதி 7

எழுத்துத் தமிழில் சொற்கள் வல்லினமெய்யில் முடிவதில்லை. அதுபோல் இன்று பேச்சுத் தமிழிலோ எந்தச் சொல்லும் மெய்யெழுத்தில் முடிவதில்லை. பொதுவாக நெடிலைத்தொடர்ந்து யகரமெய்யில் முடிவடையும் சொற்கள் இகரத்துடன் சேர்ந்து முடிவடைகின்றன. (எ.கா) வாய் – வாயி; காய்-காயி; சேய்-சேயி; பேய்-பேயி; போய்-போயி. குறிலைத்தொடர்ந்து யகரமெய்யில் முடிவடையும் சொற்கள் இரட்டித்து இகரத்துடன் சேர்ந்து முடிவடைகின்றன. (எ.கா) பொய்-பொய்யி; செய்-செய்யி; மெய்-மெய்யி. ஐகாரம் ஏறிய ஓரெழுத்து ஒரு மொழிகளும் இகரம் ஏறிய யகரத்தோடு முடிவடைகின்றன. கை-கையி; பை-பையி; வை-வையி; தை-தையி; மை-மையி.

இன்னும் இவ்வாறு சில ஒழுங்குகளுக்கு உட்பட்ட விதிகள் பல உள்ளன. அவற்றை நாம் எழுத்துவடிவத்தையும் பேச்சு வடிவத்தையும் அருகருகே காட்டியும் கேட்கச் செய்தும் கற்பிக்கலாம்.

பொதுவாகச் தனித் தனிச் சொற்களாகக் அமைத்துக் காட்டுவதைவிட, தொடர்களில் அமைத்துக் கற்பிக்கும்பொழுது மாணவர்களுக்கு அச்சொற்கள் எளிதாக விளங்கும். எடுத்துக்காட்டிற்கு "நான் கதை படித்தேன்" – நா(ன்) கத படிச்சே(ன்); அவன் படம் பார்த்தான்; அவ(ன்) பட(ம்) பார்த்தா(ன்)

எழுத்துத்தமிழில் இடம்பெறாமல் பேச்சுத்தமிழில் மட்டும் இடம்பெறும் சில அமைப்பு முறைகளும் இருக்கின்றன. இவற்றையும் மாணவர்களுக்கு அறிமுகப்படுத்த வேண்டிய தேவை உள்ளது. எடுத்துக்காட்டுக்கு "உனக்கு மூளை இருக்கிறதா?" உனக்கு மூளை கீளை இருக்கா? "ஏதாவது தண்ணீர் குடித்துவிட்டுப் போவோமா?" ஏதாவது தண்ணி கிண்ணி குடிச்சிட்டு போவோமா? இத்தகு தொடர்களைக் கொண்ட உரையாடலைப் பேச்சுத் தமிழிலும் எழுத்துத்தமிழிலும் அடுத்தடுத்துக் காண்பிக்கின்ற பொழுது மாணவர்கள் இரண்டிற்கும் உள்ள வேறுபாடுகளைப் புரிந்துகொள்ள வாய்ப்பு இருக்கின்றது. இதன் வழி எழுத்துத் தமிழில் உள்ள சொற்கள் பேச்சுத்தமிழில் எவ்வாறு மாறுகின்றன என்பதையும் அவை எவ்வாறு உச்சரிக்கப்படுகின்றன என்பதையும் மாணவர்கள் நன்கு அறிவர். திரைப்படம் அல்லது நாடகத்தில் வரும் காட்சிகளில் ஒருசில குறிப்பிட்ட பகுதிகள் மட்டும் பேச்சுத்தமிழ், எழுத்துத்தமிழ் இரண்டிலும் அமைவது போல் காட்டுவதன் வழி மாணவர்களின் ஆர்வத்தைத் தக்க வைத்துக்கொள்வதோடு விரைவில் அவர்களுக்குத் தமிழிலுள்ள இரு வழக்குகளையும் கற்பிக்கலாம். மேலும் பேச்சுத்தமிழில் பேசுகின்றபொழுது ஒருவருடைய முகபாவம் எவ்வாறு மாறுகின்றது என்பதையும் ஒருங்கே காட்டலாம்.

இவ்வாறு அனைத்து விதிகளுக்கும் எடுத்துக்காட்டுகளைக் காட்டி அவற்றைக் கணினிவழிப் பல்லூடகத்தின் (Multi media) உதவியால் பலவிதங்களில் வெளிப்படுத்திக் காட்டுகின்றபொழுது

மாணவர்களுக்குப் பேச்சுத்தமிழையும் எழுத்துத்தமிழையும் ஒருங்கே கேட்கவும் பார்க்கவும் வாய்ப்புகள் கிடைக்கின்றன. மேலும் பேச்சுத்தமிழில் "கருத்தாடல்" நிகழும் இயற்கையான சூழலையும் மாணவர் கண்முன் கொண்டுவந்து நிறுத்தமுடியும். பேசுபவரின் முகபாவம், உடல் அசைவு முதலிய மொழிசாராக் கூறுகளோடு பேச்சுத்தமிழ்த் தொடர்களை எவ்வாறு இணைத்துப் பயன்படுத்தலாம் என்ற அறிவும் மாணவர்களுக்குக் கிடைக்கும்.

மேலும் மாணவர்கள் சொந்தமாகப் பலமுறை இப்பாடத்தை மீண்டும் மீண்டும் இயக்கிக் கேட்பதன்வழி மாணவர்கள் மனத்தில் அவை நன்கு பதியும். ஆர்வமும் பெருகும். மேலும் மாணவர்கள் பொருள் உணர்ந்து படிக்கவும் கேட்கவும் வாய்ப்புகள் அதிகரிக்கும்; இந்த முயற்சியில் கணினி வல்லுநர்களும் மென்பொருள் தயாரிப்பாளர்களும் ஈடுபட்டு உதவுகின்ற பொழுது பல்வேறு பரிமாணங்களில் கணினித் திரையில் ஆர்வமூட்டும் பேச்சுத்தமிழினைக் கேட்க இயலும். இவை தாய்த் தமிழகத்திலிருந்து வெளியேறிப் பேச்சுத்தமிழ் சூழல் அமையப்பெறாமல் வாழ்கின்ற வெளிநாட்டுத் தமிழ்க் குழந்தைகளுக்குப் பேச்சுத் தமிழைக் கற்றுக்கொள்ளப் பெரிதும் உதவும்.

கட்டுரை ஆக்கத்திற்கு உதவிய நூல்

ந. தெய்வசுந்தரம் (Diglossic Situation in Tamil - A Sociolinguistic approach , Ph.D. Thesis submitted to the University of Madras, 1980 , Chennai)

Development of E-Content For Learning Tamil Phonetics

Dr.R.Velmurugan
Singapore

Introduction

It is a well established fact that the process of e-learning is endowed with a lot of advantages, of which the same are not at all available in the human-enabled teaching and learning process. People enlist a lot of advantages and benefits of e-learning. Some of the important advantages of e-learning are; it is not just learning but sharing, the content in the e-content is not static rather dynamic, it is any where and any time learning, it has a global audience, the content to be delivered through the process of e-learning has certification, it is indeed dam cheap, the instructional design in e-content will be learner centric, it invites structured feed back, it is self paced, it can be used in real time and many time, it can present the content through multimedia presentation, it will have Scientific evaluation method, the content would be authentic, and it may have the provisions like interactivity, Book marking, white board, Hot spot, Hypertext and Hyperlinks etc.,

Apart from these, the e-content will present the content in multiple formats; complete technical contents are explained with suitable graphics and Animations. The e-content will generally be in self directed and paced instructional format and smooth instructional strategies will be chalked out in such a way that learners never lose the interest. It presents the content in a simple text with unambiguous graphics and with relevant supportive headings.

Sometimes, it will have its own reference materials which generally do not burden the learners and those can appear on demand with optional frames. Some e-content has certain striking features like automatic retaking of the lessons. Wherever learners are not satisfactorily able to perform, it may have automatic learning path. Sometimes, it offers room for selection of the quantum of information considering the learners requirements. Some advance level e-content provides 3D virtual reality-synchronous and asynchronous interactivity – chat, conferencing, etc.,

The above details establish that e-learning is a mixture of different learning methods, delivered to the learners through information technology supported with educational instructional design and relevant content. The e-learning is, as a universe, comprising of three basic elements viz.

1. Content
2. Services
3. Technology

The content forms the back bone of the e-learning, services and technology forms the rider on which the content travels.

e-education and language learning

In the domain of education, a lot of metamorphoses have occurred because of the social needs and scientific advancements. The traditional means of education may not be suitable in modern days. Thanks to electronic devices which are being used in the field of education and which facilities and of course accelerates the learning pace of our learners. The use of electronic device in the domain of language teaching is quite significant and unique. There is no doubt that human enabled language teaching is powerful and the learners are comfortable enough in learning language in it. But the

machine or computer enabled learning or teaching is the need of hour as it possesses a number of advantages besides the advantages attached with the human instructor involved in teaching-learning process.

For example a computer mediated language teaching / learning process will supply rich and accurate linguistic corpora which will certainly mould the learners to the greater extent by providing them ample room of opportunities to freely mingle with the relevant and original linguistic data of language, whether it is a second language or foreign language or even first languages.

Similarly learning a language from the mouth of native speakers of the language has some added advantages especially in the second language learning situations. But a native speaker in the traditional teaching / learning process cannot address to the global learners. But the e-content travels across the world and caters to the varying needs of the learners of heterogeneous nature.

Learning Tamil

It is a known fact that Tamil is now-a-days learnt globally. Appointing native Tamil teachers for teaching Tamil across the globe is indeed practically not possible. But through e-learning it is possible to teach or learn Tamil from the native Tamil teachers since e-learning is any where and anytime learning and also has its own power and strength.

It is needless to mention that Tamil being one of the living classical languages; it maintains its tradition and obtains modernity without sacrificing its original colours, for meeting both the classical and contemporary needs of the society. Having declared it as a classical language, there is a global acclaim among the Tamils as well as non-Tamils.

Since Tamil is 20,000 years or so old, it gained a lot finesse and richness in terms of its linguistic nuances and intricacies in articulatory and auditory aspects, sequencing the allophones and phonemes, formation of words, invention of grammatical features and elements, formation of sentences / utterances and other advance level of communicative strategies. Of course, to a linguist no language in superior than other languages and inferior either, and no language is easy to learn or hard to learn. But, if a language has a rich tradition with a lot of linguistic nuances, it is, of course, in a way superior to other languages and harder to learn. In this way, as Tamil is rich and powerful, one has to take special effort in learning certain subtleties of Tamil language in order to master the Tamil as if a native Tamil speaker.

It is a matter of importance that Tamil has a lot of unique properties which are not easily get-at-able to the neo-learners of Tamil language. The uniqueness is found to exist in all levels of language viz, Phonology, Graphology, Morphology, Morphophonemic, Syntax, Semantics, Beyond syntax (Discourse and pragmatics). To learn all those peculiarities, learner has to move the heaven and earth.

Tamil e-content

Tamil, as stated above, is learnt globally. For this, Tamil is to be taught through Computer Based Teaching (CBT), Web Based Teaching (WBT) or Net Work Based Teaching (NBT). To enable this type of Tamil Teaching, various packages are prepared here and there in piece-meal. But no exhaustive work has so far been done for Tamil. The present paper tries to give some guidelines for developing e-content to teach Tamil phonetics to the learners who wish to learn Tamil as second language.

The Tamil phonetics has been studied by different scholars. Although Tamil has a number of regional dialectal variations and sociolectal variations with a lot of sound changes, there is a standard spoken Tamil spoken by the majority of the people and which is intelligible to majority of the people as well. The package to be prepared will use those standard phonemes and allophones. Since it is a pioneering attempt, only standard phonemes and allophones of Tamil can be used. But in the later stage, as this package in dynamic; dialectal and sociolectal sound variations can be used for introducing them to the learners through hypertext which will appear on demand. So, this package will cater the needs of all the learners of Tamil in all times.

e-content for Tamil phonetics

Introduction

A brief but technical introduction about Tamil phonetics will be presented through voice over and electronic text. Then it will spell out the objectives of this package besides detailing the uniqueness and merits of this package. Since this package is meant for global audience and for the audience of different nature, it will detail the rational of grading the corpora. Accordingly, the users or learners can select the options to directly go to the given frame.

Corpora

For the package, the following phonemes and allophones (as proposed by S. Rajaram) will be taught.

Vowel	:	i, ii, e, ee, a, aa, u, uu, o, oo,
Vowels allophones	:	I, E, ɛ, ʌ, æ, ɔ, Ω, ʊ, ɨ
Consonants	:	k, ŋ, c, ɲ, ʈ, ɳ, t, n, p, m, y, r, l, v, ɭ, l̲, r̲, n̲.
Consonants allophones	:	g, ɣ, ɟ, s, ɖ, ɾ, d, ð, ɸ, β.

Frame

For teaching each phoneme a frame will be spared. A learner can at the outset have some basic idea about Tamil Phonetics and its peculiarities by looking at the main frames. Then he can move to the frame of Corpora, in which he can select a particular phoneme by clicking it, and then he can move to a specific frame which tells all about a particular phoneme. If, for example, a learner comes to the frame of /p/ he will see the following type of e-text.

Model lesson for a phoneme /p/

This model lesson will tell the phoneme first and then it describes its point and manner of articulation with a Graphic and Animation that directs the way in which the particular sound can be produced. Native speaker's standard pronunciation of this sound in isolation will be given. A voice will appear producing the sound in a list of words wherein the particular sound appears. After these, a dialogue box will appear directing the learners to produce the same sound by looking at the graphics / Animation and by listening to the voice over. Then, learner's voice quality will be checked and quantified using Sonographics.

Based on the performance of the learner, the sonographic pictures will appear on the screen and the score will also appear. The learners will be directed to repeat the sound by giving some guidelines. The learner will not be allowed to move on to next frame until he produces the particular sound with the expected quality. Then, if the learner wants, he can explore the exhaustive list of words which has the given phoneme in different distribution and combinations.

Model Frame for Consonant Phoneme

1. Phoneme : / p / 2. Phonetic Description : Voiceless bilabial stop
3. Manner & point of Articulation (In the production of [P] the lips are closed and the soft palate is raised to close the nasal passage, when the lips are opened the air suddenly comes out without explosion. There is no vibration in the vocal cards.)
4. Graphics (A picture will appear to help the student produce particular sound)
5. Native speaker's voice of this sound in isolation
6. List of the words wherein this sound appears.

pakal	-	'day time'	a<u>r</u>pan	-	'a mean fellow'	
paavam	-	'sin'	ve<u>t</u>pam	-	'hotness'	
puli	-	'tiger'	kappal	-	'ship'	
pul	-	'grass'	tappu	-	'mistake'	Hot spot

More
7. Dialogue box to direct the learner to produce this sound Produce this sound / can you produce this sound?
8. Picture of sonogram Correct / incorrect
9. Exhaustive list will appear on the screen on demand from Hypertext
Go there

Model Lesson for Allophone

After the successful completion of this frame, the learners will be allowed to go to allophones of a particular phoneme one by one. For each allophone IPA notation will appear. Then, Phonetic description will appear and point and manner of articulation for the particular allophone will also appear on screen with either graphics or Animation. Allophonic distribution will appear with a list of words. Learners will be advised to produce them repeatedly looking at the list of words.

Model frame for Allophone

1. Allophone [β]
2. IPA Symbol [β]
3. Phonetic Description : voiced bilabial fricative
4. Point and manner of Articulation
   In the production of [β], the lips are closed slightly and the soft palate is raised to close the nasal passage. When the lips are opened the air stream is pushed through with a weak plosion. There is slight vibration of the vocal cards during its production.
5. Allophonic distribution
   aβayam 'shelter'
   uβaayam 'trick' Hot spot
   laaβam 'profit' More

Consolidated lesson

After seeing all the allophones of a particulars phoneme, a consolidated frame will appear. In this frame, all the allophones of a phoneme will appear with suitable examples. The learners will be directed to produce those words and to observe the differences between and among the sub members of a particular phoneme with a comparative perspective. After the consolidated frame, next frame for next phoneme will appear. The vowel phonemes and their allophones will appear at first and then consonant phonemes and their allophones will appear. After introducing all phonemes, a specially devised text bearing all the phonemes and allophones of Tamil will be displayed coupled with a native speaker's voice in a natural manner. Having listened to it carefully, the learners will be advised to read it as the model guides.Then, there will be an option for exploring the social and regional

variations of certain selective words. This link will appear only on the demands of the learners. This will enable the advance level of learners to learn the social and regional variations of those certain selective words.

Conclusion

This package will give the learners all minute details about the phonemes and allophones of Tamil, like; Phonetic and phonemic qualities of Tamil sounds, Allophonic distributions and possible combination in three positions viz. initial, medial and final, vowel- short and long, diphthongs, consonant clusters-both identical and non identical, and exhaustive list of minimal pairs and text for natural flow of sounds. In this package, the content delivery is in multiple formats i.e. through Voice, Animation, Graphics, and Electronic Text etc. This package is highly interactive. That is, the learner can interact synchronously and asynchronously with this package. So the management of learning experience is possible. This will help the learners to accelerate their learning pace. Since this package simultaneously employs testing technique which is one of the important processes of teaching; it enables the learners to go to the right path of learning by conforming and ensuring the learning achievement with a sense of self confidence.

In the process of evaluation, it gives a positive as well as negative reinforcement by giving score. In certain frames, this package will not allow the learners to go further until they do not gain the expected level of competency in a particular phoneme. This type of periodical check-up will help the learner's to progress in a slow and steady manner with the comfortable pace of learning. This package will avail a lot of linguistic data which will in turn help the learners to improve their understanding and performance in the aspects of Tamil phonetics. The data will be selected, graded and presented following the linguistic principles, educational psychology, instructional design, and technological advantages and constrains. More number of hot-spots will be given so as to help different levels of learners. So a lot of hypertexts and hyperlinks will be given. For this purpose, all the findings of linguistic researches done so far on the phonetics and phonology of Tamil language will be used especially for corpora creation and for forming data base of this package.

In total, this paper suggests only the linguistic technical know-how of developing e-content for learning Tamil phonetics. These ideas and suggestions can be fruitfully used only when right types of computer software are employed to prepare the package. It is a joint venture that both linguists and computer scientists have to use their technical knowledge together to produce a fool proof packages for teaching/learning the Tamil phonetics.

Reference

1. Bloch, B and Trager, C **Outline of Linguistic Analysis.** Baltimore, Waverly Press, 1942.
2. Downes, S **E-Learning 2.0.** http://www.downes.ca/post/31741 2005
3. International Phonetic Association, **The Principles of the International, Phonetic Associations**, London, University College, 1949.
4. Karrer, T **Understanding eLearning 2.0** http://www.learningcircuits.org/2007/0707karrer.html 2007
5. Nichols, M **E-Learning in context.** http://akoaotearoa.ac.nz/sites/default/files/ng/group-661/n877-1---e-learning-in-context.pdf 2008
6. Rajaram, S **Tamil Phonetic Readers** CIIL, Mysore, 2000

Infusing Media-Literacy to Help Learners Construct and Make Sense of their Learning

Sivagouri Kaliamoorthy
Beacon Primary School, Singapore
sivagouri_kaliamoorthy@moe.edu.sg

Abstract: Our pupils live in a technology and media-driven environment. They are also surrounded by wealth of information. Constructive learning takes place when they are able to connect, construct and relate this information to the situated context. In preparing our learners for the 21st Century, it is essential for our pupils to be able to gather information, analyse them, and relate it to the situated context. Hence, there is a need to move beyond a focus on basic competency in the core subjects to promoting understanding of content at much higher levels by weaving media literacy into curriculum and providing a meaningful experience in language literacy. Infocomm technologies could act as a powerful tool for pupils to get connected. By leveraging on technology, pupils take an active role in searching for relevant information via the Internet to substantiate their understanding of the information presented in the newspaper articles. This process helped pupils to be independent learners situated within an authentic context. By tapping on technologies and infusing media literacy into the curriculum, pupils use their four basic language skills effectively and started to take ownership of their learning.

Keywords: Information Communication Technology, Media Literacy

Introduction & Purpose

Pupils are surrounded by wealth of knowledge. Today, at the click of a button pupils can view the events happening around them in just seconds. Information is transported within seconds and it is important that our pupils are equipped with the skill to search for the information, be critical in selecting information and make sense of the information presented.

In this information age, education is mandated to respond to demands in two directions: on the one hand, it has to transmit an increasing amount of constantly evolving knowledge and know-how adapted to a knowledge-driven civilization; on the other hand, it has to enable learners not to be overwhelmed by the flows of information, while keeping personal and social development as its end in view. Therefore 'education must ... simultaneously provide maps of a complex world in constant turmoil and the compass that will enable people to find their way in it' (Delors *et al.*, p85). This translates in a shift in focus on the amount of content to be taught in schools. It calls for greater emphasis in equipping our pupils with skills to search for the relevant information independently supporting the nation-wide 'Teach Less Learn More'[1] initiative.

[1] 'Teach Less; Learn More' (TLLM) is a call for schools and teachers to focus more on the active learning of students and the construction of their own knowledge.

The nature of learning by our young digital natives has also transformed. The nature of and type of skills has also changed. They are surrounded by information. World Wide Web can be assessed at a click of the button. In their world, knowledge can be shared and co-constructed. Thus, there is an urgent need to equip them with skills and lenses to handle this influx of information.

Understanding the needs of the young learners, information communication technology is integrated in their learning processes. Information communication technological tools are applied as constructive tools. "Constructive tools are general-purpose tools that can be used for manipulating information, constructing one's own knowledge or visualizing one's understanding" (Lim & Tay, 2003). Jonassen and Carr (2000) purport a constructivist approach, "ICT as mind tools for the construction evaluating, analysing, connecting, elaborating, synthesizing, imagining, designing, problem-solving, and decision-making." The term "constructive" stems from the fact that these tools enable students to produce a certain tangible product for a given instructional purpose.

This paper takes a reflective, narrative approach in documenting my attempt to integrate media literacy into my daily lessons.

My Reflections

As a daily assembly program, the school Principal shares important news that appears in the newspaper. As an extension to the daily assembly program during the Mother Tongue Language lessons, pupils are also engaged in classroom discussion. During these discussions, pupils were observed to be very engaged and used the language appropriately. Pupils showed great interest in the issues and expressed that they would like to find out more regarding the news read to them during the morning assembly.

In the school, all Tamil pupils work in a one-to-one computing learning environment. Pupils were introduced to search engines and were guided in searching for the relevant information. Pupils were taught cyberwellness and precautionary measures were taken when pupils browse the given website. Age would not be a barrier in understanding world issues if it is tailored to meet the needs of the young learners. What really matters is whether pupils are equipped with skill to understand the implication and impact of the issue discussed.

As a start pupils start to discuss issues closer to their homes. For instance, there was an article of fighting amongst teenagers. Teacher selected this article to discuss but realised that the need to set the context before broaching and discussing the issue. During civics and moral education, a big book entitled "who can watch the television?" was introduced. The story elates about how two siblings will fight over to watch a program in television and neither would give in to the other. The mother would come and off the television set. The teacher then posed questions as to what are the consequences of these actions. The pupils then worked in their respective groups and presented moral reasoning for the action. They were able to relate chain actions that would take place if the siblings were to continue with their behaviour. Following this lesson, pupils were introduced to the article. There was an intense discussion amongst pupils and what were the implications to the society and country. Pupils related the probable consequences.

After the introduction of the Australian bush fire. Tamil pupils expressed that they wished to know more about this problem. Pupils used the Internet search engines to look up for latest update on the Australian bush fire. In the hope of searching, pupils watched the bush fire live at BBC news website. They then took upon themselves to update one another on the latest on this bush fire. Pupils

expressed civic mindedness and sympathy for those who have been affected. They discussed and evaluated the situation and thought about the things that the victims might have lost and the possible implication on their lives. It was heart warming to note pupils expressed concern and empathy for those affected. In conjunction with Total Defence day pupils had to go online and search for relevant information. The search helped them to investigate the rational for celebrating Total Defence Day and the five different defences in Singapore. It is important in ensuring the safety and security of our country and its people. This was discussed and created an awareness and understanding of issues that surrounded them. They used presentation tool to express their findings.

In the later part of the year, there was a topic on advertisement. Pupils gathered different types of advertisements and analysed the information presented in the advertisement. They discussed and brought out the underlying catch in the advertisements. They did a search online to find out the market price of those goods advertised and critically evaluated the advertisement. They reasoned whether they it was cost-effective to purchase those advertised. They presented their views to the class. Pupils used presentation tool to do up the advertisements. The computer was used as a constructive tool to construct their advertisements. They presented their advertisements and the peers evaluated and analysed the information.

All Tamil pupils were also introduced to Malay martial arts,[2] Silat. By enabling the pupils to synthesise their ideas for the creation of multimedia productions using tools such as Microsoft PowerPoint and Photostory 3 they were able to hone their information and media literacy skills. Microsoft PowerPoint was used to scaffold pupils' learning of oral skills through well placed images and sound clips. In the process of many lessons, pupils actively formulated and shared their understanding of the required curricular objectives. Tamil language pupils were actively engaged in storyboarding and scripting. This year the Primary 2 pupils have extended their exposure to interview skills when they scripted questions, video-graphed interviews and subsequently edited them via Windows Moviemaker. Technology was leveraged when pupils used the search engine to gain in-depth understanding of the culture after the lesson on martial arts. Pupils worked together in groups of four and brainstormed possible interview questions to ask the instructor to address/supplement the gaps in their search. The pupil editor collated the responses from the team members and used Microsoft Word to type the questions and prepare the template for the reporter. The templates were then emailed to the other team members for feedback. The editor then incorporates the changes and finalises the interview questions. During the hands-on sessions, pupil cameramen took mug shots (photo coverage) and passed it to the pupil producer. After the hands-on session, the reporter interviewed the Silat instructor in Tamil language and the entire process was video taped. The producer cum newscaster with the help of the other team members used Microsoft Windows Movie Maker and edited the interview segments, selected and inserted the pictures and the edited movie clip on to a PowerPoint slide presented it as a and presented the news. Pupils were given a flow chart of organisers as guide to them in the editing process.

The selection of technology was used as a constructive tool to bring out the learning and appreciating the Malay culture. Pupils used their experience in the situated context infusing the cultural

[2] Silat is a Malay Martial Arts and is originated from the Malay Archipelago thousands of years ago. It is an art of fighting and defense of the Malays.

transmission through internalizing the desirable values of respect of another culture by understanding the significance of the Malay cultural heritage. These values permeate the environment as they learn and appreciate the rich Malay cultural martial arts, Silat.

Discussion & Conclusion

Technology is used as a constructive tool to facilitate pupils learning and making sense of their learning. Pupils' engagement was evident throughout the project. They were critical about their work and had done numerous editing before submitting the project. Pupils were actively using the net to search for information to enhance their learning. The project had benefited even the weaker pupils, who was observed to be actively contributing ideas and was working towards completing their group project. There was such joy when the pupils presented their project. As the project helped to bring out the best in each pupil, pupils gave positive feedback that they would like to do more of such projects. Every team member had contributed and has equal share in the project, thus the ownership was very strong amongst them. Pupils were seen interacting, playing with the Malay pupils even after the project.

In terms of skills, all pupils had learned basic photo-taking skills and are able to use the questioning techniques to generate interview questions. Through this project it was observed that pupils had tapped on prior knowledge and experience in developing the interview questions more confidently. (e.g., interview with a journalist from the Singapore one-off Tamil Channel News Segment held in Term 1, 2009). Pupils learned to use the information and ideas presented in a graphical organiser format to organise ideas and create the end product. Pupils learned about the different job scopes/roles (e.g., producer, director, editor and reporter in a press crew and was able to practice the skills. Pupils initiated role-play not only polished the respective skills it also brought the independence in them. Pupils exhibited strong bonding and collaboration during the various collaborated sessions. The usage of technology was pervasive and pupils creating media worthy products were a big step. As Burn.,A. (2009), had pointed out "the new ability to digitally undo and reconstruct still and moving image (and audio) enables the students to become writers as well readers of the visual … the literacies of the visual semiotic they have required become extended in the digital manipulation of image, and in the trans-coding of image to word and back again, in group discussion and written commentary"

This paper is my attempt to share possible strategies in integrating digital media into our daily lessons. It is through such sharing and exchanges where ideas could build upon ideas to further push the boundaries of our pursuit for pedagogical break throughs in this fast changing world.

References

1. Burn., A. (2009). *Making New Media. Creative Productions and Digital Literacies*. New York: Peter Lang Publishing, Inc.
2. *Teach Less; Learn More- Transforming Learning From Quantity To Quality*. Singapore. Education Milestones 2004-2005 http://www.moe.gov.sg/about/yearbooks/2005/pdf/teach-less-learn-more.pdf
3. Lim., C.P., & Tay, L.Y.,(2003). *Information and Communication Technologies (ICT) in an Elementary School: Students' Engagement in Higher Order Thinking. Jl. Of Educational Multimedia and Hypermedia* (2003) **12**(4), 425-451
4. Williams, M. D. (2000). *Integrating Technology into Teaching and Learning*. Singapore: Prentice Hall.

The Use of Technology among Tamil Medium Students

Barriers and Solutions

Prof. P. J. Paul Dhanasekaran
Principal, St.Joseph's College of Education,
St. Mary's Hill, Udhagamandalam 643 001 Tamilnadu, South India
profpjpaul@gmail.com

Abstract: The paper deals with the research carried out in a Teacher Training Institute among Tamil medium students who study Diploma in Teacher Education in Tamilnadu, South India. The use of technology among Tamil medium students is very minimal. This is due to lack of confidence, demand for the use of English in the application of technology. Even though the students appreciate the application of technology in learning and teaching of language, these two reasons prevent them from active application of technology. The paper tries to analyze the barriers and tries to find out the solutions for those barriers in the use of technology among Tamil medium students who will be teachers in the secondary education in the near future. The sample for the study comprises 100 students of Diploma in Teacher Education (D T Ed). All of them have studied their Higher Secondary Course (+2) in Tamil medium and English is one of the subjects they study in their course. In spite of studying English for nearly 7 years, they have little confidence in using English either in speech or writing. After they complete the Diploma in Teacher Education they are expected to teach English for the students in classes 6 to 8. The researcher is teaching the methods of teaching English to the sample under study. The researcher taught them how to teach English and at the same time how to speak and write also.

The researcher made a programme so that the students try to speak English in the classroom. Everyday students must prepare 5 sentences on anything and speak those sentences in the classroom. This went on for a week. The second week the students are divided into groups and each group prepares 5 sentences on any title or topic and other groups are given opportunity to rewrite or reframe the sentences spoken by a particular group. Students are asked to bring English newspapers and are asked to identify simple, compound and complex sentences which have been taught by the teacher. The same groups are engaged in dialogue. These exercises had given them courage to speak in the classroom. The students are opportunity to handle the computer. Some students have learnt typewriting and this ability is used in the use of computer. They are asked to type and print the essay they had written during their composition work. Students developed their confidence and slowly started to use English in ordinary conversation in the classroom. This way the sense of fear had been dispelled. The internal and term end examination results in English also proved that the students have strengthened their confidence in the use of English and also the application of computer. This empowerment certainly will improve the use of technology in learning language. The curriculum and teaching methodology should incorporate the use of English and also the use of technology as a practical component.

Project and Discussion

The ultimate goal of today's ESL students is to acquire the ability to communicate with others in a meaningful and appropriate ways. They must become critical thinkers who know how to apply language or convey their thoughts in a variety of situations. The paper identifies two issues. i) lack of confidence ii) demand for the use of English in the use of technology. In order to address the issues, the researcher designed his instruction that involved an active, creative, and socially interactive learning process in which students would construct their own knowledge using their prior knowledge, a process governed by constructivist approach. Breaking students in small groups provides more opportunity to practice the target language as well as reinforcing the knowledge through group discussion and collaboration. In the instructional experiment, constructivist approach is applied. In second/foreign language education, constructivism is often associated with the use of technology in the classroom (Chuang&Rosenbusch 2005; McDonough, 2001; Ruschoff & Ritter 2001) Students learn best through concrete experience, dialogue and active learning (Goldberg 2002).

A constructivist approach makes it possible to alleviate some of the obstacles to developing communication skills for second/foreign language learners. In overcrowded classrooms, where teachers have difficulty in giving personal attention, students may assist each other in understanding new information through group discussion and investigation. Thus students become active participants instead of passive learners, waiting to receive information. This experiment fosters creative and autonomous thinkers who are able to convey their thoughts in a wide variety of different situations.

References

1. Chuang, H. &Rosenbusch, M.H. (2005) Use of Digital Video Technology in an Elementary School: Foreign Language Methods Course,. British Journal of Educational Technology. 36 (5), 869-80.
2. Goldberg,M.F. (2002) 15 School Questions and Discussions: From Class Size, Statndards, and School Safety to Leadership and more. Lanham, MD: Scarecrow Press.
3. McDonough, S.K. (2001) Way Beyond Drill and Practice: Foreign Language Lab Activities in support of Constructivist Learning. International Journal of Media. 28 (1), 75-81.
4. Ruschoff, B. & Ritter, M. (2001) Technology-enhanced Language Learning: Construction of Knowledge and Template-based Learning in the Foreign Language Classroom. Computer Assisted Language Learning. 14(3-4), 219-32.

Learning Tamil The Fun Way!

Mrs Kanmani Shunmugham
Tamil Teacher / Discipline Mistress
Pei Tong Primary School, Singapore

Abstract: The intent of this paper is to focus on teaching and learning Tamil language using video presentations and E-book creations using KooBits software and online blogs. These lessons have been carried out for and by Upper Primary Tamil pupils from Pei Tong Primary School, Singapore.

The facilitation of teaching and learning such that pupils learn academic content as well as learn 'how-to-learn' in a constantly changing environment is crucial in today's world. The challenges of today's World necessitate the need to package skills learned by pupils such that it becomes a useful tool for the present and future generations.

For real learning to be effective, it is important for teachers to accommodate unique learning needs of every learner. Real learning must also take place in contexts that promote interaction, and enable formal and informal learning. This can effectively take place in an environment where the pupils are allowed to enhance visual and digital literacy skills and to develop critical thinking skills through the creation of multimedia presentations.

Introduction

The 21st Century is posing many challenges for teachers and students. It is widely popular that the present day teacher has to include essential skills of the 21st Century to facilitate learning such that it becomes effective. These essential skills are Critical Thinking and Problem Solving, Communication, Creativity and Innovation, Collaboration, Information and Media Literacy, and Contextual Learning Skills. Teachers have used these skills for many years now but the real challenge of today's world is the need to package it such that it becomes a useful tool for the present and future generations. These skills prepare students for an increasingly complex life and future work environments. In order for us to be able to understand the essential skills, we have to first identify the present learning environments in place for our students. The learning environment for the present generation has to be tailored to suit their needs. This is important for real learning to be effective. If we want our students to have sound and agile minds, we need to help them achieve sound and agile bodies. It is a proven fact that a strong mind in a strong body make a better child. The child's physical needs are to be met together with socio-emotional needs.

Change is the only constant factor. Change in the education scene happens because of changing needs in an ever-changing world. Technological advancements are very rapid in the present age. Before a product or software has spent a little time on a shelf, a new updated version is ready to enter the market. That's how rapid the advancements are. In order to interact in such a dynamic environment, our students need to learn how to interact, identify and react to changes. They need to analyze new conditions and deal with these conditions. Conditions here refer to situations that students find themselves in.

Our students need to be taught the skills in order to be independent and self-directed learners who can take charge of their own learning process. We teachers need to show them how to 'feed themselves' by stopping the act of 'spoon-feeding'. When lessons are created to get them thinking for themselves, the process has started. It is very important that our students understand the need for life-long learning to keep abreast with the present day and preparing for future needs. As such lessons need to be structured in such a manner that the learning of these essential skills takes place effectively.

E-Books Creation

One of the challenges faced by Tamil teachers in Singapore is getting storybooks with a local context for our pupils' reading pleasure. Pupils get excited about stories with a storyline that they can relate to. However, there are not many books out in the market that captivate their interest.

In order to get them interested in reading, we started to create e-books written in the local context. This proved to be a success as pupils started looking forward to these story times. They even began to add in ideas to elaborate on the original stories. To capitalize on this interest, we started ICT based lessons for the pupils to create e-books on their own. Our school, Pei Tong Primary, uses KooBits Software to create the e-books.

KooBits enables the creation of E books with videos and presentations with engaging animation and interactive content. It encourages children to write spontaneously, think critically and create passionately, which in turn helps to nurture them into confident and self-motivated young writers. Before using Koobits, the pupils need to plan the book. They first create the plot and characters of their story. Then they source for pictures to fit the plot. KooBits also has a range of video and audio clips, clipart, animations and various background designs for the pupils to choose from. Next, pupils create a new book by inserting background, pictures, animations, audio clips or video if any. The pupils enjoyed watching their stories take the shape of an e-book. It excited them further to see their stories in Tamil.

All the Primary4 pupils took turns to present their stories in class. It was a successful attempt and they decided to show and tell their stories to the Primary1 pupils as well. Their common timeslot for lessons allowed the cross-level presentation of their stories. Not only were the pupils able to create e-books in Tamil, they also had a range of stories to read (and their own creations!). Furthermore, they had the confidence to present their stories to the younger children in Primary 1. It was heartening to see the Upper Primary pupils telling their stories in Tamil and encouraging the younger children to read as well.

The learning points for the pupils start with the basic skills of speaking, listening, reading and writing. They enhance their thinking skills and creativity as they source for resource materials and create their stories. The lesson also helps to enhance their visual and digital skills. When the pupils are searching for materials, they are taught how to source for materials from reliable sources.

Online Blogging

Pei Tong Primary uses an e-learning portal named AskNLearn. This portal has an educational blog, edublog, on it. Teachers at Pei Tong Primary prefer to use this blog site for pupils. The safety of the regulated site is of primary concern for the teachers. Last year, the Tamil teachers tried using edublog to start a forum for reflections. The question posed on the site was related to National Education, "What would you do or say to attract tourists to Singapore?". The pupils' responses were very encouraging as that could speak their mind freely when they keyed in their thoughts. Although they

knew that their teacher is going to read their posted comments, the pupils were more comfortable with keying in their responses as compared to saying it face-to-face. Their responses ranged from Singapore's economical progress to the F1 race. The pupils were proud of producing a blog in Tamil. Finally they were able to connect with the digital age and Tamil was no longer the language which was destined to be obsolete.

Many of the Upper Primary pupils started posting blogs to their friends in Tamil. Some of them even posted blogs in Tamil to non-Tamil speaking friends just to show them that they were able to blog in Tamil. The sense of pride in using their mother tongue served to enhance their interest in the language. As a by-product, their grades improved as well.

Conclusion

The main objective of changing the mindset of Tamil being a dormant language to Tamil the fun-to-use language has been achieved. Now, to sustain the interest is the real challenge. New and innovative methods have to be sourced and implemented so as to sustain the pupils' interest in using the Tamil language.

Many pupils find it increasingly difficult to speak in Tamil as they are raised in an English-speaking environment. As such reading and writing is affected as they have difficulties in using the language in context. In order to alleviate this problem, IT is being used as the "carrot stick" to entice children to using Tamil the fun way so as to sustain it as a living language.

Alvin Toffler wrote that "the illiterate of the 21st century will not be those who cannot read and write, but those who cannot learn, unlearn, and relearn". Therefore, we must teach our children to learn, unlearn and relearn the Tamil language so that it continues to live with the future generations.

References

1. Educational E-Learning Portal - http://primary.asknlearn.com
2. Koobits Author Software - http://www.koobits.com/
3. Partnership for 21st Century Skills - http://www.21stcenturyskills.org/

The Use of Multi Media in Teaching Tamil through Internet

R.Subramani, Ph D
Assistant Professor, Department of Journalism and Mass Communication
Periyar University, Salem-636 011, Tamil Nadu, India
Mobile: 9444204387
Email: subbu_mathi71@yahoo.co.in/erasubramani@gmail.com

Abstract: The tremendous growth that we witness in the present media scenario is the transformation of print media to audio media and then its evolution to the visual media. It is the onus and obligatory as well to exalt the media language and flourish it according to its usage. As the advancement in technology has reached its crescendo, the need for Tamil language teaching has augmented drastically. The traditional method of teaching Tamil demands more of human labour and economy. So, keeping in mind the growth of Computer, Internet and Multimedia can be used for imparting Tamil language education. This research analyses, how in this state of affair, two dimensional and three dimensional animations, audio , video, comics, comic strips and motion pictures can be used for teaching Tamil through internet medium. This paper has made an attempt to identify the effective use of open source software's and operating system for imparting Tamil language skills.

Introduction

The need for imparting Tamil through various methods has become an upsurge for Tamilians and the people who love Tamil who dwell all over the world. To perform this task, the assistance of internet is required to expand the conventional media content all over the world. Therefore, the need has arisen to increase the usage of Tamil through internet and to upgrade its service. The grammar and science and technology books in Tamil which are only in print as far as to move beyond its medium for which computer's programmes, fonts, keyboard, operating system, and usage of Unicode are the prerequisites. To convert printed magazines and books into e-books, portable document format, and dot net, the compatibility among these formats are needed at this hour.

Internet which is called as media comprises in itself innumerable features. It is the need of the hour to find out the means of using all those features for spreading the Tamil Language. A number of hindrances exist in the medium which prevent the imparting of Tamil language. But today technologies such as 2D and 3D animations, Video, Audio, Comics, Comic stripes, Motion films have made the teaching of Tamil language incredibly easy. In internet, software's and operating systems like Ubuntu(Tamil Linux),Microsoft, Microsoft Tamil office, suratha Unicode writer and converter, Google search engine and guruji search engine greatly help for the development of Tamil language. 2D and 3D software's and audio and video editing software's like Photoshop, Coral draw, Macro media, Flash, Dream weaver, Audacity, Window moviemaker, etc. serve the purpose of teaching Tamil to great extent. This research explores what are practical difficulties involved in making use of these software for teaching of Tamil language. In Tamil teaching through computer, first, shows on the screen simple letters and then using the mouse arrow shows them how to write those letter followed by the pronunciation. This type of leaning can be done through 2D and 3D animations and

motion pictures which will create a greater impact on the learners. Teaching Tamil grammar and literatures can be effectively done through audio visual media.

Basics of Tamil grammar can be shared in the form of audio, video, comics, comic strips and portable document format (PDF) in the internet. If we have basic Tamil fonts with various options like Unicode writer and converter, compatibility with 2D and 3D software's may have greater impact in Tamil teaching in the internet. With the help of open source 2D software like Inside Point, Art Range and 3D software like Motion Builders, Bryce 5.5 and with assistance of social networking websites like Back Flip, Blog Mark, Dig, Stumble Upon and with the help of animation software Tamil education related contents can be shared through audio files, photographs and videos.

Role of Multi Media in teaching Tamil

Internet comprises of various media within itself. The significant of the internet is that we can find audio, audio visual and text all under one umbrella. This multi-benefit internet has created a new dimension in the present world. Recent advances in computer technology now allow the delivery of digital audio and video in the same interface of a written script (Brett: 1997). As it is in the case of text, studies of motivation and the use of Multimedia or interactive video have demonstrated positive effects (Brett, 1996, Watts: 1989). In the same note the usage of media content and the capacity of video in an interactive manner help for better understanding among the learners.

As our life under the captivity of audio visual media, the same medium can be used as the methodology for teaching the language. Researchers have concluded that this sort of learning has crafted admirable impact on the learners. Using computer, video, and Internet-based materials ,in educational activities, eases teachers' class-management problems, increases students' and teachers' attention levels, and enhances the learning-and-teaching process's effectiveness (Melvin & Horton, 1996; Deborah, 1998; Christine, 1999; Beers, Paquette, & Warren, 2000; Kablan, 2001). This type of learning can create a different experience. Videos are compatible with constructive education due to their potential to bring real-life situations and problems into classrooms, where they are widely used (Hult & Edents, 2003; Friel & Carboni, 2000; Daniel, 1996; Cannings & Talley, 2003; Bucalos, 2003). When we are teaching a language through various medium, we can utilize text, audio and video. Through this type of learning a close relationship can be maintained.

Teaching Tamil Language through Audio Medium

The practice of oral tradition in Tamil language is in existence for thousands of years. From generation to generation our traditional art forms like Theru Kutthu, Tholpavai and Mayilattam, and oyilattam exist purely through the practice of oral tradition. The literatures that we read today were once learned through oral communication. And Tamil grammar itself was taught through oral language. Thousands of years ago, Tholkapiam was recited only through oral language. Only later it took on the written form. Today, various medium helps to maintain the oral tradition for teaching Tamil language. Sound recording in these days is very simple and hardly costs much. Recorded audio either by mobile phone or computer can be edited and modified using the software like Audacity, Hammer Head Rhythm Station, Audio book Cutter Free Edition, Eca sound, Free cycle, Flexi Music, Wave Editor, Gold wave, Jokosher, Media Digitalizer, Mp3splt,MP3 Stream Editor, Aviary Myna, n-Track Studio, NU-Tech, Pyramix, Quick Audio, Reaper, Sample Wrench, Sound booth , Sound scape 32,SoX,Total Recorder, Wave Lab, Wave Pad and Wave Surfer. If texts are given in audio file format it will have greater impact on the listeners. This software can be used for teaching Tamil through high quality audio files.

Using audio medium, grammar, pronunciation, tongue movements can be accurately recorded. With the help of listeners' own language audio files can be produced. And these audio files can be transferred into the modern gadgets like cell phone for further usage. At the same time, our languages oral tradition like proverbs, sayings, folk stories, folk songs, traditional epics, Tamil medical news and traditional agriculture can also be made available into this audio format. Through this the prosperity of Tamil language can be made known to the world. It is believed that this type of audio file format helps in recording the apt traditional pronunciations, local slang, rising and lowering tones and expressions, thus bringing life to the language. But these can be done only as document. The significance of the oral tradition is that it has undergone changes in different stages and still it has not lost its importance. Therefore to teach a language using this format will always be appropriate.

Teaching Tamil through Video

The teaching of visual communication has got immense importance and it creates a greater impact. Today almost every one of us has witnessed still camera and video recorder. In recent times the number of people who use camera in their mobile phones has considerably increased. Now it has become possible to teach language using video camera to those who know Tamil and to those who do not know Tamil. If one appears in the video camera and starts reading the lessons like traditional method, without any doubt it is not going to bear fruit in any way. But if video combined with text created by using software's like flash, truly it will have heaps of benefits. This type of video can be easily edited using open source software. For example, the software Wax has the capacity to add 2D and 3D effects to the video during the editing process. This software is user-friendly and even beginners can work on it. AVI WMV MPEG MP4 MOV Converter can be used for converting the formats like FLV, AVI, MP4, MPEG, WMV, ASF, MOV to formats like AVI, MP4, WMV,VCD, SVCD, DVD, MOV format Abcc FLV without much difficulty. This software can also be used for creating contents for the video.

Adobe Media Player is another type of video editing software. This software helps view internet video contents and television programmes in internet. Combi Movie is another player which helps MPG/MPEG files to arrange them into a single sequence of all MPEG files. This software works fast. Using video for teaching Tamil will create transformation on the learners. Therefore text and audio visual should be blended together for the effective learning.

Role of Comics and Comic strips in teaching Tamil

In teaching Tamil language, more than merely dispatching information, if it is combined with 2D and 3D pictures and motion pictures then it reaches the learners powerfully and might have greater impact on the learners.

For example, software helps us in a great in the following works.

- creating models to describe a hypothesis found in Tamil education
- demonstrating science and technology through explanatory pictures
- to draw diagrams related to mathematics and economics
- to explain an activity through motion pictures
- to narrate a comic stories
- to draw cartoons
- to create 2D and 3D animation movies according to the story

According to the medium open source software can be used for creating 2D, 3D and motion pictures.

Motion Builders (Personal Learning Edition) - helps professionals and 3D animators, Even this software can be used for teaching Tamil language through 3D animation.

Bryce 3.5 is another kind of 3D animation software which goes very handy even for the new users, thus making possible use of 3D models. 3D models too can be easily created for teaching Tamil language.

Scribers Desktop Publishing is a commanding software. Using this software, all sorts of texts can be created. This software will be of great help in educating Tamil through 2D animation.

Smooth Draw NX is 2D animation software. High quality drawing and hand drawings can be made from this software. A number of tools like brush, pencil, pen tool are available in this software.

Pixia is used for drawing. In this software there are various tools layers, brushes, masking tools, light correction included for better quality outputs.

Insight Point can be used for internet and graphics. Using this software our thought can be brought into texts and graphics. High quality bit map pictures and internet related visual trips can be created. This software can also be of immense assistant to make 2D and 3D graphics for teaching Tamil language. Software as these mentioned below serves good purpose like creating 2D and 3D animations and pictures. Along with audio, writing system can be taught in step by step to create interest in the mind of listeners.

Teaching Tamil through internet

Internet acts as bridge for connecting all the media together. In today's scenario, web blogs and social net working websites works as a major source of alternative mass media. Anyone can create his/her own blog or become member of any of the social networking websites and share his/her opinions with much ease. Breaking the boundary walls, internet connects the whole world and it has opened a large way for Tamil education worldwide.

Web blogs and social networking web sites play a significant role in imparting Tamil education all over the world in different forms-audio, video, 2D, 3D and motion pictures, etc. One can easily share what he has in his mobile phone, computer and recorder to a large number of people through web blogs and social networking websites. Even more, using third generation technology cell phones, one can share Tamil education contents through internet. Now, there is more scope for web blogs and social networking websites being translated into other languages. Following are the some of the blogs and social networking websites which can be used for teaching Tamil language.

Aim- this social networking websites can be used for sharing information on line. For the purpose of Tamil education online quiz can be conducted and shared.

Delicious- This website is widely used for compiling and sharing social book marks. Through this web site Tamil education can be effectively done by compiling and sharing Tamil education related book marks worldwide.

Ask-This website acts as search engine for finding out websites, pictures, news, maps, local and business related news. This website is also helpful in finding out Unicode in Tamil.

Back file-Tamil teaching related texts, audio files, photographs, and video can be easily shared through this website. A visitor to this website has the option to leave their comments and in turn the website authorities reject or edit or reply to the received comments. Since the content of this web site is

arranged according to the year and topic it is easy for the visitors to collect information. These days, many compilers edit and publish the Tamil blogs which are famous worldwide. All these will help in large for Tamil teaching. Anybody who has basic knowledge about computer can access internet Tamil content with the help of Unicode.

Through social networking websites like buzz, yahoo.com, Digg, digg.com, diigo.com, fark.com, faves.com, friendfeed.com, kirtsy.com, linkagogo.com, linkedin.com, live.com, mister-wong.com, mixx.com, multiply.com, myspace.com, netvouz.com, facebook.com, stumbleupon.com, texts, videos, audio files, photographs, sports and feed backs can be easily accessed and shared using these websites. Tamil education related World Wide Web blogs can be created with the help of this software. This website has the facility to give feedbacks on the contents of the web site. Online opinion can also be shared using this type of web site.

Challenges in using Multimedia in the internet

The compatibility among the operating system, fonts, key board, and usage of Unicode are the prerequisites for the successful operation of Tamil in internet medium. To convert printed magazines and books into e-books, portable document format, and dot net, the compatibility among these formats are needed at this hour. The compatibility among the format of audio files, players, and speed of the internet browser are ought to be analyzed. The speed of the computer is highly dependent on the speed of the internet browser, because of this reason; we are facing trouble in sharing the high-quality video files.

The capacity of the broadband, wireless, and wired network are the deciding factors of the Data exchange. Like English we do not have sound recognizer, signature recognizer, optical character recognizer, dictation writer, and live spell checkers in Tamil. This inadequacy is highly preventing the effective use of Multimedia in teaching Tamil through internet. Users are encountering problem in downloading and installing the Tamil conventional fonts in the system. We can solve this problem by using portable document format.

We ought to have titles, symbols or tags in Tamil and English search engines. If Tamil contents are inserted in the search engine, the users can easily avail the required information. If we offer the same content in many media the users may avail it easily. The cell phone technology and internet technology are acting as an interface for Multimedia. We can make use the features of these available facilities. We may think about the effective utilization of Mobile streaming, mobile video streaming, and internet video streaming for teaching Tamil.

Conclusion

Many Scholars have articulated their research discourse pertaining to Tamil teaching through Internet, and submit their recommendations and shortcomings in the usage of Tamil in the new media. At the onset this paper has also made an attempt to identify the existing Multimedia open source software to teach Tamil through internet. Based on the exploratory approach some suggestions have been codified in the effective use of two dimensional and three dimensional animations, audio , video, comics, comic strips and motion pictures in teaching Tamil through internet medium, but certain guidelines ought to be laid down in the preparation of lesson plan. If individuals start devising lesson plan, learners may encounter many troubles. Many English language portals have been offering basic grammar with multimedia features to the global learners. They are effectively making use the multimedia software's for teaching English to the native and non native learners. We can also replicate the successful models

in our Tamil portals. Multimedia to be an effective tool for the Promotion of learning which provides for constructive source of knowledge of language, Better learning is not an outcome of better ways of instruction; rather, it is a product of more opportunities to construct knowledge of language (Papert: 1993). We must make sure the difference between the learning from technology and Leaning with the technology. Device can offer information, but it has to be used to posses as a cognitive tool in the construction of knowledge acquisition process. The academics and technocrats have jointly initiated discussion and put forward constructive suggestions to the Tamil community.

References

1. Bennett, *Are video lectures an effective technology tool?* Department of creative technology, University of parts mouth, UK
2. Claire Kramsch and Roger W.Anderson (1999), *Teaching text and context through Multimedia*, Language learning and technology, Vol.2.No.2, www.IIt.msu,ed
3. Devaki.L, *Language learning from or with Multimedia*, International journal of Dravidian linguistics, Vol.xxxi.No.2
4. Papert.S(1997) *The children's machine: Rethinking school in the Age of computer*, New York, Basic books
5. Paul Brett, (1997). *A comparative study of the effects of the use of Multimedia on listening comprehension*, Elsevier science ltd
6. Yahhui liv, (2007), *Pragmatic awareness in Multimedia based language Teaching*, school of foreign language &culture, Ningxia University, China
7. Ozdener, N. & Esfer, S. (2009). *A comparative study on the use of information technologies in the development of students' ability to comprehend what they listen to and watch*, International Journal of Human Sciences, Available: http://www.insanbilimleri.com/en

TAMIL DIASPORA:

TEACHING TAMIL

AS A SECOND LANGUAGE

AND IMPACT OF TECHNOLOGY

தமிழ் இணையப் பல்கலைக்கழகம்

செயல்பாடுகளும் - சவால்களும்

முனைவர். ப.அர. நக்கீரன்
இயக்குநர், தமிழ் இணையப் பல்கலைக்கழகம், சென்னை

சுருக்கவுரை: உலகு தழுவி வாழும் தமிழர்கள் தமிழோடும், தமிழ்ப் பண்பாட்டோடும் தொடர்பறாது வாழ வேண்டும்; தமிழ்க் குழந்தைகள் தமிழ் கற்க வேண்டும்; தமிழில் பேச வேண்டும் என்ற நோக்கில் தொடங்கப்பட்டதுதான் தமிழ் இணையப் பல்கலைக்கழகம். அது தமிழை, சான்றிதழ்க் கல்வி, பட்டயம் / பட்டக்கல்வி, மேற்பட்டக் கல்வி, ஆராய்ச்சி என்ற நிலைகளில் கொண்டு செல்ல முயன்று வருகிறது. இம்முயற்சியில் அது சந்தித்த சிக்கல்களும், அதற்குரிய தீர்வுகளும் இக் கட்டுரையில் எடுத்துக் கூறப்பட்டுள்ளன.

முன்னுரை

உலகு தழுவி வாழும் தமிழ் மக்களும், தமிழில் ஈடுபாடு உள்ள ஏனையவர்களும், தமிழ் மொழியைக் கற்பதற்கும், தமிழர் வரலாறு, கலை, இலக்கியம், பண்பாடு பற்றி அறிந்து கொள்வதற்கும், வேண்டிய வாய்ப்புகளை இணையம் வழியாக அளிப்பது தமிழ் இணையப் பல்கலைக்கழகத்தின் நோக்கமாகும்

தமிழ் பயில்விக்கும் பள்ளிகள் இல்லாத ஊர்கள் / நாடுகளில் வாழ்கிற மழலைக் குழந்தைகள், பள்ளிக் கல்வி முடித்தவர்கள், ஓரளவு தமிழ் அறிவு பெற்றிருந்து, மேலும் படித்துத் தமிழில் பட்டம் பெற விரும்புபவர்கள், தமிழர் பண்பாடு, கலை, பாரம்பரியம் முதலியவற்றை அறிந்து கொள்வதில் ஆர்வமுள்ளவர்கள் ஆகியோர்தாம் இப்பல்கலைக்கழகம் எதிர்நோக்கிய பயனாளர்கள்.

இதன் நோக்கங்களையும், பயனாளர்களையும் மனத்தில் கொண்டு தமிழ்க்கல்வி, குழந்தைகளுக்கான மழலைக்கல்வி என்றும் ஒன்று முதல் ஆறாம் வகுப்பு வரை சான்றிதழ்க் கல்வி என்று மூன்று நிலைகளிலும், ஏழாம் வகுப்பு முதல் - பன்னிரெண்டாம் வகுப்பு வரை மேற்சான்றிதழ்க் கல்வி என்று மூன்று நிலைகளிலும் பட்டயம், மேற்பட்டயம், பட்டம் என்று ஒருங்கிணைந்த இளநிலைத் தமிழியல் கல்வியும் இப்பொழுது வழங்கப்படுகின்றன. மேற்பட்டக் கல்வி வழங்கவும் திட்டங்கள் உள்ளன.

சான்றிதழ்க் கல்விப் பாடங்கள் அனைத்தும், கேட்டல், பேசுதல், படித்தல், எழுதுதல் ஆகிய திறன்களை வளர்க்கும் வகையில் வடிவமைக்கப்பட்டு இணையத்தளத்தில் கொடுக்கப்பட்டுள்ளன. இதே போன்று பட்டக்கல்விப் பாடங்களும் பாடச் சுருக்கம், பாடப்பனுவல், பாடல்கள் ஒளி-ஒலிக் காட்சிகள், தன்மதிப்பீட்டு வினாக்கள் ஆகிய பகுதிகளுடன் வடிவமைக்கப்பட்டுள்ளன. இப்பாடங்கள் எல்லாம், ஆசிரியர் துணையின்றி மாணவர்கள் தாமே கற்கும் வகையில், அசைவுப் படங்கள், இசைப் பாடல்கள், ஒலி-ஒளிக் காட்சிகள் ஆகிய பல்லூடக வசதிகளுடன் தரப்பட்டுள்ளன. சான்றிதழ்க் கல்வி வாய்மொழித் தேர்வு, காட்சித் தேர்வு, இணையவழித் தேர்வு, எழுத்துத் தேர்வு என்பவை மூலமும் பட்டக்கல்வி, இணையவழித் தேர்வு, எழுத்துத்தேர்வு என்பவை மூலமும் மதிப்பீடு செய்யப்படுகின்றன.

இணையம் வழிக் கல்வியின் நன்மைகள், குறைபாடுகள், வாய்ப்புகள், சவால்கள்

நன்மைகள் **(Strengths)**

1. உலகின் எந்த மூலையில் உள்ளவர்களும் இக்கல்வித் திட்டங்களில் சேர்ந்து படிக்கலாம்.
2. ஆசிரியர் துணையின்றி கற்கும் வகையில், பாடங்கள் அனைத்தும் அசைவுப் படங்கள், இசைப் பாடல்கள், ஒலி-ஒளிக் காட்சிகள் போன்ற பல்லூடக வசதிகளுடன் வடிவமைக்கப்பட்டுள்ளன.
3. அவரவர் வீடுகளிலிருந்தே படிக்கலாம்; எப்பொழுது வேண்டுமானாலும் நேரம் - காலம் பார்க்காமல் படிக்கலாம்.

4. சிறுவர் முதல் பெரியோர் வரை, பணிக்குச் செல்வோர், செல்லாதவர், வீட்டில் இருக்கும் பெண்கள் என்று யார் வேண்டுமானாலும் படிக்கலாம்.
5. வயது வரம்பு இல்லை.
6. அவரவர் திறமைக்கேற்பக் கல்வித் திட்டங்களில் சேர்ந்து தொடர்ந்து படித்துக் கொண்டிருக்கலாம்.
7. மாணவராகப் பதிவு செய்து கொண்டபிறகு எப்பொழுது வேண்டுமானாலும் தேர்வு எழுதிக் கொள்ளலாம். காலக்கெடு எதுவும் இல்லை.
8. சான்றிதழ்/பட்டம் என்றில்லாமல் அறிவு வளர்ச்சிக்காகவும் படிக்கலாம்; படிப்பதற்குக் கட்டணம் ஏதுமில்லை.
9. தன் மதிப்பீட்டு வினாக்கள் மூலம் கற்றதை மதிப்பீடு செய்து கொள்ளலாம்.

***குறைபாடுகள்* (Weaknesses)**

1. இணையத் தொடர்புடன் கூடிய கணிப்பொறி தேவை.
2. ஐயங்களைத் தீர்க்க ஆசிரியர் இல்லை.
3. இணையத் தொடர்பின் வேகம் குறைவு.
4. தேர்வுகளை ஒரு தொடர்பு மையத்திற்குச் சென்றுதான் எழுத வேண்டும்; தொடர்பு மையங்கள் பக்கத்தில் இருக்க வேண்டும். அல்லது ஒவ்வொரு ஊரிலும் ஒரு தொடர்பு மையம் ஏற்படுத்த வேண்டும்.
5. பள்ளிக்கூடங்களில் உள்ளதைப் போலக் 'குழுவாக சேர்ந்து கற்றல்' என்பது முடியாது.
6. ஆசிரியர் தேவையில்லை என்று சொன்னாலும், குழந்தைகளை ஊக்கப்படுத்தி படிக்க வைக்க பெற்றோர்களின் துணை தேவைப்படும்.
7. தமிழை மட்டும் சொல்லிக் கொடுப்பதால் பள்ளிக் கல்வி முழுமை அடைவதில்லை.
8. கணிப்பொறித் திரையில் தொடர்ந்து பாடங்களைப் படிக்க முடியாது; தேர்வுக்குத் தயார் செய்ய முடியாது; அதனால் பாடப் புத்தகங்கள் வேண்டும்.
9. ஆசிரியர் இல்லாமல் கல்வி முழுமையடையாது.
10. தமிழர்கள் வாழும் நாடுகளின் அங்கீகாரமும் ஒப்புதலும் பெற வேண்டும்.

***வாய்ப்புகள்* (Opportunities)**

1. வளர்ந்து வரும் தகவல் தொழில் நுட்பத்தால், கணினியின் விலை குறையும்; எளிதில் அனைவருக்கும் கிடைக்கும்; பயன்பாடு பெருகும்.
2. இணையத் தொடர்பும் (Internet connection) செயற்கைக்கோள் வழியாக, கம்பி இணைப்பு இல்லாமல் எளிதில் கிடைக்கலாம்.
3. அதனால் செயற்கைக்கோள் வழியாக 'இணைய வகுப்பறைகளில்' பாடங்கள் நடத்தலாம்.
4. இதனால் சிறந்த ஆசிரியர்களின் பயன் ஒரே நேரத்தில் பல பள்ளிகளுக்குக் கிடைக்கலாம்.
5. பயிற்சி பெற்ற ஆசிரியர் இல்லை என்ற குறை நீக்கப்படலாம்.
6. வருங்காலத்தில் பள்ளி மாணவர்கள் மடிக் கணினிகளையும், கைக் கணினிகளையும் பயன்படுத்திக் கற்கலாம்.
7. மின்நூலகங்களில் உள்ள நூல்களை மாணவர்கள் தேடிப் பார்க்கலாம்.
8. மின்னஞ்சல் மூலம் ஐயங்களுக்குச் சிறந்த ஆசிரியர்களிடமிருந்து விளக்கங்கள் பெறலாம்.
9. மொழி கடந்து, பார்வை நூல்களையும், மற்ற தகவல்களையும் பெறலாம்.
10. கணிப்பொறியிலேயே சோதனைகள் செய்யலாம்.
11. தேர்வு முறைகள் இல்லாமல் போகலாம்; மதிப்பீட்டு முறைகள் மாறலாம்.

***சவால்கள்* (Threats)**

1. ஏட்டுக்கல்விக்கு அப்பாற்பட்டு மானுட ஆக்கத்திற்குத் தேவைப்படும் ஆசிரியரின் அறிவுரைகள், அனுபவங்கள் இல்லாமல் போகலாம்.
2. அன்பு, பாசம், நட்பு, விளையாட்டு, சிறுவயதுக் குறும்புகள், உறவுப் பாலம் ஆகியன குறையலாம்.
3. மனித நேயம் மறைந்து போய், எந்திரங்களாய் மனிதன் மாறலாம்.
4. கற்றதைப் புரிந்துகொண்டு பயன்படுத்தும் அறிவு இல்லாமல் போகலாம்.
5. ஒவ்வொரு மனிதனும் தனித்தனித் தீவுகளாய் மாறிப் போகலாம்.

த.இ.ப. எதிர்கொண்ட சிக்கல்களும் தீர்வுகளும்

கல்வித் திட்டம்

1. பாட ஆசிரியர்களிடமிருந்து பாடங்களைப் பெற்று, சரிபார்த்துச் செப்பனிட்டுத் தளத்தில் இடுவதற்கு நீண்ட காலம் ஆனது.
2. ஒவ்வொரு பாடமும் தனித்தனியாகத் தணிக்கை செய்யப்படுவதால், ஒரு பாடத்திற்கும், மற்றொரு பாடத்திற்கும் இடையில் ஏற்படும் கருத்து முரண்பாடுகள், கூறியது கூறல், பொருள் குற்றம், சொல் குற்றம் ஆகியவை நீக்கப்படுவதில்லை.
3. பாடங்களில் சீர்மையில்லை என்பதோடு, பாடங்களைத் தளப்படுத்தியதிலும் சீர்மை காணப்படவில்லை. ஒரே நேரத்தில் பல நிறுவனங்கள் இப்பணியைச் செய்ததும் ஒரு காரணமாக இருக்கலாம்.

எனவே இக்குறைகளைக் களைய வேண்டியது உடனடித் தேவையாக இருக்கிறது. இதன் முதல் படியாக எல்லாப் பாடங்களும் சீராய்வுக்கு உட்படுத்தப்பட்டன. கூறியது கூறல் போன்ற குறைகளைக் களைந்த பிறகு முதலில் 27 தாள்களாகவும் பின்னர், 24தாள்களாகவும் குறைக்கப்பட்டன. இந்த 24 தாள்களும் மீண்டும் செம்மைப்படுத்தப்பட்டுக் கொண்டிருக்கின்றன.

பாடப்புத்தகங்களின் தேவை

இணையவழிக் கல்வியின் ஒரு நன்மையாகக் கூறப்படுவது பாடப்புத்தகங்கள் தேவையில்லை என்பதுதான். பாடங்கள் எல்லாம் இணையத் தளத்தில் இடப்பட்டிருக்கும் என்பதால், யார் வேண்டுமானாலும், எப்பொழுது வேண்டுமானாலும் அவற்றைப் படிக்கலாம். பல்லூட வசதிகளோடு பாடங்கள் போடப்பட்டிருப்பதால் ஆசிரியர் தேவையில்லை என்பது இன்னொரு நன்மையாகக் கூறப்படுகிறது.

தளத்தில் பாடங்களைப் படிப்பதற்குத் தேவையான முக்கியக் கூறுகள்:

1. கணிப்பொறி
2. இணையத்தள இணைப்பு
3. தொடர்பறாத, எப்பொழுதும் கிடைக்கும் வகையில் இணையத்தளத்தின் வேகம்

சில முன்னேறிய நாடுகளில் கூட அனைவருக்கும் கணிப்பொறி இல்லை; இணையத்தள இணைப்பு இல்லை; அப்படியே இருந்தாலும் வேகம் இல்லை. அப்படியென்றால் மற்ற நாடுகளின் நிலைமை எப்படியிருக்கும். மேலும் எவ்வளவு நேரம் கணிப்பொறியில் பாடங்களைப் படிக்க முடியும்? அதற்கான கட்டணம் என்ன? என்ற கேள்விகளும் எழுகின்றன.

எனவே பாடங்களை ஓரளவுக்குப் புரிந்து கொள்ள இணையத்தளப் பாடங்கள் பயன்பட்டாலும், தொடர்ந்து படிக்கவும், தேர்வுக்குத் தயார் செய்யவும் புத்தகங்கள் மிகவும் தேவை என்பது பலருடைய கருத்தாகும். எனவே த.இ.ப கல்வித்திட்டப் பாடங்களுக்குப் பாடநூல்கள் தயாரித்து, PDF வடிவில்

தளத்தில் இடும் பணி தொடங்கியுள்ளது. தேவைப்படுவோர் இவற்றைப் பதிவிறக்கம் செய்து கொண்டு படிக்கலாம்.

இணைய வகுப்பறை

ஆசிரியர் துணையின்றி, மாணவர்கள் தாமே படித்து புரிந்துகொள்ளும் வகையில் பாடங்கள் வடிவமைக்கப்பட்டிருக்கின்றன என்று கூறினாலும் அது முழுமையாகப் பாடங்களைப் புரிந்துகொள்ளத் துணைபுரிவதில்லை. கேள்விகளுக்கு விளக்கங்கள் தேவைப்படுகின்றன எனவே இதற்கு ஒரு ஆசிரியரின் துணை நிச்சயம் தேவைப்படுகிறது. ஆசிரியரின் அனுபவங்கள் இல்லாத கல்வி முழுமை அடைவதில்லை. இக்குறையைப்போக்க இப்பொழுது த.இ.ப. இணையவகுப்பறையைத் தொடங்கியுள்ளது. இவ்வகுப்பறைகளில் சிறந்த ஆசிரியர்கள் பாடங்களை நடத்துகிறார்கள். அவை த.இ.ப. இணையத்தளத்திலும் போடப்பட்டுள்ளன. யார் வேண்டுமானாலும், எப்பொழுது வேண்டுமானாலும் பாடங்களைக் கேட்டுப் பயன்பெறலாம். கேள்விகளையும் விளக்கங்களையும் மின்னஞ்சல் மூலம் கேட்டுத் தெளிவு பெறலாம்.

தேர்வுகள்

இணையவழிக் கல்வி எதிர்கொள்ளும் இன்னொரு சிக்கல் தேர்வுகள், இன்றைய தேர்வு முறைகளின்படி, இணையவழி மாணவர்கள் இரண்டு நேரடி கணிப்பொறி வழித் தேர்வுகளையும், ஒரு எழுத்துத் தேர்வையும் ஏற்க வேண்டும். இவற்றை அவர்கள் தங்களின் இல்லங்களில் இருந்து செய்ய முடியாது. ஒரு தொடர்பு மையம் வேண்டும். தொடர்பு மையங்களுக்கு வந்துதான் அவர்கள் தேர்வுகள் எழுத வேண்டும். அப்படியென்றால் ஒவ்வொரு ஊருக்கும் ஒரு தொடர்பு மையம் வேண்டும்; அதற்கு அந்தந்த ஊரில் உள்ள சமூக ஆர்வலர்கள், ஆசிரியர்கள் முன்வர வேண்டும்.

ஒரு நாட்டிற்கு ஒரு மையம் என்பது ஏற்கத்தக்க ஒன்று அன்று. எனவே ஒன்று எல்லா ஊர்களிலும் தொடர்பு மையம் ஏற்படுத்த வேண்டும் அல்லது, தேர்வுகள் தேவைப்படாத ஒரு மதிப்பீட்டுத் திட்டத்தை உருவாக்க வேண்டும்.

அரசுகளின் ஒப்புதல்

தமிழ் இணையப் பல்கலைக்கழகத்தைத் தொடங்கியபோது, மலேசியா, சிங்கப்பூர் போன்ற தமிழர் அதிகம் வாழும் இடங்களில் இருந்து இதற்கு பெரும் வரவேற்பு கிடைக்கும் என்று எதிர்பார்க்கப்பட்டது. ஆனால் அது நடக்கவில்லை. அதற்கு முக்கியமான காரணம், இக்கல்வியை அந்த நாடுகள் இன்னும் அங்கீரிக்க வில்லை. எனவே அரசியல் சார்ந்த, அரசு சார்ந்த சில முடிவுகளையும் எடுக்க வேண்டியுள்ளது.

முடிவுரை

இணையக் கல்வி என்பது உலகின் முன்னோடித் திட்டம். எதிர்கால முன்னேற்றங்களையும் தேவைகளையும் கருத்தில் கொண்டு, மிகச்சிறந்த சிந்தனையாளர்களால் முன்வைக்கப்பட்ட ஒரு திட்டம். இப்படி ஒரு கல்வித்திட்டம், த.இ.ப. தொடங்கப்பட்ட காலத்தில் இல்லை. எனவே எந்த ஒரு வழிகாட்டுதலும் இல்லாமல் மிகச் சிறந்த முறையில் உருவாக்கப்பட்டுள்ள இப்பாடத்திட்டத்தில், ஏற்பட்ட சில சிக்கல்கள் கால ஓட்டத்தில் கண்டறியப்பட்டு, அவற்றிற்கான தீர்வுகளும் மேற்கொள்ளப்பட்டு வருகின்றன. இது முடிவுள்ள பணியன்று, தொடர்ந்து சீராய்வுகள் செய்யப்பட்டு, திருத்தங்கள் மேற்கொள்ளப்பட்டுச் செம்மைப்படுத்தப்படவேண்டும்.

Enhancing creativity through Multimedia
A study in malaysian tamil schools

Paramasivam Muthusamy, PhD
param@fbmk.upm.edu.my
University Putra Malaysia

&

Kanthimathi Letchumanan, MA
kanthi65@hotmail.com

Abstract: Communication and Information Technology (ICT) have made deep inroads to teaching and learning among the students. In Malaysia there are 524 Tamil schools and 90% of these schools have been equipped with computer laboratories. School Curricula have been modified to include ICT in order to upgrade teaching and learning. The computer with its internet and hypermedia capabilities is a powerful tool to enhance learning. With its unlimited collection of text, sound, pictures, video, animation and hypermedia provides meaningful context to facilitate comprehension (Bruner, 1986). The implementation of ICT in the classroom are both an innovation in technology and teaching (Scrimshaw, 2004).The study for this paper is being conducted in 10 Tamil schools in Klang Valley, Peninsula Malaysia to examine how far incorporation of ICT could promote students' creativity in their performance and suggest how teachers could use multimedia effectively. Questionnaire and interview methods will be used to collect data for the study. 100 students are selected by the respective schools. 50 students are grouped as experimental group and another 50 as the control group. The experimental group will be exposed to the multimedia during the process of language learning where students will be encouraged to play computer games in Tamil in the classroom. Meanwhile, the control group will be exposed to traditional approach while learning Tamil language. Both the group's performance will be analyzed based on their skill to write essays, richness in their vocabulary and appropriateness in pronunciation

Introduction

One of the most significant changes in education in recent years has been the availability of a range of Information Communication Technologies (ICT). Thus, ICT is no longer a new terminology nowadays. Almost everyone is familiar with ICT not only at work but also in schools as it encompasses the Internet and multimedia. The power of ICT can be used effectively in language teaching and leaning as there is a paradigm shift from traditional teaching to using ICT in classrooms. Besides that, the generation born after 1980 are named as digital mind and also known as N-Gen - Net Generation (Tapscott, 1998). These groups of students are highly influenced with Internet and have changed their learning attitudes and abilities (Adone at el., 2007). Computer is not an unfamiliar gadget to this group of students. Students feel that computers with the help of internet have helped to produce a good work. Furthermore, the computer with its multimedia effects, in the form of its unlimited collection of text, sound of pictures, video, animation and hypermedia provide meaningful context to facilitate comprehension (Bruner, 1986). Furthermore, multimedia tool is believed to

provide the possibilities of multiple perspectives and a realistic learning environment. The real power of multimedia to improve education may only be realized when students actively use them as cognitive tool. Furthermore computer based learning is more motivating for students and this is generally accepted by educators and by administrators.

ICT in Malaysian Schools

Under the Smart School Project, about 8000 schools in Malaysia have been equipped with computer facilities in 2005. By the year 2010, it is projected that about 10,000 primary and secondary schools will have computer facilities and more schools will obtain computer with Internet connection and teachers will be encouraged to use it in their classroom teaching (Malaysian Ministry of Education, 1997). ICT is aimed at producing students with knowledge, thinking skills and innovations, which eventually contribute to the knowledge-based economy (Economic Planning Unit, 2001). It can be said that the Malaysian government is spending a huge amount of money for the advancement of ICT use in schools. Thus, it is important to see that the ICT has been adopted in the schools.

ICT in Malaysian Tamil Schools

In Malaysia there are 524 Tamil schools and 90% of these schools have been equipped with computer laboratories. With this facility, Tamil school administrators have made it compulsory for teachers to inculcate ICT into their teaching especially in Science and Mathematics. Some teachers also included ICT in teaching Tamil. The advancement and usage of ICT through computer in schools are also due to the support from Non-Governmental Organizations (NGO) other bodies such as, Parents-Teachers Association (PTA) various community movements etc.

Although, the implementation of ICT in schools have been highly encouraged and publicized according to Mullai Ramaiya & Sudandra (2001), the impact of using computer in teaching and learning in Tamil schools in Malaysia is comparatively lower as against Chinese schools (p12). In the mean time, according to Paramasivam (2002), even in those Tamil schools where the computer laboratories are available there was no systematic teaching by using computer as an instructional tool. This was mainly because of the fact that the teachers/instructors did not have enough computer literacy in imparting computer skills to students.

This scenario slowly changed and a positive shape took place since 2003, when the Malaysian government has a change in curriculum policy. It was made compulsory for Science and Mathematics to be taught in English. Due to this concept, a lot of money was invested in the training of teachers to impart ICT knowledge in them. This gave teachers more opportunities to use the computer more effectively. Tamil teachers also took this golden opportunity to bring in changes into the teaching and learning of Tamil language using computers. Students are now exposed to use multimedia in learning. Multimedia language games have prompted students' interest in learning.

Multimedia and Student's Creativity

In a traditional classroom, teachers play a dominant role as they provide all the information to students. Students on the other hand, are passive learners and follow teacher's instruction. But today's world has changed so much when ICT is incorporated into teaching and learning. Students are now more active and they play an important role in learning with the help of the computers. Teachers are no longer dominant but mere facilitators in providing education. According to Papert (1996), young people's access to information is more interactive and non-sequential and they learn for the pleasure

and benefits of discovery. Due to the endless access to Internet they obtain a wide range of information and facts. Therefore, the penetration of ICT cannot be ignored and given the cost, should be used to support learning and teaching (Livingston and Londie, 2007). Besides that, Scrimshaw (2004) also points out that the implementation of ICT in the classroom is "both an innovation in technology and teaching" (p.9). On the other hand, multimedia is a combination features of text, graphic, art, sound, animation and video elements with facilities for interaction. Thus, multimedia is a powerful presentation tool, which can be effectively used for teaching. Studies showed that if students are stimulated with audio, they will have about 20 % retention rate, audio-visual is up to 30 % and in interactive multimedia presentation, the retention rate is up to 60% (Vaughan, 1997; p10). Hence, multimedia tools can enhance many skills such as, functional communication as a result of enriched vocabulary, critical and creative thinking. This article looks at how multimedia based language games obtained through several websites or in the form of Compact Discs (CD) had contributed significantly in the remarkable enhancement of language development among the Tamil students. Further, this article shows empirical evidence the differential achievement rate when compared to the students who have not used multimedia games in the learning process.

Multimedia Language Games

Online computer games are not only potential for engaging and entertaining the users, but also in promoting learning. Simon (1996) has noted how our outlook of learning has been changed from being able to recall information to being able to find and use information. Thus, computer can be an effective tool for enhancing learning even though the present generation students pass much time by playing online games (Turgut, 2009; p761).

Role play and games are used in language classrooms to let students practice language before they use it in the "real world". Video games are another avenue for "experimentation in a safe 'virtual environment'' (Kirriemuir, 2002) Learners may be hesitant to participate in language classes because of not wanting to make mistakes in front of their peers, but may be more willing to interact with video game in order to gain valuable linguistic feedback and practice with language before applying their knowledge in the "real world". As online games and CDs are highly interactive, they are able to give valuable linguistic feedback. For example, (moZhiviLaiyaaTTukaL-**Senthamizh**), Wordsmith and Oarsman (2003).

In some games, the players must vocally interact with the characters of the games via a microphone and use correct vocabulary, pronunciation or grammar as well as speak appropriately which can suit in the game's context. If the player's utterance is incorrect, these games will prompt the player to alter his/her pronunciation. This gives the player many opportunities to improve his/her speaking ability and pronunciation through implicit feedback.

Methodology

The study is being conducted in 10 Tamil schools in Klang Valley, Peninsular Malaysia. Ten students from Year Five, from each Tamil school will be taken as the subjects of the study (n= 100). Five students from each Tamil school will be exposed to traditional teaching while another five students will be exposed to multimedia based teaching strategies. Both the groups will be exposed to the identified teaching methodology for three months. The teachers will be trained in such a way that they can appropriately teach the experimental classes according to the objectives of the present research.

Framework

The study will look at how multimedia in the form of games is incorporated in teaching and learning of Tamil. The games can be accessed through the internet or played through CD. The differences in the students' performance between the two groups will be observed in the use of vocabulary, ability to write essay and confidence in pronunciation. The result will be given by way of comparing the performance of both groups.

Control Group (traditional setting) N=50	**Aspects**	**Experimental Group (multimedia setting) N=50**
Board + Chalk	← Pronunciation →	Games through CD
Teacher centered	← Vocabulary →	Student centered
Formal, passive leaner	← Essay →	Informal, active learner
Monotonous		Interactive

Conclusion

Since multimedia is basically a medium of entertainment, it is a misnomer to think that this media is doing more harm to the students rather than contributing in their studies. In fact, this research has given a clear indication that the students who are exposed to teaching through multimedia based methodology could excel in many ways while engaged in language discourse. Apart from this, the research has also shown that the exposure to multimedia based games could make the students excel in all the basic language skills.

Reference

1. Adone, D., Dron, J., Pemberton, L. & Bagne, C. (2007) E-Leaning environments for digitally-minded students. Interactive leaning Research, V 18(1) pg 41-53
2. Condle, R., & Livingstone, K.(2007). Blending online learning with traditional approaches : changing practices. British Journal of Educational Technology V38 (2) pg. 337-348
3. Mullai & Sudandra. (2001). Factors that impaction the use of computers in teaching/leaning in Tamil Schools. TI 2001 Conference Proceeding. p12
4. Paramasivam. (2002). Malaysia Tamil School and ICT Usage. TI 2002 Conference Proceeding p192-197
5. Mohanlal,Sam (2003) Wordsmith and Oarsman- Language games in Vocabulary Development in Tamil., Central Institute of Indian Languages,Mysore,India.
6. Senthamizh II, A Multimedia title for teaching/learning Tamil as first language.(Based on the Curriculum of Schools in Singapore),Apple Soft, Bangalore and Singapore Education Society,Singapore,2000
7. Simon, H.A.(1996). Observation on the sciences of science learning. An Interdisciplinary Discussion, Carnegie Mellon University, Department of Psychology, Washington. DC.
8. Tapscott,D.(1998). Growing up digitial: How the web changes work, education, and the ways people learn. Change Magazine, pg 11-20
9. Turgut, Y., & Irgin, Pelin (2009). Young learners' language games learning via computer games. Procedia Social and Behavioral Sciences 1 (2009) pg. 760-764
10. Vaughan, Taay.(1997). Multimedia making it work. (3rd edition) New Delhi: Tata McGraw Hill.

Use and the Impact of Information Technology in the Teacher Training

Dr Seetha Lakshmi
Asian Languages & Cultures
National Institute of Education, Singapore.

Abstract: Technology ought to be harnessed to enhance a lesson rather than be essential for teaching. It can assist students, especially those learning Tamil as a second language, to realize the beauty and joy of speaking the language. The media and IT based unique technological devices that have been used for second language teaching and learning proved as potential and effective tools (Rafael Salaberry M., 2001 & Zongyi Deng et al., 1999). This paper highlights the IT and pedagogical based research initiatives carried on at the National Institute of Education (NIE) on Teaching and Learning of Tamil Language in Singapore.

National Institute of Education and its Parent organization Ministry of Education in Singapore are enhancing and harnessing new ways of using IT to improve the quality of Education in Singapore. The convergence of interest shown by researchers in implementing new methods of teaching Tamil using IT has been welcomed by educationists here.

By adapting and encompassing IT resources and software, the National Institute of Education is constantly improving the quality of IT teaching in Tamil. IT has been used widely in Teaching Pedagogical, Literature Modules and in a module on the Use of Language in Singapore. For example, Web quest, Video conferencing, Multimodal resources creation, Learner based curriculum production, Group investigation, Digital Story telling and Corpus data bank is in use. By looking at some of the good practices developed in the field of this technology, the institute is creating new materials which will help students learn Tamil in a fun and interesting manner.

NIE also focuses in the design and development of new methods through research and monitors the problems that arise. By conducting pre and inservice courses for Tamil teachers, the feedback from their set of practices also help to set a new strategic position for improvement in the system of teaching which will help in the progress of the Tamil Language worldwide.

Since IT has boomed into many aspects of our lives and education, it is necessary that it covers the vast areas of our teaching in Tamil as well. Thus all this research initiatives and journals will help us in keeping Tamil abreast with IT.

Introduction

It is no doubt that technology is a communication tool in our lives today. What is amazingly most amazing is that this type of technology is not only modernized but also provides us with information that we do not know, and hence, benefits us. Amongst the communication tools created by Man, IT related tools rightly accomplish the network goals such as announcing, knowledge feeding and inner

happiness. In countries like Singapore, not only is the dominance of technology significant but also the impact. In Singapore's education system, IT has been playing an important role in various levels. Definitely, Tamil language is no exception.

In Singapore, students take English language as their first language and take Mandarin, Malay or Tamil as their second language. How is this technology used in Tamil then?

After the stage of memorization and teaching through class representative or leader, blackboard and chalk pieces came about. After which, teaching tools; such as keyboard, computer, smart board, and Tablet PC that consist of computer and mobile phone provide students with the language benefits in class. Tamil letters, Tamil songs, Tamil vegetables, Grandmother stories, are all being sold in the form of CDs/DVDs even in today's commercialized level, and all these have; Tamil's nuances, the beauty of pronouncing in Tamil, vocabulary building in Tamil, India's nature as well as the beautiful Tamil spoken by qualified hosts in their native language that provides a feast for students who hear and view them. Here, the beauty of the language and the benefits of its nativity are displayed in a manner that students can know about. In this stage, we shall see how information technology is used in teaching and learning, at National Institute of Education that trains teachers, who teach Tamil.

Tamil Language Division at the National Institute of Education

NIE, the only training college that provides training for teachers' of Ministry of Education, has 13 academic groups in which Asian Languages and Cultures is one such division. Here, there are Mandarin, Malay and Tamil divisions that teach the respective languages. In the Tamil section, there is a two year training programme for Tamil teachers, who are under the Diploma in Education classes, and also a ten-month post-graduate Degree in Education is being conducted for them. Other than these, there are Foundation Programmes for students, who excel in their Mother Tongue, teachers taking four years training also have special training curriculum for the first two years. This is where students will spend their two years in in-depth knowledge enhancement and in the next two years, they will join the students in the Tamil section taking Diploma in Education.

Under the program that encourages a specific percentage of teachers of the Ministry of Education to study Masters, teachers join in the evening classes and even attend classes during their holidays. Masters and PhD classes are also being conducted. To further enhance the talent of current teachers in pedagogy, trainings are also conducted in between work. We shall now view the significance played by IT in all these programmes.

Pre-service courses and Information Technology

Student teachers who study for two years have their content-filled lessons saved in the computer and used during curriculum. Also, computer related fundamental training, lesson related internet searches, those festivals celebrated there, are recorded and saved, and all these are used during discussions in classes. Speeches by both foreign and local speakers, as well as Literature, education and culture related presentations that are available in the market are used as additional lesson materials.

Through the means of computer, students are able to produce their studies related assignments. A good example would be creation of their own websites. Student teachers under the Tamil section, learn their content-based subjects such as Literature and Language, in a manner to also receive related explanations. The lecturers also use computer and IT in order to help the student teachers to enhance

their talents as a teacher. Tamil Language, Civics and Moral Education, and Tamil literature- the related teaching and learning of subjects allow Tamil to be known, understood and criticized by the usage of textbooks along with IT. Here, IT has attained a vital state in teaching and learning. Besides, IT is greatly used by teachers to realize how teachers can enhance Tamil in their career, through the four stages – listening, speaking, reading and writing. Below are some examples:

*Web quest *Multiple Intelligence *Multimedia functional approach *Multimodal approach	*Action research approach *Task based approach *Group investigation *Digital storytelling	*Student package for students' self paced learning *Learning package to learn with teachers' guidance *Assignments produced and submitted via computer

In order to capture the students' attention in the best way, at the start, middle and end of the lesson, and to channel their thinking, so as to attract them to the lesson, both old and new movies are segmented and compiled into small clips so as to be saved. The use of IT can be seen here too.

Currently, NIE's blackboard that is a computer tool, allows students to download their lectures along with; blogging, podcasting, webbing and chatting. The week before the student teachers start to teach, other than students discussing lesson related issues, the Safe Assignment method in use, also helps to keep track of the commitment of the students, and how they will produce and submit their assignments according to the guidelines provided. However, the worrying issue here is that Tamil does not have Safe Assignment method. When comparing Tamil with other Mother Tongue Languages in Singapore, for international human language, there are facilities; such as OCR and voice recognizer. It is needless to say for Malay language as their font provides them with a great opportunity. For Tamil, it has the newly provided and introduced Unicode font, Murasu Anjal Version that is now either showing numerous new faces or dimensions. Amongst the few is that in Singapore, either each individual or each department used to have one computer input system. However, it has now changed to everyone using T99 keyboard and Unicode fonts to type. It allows to be used in various ways; previously used insertion of typed documents to be viewed through Unicode fonts, downloading of data in Tamil from the internet, and to know the impact of Tamil. It also has a dictionary feature that allows one to find the meaning of Tamil words in either Tamil or English. So far, in this small island, Tamil teachers who have been separated in various ways in Tamil typing will now have the Tamil society using one computer language to converse and to socialize. Moreover, the newly standardized software will allow students to use Tamil conveniently. This invention that came about after three years of hard work is now used for training in NIE.

For students to excel in their second language, it is essential that they build their vocabulary list and use language in the motive of using it. In that way, software containing vocabulary games can be created and demonstrated in a manner that is suitable for students. This software can be also used to enhance students' enjoyment in listening to spoken Tamil. The students can listen at home to an edited passage that is recorded in spoken Tamil and make use of known and unknown words from that passage to replace the words they use every day in their speech, composition, and other assignments. They can also use it for writing for various levels of daily life, by sharing it with many

through the computer; and presenting Tamil assignments through computer. Receiving assessment by recording views on a particular occasion, through podcasting, use and view numerous creations through YouTube, accessing them, thinking of how they can be partly or fully be used, how to create one that is better off, which all shows computer's use.

At NIE, under the method of resource development, students and teachers produce many new creations that are in effect by working in conjunction with schools. Here, we can see this occurring in the background of Singapore kids, for them. At the same time, this is of much help for students living in countries like Singapore that is multi-racial, multi-cultural and multi-language. This is due to us having known more of others rather than ourselves, in many instances. That is not wrong. However, it is greatly wrong of us not knowing ourselves. That too, in a country like Singapore, when children think that their previous generation has prepared everything for them and has laid the path for them too, and when they grow into teenagers, if the facilities they require are not there, then the fault is ours. Hence, in order for us not to face a similar state, it is true that IT does help.

Lee K Y (2004) states that "English was necessary, given Singapore's multiracial makeup and given the access it provided to international trade and technical know-how. The mother tongues on the other hand, anchored Singaporeans in their Asian roots and values" (Laurel Teo, 2004: 1). This is the emphasis, our Tamil teachers would like to stress among the students. At the same time, the students who are the netizens of 21st century compare their Tamil class with their English class and the same goes to the teaching materials. Although English has many innovative IT resources, it is time to build in Tamil too with the limited financial, manpower and professional support.

Based on this, currently our trainees are producing the resource bank and add on with the existing resources. When using IT, it is vital to know which, is effective in it. Instead of transferring a word document into PowerPoint and make it as a powerless point, teachers can use it with pure effective engagement. Here, the September 11th World Trade Centre Crash is a good example for using media / IT to its highest stage. That particular strength is compatible for TV and computer than the newspaper or radio (Mahizhnan Arun, 2002). Hence using it in an innovative and influencing way is very impactful and useful.

In second language learning, corpus data plays a critical role. Recording the native, first and second language learners' voices or conversations and use it to teach or give it to the student as a homework and listen to it at home, will provide tangible benefits.

For building up the vocabulary, to understand the culture, identity and grammar, digital story telling is a suitable form of teaching materials. This has multimodes for the senses of the learner and to capture his attention towards it. Bringing the cultural artifacts, discuss about it and use to build their digital story telling, the Dip Ed II trainees are currently involved in it. It is because; teachers have to capitalize the digital tools to capture themselves and their students' prior and current knowledge to develop themselves as global elites.

Due to video conferencing, teaching of lessons in the class has changed; lessons are now in a manner whereby individuals can sit at a preferred corner and study. At NIE, two lessons were conducted in this method, in a manner of studying during a lesson, where the lecturer sat in one corner as the student teachers sat at their homes. Later, changing from the usual accessing method, alternative assessment is adopted or is in a producing manner, where educational tour related assessment tools

are in the process of being created. These are all some of the examples. The explanations of photos related to these will take place after photo shoot.

In times of nationwide health threats such as SARS and H1N1, students can learn from home, through the means of IT. This method also took place in Tamil, along with the school education system, several computer companies provide students with education tools that allow them to learn Tamil through the computer. For this, money is deducted from students' Edusave for education to take place.

Conclusion

Tremendous amount of literature (Gopinathan S., 1999; MOE, 2005; Seetha Lakshmi et al., 2005; Klein, R.R. Rogers, P.C. and Zhang Yong 2006.) has argued for the pivotal role of IT in second language education. As Gopinathan S and Saravanan V., pointed out, Globalisation has changed the economic activity forms and provided new opportunities. Especially countries like India showcased its IT talents and created new wave in the job market and in the other domains with their talents in biotechnology, banking and biomedical (2000). This goes true with the Tamil trainee teachers who are IT savvy. With their bilingual talents, I have used them to create digital story telling which are very essential for the learning of Mother Tongue in a multicultural and multilingual society. These trainees will be using their language and IT based talents to make the school students adapted to the language learning. Ministry of Education has invested heavily in Information Technology through its Three Master Plans for ICT in Education (1997-2002, 2003-2008, and 2009-2014) and we could witness the impact of it even among the Primary one students. Schools provided the IT support to the students with blogs and face book facilities and it is true that these kids are conversing through IT than anything else.

Due to the above stated various reasons or initiatives; from those who now study pre-school till to those studying Masters, and PhD, they have received many contacts by themselves in Singapore – blog, website, YouTube picture and face book. At the same time, there is no day that the heart yearns to know when will the day come, when there will be a way to install Tamil into our mobile phones and use it carefree.

These are a few, and it is important to understand that creating and using IT will not produce a better second language learner unless the teacher has the passion, background knowledge of IT and the customers i.e. students. No matter what happens, though it is acceptable for one student to use either one computer or one machine in their individual life, when it comes to a class, one machine is used by many, who jointly speak, work and use it to learn knowledge, discussion and engagement that has to do with enhancing education, human progress, social understanding, and multi-understanding. It can be said that positive interdependence based lifestyle is strengthened.

References

1. Arun Mahizhnan. (2002). The Mass Media: Competence and Impotence, Keynote speech delivered at the NIE Tamil Seminar, Singapore on 12. 01. 2002.
2. Gopinathan S., (1999). Preparing for the Next Rung: economic restructuring and educational reform in Singpaore, Journal of Education and Work, Vol. 12, No 3, 1999. Pp295-308.
3. Deng Zongyi and Gopinathan S,.(1999). Integration of Information Technology into Teaching: The Complexity and Challenges of Implementation of Curricular Changes in Singapore. Presented at the Annual Conference of the Australian Association of the Research in Education,

Fermantle, Australia. 2-6 December 2001. http://www.aare.edu.au/01pap/den01468.htm accessed on 15 September 2009.

4. Gopinathan S., and Saravanan V.,(2003). Education and Identity Issues in the Internet Age: The Case of the Indians in Singapore. In Asian Migrants and Immigration: The Tensions of Education in Immigrant Societies and Among Migrant Groups. Charney Michael and Yeoh Brenda S A., and Tong Chee Kiong(eds), Amsterdam: Kluwer Academic Publishers. Pp.32-51
5. Klein, R.R. Rogers, P.C. and Zhang Yong. (2006). Technology for Education in Developing Countries, 2006. Fourth IEEE International Workshop. 10-12 July 2006. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?tp=&arnumber=1648403&isnumber=34561 accessed on 15 September 2009. Pp.36-37.
6. Ministry of Education(2005). Tamil Language Curriculum and Pedagogy Review Committee Report, Singapore.
7. Rafael Salaberry M., (2001). The Development of Past Tense Morphology in L2 Spanish. Studies in Bilngualism. USA:John Benjamins.
8. Seetha Lakshmi, Viniti Vaish and S. Gopinathan (with the assistance of Vanithamani Saravanan) (2005). A Critical Review of the Tamil Language Syllabus and Recommendations for Syllabus Revision, CRPP Project 36/04SL, Centre for Research in Pedagogy and Practice, National Institute of Education, Singapore.
9. Seetha Lakshmi and Jarina Peer.(2009). Use of Tamil Language and IT in Tamil Language Education Redesigning Pedagogy International Conference, National Institute of Education, Singapore.
10. Teo Laurel.(2004). MM relates his personal struggles with language. The Straits Times, 8.3.2004. P.1

Tamil Language Teaching in UK.

A case study of one supplementary school's experience of working in partnership with a mainstream school to promote a community language.

Siva Pillai
Educational Studies,Goldsmiths, University of London SE 14 6NW. UK
Principal Examiner for Cambridge ASSET Examination (Tamil Language)
Chief Examiner for London Edexcel Examination (Tamil Language)

The focus of this paper will examine how a small project run in collaboration with a mainstream school achieved much more than it set out to and has used the successes achieved to review the work of their organisation.

Supplementary schools in the UK provide a vital service to the communities they serve, not least in the development of community languages but more important is the opportunity for young people to learn about their own cultural identity and speak in their mother tongue in a safe and stimulating environment.

Many supplementary schools are weekend community schools, run by community members representing their particular cultural / ethnic group. These schools have been set up with the aim of supporting members of that group who reside in the local community, through providing cultural and educational programmes for its members.

A more recent development within some supplementary schools is the link being made with mainstream government funded schools that all follow the National Curriculum (NC). In particular the syllabus for NC Language teaching that gives recognition to the importance of community languages alongside the traditional modern languages of French, German and Spanish.

There are more than 50 Tamil teaching supplementary schools in the UK.

One such supplementary school in South London in the United Kingdom,

The Tamil Academy of Language & Arts has made such links providing the opportunity to work in partnership with mainstream schools in the area.

The emphasis of partnership is important here as the opportunity to share language teaching alongside cultural activities has not only been well received by the Tamil community but also members of the wider community linked to the school.

The collaborative project with mainstream schools (Primary Key Stage 2) has been so successful that it is now part of the school's annual programme for language development and more recently achieved outstanding recognition through the achievement of a European award for language teaching.

The project approach combines language teaching with South Indian Classical dancing lessons and is described by the Local Education Advisor as - an inspirational project for other teachers of community languages, a brilliant example of a cross-curricular approach and infusion of language learning with creativity embedded it in a rich context.

The template for this project outlines how the language sessions will work in tandem with the dance classes and is a unique feature of the work we continue to develop in partnership with the school building on best practice.

The unit of work is clearly structured and is spread out over 10 weeks of teaching. All activities are carefully sequenced and sufficient detail for the activities are given within the teaching sequence.

The assessment criteria are given to pupils to encourage self and peer-assessment and to help them develop independence.

Progression through the weeks is achieved by linking core activities to previous learning. All the work culminates in a final presentation and the production of the portfolios.

Evaluation of the approaches used has led to an increased confidence in the use of the target language and greater consideration given to the linguistic objectives for classroom interaction.

None of these achievements could have been progressed without the collaborative approach in working in partnership with mainstream colleagues and the support we have received from the Local Education Authority. The vital resource of ICT and the Academy's strong focus on the promotion of the vitally important cultural and linguistic aspects of what we do has enabled us to sustain and strengthen our work over time, to become much more innovative in our approach to language teaching.

What are our overarching aims?

The voluntary aided Tamil Academy of language Arts, aims to meet the needs of children of Tamil heritage and their parents in the Borough of Lewisham, South London.U.K.

We achieve these aims by providing a range of educational activities and opportunities in a safe and welcoming environment. We measure the impact of our service by monitoring the numbers of students attending and through our quality assurance procedures.

In the interest of social cohesion we wish to develop our outreach work to promote better understanding of the Tamil community in Lewisham. UK.

What we have developed

We have built strong links with local schools and continue to develop the partnership and strive to ensure that our work supports the wider community.

Over many years we have worked hard to build a strong base within the community to support children and their families in being able to access and benefit fully from the education system in the U.K.

A philosophy rooted in the promotion of equity and social justice in the interest of social cohesion is at the heart of what we do.

The Academy provides a wide range of activities including ICT, language and cultural development activities but more important a safe and secure environment in which young people can develop their language and engage with cultural activities that they would not normally have the opportunity to be exposed to.

The Academy achieved 'European Award for Languages' in 2007 and shines as the beacon school for Supplementary Education. Our partnership is with main stream education through our work with Downers School in South London. U.K.

The outcome of our successful partnership with mainstream education is the publication of a Tamil Teaching Guide in the UK. (Curriculum guide for the Tamil Language) 2003 and is designed to meet the requirements of the National Curriculum. The TALA context recognises the equal status of languages and this Tamil Language Framework will provide support for the under-resourced group of Tamil Language teaching contexts.

The guide highlights good practice in language teaching in particular the need to encourage the use of stories, visual aids and other text in the target language.

The core concern of the guide is to ensure the recognition of students' achievements as well as give a practical advice in delivering the teaching of the Tamil Language in U.K. schools.

Collaborating with a mainstream school has meant that staff at the Academy have been able to engage in professional development and as a result understand and engage with initiatives introduced by the government. This has provided an excellent opportunity to implement the policy for Every Child Matters to ensure that it underpins all of work.

The work that we do at the Academy is designed to meet the 5 outcomes of the Every Child Matters (ECM) agenda.

1. Being healthy
2. Staying safe
3. Enjoying and achieving
4. Making a positive contribution
5. Economic well-being

Our curriculum at the Academy now reflects these elements and we have designed activities, some IT based to ensure that our students understand their importance.

This was not a planned outcome of our work with mainstream schools but one which we feel has forced us to rethink the way in which we are teaching languages and the all important cultural aspects of what it means to be a Tamil by looking more deeply at the purpose of and need to learn the Tamil language.

These developments could have been achieved without the use of ICT which not only plays a vital role in the teaching and learning of languages but provides a more effective mode of communication with our partners.

Community languages such as Tamil have benefitted from the ICT in the following ways:

- We have been able to share resources with our partner;
- highlight links to the net providing opportunities to communicate across the World Wide Web.
- Using IWB-Interactive white board in teaching the Tamil language. Children use IWB in their normal lesson in the mainstream school. Create activities and games using ICT. Most of the children have access to computer and WEB. All teaching resources are created using ICT

Our students are familiar with technology as a learning tool in other subjects as well as languages and use ICT on a daily basis.

Students at the Academy and our partner school are able to access resources that we continue to construct and develop. These are designed to ensure that maximum benefit is derived not only in developing language skills but also a wide repertoire of IT and communication skills that will equip our students for the world of work.

Currently we use the following IWB- Interactive Whiteboard:

Email and blogging are encouraged using the target language and a bank of visual aids is designed to encourage participation by younger members of the Academy.

Finally we will focus over the next year on the outcomes of the National Language Strategy. By 2010 it is expected that the opportunity to learn another language will be provided in all primary schools. It is up to the school what language(s) to offer. Some schools are already offering Community Languages.

We are also keen to work towards supporting our students in achieving the ASSET Languages qualification (http://www.assetlanguages.org.uk) and look forward to developing this sin partnership with schools.

Cambridge - Asset Languages is a voluntary assessment scheme. This support the National Languages Strategy by providing recognition of achievement. (OCR-Oxford, Cambridge and RSA Examination board) and associated accreditation options against DCFS criteria.

Conclusion

The work in partnership with a mainstream school has provided an exciting challenge for the academy and stimulated our thinking about the way in which language is learnt and the vital role it plays in cultural identity.

We look forward to further developments in the future and welcome the sharing of ideas and best practice in the interest of promoting the Tamil language and arts.

Enhancing the Process of Learning Tamil with Synchronized Media

Vasu Renganathan
University of Pennsylvania
(vasur@sas.upenn.edu)

Introduction

The process of learning Tamil as a second language has been as complex as the language itself. With the advent of new technological resources that augment the process of learning a second language, it becomes more of a challenge as to how one would make use of technology for language learning in a very prudent manner, so one can make sure that technologically advanced lessons are pedagogically sound in all manner. The term 'multimedia' has been a buzz word in the second language pedagogy ever since technology was incorporated into second-language curriculum. What are all the media that constitute multimedia? Obviously, audio and video constitute the two primary media, but they have been part of language learning process ever since the analog tapes of both audio and video were invented. Presumably, the web technology gave a fresh new dimension to how these two media can be used for the purposes of language learning. Besides, there are also other methods of learning using web media including the use of web forms, chat rooms, message boards etc., which should also be considered to be part of 'multimedia'. This paper attempts to study such web enhanced media that are popular in language learning curriculum and discuss how each of such media are pedagogically sound, particularly in the context of Tamil learning. Suitable illustrations are made from the website "Tamil Language in Context" accessible at: http://www.thetamillanguage.com/ or http://www.southasia.upenn.edu/tamil/.

Synchronized media and Diglossia

As part of the language learning task, a learner prefers to have access to multiple media namely audio, video, glosses, grammar notes and cultural notes in a user-friendly environment. Keeping this in mind, Tamil lessons are provided in this site for the learner to learn the language in a graded fashion from simple to complex form of the language. Tamil being a diglossic language, any Tamil learner is obligated to learn multiple registers of the language along side of the distinctions based on formal literary variety and informal spoken variety; standard spoken form versus dialect forms and so on. Further, learning and using any diglossic language entails understanding of both culture as well as speech contexts on the part of the learner. Schiffman (1978:100), for instance, claims with suitable illustrations that learning a diglossic language presupposes a competence on the part of a speaker to evaluate each situation and come up with socially appropriate linguistic forms. The screenshot provided as below (fig. 1) illustrates how this is achieved by presenting the lessons in such a manner that the student uses the same page to get access to various dimensions of the language using a number of media namely audio, video, glosses, ability to switch between spoken and written form of the script, and access to corresponding grammar and cultural notes. Particularly, use of videos shot with conversations of native speakers in real contexts capture not only cultural information, but also socially appropriate linguistic forms. Another advantage of using videos for learning any language is that it allows learners to get to know both linguistic as well as non-linguistic information like body language, which plays an important role in using the language in an authentic manner.

Any curious language learner can come up with a variety of questions during the process of their learning of a language and such questions may range from grammatical explanations, phonetic information, glosses for vocabularies, pronunciation issues and so on. In the case of Tamil, though, the type of questions tend to be more due to its supplemental complexities in terms of its diglossic nature based on informal versus formal distinctions, agglutinative nature of word formation and so on. Keeping these complexities in mind, the lessons in this site are prepared in such a way that answers for such questions are presented to the learner in a user-friendly manner. The mouse-over gloss allows the learner to place the mouse in any word of their choice and get the meaning of the word while reading dialogues. This process, which is called "incidental vocabulary learning" by Hulstijin et al (1996), allows the learner to build their vocabulary during the process of their listening and reading activity. This is as opposed to memorizing vocabularies in isolation, which does not give the learner an opportunity to correlate the use of vocabularies in contexts. Further, glosses are supplemented with suitable links to let students get the translations of dialogues along side of each line, if they prefer. Navigation bar in the QuickTime video window allows students to listen/watch any piece of the dialogue multiple number of times at their own convenience. Grammar and cultural lessons that correspond to each dialogue in the lessons allow them to browse the respective information synchronously while listening to any particular dialogue. Especially, this ability allows students to pace their learning on different stages so they can learn the lessons with or without such pedagogical aids on a gradual fashion.

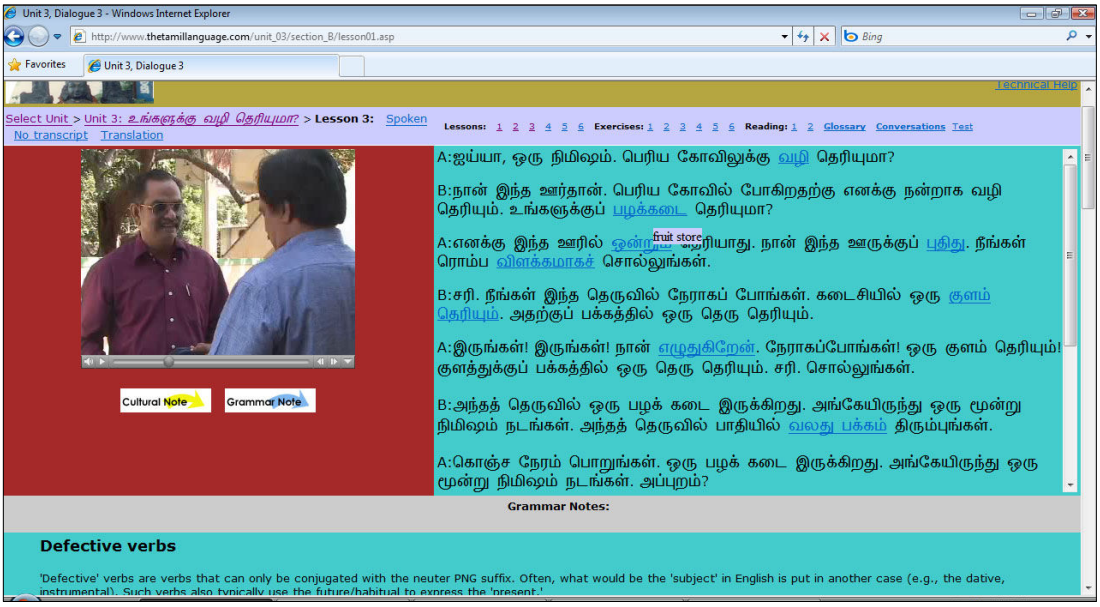


Fig. 1. Synchronized media

This site offers a total of thirty six lessons for first year and another thirty six lessons for second year, and every lesson is built in a synchronized fashion with video, audio, glosses and so on as discussed above.

Preparation of Grammar and Culture Lessons for Diaspora students

Grammar and cultural lessons are presented in a graded fashion so a progress in learning can be monitored suitably with reincorporation of both vocabulary and grammatical information. Subsequently, each of these lessons is followed by suitable comprehension exercises that are designed

to test student's skill that is acquired from each lesson. Each video lesson comprises of a dialogue, depicting a authentic speech situation, like conversations in a store, conversing with auto drivers, asking for directions etc. Grammatical information presented in each dialogue would focus on a particular topic in grammar, and a piece of cultural information is included along with it based on the nature of conversation. A detailed explanation on cultural information is included with illustrations on the same page, so it gives an opportunity for students to watch the video paying special attention to such information, either grammar or culture.

Thus, imparting both cultural and grammatical information is necessary in language classrooms in general and Tamil language class rooms in particular for the reason that Tamil classrooms in any diaspora setup are always mixed with both true learners as well as heritage learners with different proficiency levels. In an earlier paper (Renganathan 2008) I discussed how the proficiency levels of students who take Tamil classes in the US context are always heterogeneous in nature and how defining a single profile encompassing all of the students in any particular level is impractical. Unlike in the countries like Singapore and Malaysia, exposure to Tamil language for heritage learners in the US is limited only to their homes where their parents are the only source for Tamil speech context. Also, in a multi-ethnic country like England and US, many families have parents whose mother tongues are different – with Tamil and Hindi, Tamil and English, Tamil and Kannada and so on. In those cases, the students who come from such families are left with even less exposure to contexts for Tamil speech situations. Often, despite their heritage background, their understanding of Tamil is like any true learner, who do not have any background knowledge whatsoever related to both culture and grammar of Tamil. In this sense, making any technologically-enhanced language lessons should not only be self-explanatory in all respects, and they should also be able to cater the needs of all students whose proficiency level falls in various degrees. Often, Tamil lessons developed, keeping in mind the learners of Tamil classes in Tamil Nadu, fail to address the needs of diaspora students adequately. In any diaspora context, introducing vocabularies, grammar and cultural information should be done in a gradual fashion from simple to complex forms with a step-by-step instruction on how to comprehend and use them in actual context. This is otherwise called a bottom-up learning, as opposed to top-down learning where lessons are made with a random choice of vocabularies and grammar, and make the students decipher them gradually.

Use of the Tamil Website in Smart classrooms

Smart classrooms are equipped not only with playing audio and video files, they also have enough resources to connect the class room computers to networked labs, which allow the faculty to incorporate lab resources in their instruction. In this sense, the classrooms that are used for language teaching purposes at the University of Pennsylvania are furnished with enough technology including big-screen projectors connected to computers, DVD players, audio players, besides the university wide local network. All of the resources provided in the Tamil website are made use of effectively in this type of classrooms, and in this respect, an ideal environment for multimedia-enhanced teaching can be achieved. This is as opposed to traditional classroom teaching where students and teachers are left with very limited resources for learning and teaching respectively. Story board, text as well as audio message boards, synchronous chat environments etc., are some of the technological advances that enable language learning a more convenient process than before. Particularly, these resources can be best implemented only in the context of smart classrooms.

Research has shown that CMC motivates learners to engage themselves in meaningful communication in the target language and leads to effective language learning (Brown, 1994; Hanson-Smith, 2001; Meskill & Ranglova, 2000). CMC can be synchronous or asynchronous; it can be text-based (email, online discussion forums, chat rooms and so on) or voice-based (voicemail, audio-enabled chat rooms and message boards). One of the advantages of smart classrooms is that audio-enabled Wimba board (http://www.wimba.com/) conversations can be played-back in classrooms and can thus be integrated with language lessons. In this respect, CMC and class room teaching can be integrated successfully. Typing in Tamil has been a problem among students as they are not used to any of the typing methods as well installation of appropriate software in their computers. In this respect, the Transliteration based Tamil typing that is implemented with Unicode characters and Javascript application in the Tamil website enables students to type conveniently in Tamil in order to participate with other students in CMC environments both synchronously and asynchronously. Thus, both voice and Tamil text chat can be integrated successfully for Tamil.

Thus, any technology-enhanced language lessons must be developed in such a manner that various media are synchronized to provide a pedagogically sound environment. In this respect, the Tamil website at the University of Pennsylvania makes use of audio, video, glosses, comprehension exercises, transliteration application and a host of other pedagogically relevant methods in order to bring forth a synchronized media for Tamil language learning.

References

1. Brown, H.D. (1994). *Principles of Language Learning and Teaching*: Upper Saddle River, NJ: Prentice Hall Regents.
2. Hulstijin, J. H. Hollander, M., & Greidanus, T. (1996). "Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words." *The Modern Language Journal*, 80(3), 327-339.
3. Meskill, C., & Ranglova, K. (2000). Sociocollaborative language learning in Bulgaria. In M. Warschauer & R. Kern (Eds.), *Network-based language teaching: Concepts and practice* (pp. 20-40). New York: Cambridge University Press.
4. Renganathan, Vasu. (2008). "Formalizing the knowledge of Heritage Language Learners: A technology-based approach." South Asia Language Pedagogy and Technology, Vol. 1., University of Chicago. (http://salpat.uchicago.edu/index.php/salpat/article/view/30).
5. Schiffman, Harold F. (1978). "Diglossia and Purity/Pollution in Tamil." In *Contributions to Asian Studies, Vol. II: Language and Civilization Change in South Asia*. Clarence Maloney (ed.) Leiden: E.J. Brill. Pp. 98-110.

TAMIL WEB PORTALS, E-CONTENT, WIKIPEDIA, BLOGGER AND AGGREGATOR

எதிர்கால இணையத்தில் தமிழ் யூனிகோடு

வளர்தமிழ் ஆய்வில் தரவுத்தளங்களும், திரட்டிகளும்

முனைவர் நா. கணேசன்
ஹ்யூஸ்டன், அமெரிக்கா

கணினி வலையில் தமிழ்

இந்திய மொழிகளிலேயே முதன்முதல் கம்ப்யூட்டரிலும், பின்னர் இணையத்திலும் ஏறிய மொழி தமிழ். தமிழ் எழுத்தின் எளிமை மற்ற இந்திய எழுத்துக்களை ஒப்பிட்டால் விளங்கும். ஏனை இந்திய எழுத்துப் போலன்றி, விராமப் புள்ளி தமிழில் கூட்டெழுத்துக்களை உடைத்தெறிகிறது. அச்சு, தட்டச்சு, கணியச்சு, இணையக் குழுக்கள், … தமிழுக்கே முதல்-வரவான இவை *புள்ளியின்* நன்கொடை! கணிப்பொறிகளில் தமிழ் எழுத்துக்கள் 1980-களின் ஆரம்பத்திலே தெரிந்தன. அப்போது ஆப்பிள், மைக்ரோசாப்ட் கணினிகளில் பல எழுத்துருக்களும், எழுதிகளும் (editors), ஆர்வலர்கள் தோன்றவைத்தனர். ஜார்ஜ் ஹார்ட் (பெர்க்கிலி), ஆதமி (கே. சீனிவாசன்), மயிலை (கல்யாண்), நளினம் (செல்லையா), கம்பன் (வாசுதேவன்), அணங்கு (குப்புசாமி), துணைவன் (ரவி), அஞ்சல் (முத்தெழிலன்) … போன்ற எழுதுரு, செயலிகளைக் குறிப்பிடலாம். இவை யாவும் இலவசமாகப் பயனர்களுக்குக் கிட்டியதால் பலப்பலரும் தத்தம் கணினிகளில் பொருத்தித் தமிழைப் பார்த்தனர். இருப்பினும், ஒவ்வொன்றும் தட்டெழுதத் தனியான எழுதிகளும் செயலிகளும் வேண்டியிருந்தன. வடிவான எழுத்துருக்கள் வேண்டித் தமிழ்நாட்டில் பத்திரிகை நிறுவனங்கள் தனித்தனி எழுதுருவாக வடித்துக் கொண்டனர். பத்திரிகை, புத்தகப் பதிப்பகங்களில் கணிஅச்சிடல் வேரூன்றியது. முதன்முதலாக தினமலர் பத்திரிகை முழுமையும் கணிஅச்சில் இயங்கத் தொடங்கியது. பின்னர் கணினிகளால் படிப்படியாக ஈயஅச்சுக் கோர்ப்பதே ஒழிந்து இன்றைய நிலைக்கு உயர்ந்தது.

இணையம் உதயமாகி வேகமாக வளரத் தொடங்கியது 1994 முதல் என்று சொல்லலாம். அதற்குமுன் அமெரிக்காவின் பெரிய பல்கலைக் கழகங்களிலும், ஆய்வுக் கூடங்களிலும் மட்டுமே இணையம் புழங்கியது. இந்தியாவில் இணையம் 1990-களின் மத்தியில் பரவ ஆரம்பித்தது. இணையத்தின் வருகையால் தமிழ்க் கணிமை புதிய விசை பெற்றது, வைய விரிவு வலைப்பக்கங்கள் பிறந்தன. மின்னஞ்சல், அரட்டை, குழுமங்கள், … என விரிந்தது. *தமிழ்.நெட்* குழுவைப் பாலாபிள்ளை தொடங்கி நடத்தினார். 8-பிட் குறியீடு ஆகிய திஸ்கி உருவாகி யாகூ குழுக்கள் தொடங்கப்பட்டன. தொன்னூறுகளின் கடைசியில் இருந்து ஆறாம்திணை, இன்தாம், திண்ணை, அம்பலம், திசைகள், … ஆதி மின்னிதழ்கள் குறிப்பிடத்தக்கன. பின் வணிக இதழ்கள் - குமுதம், விகடன், … - இணையத்தில் வலம்வரத் தொடங்கின. விசைப்பலகைகள், எழுத்துக் குறியீடுகள் பலவாகிக் குழப்பங்களும் மிகுதியாக இருந்த சூழல் என்றும் குறிப்பிட்டாக வேண்டும்.

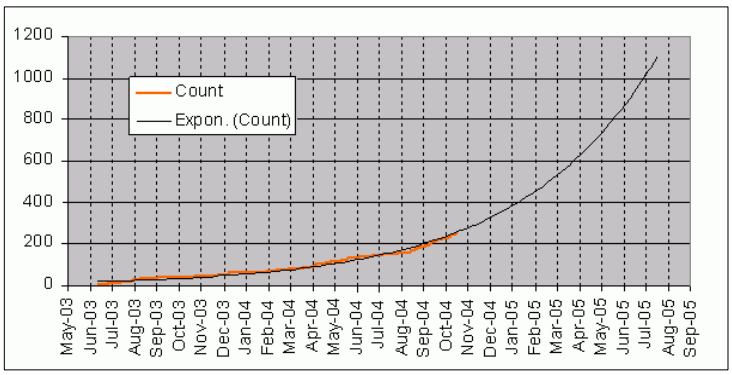
இணைய மாநாடுகளும், பல்கலை தரவுத்தளங்களும்

சிங்கப்பூரில் நா. கோவிந்தசாமி போன்றோர் ஒருங்கிணைந்து 1997-ல் முதல் இணைய மாநாடு நடத்தினர். பின்னர் சென்னை மாநாட்டில் தரக்கட்டுப்பாடாக தமிழ்99 விசைப்பலகை, டாப், டாம் குறியீடுகள் அரசாங்கம் அறிவித்தது. ரவீந்திரன் பால் வடித்த துணைவன் விசைப்பலகை தமிழ்99 பலகையின் முன்மாதிரி ஆகும். துணைவன் விசைப்பலகை இந்திய மொழிகளுக்கு முதல் ஒலியியல் (phonetic) விசைப்பலகை என்பதறியலாம். மலேசியா 'நயனம்' இதழாளர் இராஜகுமரன் இணையம் என்ற சொல்லை 1996-ல் தந்தார். வைய விரிவு வலை, உலாவி, பின்னூட்டு, தொடுப்பு, ஒருங்குறி (இராமகி), மட்டுநர் (moderator), வெள்ளுரை (plain text), செழியுரை (rich text), திரட்டி, … போன்ற கலைச்சொற்கள் கூகுள்குழுக்கள், பதிவுலகில் பரவிவிட்டன.

முகுந்தராசு அளித்த எ-கலப்பை, திஸ்கி, யூனிக்கோடு இரண்டையும் உள்ளடக்கிய உமர்தம்பியின் தேனீ எழுதுரு, ஏதோவொரு குறியில் இருப்பதை ஒருங்குறிக்கு மாற்ற சுரதாவின் பொங்குதமிழ் மாற்றி, ... வலைப்பதிவு எழுதுவதை மக்களுக்கு எடுத்துச் சென்றன. இணையம் உருவாவதன் முன்னமே கோபர் வழங்கியில் சங்க, பிற்கால இலக்கியங்களைத் தந்த கொலோன் பல்கலை மால்ட்டனின் முயற்சி முன்னோடியானது. இணையப் பல்கலை, மதுரை முன்னியம், தமிழ்மரபு அறக்கட்டளை, ஈழ நூல்களை அழிவில் இருந்து காக்கும் நூலகம் திட்டம், எண்மியத் (digital) தேவாரம் (புதுச்சேரி), அகராதிகளுக்கு சிக்காகோ பல்கலை ரோஜா முத்தையா தளம், ... தமிழாய்வில் தினமும் பயன்படும் தளங்கள். இணையப் பல்கலை ஒருங்குறிக்கு மாறவிருப்பதாக அறிகிறோம். 2003-ல் தொடங்கிய விக்கிபீடியா, கலைச்சொல் விக்சனரி உபயோகமான தரவுகளைத் துழாவுவோருக்கு உடனுக்குடனே அள்ளித் தருகின்றன. இவை எல்லாமும் ஒருங்குறியில் இயங்குவதால் தேடிகளில் துழாவல் எளிது, எனவே சிங்கப்பூர் தன் ஒரே குறியேற்பாக அதனை அறிவித்துள்ளது.

இந்திய மொழிகளின் முதல் வலைத்திரட்டி 'தமிழ்மணம்'

"தமிழ்மணத்தில் என்னையும்வை என்றேனே – அவன் தனிமனத்தில் இருநினைவா என்றானே" என்று காதலி கூற்றாக அன்றே சொல்லிப் போந்தார் புரட்சிக்கவிஞர். கணினி முன்னால் அமர்ந்தால் ஆகும் காலச் செலவைப் பார்த்தால் அந்தத் தீர்க்கதரிசன வாக்கின் உண்மை புலப்படும். **2003** முதல் தமிழ் வலைப்பூக்கள் மலர்ந்தன. பதிவுகளுக்காக நியூக்லியஸ் நிரலைத் தமிழ்ப்படுத்திய காசி ஆறுமுகம் தமிழ்மணம் (**http://tamilmanam.net**) திரட்டியை நிர்மாணித்தார். வலைத்திரட்டிகள் மாற்று ஊடகமாக இணையத்தை நிலைநிறுத்தின. சிற்றிதழ்களுக்கும், தனி வலைப் பதிவர்களுக்கும் தனி வாசகர் வட்டத்தைத் திரட்டிகளே நல்கின. தமிழ்மணம், மறைந்த தேன்கூடு, தமிழ்வெளி திரட்டிகளும், தமிழிஷ் பதிவிணைப்பு தளமும் வலைப்பதியும் புதுமுகங்களுக்கு உற்சாகம் ஊட்டுவன. ஐந்து ஆண்டுகளாய் இயங்கும் தமிழ்மணத்தில் **6000** பதிவுகளும், நாளும் **300** இடுகைகளும் வெளிவருகின்றன, ஒரு நாளுக்குச் சுமார் **2500** மறுமொழிகளையும் திரட்டுகிறது.



*தமிழின் – தமிழ்மணத்தின் வலைப்பதிவு வளர்ச்சி (**தொடக்க காலம்**)*

பரந்து கிடக்கும் உலகத் தமிழரை இணைத்துப் பயன்படும் தகவல்களை உடனுக்குடன் பரிமாறத் திரட்டிகள் துணை இன்றுண்டு. பிரச்சினைகளைச் சுயநலம், வணிக நோக்கம் இன்றி எடுத்துச் சொல்லும் செய்தி ஊடகங்கள் தமிழிலில்லை. அனைத்து நாடுகளிலும் தமிழர்கள் சிறுபான்மையர் என்பதால் அரசூடகங்களின் ஆதரவில்லை. **PBS** தொலைக்காட்சி, **NPR** வானொலி போன்றவை ஏற்படுத்தும் நிதிவசதியும் இல்லை. இக்குறையை நீக்க வலைப்பதிவுலகம் நிறைய உதவும். ஒரு பிரச்சினைக்கு எல்லாப் பரிமாணங்கள், பல்தரப்பு விவாதங்களை அளிக்கும் பதிவுகளால் உண்மையைப் புரிந்துகொள்ள முடிகிறது. தமிழ் செம்மொழி இலக்கியங்களைக் கற்கும் மாணவர்களுக்குப் இலக்கண, இலக்கியங்களை விளக்கப் பதிவுலகில் இயலுகிறது. இணையத்தில் பாரிஸ் பேரா. செவ்வியாரும் நானும் **15** வருடமாகத் தொடர்ந்து தமிழ்பற்றி வலையாடுகிறோம்.

தரவுத்தளங்களில் எழுத்து வரிவடிவங்கள்

எதிர்காலத்தில் இந்தியாவில் எழுத்துப்பெயர்ப்பும், கணிமொழிபெயர்ப்பும்

ஐரோப்பிய நாடுகள் காலனி ஆட்சியில் சிக்கவில்லை. சீன நகரும்-அச்சு நுட்பம் ஐரோப்பாவில் அறிமுகமாகி குட்டன்பர்க் போன்றோரால் பதிப்பகங்கள் பெருகின. ஒவ்வொரு மொழியும் விவிலியம், சர்ச் போதனை நூல்களாக முதலில் வளர்ந்தது. கல்விப் பரவலாக்கம் மக்களிடையே மறுமலர்ச்சி, விழிப்புணர்ச்சி, விடுதலை உணர்வுகளை உள்வாங்கித் தனிநாடுகளைக் கட்டமைத்தன. அந்த இயல்பான வளர்ச்சி காலனி ஆதிக்கத்தால் இந்தியாவில் தடைப்பட்டது. ஓரளவுக்கு இந்திய நடுவண் அரசு வேற்றுமைகளை அங்கீகரிப்பதன் வடிகால்களாக மொழிவாரி மாகாணங்களை அனுமதித்தது. இந்திய ஒருமைப்பாட்டு நிலையான அரணுக்கு இந்திய ரூபாய் நோட்டில் காணப்படும் எல்லாத் தேசிய மொழிகளுக்கும் சமமான ஒரே அந்தஸ்து நடைமுறையில் வழங்கப்படல் வேண்டும். **19-**ஆம் நூற்றாண்டின் 'தாய்மொழி' **(mother tongue)** கோட்பாட்டின் விளைச்சல் மொழிவாரி மாநிலங்கள் என்ற தகுதியேனும் கிடைத்தது எனலாம். இனி, 'தாய்எழுத்து' **(mother script)** பயன்படும் வகையைப் பார்ப்போம்.

தமிழ்நாட்டிலிருந்து கல்லாத மக்களும், தொடக்கக்கல்வியே உடையோரும் வடஇந்தியாவில் வாழச் சென்றால் இந்தியைத் தானாகக் கற்றுக்கொள்கின்றனர். அதேபோல், தமிழ்நாட்டுக்கு வேலைக்கு வருவோரும் தமிழைப் பேசுகின்றனர். இவர்கள் செல்பேசிகளில், கணியுலா மையங்களில் அவரவர் 'தாய்எழுத்திலே' மற்ற மாநில எழுத்துக்களைப் படிக்கத் தொழில்நுட்பம் உதவுகிறது. மத்திய, மாநில அரசுகளின் மடலாடல், தகவல், அரசாணைகள் அந்தந்த மாநில எழுத்துப்பெயர்ப்பிலும், மொழிபெயர்ப்பிலும் கணிநுட்பங்களால் தானியங்கியாக உருவாகும் நாள் தொலைவில் இல்லை. நிதி, கணிஞர்களை அளித்து கூகுள் போன்ற நிறுவனங்களுடன் அரசு வழிகாட்டினால் இந்திய மொழிகளுக்கு இடையில் (உ-ம்: தமிழ் **< >** இந்தி) கணிமொழிபெயர்ப்பு வசதி நல்ல வளர்ச்சிபெறும். நகர்கணிகள், செல்பேசிகளில் தமிழ் இல்லாத பட்சத்தில் தரக் கட்டுப்பாட்டுடன் ஆங்கில எழுத்தில் காட்டும் நிரலிகள் வேண்டும். உத்தமம் **–** தமிழ்நாடு அரசு நிபுணர்கள் கூடி, தமிழ் **>** ஆங்கிலம் எழுத்துப்பெயர்ப்பை **(transliteration)** நிர்ணயிக்க வேண்டும். உதாரணமாக, உயிரெழுத்து இரண்டின் நடுவே உள்ள ககர எழுத்தைத் தவறாக **'g'** என்று எழுத்துப்பெயர்ப்பதை இணையத்தில் காணலாம். மொழியியற்படி இது பொருந்தாது: **Intervocalical க் in Tamil has a voiceless fricative sound, and usually it is rendered as k̲ or ḥ in world's languages.** அழகு, முருகன், அகம், தூரிகை போன்றனவற்றின் உராய்வொலி **azak̲u, muruk̲an, ak̲am, thuurik̲ai** அல்லது **azaḥu, muruḥan, aḥam, thuuriḥai** என்று பெயர்க்கப்படல் தமிழின் ஒலிப்புக்குச் சிறப்பு. அன்றேல் கன்னடம், தெலுங்கு ஒலிப்பாகி, தமிழின் தன்மையை இழக்கும். எழுத்துப்பேர்ப்புத் தர-உறுதிப்பாட்டை இணைய மாநாடுகளில் தமிழ்நாட்டு அரசாங்கம் அறிவிக்கவேண்டும். தமிழ் எழுத்து அறியாத் தமிழ்க் குழந்தைகளும், செம்மொழி கற்க வரும் மேலை நாட்டாரும் தமிழின் சிறப்பு ஒலிகளை அறிந்துகொள்ளக் கணினியில் சரியான ரோமன் எழுத்துப்பெயர்ப்பு துணைநிற்கும்.

எதிர்காலத்தில் எழுத்துச் சீர்மையும், தனித்தமிழ் எழுத்தில் மீக்குறி ஏற்பும்

தமிழ் எழுத்தைக் கல்லாத மக்கள் லட்சக் கணக்கில் உள்ளனர். ஆங்கிலக் கல்வியே கற்கும் பல மில்லியன் மாணவர்களும் தமிழ்நாட்டிலும், வெளியேயும் வாழ்கின்றனர். அரசியல்வாதிகள், தமிழாசிரியர், பத்திரிகைத்துறை, **...** என்று சிறு விழுக்காட்டினருக்கே தமிழ் வருமானம் தரும் தொழிலாக உள்ளது. ஏனையோர் தமிழை விரும்பிப் படிக்கத் தமிழ் எழுத்தின் வரிவடிவம் சீர்மையுற வேண்டும். எழுத்துச் சீர்மையின் தேவையை விளக்கும் அறிஞர் வா. செ. குழந்தைசாமியின் சொற்பொழிவை தமிழ் இணையப் பல்கலை வைத்துள்ளது: **http://www.tamilvu.org/esvck/index.htm**

இலங்கை, இந்தியத் தமிழரின் அடுத்த தலைமுறை தமிழில் நுண்பதிவிட **(microblogging e.g. Twitter)** உயிர்மெய்ச் சீர்மை அவசியம். ஒருங்குறியின் சிறப்பம்சத்தை இங்கே குறிப்பிடலாம்: நன்னூல் இலக்கணம் சொல்வதுபோல சார்பெழுத்தாக, பிணை விலங்கை உடைத்துத்தான் உ/ஊ உயிர்மெய்கள் ஒருங்குறியில் ஏற்றப்பட்டுள்ளன. உ/ஊ உயிர்மெய்களைப் பிரித்தும் விரும்புவோர் இணையத்தில் படிக்கலாம், எழுதலாம் என்ற அரசாணை வந்தால் தமிழ் கற்பித்தல் எளிமையாகும். செம்மொழி படிக்கத் தமிழினத்தில் பிறவாதோரும் வருகின்றனர். அவர்களுக்கும் உதவும்.

வீரமாமுனியின் பெரும் சீர்திருத்தமும், **1978**-ல் அரசு கொணர்ந்த சிறு மாற்றமும் செய்த நன்மை குன்றின் மேலொளிரும் விளக்கு. இனிக் குறைந்தபட்சச் சீர்மையாக உ, ஊ உகரக் குறிகள் திருத்தம் தேவை. பழக்கமான ஊகார கிரந்தக் குறியும், அழகியல் பொருத்தப்பாடுடைய, வரிநீளத்தை அதிகரிக்காத உகரக் குறிக்கான பரிந்துரை (முதல் **4** மெய்யெழுத்து வரிசை, அ-ஒ உயிர்மெய்கள்) எடுத்துக்காட்டு:

◌	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ
க்	க	கா	கி	கீ	கு	கூ	கெ	கே	கை	கொ
ங்	ங	ஙா	ஙி	ஙீ	ஙு	ஙூ	ஙெ	ஙே	ஙை	ஙொ
ச்	ச	சா	சி	சீ	சு	சூ	செ	சே	சை	சொ
ஞ்	ஞ	ஞா	ஞி	ஞீ	ஞு	ஞூ	ஞெ	ஞே	ஞை	ஞொ

மேலதிக விவரமாக, தமிழ் நெடுங்கணக்கும், சீர்மையில் கட்டுரையும்: **http://nganesan.blogspot.com/2009/08/ezhuttu-korppu.html**

ஐரோப்பிய மொழியினர் உலகின் எல்லா எழுத்துக்களையும் ரோமன் எழுத்துக்கள் மீது மீக்குறிகளை **(diacritical marks)** ஏற்றிக் காட்டுகின்றனர். (அ) அம்முறை தனித்தமிழ் இலக்கணம் கூறும் **18** மெய்யெழுத்துக்களின் மீதும் ஏற்பட்டால் கிரந்த எழுத்துக்கு ஒரு மாற்றுவழியாகும். (ஆ) ஒலிப்பு பிசகாது இருப்பதால் துணைக்குறிகளின் தேவை - பிறமொழி எழுத்துக்களைத் தமிழ் எழுத்தில் பெயர்க்க - இருக்கிறது. இதற்குத் தரக்கட்டுப்பாடும், யூனிக்கோடு வரைபொறி **(rendering engine)** அனுமதிக்க வைக்கவும் உத்தமம்-இணையப் பல்கலை குழுவமைக்க வேண்டும். துணைக்குறிகளில் பொருத்தமானவற்றைத் தேர்ந்து அறிவிக்கலாம்.

முடிவுரை

இதுகாறும் தமிழ்க் கணினி, வலைப் பதிவு, தரவுத்தள வளர்ச்சிகளின் முக்கிய மைல்கல்களைப் பார்த்தோம். தமிழ் ஆய்வு சிறக்கவும், கற்பித்தல் எளிமையாவதற்கும் தேவையான **(1)** எழுத்துப்பெயர்ப்பு விதிமுறைகள் **(2)** எழுத்துச் சீர்மையில் குறைந்த பட்ச அரசாணை **(3)** தமிழ் எழுத்துக்களில் ஐரோப்பிய எழுத்துக்கள் போல மீக்குறி ஏற்றம் பற்றிய அறிமுகத்தையும் இக்கட்டுரை வழங்கியுள்ளது. விரிவாக, என் வலைப்பதிவில் **(http://nganesan.blogspot.com)** கருத்துத் தெரிவிக்கலாம்.

Automatic E-Content Generation

E.Iniya Nehru
National Informatics Centre, Chennai

and

T.Mala
Anna University, Chennai

Abstract: Automatic E-content generation is an innovative idea in the field of Natural Language Processing. It aims on developing an intelligent tutoring system in Tamil language. This system focuses on delivering personalized content in Tamil language to an individual user needs based on their learning abilities and interests. The system is divided into two parts. The first part involves the domain corpus collection and construction of the knowledge base. The knowledge base is constructed using text categorization and information extraction techniques. Categorization is responsible for classifying given Tamil documents to their target class. This is performed using the Naive Bayes (NB) algorithm. Information extraction system was constructed by developing information-extraction rules by encoding patterns (e.g. regular expressions) that reliably identify the desired entities or relations. The second part involves identifying the interests of specific user's and then user profiling the identified interests by looking through their browsing history. The topic analyzer is constructed to analyze the user's profile and evaluate the user's knowledge using intelligent evaluator system. Then the personalized content will be generated based on the knowledge level of the user.

This paper is organized as follows. Section 1 discusses the literature survey in the area of content generation, section 2 talks about the overall system architecture, section 3 talks about the categorization, section 4 gives details on information extraction system and section 5 gives the performance evaluation and conclusion of the proposed system.

Literature Survey

The bulk of the text categorization work has been devoted to cope with automatic categorization of English and Latin character documents. El-Kourdi et al. used Naive Bayes algorithm to automatically classify non-vocalized Arabic web documents to one of five pre-defined categories [2]. Taeho Jo and Dongho Cho proposed an alternative approach to machine learning based approaches for categorizing online news articles [6]. Chinglai Hor et al. suggest the use of a hybrid RS-GA method to process and extract implicit knowledge from operational data derived from relays and circuit breakers [1]. S.Iiritano and M.Ruffolo described a prototype of a vertical corporate portal that implements a KDD process for knowledge extraction from unstructured data contained in textual documents [4]. Jose M.Alonso et al. presented a user-friendly portable tool designed and developed in order to make easier knowledge extraction and representation for fuzzy logic based systems [5]. Harith Alani et al. developed a tool called Artequakt. Artequakt automatically extracts knowledge about artists from the Web, populates a knowledge base, and uses it to generate personalized biographies [3].

Overall System Architecture

The overall system architecture as shown in the figure 2.1 explains the various functionalities required to perform the automatic content generation. Automatic content generation focuses on delivering personalized content in Tamil language to an individual user needs based on their learning abilities and interests. Knowledge base is constructed using document categorization and tourism related information extraction. This phase uses Naive bayes classification algorithm to classify the tourism documents as temples, hill stations and hotels. After the classification process, information should be extracted from those documents. The extracted information should fill the template of each document of a particular category. Thus the knowledge base was constructed using text categorization and information extraction.

An individual user's interests are identified and recorded to create a user profile. A user profile is specific to a user and is subjected to change over time. The Topic categorizer is used to categorize the topic based on user's query. The topic analyzer is used to analyze the user's profile and evaluate the user's knowledge using intelligent evaluator system. Based on the user's knowledge, intelligent evaluator system makes a decision to suggest the same topic or to suggest a new topic retrieved from the knowledge base. Then the personalized content will be generated based on the knowledge level of the user.

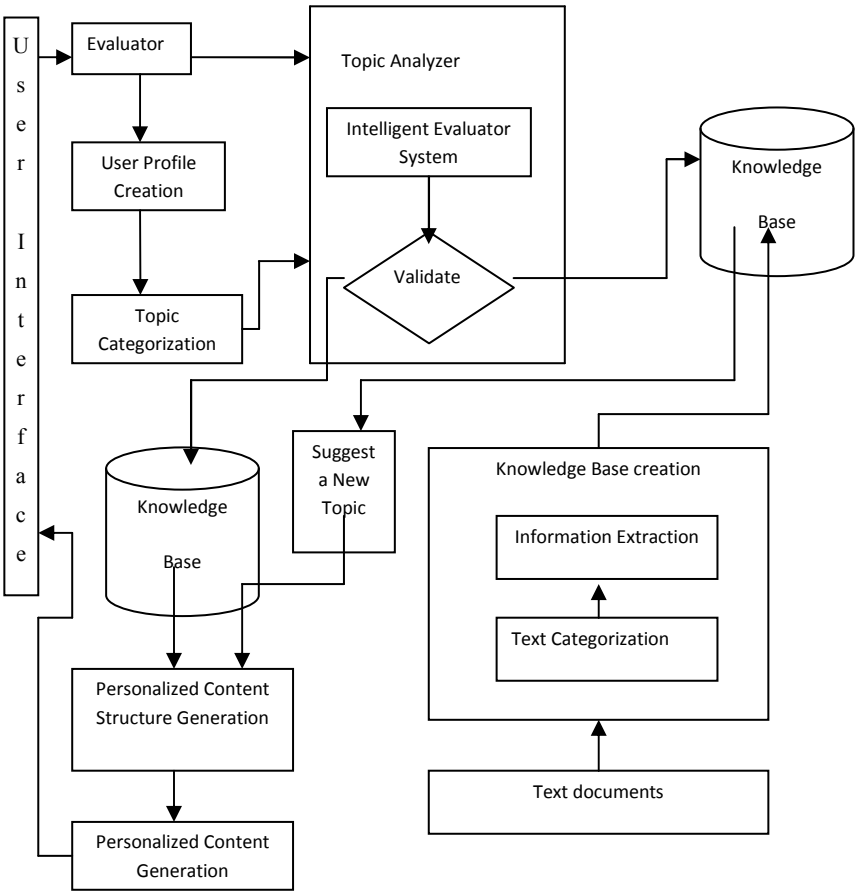


Figure 2.1 Detailed block diagram

Document Categorization

Document categorization is the task of automatically sorting a set of documents into categories from a predefined set. In the training phase, a set of documents are given with class labels attached, and a classification system is built using a learning method. Once the categorization scheme is learned, it can be used for classifying future documents. The following section gives a detailed view of preprocessing and classification of documents.

A data preprocessing phase is required to weed out those words that are of no interest in building the classifier and also to reduce the processing time. Preprocessing phase comprises of two stages namely stop word removal and term pruning. In order to remove the stop words from the document, stop word list is prepared and the input tourism document is compared with the list and then stop words are removed. After removing the stop words, each document must be transformed in to a feature vector. This text representation, referred to as the bag-of-words.

The problem of finding a "good" subset of features is called feature selection. Feature selection can be done by term pruning method. Term pruning is done according to the term frequency values. Here feature selection refers to relevant words. This would eliminate the very rare words that occurred and also the very common words that occur in almost every text document. Relevant words are then passed to the classification phase.

This phase uses naive bayes classification algorithm to classify the tourism documents as temples, hill stations and hotels. The NB classifier computes a posteriori probabilities of classes, using estimates obtained from a training set of labeled documents. When an unlabeled document is presented, the posteriori probability is computed for each class and the unlabeled document is then assigned to the class with the largest a posteriori probability.

Information Extraction

Information extraction (IE) distills structured data or knowledge from unstructured text by identifying references to named entities as well as stated relationships between such entities. The document is tagged to identify the location, name of the place. Identify the Domain Events using rules i.e., retrieval of sentences based on heuristic rules. The heuristics rules are formed manually by analyzing the documents. Extraction of syntactic patterns from the sentences is done using part of speech information. Compare the syntactic patterns from the sentences with the predefined pattern. If it matches, then the corresponding template should be filled. The templates for various domain events are filled and then they are merged to form the complete filled template for the document.

Performance Evaluation and Conclusion

From the 50 Tamil Nadu tourism documents, the first 30 are used for training and the next 20 are used for testing. The classification task considered here is to assign the documents to one the two categories. The Precision/Recall is used as a measure of performance for document categorization. Recall is the percentage of total documents for the given topic that are correctly classified.

Precision is the percentage of predicted documents for the given topic that are correctly classified. The recall and the precision measures are calculated for each category and the average values are shown in the following table.

Performance Evaluation Results Table

Category	Average Precision	Average Recall	Average F-measure
Kovil	100.00	80.00	88.89
Malai	83.33	100.00	90.71
Macroaverage	91.67	90.00	**89.80**

The result shows that the Naïve bayes achieves a macro average of 89.8. Thus it shows that the naïve bayes classifier performs well and its effectiveness is better in classifying the Tamil documents. The current work is confined to tourism documents of Tamilnadu state, but it can be extended to any category of documents. Categorization work can be further enhanced by adding more categories.

References

1. Chinglai Hor, Peter A. Crossley, Dean L. Millar "Application of Genetic Algorithm and Rough Set Theory for Knowledge Extraction" IEEE transactions on Power Tech, pp. 1117 - 1122 , July 2007.
2. El-Kourdi M., Bensaid A. and Rachidi T. "Automatic Arabic Document Categorization Based on the Naïve Bayes Algorithm". Proceedings of COLING 20th Workshop on Computational Approaches to Arabic Script-based Languages, pp. 51-58, August 2004.
3. Harith Alani, Sanghee Kim, David E. Millard, Mark J. Weal, Wendy Hall, Paul H. Lewis, and Nigel R. Shadbolt "Automatic Ontology-Based Knowledge Extraction from Web Documents" IEEE transactions on Intelligent systems, Vol.18, Issue no.1, pp-14-21, January 2003.
4. S.Iiritano M.Ruffolo "Managing the Knowledge Contained in Electronic Documents: a Clustering Method for Text Mining" Database and Expert Systems Applications, 2001. Proceedings of 12th International Workshop on 3-7 September 2001, pp. 454 – 458, September 2001.
5. Jose M. Alonso, Luis Magdalena, Serge Guillaume "KBCT: A Knowledge Extraction and Representation Tool for Fuzzy Logic Based Systems" Proceedings of IEEE International Conference on Fuzzy Systems,Vol. 2, pp. 989-994, July 2004.
6. Taeho Jo and Dongho Cho, "Index Based Approach for Text Categorization", International journal of mathematics and computers in simulation, Issue 2, Volume 1, 2007.

ICANN and IDN

Internet Corporation for Assigned Names and Numbers International Domain Names

S.Maniam
Singapore
Email: maniam@i-dns.net

Introduction

The new institutional economics (NIE) asserts that institutions are the "rules of the game;" although it does address the problem of how individuals and organizations try to change the rules, it tries to maintain a sharp distinction between the rules and the players. But in many cases the organizations that operate regulatory institutions and create and enforce the rules can be considered players in the game as well. This is especially true when the regime and the organization is a new entity seeking to establish the legitimacy and universality of its regulatory scheme.

IDNs are an important but neglected topic in Internet governance studies. The original domain name system used a simplified character set based on the Roman alphabet, known as "restricted ASCII."

1 This meant that languages that relied on non-Roman scripts, such as Arabic, Korean, Chinese, Hindi, Tamil Russian or Japanese, could not be represented as domain names. The geographic and cultural bias introduced by such a standard should be evident. It required the development of a new domain name standard based on Unicode to enable representation of Internet addresses in other language scripts.

2 After ten years of agitation by advocates of non- Roman scripts, it is now possible to have domain names in any alphabet. These are known as internationalized domain names or IDNs. ICANN has (finally) announced its willingness and readiness to distribute IDN top level domains in 2007.

3 As this happens, Internet users will witness a striking transition from Internet domain names, URLs and email addresses based on ASCII characters to character sets that include all the world's language scripts.

The creation of IDN top level domains (TLDs) has the potential to open up large new markets for the domain name registration services industry. Although it is currently based on a character set that the majority of the world either cannot read or does not use naturally, ASCII- based domain name registration services already command around $3 - $4 billion in annual revenues; the number of registrations (not the revenue) has been growing at a rate of about 7 –10% annually

This paper examines contention over IDN policy among ICANN, the gTLD interests and the ccTLD interests. As noted above, we try to explain ICANN management's choice of policies in terms of a bargaining perspective. Our analysis emphasizes how the policies regarding new IDN top level domains produced by ICANN are strongly influenced by a calculus reflecting its organizational self-interest, and specifically its desire to gain economic and political forms of support from countries outside the United States.

What Needs to be Done

a. One particularly important aspect of ICANN's launch of new generic top-level domains (gTLDs) will be the availability of Internationalized Domain Names (IDNs) at the top level. That eagerly anticipated enhancement to Internet participation has also raised some issues.

For example, current practice dictates that gTLDs contain at least three characters – two-character Latin gTLDs are reserved for country-code top-level domains (ccTLDs). However, in certain languages one or two characters commonly express a complete word – and they would not be confused with present-day ccTLDs.

b. Prohibiting the registration of names of less than three characters in certain languages may hobble IDN use in certain languages but it is difficult to fashion a uniform set of rules to govern a potential relaxation of this requirement that works universally.

ICANN's approach to this issue is similar to its approach on many issues regarding implementation of the policy for the introduction of new gTLDs.

c. Get expert advice on the matter. The use of experts allows ICANN to obtain experience and skill economically outside its core competencies and develop material for public discussion in a timely manner.

d. Use that advice to formulate some sort of model.

e. Then conduct public discussion on the issue

This process has been used effectively thus far in the new gTLD implementation. ICANN has consulted with: technical, DNS, risk management and linguistic experts, dispute resolution providers, and others. In this case of character limits and IDNs, ICANN is engaging a small team to evaluate this problem and provide expert advice from both sides of the problem: that IDNs must be effectively engender regional participation and that the rules must provide stability, i.e., that the domain name system (DNS) work in a way predictable to users.

Again, that process for reaching implementation: identify issues, get expert advice, create a model for public discussion, discuss, iterate the model, and so on.

The idea is that the experts crystalise the discussion in a timely way and therefore encourage meaningful participation. We are at step number two of this process that will include all interested parties. The process for developing a preliminary set of assumption will be publicly reported so the ensuing public discussion can be informed and timely.

Everyone at ICANN appreciates the comments made on this particular issue and other IDN issues – all going toward an effective way to increase effective regional participation in the Internet.

Implementation Update

DNS Stability Panel: Interisle has been contracted to form the DNS Stability

Panel that conducts the technical string requirement evaluations for requested IDN ccTLDs. This includes a verification that the delegations of new TLD strings will not result in user confusion with any existing strings in the DNS. Interisle is currently in the process of forming the DNS Stability Panel for Fast Track Process String Evaluation and providing on-boarding material to panel members.

Online Request System

The online system through which IDN ccTLD requests will be submitted is in the final stages of development and testing. The next three weeks will be used:

a. finalize the online content

b. perform a legal review

c. undertake a live test.

As part of the live test ICANN has consulted with representatives of five countries and invited them to participate in testing of the system. Testing includes: submitting a test request in the system, processing / qualification by staff, providing feedback on the test to participants. These tests will run from the end of September through 9 October after which a new status report of overall Fast Track implementation will be released.

Online IDN area

The IDN area online will be revised with an FAQ, factsheets, and a manual reference for use of the online request system. The new and improved site will be released prior to the Fast Track Process launch time.

Linguistic Processes

There are a few aspects of the Fast Track Process that are related to linguists or require the advice or statements from experts in writing systems. An important piece of this relates to the requested string(s) as a meaningful representation of a country or territory name. UNGEGN has agreed to support Fast Track participants as needed with referrals to such expertise. The referrals will be provided through an ICANN point of contact and the method for requesting such expertise will be described in the proposed Final Implementation Plan. ICANN plans to support to those requiring linguistic assistance.

Outstanding Issues

Several topics that has been discussed in public comment on implementation proposals of the Fast Track Process since the initial draft implementation plan was released on 23 October 2008. These include: (i) the form of relationship between an IDN ccTLD manager and ICANN, (ii) cost considerations regarding contribution to processing and TLD support costs, (iii) management of variant TLDs. Solutions to these issues have been discussed, and it is believed that current opinions or positions of each community segment is well understood.

INFITT and ICANN

In various ICANN meetings on IDN language has been the main area of concern and expertise is essential. These are the following areas on which INFITT could offer:-

1. Dealing with variant issues in Tamil
2. Selection of TLDs (Both GTLDs and CCTLDs)– to ensure there misconception
3. Coordination on any recommended restrictions on names registered across several Tamil speaking countries/diasporas Spoofing

INIFTT could be the catalyst to ICANN on IDN issues. This will be a precedent for others like Arabic, Cyrillic or Chinese organizations for their contribution to ICANN.

Organisation such as INIFTT which has representation from Europe to USA has the global representation in its activities thus making the best agency to offer linguistic expertise to Internet Organisations.

Conclusion

This paper analyzed how ICANN's policy response to the problem of introducing IDNs between ccTLD registries and gTLD registries. As the result of adopting an IDN fast track for country code registries, country code registries will be able to offer multilingual domain names earlier than gTLDs, assuming that ICANN and each country are able to reach consensus on the details of an IDN ccTLD contract.

While the current IDN market is a lot smaller than the ASCII market, this is due to the lack of IDN top level domains, which limits IDN names to the second-level. This restricts service to an inconvenient, hybrid combination of ASCII and other scripts. The full IDN service enabled by the new IETF standard and new IDN top level domains will probably realize the earlier, high expectations regarding strong demand for IDNs in those parts of the world that use non-Latin scripts. The size of the current domain name market could easily double or triple once IDN TLDs are introduced. Because of the high switching costs associated with domain name registrations, whoever enters this market first will get the lion's share of the market in any given language group like .COM.

Both ccTLDs and gTLDs want to operate IDN top level domains. For both market actors, translation of their current TLD strings into IDNs is seen as the key to future growth as of 2009. ICANN does not actively respond to the efforts of the gTLD registries to extend their top level names into IDNs under the concerns expressed by some governments in such practice, while it accepts ccTLDs' claim to do so mainly because their corresponding governments support such consistency.

What accounts for this difference, the policy distinction between gTLDs and ccTLDs? In our view, the critical factor is ICANN's own organizational self-interest. ICANN has little hierarchical authority over ccTLDs and it needs to lure them into full contractual participation in its regulatory scheme if it is to solidify its position as the global domain name regulator. ICANN also has a longstanding problem with its legitimacy and support among national governments outside the United States. Because national governments strongly support privileged access to resources for their own country as fundamental rights, and in particular want to support the market position of their own national registry against the (mostly U.S.-centered) gTLD operators, ICANN can please national governments by doling out new IDN top level domains to their ccTLDs. By doing so, ICANN shows that it can deliver benefits to national governments.

Some national governments (e.g., China India or Korea) try to be careful about the use of "their" language script in the domain name space. Here again, ICANN may garner support from countries that have been reluctant participants or non participants (such as China) by acceding to such demands.

Ironically, under governments' strong objection to ICANN's IDN ccTLD contracts, this first-mover advantage ticket in the emerging IDN market, ccTLD IDN fast track, may give its opportunity to gTLD IDN fast track since gTLDs do not have such political tension with ICANN. ICANN's strategy to enforce ICANN's strong regulatory power over ccTLDs through IDNs will provoke the debate on role of ICANN, USG and ccTLD infrastructure management in the GAC.

References

1. IDN ccTLD Fast Track Process - http://www.icann.org/en/topics/idn/fast-track/
2. ICANN Meetings in Lisbon Portugal Transcript - IDN - GAC - GNSO & ccNSO Working Groups Workshop 28 March 2007 - http://www.icann.org/en/meetings/lisbon/transcript-idn-wg-28mar07.htm
3. IDN ccTLD Fast Track Program Proposed Implementation Details Regarding Arrangement between ICANN and prospective IDN ccTLD Managers May 2009 - http://www.icann.org/en/topics/ idn/fast-track/proposed-implementation-details-dor-29may09- en.pdf
4. IDN ccTLD Fast Track Program Proposed Implementation Details Regarding Development and Use of IDN tables and Character Variants for Second and Top Level Strings (revision 1.0) May 2009 - http://www.icann.org/en/topics/idn/fast-track/proposed-implementation-details-idn-tables- revision-1-clean-29may09-en.pdf
5. IDN ccTLD Fast Track Program Cost Analysis of IDN ccTLDs Focus on Program Development and Processing Costs June 2009
6. http://www.icann.org/en/topics/idn/fast-track/analysis-idn-cctld-development-processing-costs-04jun09-en.pdf

Development of Social Networking Website and Tamil E-learning Software

Using Unicode / ISCII standards

Dr.A.Muthukumar,
Professor in Computer Applications,
Kumaraguru College of Technology,
Coimbatore-641006, India.
Email: mail@profmuthu.com

Introduction

The project involves development two portals namely a Tamil Social Networking site and a Tamil E-learning Website. Since Tamil is spoken by Tamil Diaspora in various countries and spoken or used in Fiji (Republic of), Guyana, India, Malaysia, Mauritius, Singapore, Sri Lanka (Ceylon), Trinidad & Tobago, Zanzibar & Pemba (Tanganyika), the author proposes to develop such Tamil Website which can be used by all people. Before going to discuss the details of development, let us see introduction of the topics as below.

Social Networking

A social network service focuses on building online communities of people who share interests and/or activities, or who are interested in exploring the interests and activities of others. Most social network services are web based and provide a variety of ways for users to interact, such as e-mail and instant messaging services. Social networking has encouraged new ways to communicate and share information. Social networking websites are being used regularly by millions of people.

Everyday, more and more people sign up for social networking sites like Facebook and Twitter. And it's not just individuals, some businesses and non profit organizations are using the site to their advantages. At one time Myspace, Twitter and Facebook were geared toward a specific group of people. Now those social networking sites are turning into another form of advertising for some non profit organizations. The United Way, The American Red Cross and the Mohawk Valley Chamber of Commerce, are just a few non profits using the social networking sites.

An increasing number of academic commentators are becoming interested in studying Facebook and other social networking tools. Social science researchers have begun to investigate what the impact of this might be on society. Typical articles have investigated issues such as Identity, Privacy, E-learning, Social capital and Teenage use. The application domains for social networking are government, business, marriage portals, educational and medical applications. A few sites do exist where you can talk to others in languages besides English. Our interest is designing Tamil social networking website so that people who know Tamil and basic usage of computer can use this site and communicate with others. The target audience for this portal is people who speak, read and write Tamil. The Unicode standard helps in facilitating the implementation of this website.

E-learning

E-learning (or sometimes electronic learning or eLearning) is a term which is commonly used, but does not have a common definition. Most frequently it seems to be used for web-based distance education, with no face-to-face interaction. However, also much broader definitions are common. For example, it may include all types of technology enhanced learning (TEL), where technology is used to support the learning process. Most of the Tamils living away from their homeland find it difficult to teach their children Tamil. Even if they teach themselves it is limited to teaching alphabets, numbers and few rhymes only and not more than that. Out aim is to provide E-learning through web portals by including audio, video and other computer software tools to teach the language not only from the Alphabet but also teaching the language features like the grammar, Tamil literatures etc. Using Unicode standards the tool will be developed to teach Tamil.

The Technology

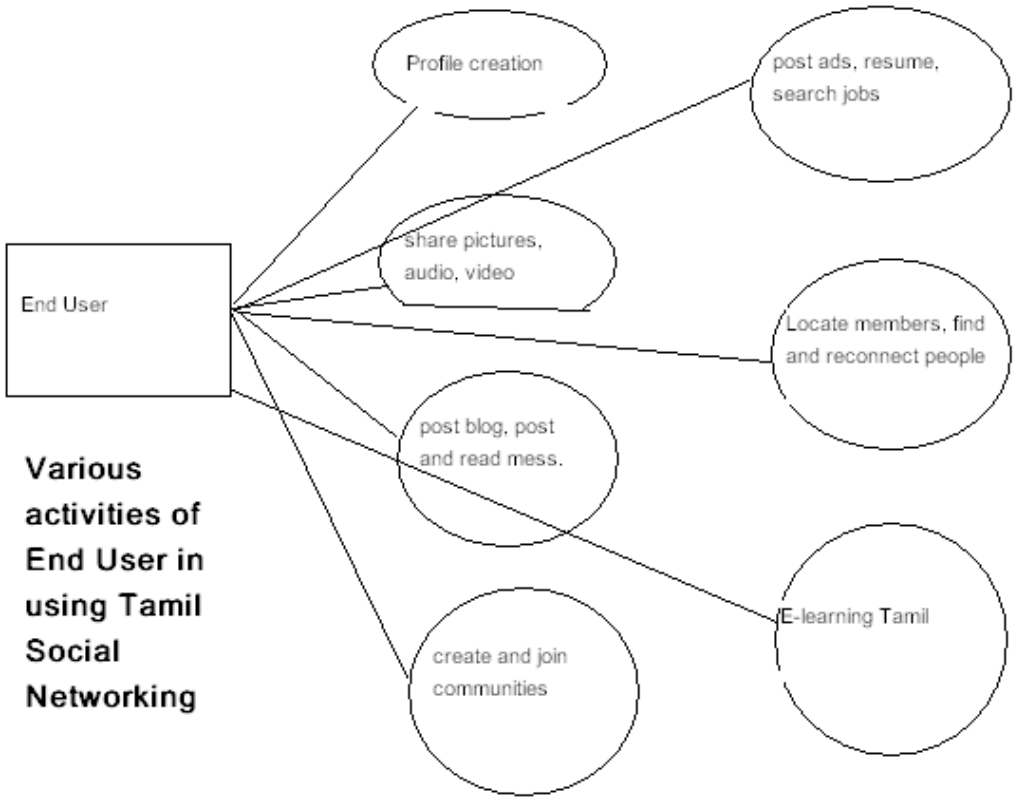
Like Windows XP or other software displaying Tamil fonts in menus, our aim is to develop websites displaying Unicode formatted texts (menu commands) displayed so that the user gets the feeling of using his language for communicating with others in his community of users. Even though it looks very simple, the development of social networking software and Tamil E-learning software requires complexity of design and implementation of the website. Hence it is well convinced that the work is worth doing and used by Tamil communities in future.

The author through the feedback given currently restricts to use Unicode and after comparing both Unicode and Iscii, he will use the right one for the implementation. Let the conference may input more on these lines.

Various Social Networking sites are compared for the features that they support. The advantages and disadvantages are compared. The ones that suit the multilingual interface are selected. The various social networking sites compared are as below.

Twitter Orkut IMBEE Sconex 43 Things TagWorld Friendster Bebo Hi5	MyYearbook Facebook Classmates.com Myspace Yorz Ecademy Ryze Xing Linkedln
Reddit Digg Del.icio.us LiveJournal Yahoo 360 Flickr	Xanga YouTube Beware Microlenders Zopa Campaigns Wikia Essembly.com and Many more….

Even though the social networking design and implementation is matured for the time, but the usages and number of new users are increasing day by day. Also applications in which the social networking sites are increasing and the table above shows only the important sites and there are many more. The advantages of comparing them would enable us to develop a world class website with many important features that are proved among other communities already. Again, only the important features that can be applied to our website is given below in the following figure.



If we notice, the features are Profile creation, sharing pictures, audio, video, post blog, post and read messages, create and join communities, post ads, resume, search jobs, locate members, find and reconnect people and using of Tamil E-learning software.

The social networking websites can be classified as communication websites, Business and professional networking sites, social content sites, Social Lending sites: Microlenders/Microfinance, Political social networkings, Mobile social software and Social networking blogs. Social networking research like "Imagined communities", analyzes privacy conerns of members and impact of those concerns of their behavior. Hence our intention is to develop a social networking website with privacy preserving concept. For this the author will use the research knowledge of privacy preserving data mining which is already there and matured.

Tamil E-learning software is being developed as web-based software and hence there is no need to buy the software and install in his computer. He can just login to the internet and the site and use the Tamil learning software. It is proposed to design e-learning software with all educational technology gadgets like audio, video, animation and other technologies.

First, it is proposed the develop Tamil Alphabet teaching using audio, video and animation. There is a separate studio used to shoot the video of Tamil teachers teaching the language and it will be put in the website. Hence the student wherever he is in the world, learns the language with the same pedagogy of class room environment and ease at his home learning. Next, the project aims in taking the Tamil text books of students from first standard till plus two level in Tamil Nadu and trying to capture the video lectures and audio lectures and other software animation tool based teaching. Even though it seems a big project, the author aims to do this job with the help of Tamil scholars and Tamil interest groups including governmental agencies and Tamil University.

When enough funding is obtained, the author proposes to host this website using a technology called cloud computing. Cloud computing is a technology in which one need not pay for the hardware and software fully, and it is enough the company pays for the time and usage of the software. Hence it is practical to host such a big ambitious plan by us.

Conclusion

Tamil language usage is reducing day by day because of relocation of people, and education in English medium etc. Hence it is the right time to enable the users to maximize his language usage during the leisure time through tools such as Social networking sites and E-learning software. Also these two technologies are inevitable for us and hence the advantages are to be given to those people who do not know languages other than Tamil. Hence, the author is developing these two technology based website. As already mentioned in the previous chapter, the first technology, namely Tamil Social Networking Website enables one to create and share profile, share photos, videos, audios, post blog, post and read messages, create and join communities, post advertisements, post resume and search jobs, locate and reconnect with people around the world. Also the second part the project namely; Tamil E-learning software is added as a feature in the first project. This enables to learning the language from Alphabet to the standard of plus-two. This is a very big project and it is ambitious to us. E-learning software is being developed with the Tamil Scholars or Teachers lectures being video tapped and used in it. Tamil dictionary and keyboard interface to search and type respectively are used in the site.

Acknowledgement: The author acknowledges the infitt.org and IITS, Germany for having selected the paper and providing necessary assistance. Also he thanks his organization Kumaraguru College of Technology, Coimbatore, India, for permitting him to attend the conference.

தமிழ் விக்கிபீடியாவும் துணைத்திட்டங்களும்

அறிமுகம், தமிழ் இணையத்தில் அவற்றின் வகிபாகம், எதிர்காலம், செயன்முறை விளக்கம்

முரளிதரன் மயூரன்
திருக்கோணமலை, இலங்கை
email: mmauran@gmail.com

இக்கட்டுரை தமிழ் விக்கிபீடியர்களையும் தமிழ் விக்கிபீடியா தொடர்பான நல்ல அறிமுகத்தைக்கொண்டவர்களையும் தனது முதன்மை வாசகர்களாக/கேட்பவர்களாக கொள்ளாமல் பரந்தளவிலான தமிழ் இணையப்பயனர்களைக் கருத்திற்கொண்டு வடிவமைக்கப்படுகிறது.

பரந்தளவில் தமிழ் இணையத்தில் திறந்த புலமைச்சொத்து, கூட்டுழைப்பு போன்ற கோட்பாடுகளுடன் விக்கிபீடியாவையும் அதன் துணைத்திட்டங்களையும் அறிமுகப்படுத்தி இப்பணிகளில் மேலும் பலரை இணைத்துக்கொள்ளும் நோக்கத்தினைக்கொண்டது.

தமிழ் இணைய மாநாடு 2009 இதற்கான சரியான காலப்பகுதியில் நடைபெறுவதாலும், அது இந்நோக்கங்களைத் தாங்கி வரும் இக்கட்டுரை சமர்ப்பிக்கப்படுவதற்கான மிகச்சரியான களமாக இருப்பதனாலும் அம்மாநாட்டுக்கென இக்கட்டுரை வடிவமைக்கப்பட்டு சமர்ப்பிக்கப்படுகிறது.

விக்கிபீடியா அறிமுகம்

விக்கிபீடியாவின் பின்னணியாய் அமையும் கோட்பாடுகள், விக்கிபீடியாவின் வரலாறு, விக்கிபீடியா சமூகம் தொடர்பான அறிமுகம், கட்டற்ற திறந்த மூல இயக்கம் , திறந்த புலமைச்சொத்து தொடர்பான அறிமுக விளக்கங்களும் அக்கோட்பாடுகள் விக்கிபீடியாவை எவ்வாறு உருவாக்க விழைந்தன, விக்கிபீடியா ஊடாக வெற்றியை நிரூபித்தன என்பனபோன்ற தகவல்கள்.

தமிழ் விக்கிபீடியா அறிமுகமும் தகவல்களும்

தமிழ் விக்கிபீடியாவின் உருவாக்கம், வரலாறு, தமிழ் விக்கிபீடியர் சமூகம் தொடர்பான அறிமுகங்களும் தமிழ் விக்கிபீடியா தொடர்பான புள்ளிவிபரங்கள், சிறு ஆய்வுரீதியான தகவல்கள்.
தமிழ் விக்கிபீடியாவின் புள்ளிவிபர அறிக்கைகளும், தமிழ் விக்கிபீடியா தொடர்பான ஏனைய கற்கைகளும் உள்ளடக்கப்படும்.
ஆங்கில விக்கிபீடியா, பொதுவாக ஏனைய மொழி விக்கிபீடியாக்களுடன் தமிழ் விக்கிபீடியாவினை ஒப்பிடலும், வேறுபாடுகள், ஒற்றுமைகள் போன்றவற்றை விளங்கிக்கொள்ளலும்.

விக்கி தொழிநுட்பம், பாதுகாப்பு, நடைமுறைகள் பற்றிய அறிமுகம்

விக்கிபீடியா கோட்பாடு ஒருபுறமும் தொழிநுட்பம் மறுபுறமாயும் ஒன்றுடன் ஒன்று ஒருங்கிணைந்தும் விக்கிபீடியா என்கிற அசைவியக்கத்தை நடத்திச்செல்வது தொடர்பான விளக்கங்கள்.
விக்கி முறைமை இயங்கும் தொழிநுட்ப அடிப்படைகள் குறித்த அறிமுகம்.

விக்கிபீடியாவில் கடைப்பிடிக்கப்படும் பாதுகாப்பு ஏற்பாடுகள், நிர்வாக முறைகள் போன்றன குறித்த ஆழமான பார்வை.
விக்கிபீடியாவின் போதாமைகள், அவற்றை நிவர்த்திக்க எடுத்துக்கொள்ளப்பட்ட, எடுத்துக்கொள்ளப் படக்கூடிய முயற்சிகள் குறித்த உரையாடல்.
விக்கிக் குற்றங்கள், பொதுவாக விக்கிபீடியர்கள் கடைப்பிடிக்க எதிர்பார்க்கப்படும் ஒழுக்கங்கள், பொறுப்புக்கள் போன்றன பற்றிய விளக்கம்.

தமிழ் இணையச் சூழலும் விக்கி அசைவியக்கமும்

தமிழ் இணைய உள்ளடக்கத்தை உருவாக்குவதில் விக்கி அசைவியக்கம் ஏற்றுக்கொண்டுள்ள வகிபாகம்.

தமிழில் திறந்த உள்ளடக்கத்துக்கான தேவை.

தமிழில் திறந்த கூட்டுழைப்பு மூலமாகவே நடுவப்படுத்தப்பட்ட செயற்றிட்டங்களை முன்னெடுக்கக் கூடியதாக இருக்கும் யதார்த்தமும் சாத்தியங்களும்.

தமிழ் இணையத்தின் வளர்ச்சிக்கும் தமிழ் மொழி, மொழிசார் மக்களின் மேம்பாட்டுக்கும் விக்கி அசைவியக்கம் ஆற்றக்கூடிய பங்களிப்புக்கள்.

தமிழ் இணையக்குடிமக்கள் விக்கி அசைவியக்கத்துடன் இணைவதற்கான, விக்கி அசைவியக்கத்துக்குப் பங்களிப்பதற்கான சாத்தியங்கள்.

விக்கி அசைவியக்கங்களில் அரசுகள் தொடக்கம் பலம் வாய்ந்த நிறுவனங்கள் வரை பங்குபற்ற முன்வரவேண்டிய தேவைகள் குறித்த உரையாடல்

சகோதர விக்கித்திட்டங்கள்

தமிழில் வெற்றிகரமாக இயங்கும், தமிழ் இணையத்துக்குப் பயனுள்ள ஏனைய சகோதர விக்கித்திட்டங்கள் குறித்த அறிமுகம்.

- தமிழ் விக்சனரி: தமிழுக்கான நடுவப்படுத்தப்பட்ட பன்மொழி அகரமுதலியாக இது இருக்கும் பாங்கினை விளக்குதலும் விக்சனரியை அண்டி ஏற்பட்டுள்ள கலைச்சொல்லாக்க அசைவியக்கம் பற்றிய சிறு அறிமுகங்களும்.
- தமிழ் விக்கிமூலம்: திறந்த புலமைச்சொத்துக்கள், தமிழின் மூல நூற்களை, அறிவு மூலங்களை இதன்வழி இணையத்தில் வைத்திருப்பதற்கான வழிவகைகள் தொடர்பான உரையாடல்
- தமிழ் விக்கிநூல்கள்: திறந்த நிலையில் கூட்டுழைப்பாக நூற்களைத் தமிழில் உருவாக்கும் வாய்ப்புக்கள் பற்றியும் அதற்கு விக்கி நூல்களைப் பயன்படுத்துவதில் ஏற்படக்கூடிய போதாமைகள், நன்மைகள் போன்றனவற்றை அறிதலும் போன்றன.

செய்துகாட்டல்

விக்கிபீடியாவுக்கு பங்களிப்பது எப்படி என்பது குறித்த அடிப்படையிலுருந்தான செயன்முறை விளக்கம்.

மாநாட்டின் அரங்கத்தின் சாத்தியப்பாடுகளைப்பொறுத்து இதனை வடிவமைத்துக்கொள்ளலாம்.

பின்வரும் படிமுறைகள் உள்ளடக்கப்படும்.

- மாதிரிக்கட்டுரைகள் உருவாக்குதல்
- திருத்துதல்
- கண்காணிப்பு நடவடிக்கைகளை நேரடியாகக் கண்டுணர்தல்
- மீடியா விக்கி நடையியல்
- கலைக்களஞ்சிய நடை பற்றிய அறிமுகம்

போன்றன இதனுள் அடக்கம்.

Tamil Localization,

Open Source Software,

Tamil Keyboard

தமிழ் திறவூற்று மென்பொருள்கள்

'தமிழா' தோற்றமும் தொடர்ச்சியும்

சி.ம.இளந்தமிழ்
(தமிழா குழு - மலேசியா)
email: elantamil@gmail.com

திறவூற்று மென்பொருள்களின் அவசியத்தை உலகம் உணர தொடங்கி நெடுநாள் ஆகியும், தமிழ் தகவல் தகவல் தொழில்நுட்ப வளர்ச்சியில் தமிழ் திறவூற்று மென்பொருள்களின் வளர்ச்சி குறைவாகவே உள்ளது. இக்கட்டுரை தமிழ் திறவூற்று மென்பொருளான 'தமிழா' மென்பொருளின் வளர்ச்சியையும்,பயன்பாட்டினையும், எதிர்கொண்ட சிக்கலினையும், சில தீர்வினையும் முன் வைக்கின்றது.

பொதுவாக கட்டற்ற மென்பொருள்களணைத்தும் இலவசமாகவே கிடைக்கின்றன.அது கட்டுப்பாடுகள் அற்றதாக, விடுதலை மனப்பாங்கை அடித்தளமாகக் கொண்டதாக இருத்தல் வேண்டும்.ஒரு கட்டற்ற மென்பொருளானது, பின்வரும் இயல்புகளை கொண்டிருத்தல் வேண்டுமென ஸ்டால்மென கூறியிருக்கின்றார்:

- எந்த நோக்கத்திற்கு வேண்டுமானாலும் மென்பொருளைப் பயன்படுத்துவதற்கு உரிமை
- மென்பொருளைப் படிப்பதற்கும் மாற்றுவதற்கும் உரிமை
- அடுத்தவருக்கு உதவுவதற்காக மென்பொருளைப் படியெடுக்க உரிமை
- மென்பொருளை மேம்படுத்தவும், அதனை சமூகத்தார் அனைவரும் பயன்பெறுவதற்காக பொதுவில் வெளியிடுவதற்கும் உரிமை

இன்று கனூ/லினக்ஸ் இயங்கு தளத்தைத் தவிர்த்து எண்ணற்ற பல கட்டற்ற மென்பொருள்கள் நமக்குக் கிடைக்கின்றன .கனூ/லினக்ஸ் கூட டெபியன், பெடோரா, ஒபன்சுசே, உபுண்டு என பல வெளியீடுகளாக உள்ளன .இத்தகைய மென்பொருள்களில் மிக முக்கியமானது ஒபன் ஆபீஸ் (www.openoffice.org). மைக்ரோசாப்ட் ஆபீஸ் போன்ற இம்மென்பொருள் மிகவும் சக்தி வாய்ந்தது. தமிழிலேயே கிடைக்கின்றது .

இக்கட்டுரை அம்மென்பொருளைக் கொண்டு உருவாக்கப்பட்ட தமிழா மென்பொருளை பற்றியதுதான் என்பதனைக் கருத்தில் கொள்ளவும்.

ஒபன் ஆபீஸ் தவிர்த்து, பயர்பாக்ஸ் (www.firefox.com), கிம்ப் (www.gimp.org), எவலூஷன் (www.dipconsultants.com/evolution/), டக்ஸ் பெய்ண்ட் (www.tuxpaint.org) என்று பல கட்டற்ற மென்பொருள்களும் உள்ளன மென்பொருள்களின் முழுமையான பட்டியலைக் காண்க: http://directory.fsf.org/

ஒபன் ஆபிஸ் (OpenOffice.org) என்ற அலுவலக மென்பொருள் தொகுப்பு ஆகும் .ஒபன் ஆபிஸ் ஒரு கட்டற்ற மென்பொருள்; இலவசமாகவே கிடைக்கின்றது; தமிழிலேயே கிடைக்கின்றது .சன் மைக்ரோசிஸ்டம்ஸ் (Sun Microsystems) என்ற நிறுவனத்தால் மேம்படுத்தப்பட்டு 2000 இல் உலக மக்களுக்கு இம்மென்பொருள் இலவசமாக கொடுக்கப்பட்டது .கடந்த ஒன்பது ஆண்டு காலமாக, பல்லாயிரக்கணக்கான ஆர்வலர்களால் ஒபன் ஆபிஸ் தொடர்ந்து மேம்படுத்தப்பட்டு இன்று உலகின் தலைசிறந்த அலுவலக தொகுப்புகளில் ஒன்றாக திகழ்கின்றது.

தமிழா மென்பொருளை முதழில் உருவாக்க எண்ணம் கொண்டவர் தமிழ்நாட்டைச் சார்ந்த முகுந்தராஜ் அவர்கள் அவர்களுடன் இணைந்து பல தமிழ்/தகவல் தொழில்நுட்ப ஆர்வலர்கள் இணைந்து தமிழா குழுவினை தோற்றுவித்து உருவாக்கியதுதான் தமிழா மென்பொருள் தொகுப்பாகும்.

சிக்கல்கள்

1. எழுத்துரு சிக்கல்
2. அமைப்பு சிக்கல் (Randering)
3. கலைச்சொற்கள்
4. Bug Report
5. Review

தொடக்கத்தில் -2002ல் இதனை உருவாக்கும் பொழுதில் யுனிகோட்டில் செயவதா அல்லது திஸ்கியிலும் செய்வதா என்ற சிக்கல் எழுந்தது .பிறகு இரண்டு குறியீட்டிலும் பயன்படுத்தும் வண்ணம் வடிவமைக்கப்பட்டது.

Indic-Keyboards

A multilingual Indic keyboard interface

A.G. Ramakrishnan, Akshay Rao, Arun S., Abhinava Shivakumar
MILE lab, Dept. of Electrical Engineering
Indian Institute of Science, Bangalore, India

Abstract: Input method editors or IMEs provide a way in which text can be input in a desired language. Traditionally, IMEs are used to input text in a language other than English. Latin based languages (English, German, French, Spanish, etc.) are represented by the combination of a limited set of characters. Because this set is relatively small, most languages have a one-to-one correspondence of a single character in the set to a given key on a keyboard. When it comes to East Asian languages (Chinese, Japanese, Korean, Vietnamese etc.) and Indic languages (Hindi, Kannada, Gujarati etc.), the number of key strokes to represent an akshara can be more than one, which makes using one-to-one character to key mapping impractical. To allow for users to input these characters, several Input Methods have been devised to create Input Method Editors (IME). Indic-keyboards provides a simple and clean interface supporting multiple languages and multiple styles of input working on multiple platforms. XML based processing makes it possible to add new layouts or new languages on the fly. These features, along with the provision to change key maps in real time makes this software suitable for most, if not all text editing purposes. Since Unicode is used to represent text the software also works on most applications. The output of each key press is sent to the current focused window. This means that the software can be used in any application that can render Unicode.

Objective

The focus has been to develop a multilingual input method editor for Indic languages. The interface should be minimalistic in nature providing options to configure and select various languages and layouts. Configurability is inclusive of addition of new layouts or languages and option to enable or disable the software. Inputs can be based on popular keyboard layouts or using a phonetic style.

Motivation

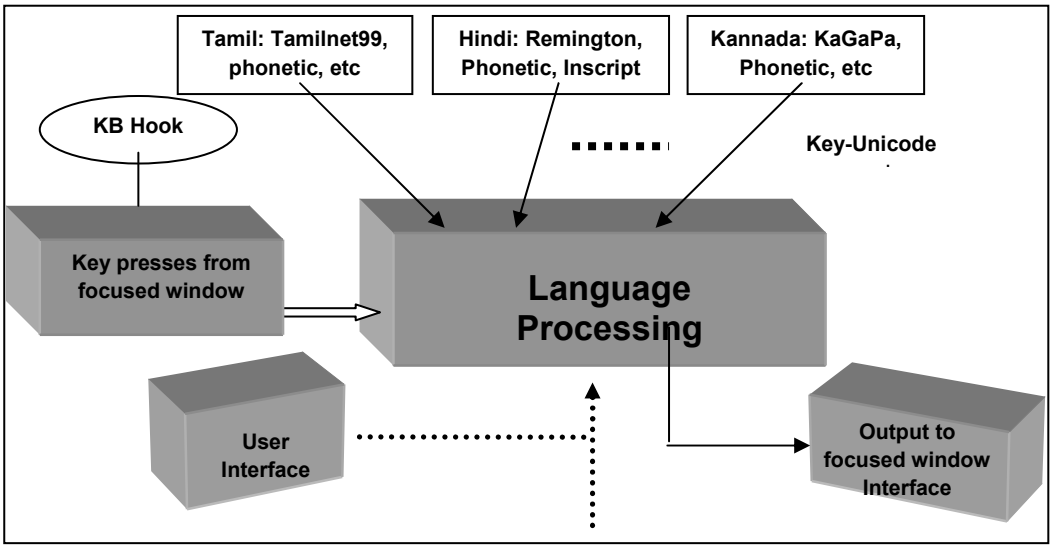
- To provide a free software with an unrestrictive license.
- Phonetic as well as popular layout support in a single package.
- Need for a single multiplatform software.
- Ease of configurability and customizability.

Introduction

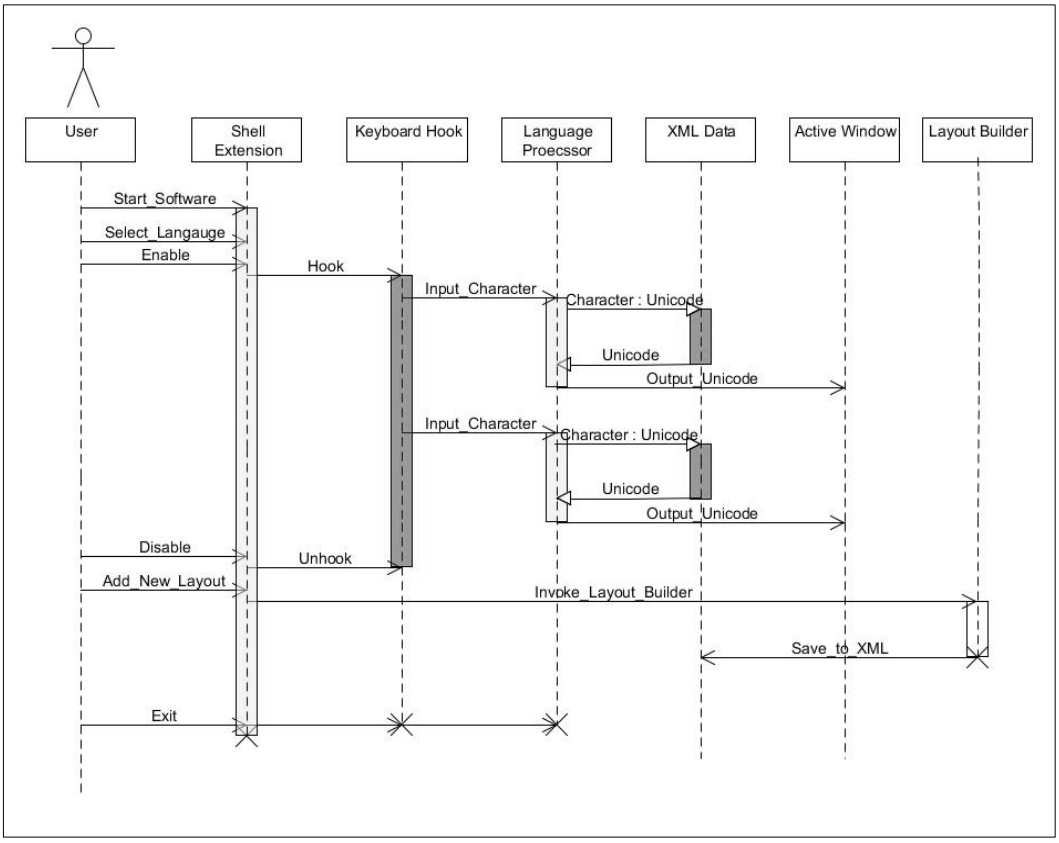
Indic-keyboards – A Multilingual Indic keyboard interface is an Input Method Editor that can be used to input text all the Indic languages under Unicode, namely, Tamil, Hindi, Kannada, Telugu, Gujarati, Marathi, Bangla, Odiya, Gurumukhi, Malayalam. The input can follow the phonetic style making use of the standard QWERTY layout along with support for popular keyboard and typewriter layouts of Indic languages using overlays. The languages are encoded using the Unicode standard. It is multi-

platform, currently designed to work on Microsoft Windows and Linux. The software uses Operating System specific Keyboard hooks to obtain characters from the keyboard and to print the Unicode to the current focused window. Eclipse SWT is used for the front end. XML is used to specify the language grammar and the Unicode equivalents for the languages.

Design



Event Sequence



Features

- Phonetic as well as popular keyboard layouts.
- XML based processing.
- Dynamic module enabling addition of new KB layouts.
- LINUX & WINDOWS.
- No installation hassles.
- System files are untouched.
- No recompile necessary to add new KB layouts.
- Phonetic key maps can be changed to suit user requirements.
- User Interface for adding new layouts.
- Open Source
- Option to show image of the current Keyboard layout.

Conclusion

In conclusion, the project has been an attempt to having a very dynamic input method editor. The flexibility of adding new Indic languages on the fly, modification of the existing layouts, changing the key press - Unicode input combination for phonetic input and a host of many other features makes for very easy to use software. The main focus has been on flexibility; ease of use and to keep things to a minimum. Keeping that in mind, we have abstained from touching or modifying any system files as well as relieving the user of all installation hassles. From the user's perspective, all one needs to do is download the files and run it. This also means the user can run the software through a pen drive, CD, DVD, hard disk or any media for that matter. The software being open source and licensed under the Apache 2.0 License, developers and users alike can modify, recompile, or rewrite the entire source and can also make these appendages closed source. Apache 2.0 license also allows developers to sell the modified code. All in all, a very dynamic, flexible, easy to use, clean, unrestrictive input method editor has been designed, which is multiplatform and multilingual.

Tamil Localisation Process – A case study

Kengatharaiyer Sarveswaran
sarvesk@uom.lk

and

Gihan Dias
gihan@uom.lk

Department of Computer Science and Engineering, Faculty of Engineering,
University of Moratuwa, Sri Lanka

Abstract: Localisation has become an active area in computer field and many organisations and individuals are localising software into their preferred languages. Different people use different localisation processes to do localisation. We cannot see much difference in the existing localisation processes. There can exist more than one localisation for a language. This difference may occur if localisers follow different glossaries and style guides. Many software are localised to Tamil language too. The aim of this paper is to describe a Tamil localisation process that is followed in Sri Lanka.

Keywords : Tamil , Localisation, Localisation Process, Locale

Background

Localisation is the process of modifying products or services to account for differences in distinct markets including language and cultural differences [3]. In software domain, converting Graphical User Interfaces (GUI) to a language while considering local policies and cultural factors can be referred to as software localisation. With the boom of Free and Open Source Software (FOSS) many organisations and individuals have started to localise a number of FOSS operating systems and Application software [2]. FOSS offers a great deal of freedom in contributing to it so that many organisations and individuals come forward from all around the world to localise FOSS. And the number of languages a FOSS support also has become marketing factor now. Not only the FOSS organisations but also the proprietary software industries also do localise their software mainly to capture the markets.

The language and other factors like date and time format, measurements, number formats, currency are collectivity referred to as locale, viz EN, en, en_US etc. A new locale is created when software is localised into a new language. Some organisations release their original software and they release the language packs separately, on the other hand some software are released directly as localised versions. These localised versions of software and languages packs are identified using locales. There are different naming conventions followed to name a locale. Mostly the locale name contains two-letter notation according to the ISO 639-1 standard [11]. However when a single language is used in two places with different parameters mentioned above, then in addition to the ISO 639-1 language code, the ISO 3166 version of country codes [7] are also joined using an underscore or a hyphen. en_US and en_UK, en-US and en-UK are some examples for this naming convention.

Different organisations follow different processes to localise software. However the main steps involving in localisation include extracting the source files, translation, testing and packaging [8][6]. Since anyone can do the localisation and also many software need to be localised there are chances for lot of confusions and inconsistencies mainly in the translations and the way the words are translated. To overcome these issues few standards such as Glossaries and localisation style guides are used during the translation phase. Glossaries are well known documents which have alphabetical list of terms in a particular domain of knowledge with the translation of those terms. GUI may consist of many elements such as menus, dialog boxes, buttons, user messages etc. There are different factors that should be considered when translating these elements. Different organisations follow different policies in translating these elements. Style guides represent how these elements should be translated. Not only these elements but also the things like how to translate acronyms, how to assign access keys, what tense to be used and where these tenses to be used, what plural rules to be used are specified in Style guide [5].

Technologically there are many tools and techniques that have been introduced to ease the tasks in each phase of localisation process. Mainly in the tedious translation phase many tools have been used based on the type of source files. For example to translate Portable Object (PO) files, the tools like POEdit, Pootle can be used [13]. During the translation the localisers may come across the terms that may repeat and also the terms that have been translated already for different software. The technique called translation memory makes it possible to reuse such terms. Many Computer Aided Translation tools support for these techniques to automate the translation to some extent. Since the localisation is not just dictionary translation, translation tools may not be useful in this context.

Localisation efforts in Sri Lanka

Localisation is very important in developing countries and it gives many benefits [1]. Specially the FOSS localisation not only lets users to use the software in local language, but also allows users to have the software free of charge [2]. Being a developing country Sri Lanka, is estimated to have an overall English Literacy of around 20% while the overall literacy rate is 90.6% [4]. Therefore the local language computing definitely reduces the digital divide that was caused due to the language barriers in Sri Lankan context as well. There are many efforts that have been made to localise Software in Sinhala and Tamil. In Sinhala there are lot of efforts to localise Operating Systems and Application Software. Presently FOSS application software such as Mozilla products, Joomla!, Moodle, GeoGebra, Squirrel Mail etc are being translated in to Tamil. Also localised versions of supporting materials are prepared in both languages [12].

Tamil localisation in Sri Lanka

Tamil is an official language in India, Sri Lanka and Singapore [10]. However Tamil people are scattered all around the world. There are many efforts that have been made to localise system software and application software in Tamil all around the world. Locales 'ta' and 'ta_IN' or 'ta-IN' were there when we entered to the localisation arena. Glossaries are identified as a key element in localisation which helps to translate the strings. India – Tamil Nadu and Sri Lanka follow different IT – Tamil glossaries and therefore for a single IT term, we may get two different translations. There was a glossary published by Sri Lankan Official Language Commission with the collaboration of Indian scholars and Sri Lankan scholars. Even in that collaborative effort for many IT terms two Tamil translations, one for Indian -Tamil Nadu and one for Sri Lanka, are given.

Style guide is another important element in localisation and we had some disagreements with the style guides that were used in other Tamil localisation.

There are some features that were not very appropriate in the Sri Lankan context. For example, in Mozilla Firefox we can define custom features like search engines, feeds etc which are local to a country. Search engines that are defined for India will help to search Indian news and matters. Therefore there was a need to go for new locale.

In addition to that, some organisations strictly follow the combined version of ISO 639-1 and ISO 3166 to define locales. For example Joomla! is one of such organisations, which defines locale in this combined format.

Due to the uniqueness in glossary, style guide, technical requirement and standard policies the Tamil localisation efforts that are being taken in Sri Lanka are identified using the locale name ta-LK or ta_LK. The ISO 639-1 language code alone is not meaningful enough to define the locale name.

Many application software are localised into Tamil language and it has been used in Sri Lanka as well as in other parts of the world. Most of the Tamil Localisations are being done at the University of Moratuwa, Sri Lanka. Not only the software GUI Localisations but also the user manuals are being prepared.

Localisation Process

At the University of Moratuwa we practice a particular localisation process similar to those who are mastering the field of localisation in other parts of the world. Here most of the phases of the process are handled using Pootle, a FOSS tool. The detail phases in the localisation process can be given as follows:

1. Identify the Software that needs to be localised and analyse the feasibility of localising it. Contact the respective organisations and inform them
2. Get the language source files and identify the appropriate tools to do the translation
3. Assign tasks to translators and get untranslated strings translated. Initially they use translation memory to translate the language strings.
4. Review the translated strings, mainly for spelling mistakes and policy mismatch
5. Package the translated files and compile the localised version of the software
6. Get the localised version of the software reviewed by people who are really going to use that software
7. Submit the localised version to the respective organisation and make the localised version available to the public
8. Prepare the required supporting materials like user manual or short guides
9. Spread local applications and take them to end users
10. Maintain the language packs and update them with the new versions and feedbacks from end users

The software is identified depending on the requirements and most of the time FOSS applications are selected. Next the localisation methodology is analysed. If it is feasible to do the localisation then it is

initiated. If there are many modules to be localised then only the essential parts of the modules are identified at the first phase of localisation. After that the tools are selected based on the format of the language source files. There are software for which we may need to use their own translation IDE. However, most of the language source files support for PO format. Therefore Pootle is being used to handle the localisation process in our context.

Once the language source files are identified then those files are added to the Pootle server and they are assigned to the translators. Then the translators download those files to their local machines and do the initial automatic translation using the translation memory. Then they carry on and translate the untranslated strings using our glossary and the style guide.

The IT- Tamil glossary published by Sri Lanka Official Language Commissions in 2000 is outdated now. Not only a lot of new terms have introduced after the year of 2000 but also some of the existing translated words in that glossary are not very appropriate. Therefore, based on that glossary we are continuously building a new set of terms and also we add new terms to the new glossary that we are building. For preparing this glossary we follow a separate process and that is not in the scope of this paper.

We also have our own style guide according to which our translators do the translations. Compared to other style guides that are used in Tamil localisation our style guide has some notable differences. These are due to our contextual need. The main differences are,

- Provide access keys in English: This is because Sri Lanka is a country where three languages including English are in use. Therefore producing a keyboard in one language is not feasible, especially for government and public sectors. The standardised keyboard is a trilingual keyboard. If we provide access keys in Tamil it may be difficult to switch to Tamil language before each time access the menu. This may increase the work for user rather than reducing it.
- Write English acronyms in English itself, but may write them in Tamil within braces if necessary: This is because the transliterated form of English acronym may give funny meaning in Tamil. Therefore we write them fully in Tamil or let them in English.
- Do not transliterate and write names in Tamil, but may write transliterated version in braces: Again we do not transliterate names as they may give wrong pronunciation. But if it is really necessary we give the transliterated version in braces.

Once the translation phase is over the translated strings are reviewed. A team is formed for this and it reviews the translated strings. Then only it is built into the final product. Again the localised product is given to users who are really going to use that software. Then with the suggestions from the reviewers the translation is committed to the relevant organisations and released for public use.

Along with the translation of the software, the supporting materials such as user manuals or very short guides are also prepared in Tamil to provide support to a novel user. The prepared localised manuals are also released to the public.

A number of roles should be played by a team of people throughout this localisation process. Translation project managers, project owners, translators, linguistics, manual authors, supervisors and end users are the main roles in our localisation process. Moreover none of these works is one time work. Hence the process continues by maintaining the translations and taking feedbacks from the end users. These feedbacks are incorporated in the successive releases of the software.

However much effort is put into a localisation work it has less or no worth until it reaches the proper audience. This crucial phase is not practiced in any other localisation process. Apart from carrying out the localisation, we also do local application awareness programs all around the country. Several such programs have been held successfully in Schools and in Universities in Sri Lanka. The aim of these awareness programs is not only to introduce the localised version of the software but also to motivate users to use the local applications. In some situations we also give rewards to motivate users.

Success stories

There are many FOSS software that have been translated to Tamil language and they are being used by a range of users in Sri Lanka, from schools kids to university students. There are people who have come up with web sites after the introduction of local applications. Also as trainers we could see the enthusiasm of the people about localised software during the awareness sessions [12]. Mozilla Firefox, Mozilla Thunderbird, Moodle, Joomla!, SquirrelMail, Horde, GeoGebra are some applications that have been successfully translated using the localisation process that is discussed in this paper.

Not only these software but also the user manuals for these applications have been prepared. These are available in electronic format [12] and also in hard copy format. We continuously revise and update the manuals as well as the software..

Another success factor of our localisation efforts would be the number of hits that these software releases get over the internet. Among the above mentioned software Mozilla Firefox, Joomla! and Moodle have got the popularity all over world. These have been used by many people in Sri Lanka as well as people living in other countries. Out of the above mentioned software, Mozilla Firefox got the highest success and it was downloaded by more than 21 000 people [9].

Conclusion

Through the experiences and outcomes we can say that our localisation process is a successful one. Specially, taking the local applications to the people is a very important phase in it. This will increase the usage of local applications as well as improve the computer literacy.

Acknowledgement: We thank the LAKapps team members of University of Moratuwa, Sri Lanka for the work that they do on localisation and for their support. We also thank ICTA Sri Lanka and LK Domain for funding these efforts.

References

1. Anousak Souphavanh, Theppitak Karoonboonyanan, 'Free/Open Source Software: Localization', Asia-Pacific Development Information Programme, 2005.
2. S.W.Q. Jaffry,U.R. Kayani, 'FOSS Localization: A Solution for the ICT Dilemma of Developing Countries', 9th International IEEE MultiTopic Conference, 2005.
3. What is localization?, http://www.lisa.org/Frequently-Asked-Que.46.0.html. Accessed : 2009-08-05
4. Wasantha Deshapriya, 'Sri Lankan Country Report on Local Language Computing Policy', PAN Localization.
5. Microsoft Language Portal, http://www.microsoft.com/language/en/us/download.mspx. Accessed : 2009-08-05
6. Localization process, http://www.project-open.com/whitepapers/localization/l10n_biz_view.html. Accessed : 2009-08-05

7. English country names and code elements, http://www.iso.org/iso/english_country_names_and_code_elements. Accessed : 2009-08-05
8. Localization process, http://developer.apple.com/internationalization/localization/process.html. Accessed : 2009-08-05
9. Mozilla addons, Tamil (LK) Language Pack 3.0, https://addons.mozilla.org/en-US/firefox/addon/6651. Accessed : 2009-08-05
10. Tamil language, http://en.wikipedia.org/wiki/Tamil_language. Accessed : 2009-08-05
11. Codes for the Representation of Names of Languages, http://www.loc.gov/standards/iso639-2/php/code_list.php. Accessed : 2009-08-05
12. www.lakapps.lk. Accessed : 2009-08-05
13. Open Office –Wiki, http://wiki.services.openoffice.org/wiki/Pootle_User_Guide, Accessed : 2009-08-05

தமிழ் மென்பொருள்களும் மக்கள் பாவனையும்

சிவா அனுராஜ்
நிர்வாக இயக்குனர் – புலம்பெயர் தமிழர் உலகம் (NRT World Inc.), வட அமெரிக்கா
Email: tamilambu@yahoo.com

கணினி மற்றும் இணையம் போன்றவற்றின் பாவனையில் தமிழ் மென்பொருள்களின் இருப்பும் பாவனையும் தொடர்பாக 'தமிழ் மென்பொருள்களும் மக்கள் பாவனையும்' என்ற தலைப்பிலான இந்த கட்டுரை ஆராயவிருக்கிறது. கடந்த 7 வருடங்களாக கணினி, இணையம் போன்றவற்றின் தமிழ் பாவனையாளர்களுடன் எனக்கு இருந்துவந்த நெருக்கமான தொடர்பு இந்தக் கட்டுரையை வரைவதற்கு தூண்டுதலாக அமைந்ததுடன் இங்கே ஆராயும் விடயங்களின் உண்மைத்தன்மையையும் நியாயப்படுத்தும்.

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 பாவிற்பதற்கான 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 பார்த்தால் பாவனையாளர்களே கருத்தில் கொள்ளப்படவேண்டியவர்கள். அந்தவகையில்,

பாவனையில் உள்ள தமிழ் மென்பொருள்கள் சரியான முறையில் பாவனையாளர்களிம் சென்றடைந்திருக்கிறதா?

தற்பொழுது பல விதமான உயர் பாவனைத்திறன் கொண்ட தமிழ் மென்பொருள்கள் உள்ளபோதிலும் அவை பெரும்பாலான மக்களிடன் 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 ஆற்றும் ஒரு சேவையாகவே கருதப்படுகிறது. இதனால் அந்த மென்பொருள்களினை உருவாக்குவதுடன் தமது கடமை முடிந்துவிடுவதாக பலர் கருதுகின்றனர். மக்களுன் இவ்வாறான மென்பொருள்களினை காட்சிப்பொருள்களாக பார்க்கிறார்களேயன்றி பாவனைக்கானதாக உணரவில்லை.

எவ்வாறு சகல தமிழ் மக்களிடமும் இந்த தமிழ் மென்பொருள்களை கொண்டு சேர்க்கலாம்

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 பாவனையாளர்களை 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.

Inside Tamil Unicode

Tamil Inayam 2009

Sinnathurai srivas

Document No: Arai/2u Version: 2.2 Date: 15/10/2009

Document Purpose

Titling the document as "Inside Tamil Unicode", the author intends to interpret, analyse and detail the inner assignment of character codes in Tamil Unicode Encoding and expand on how this understanding can be applied to the development of Tamil software and Unicode Tamil font. As the author comes from a background of Computing and Information Technology development coupled with a long term involvement in researching the ins and outs of Tamil Alphabet and phonology, this paper intends on detailing the technical and cultural merits of the current Tamil Unicode encoding and opens up thoughts on future encoding enhancements.

Audience

This paper together with a presentation on stage specifically targets Tamil Computing related software development, complex rendering processing, Tamil Unicode and Indic Unicode font development, speech recognition using Tamil alphabet system, Tamil pronunciation dictionary, spell and grammar check for Tamil, and any one interested in Tamil as a language. This document also touches on the authors' long proclaimed interpretations of the definitions of Tamil alphabet/ezuththu and thoughts on Ezuththuch chiirmai.

Background

INFITT is the technical institution that works hand in hand with Unicode Consortium, Tamil Nadu government, Tamil IT related government organisations of the world, Tamil and other universities, and Tamil computing related software and hardware developers, along side Unicode Consortium. This paper is intended for describing technical and cultural information to anyone who chooses to use it.

Unicode Consortium is the international institution tasked with standardising all languages of the world for enabling the standard international multilingual computing. Tamil Unicode is already a working entity in Computing and the works in progress are for enhancing and facilitating a totally empowering Tamil computing environment.

The following topics are covered in this presentation.

- Unicode as a Linear Encoding
- Introducing Fallback Unicode characters to match the linear encoding
- Linear encoding as an extension of Tolkappiyam
- The need for complex rendering and the displaying/printing of Tamil
- Tamil in cut down versions of OS and domestic appliances
- Simplifying sorting process and code point for inherent "a".
- Pulli vs Virama

- Matra vs Matrai
- Extra long vowels, kuTTiyal, Matrai
- Deprecating the duplicate "au" marker
- Eliminating the alternative aravu usage with combining ee and oo.

Inside Tamil Unicode

Tamil Grammar and Liner Unicode Encoding

Ancient Tamil grammar Tolkappiyam, which is also the contemporary Tamil Grammar, defines 30 alphabet (ezuththu) and 3 markers as the entities used in Tamil writing system. It is important to note that unlike any other languages of the world, Tamil writing system names the "Places of Articulation/Birth" (Pirappidam) in human organs, which generate speech sounds/phonemes, and defines those Places of Articulation (PoA) as alphabet. It is also important to note that each alphabet represents a spectrum of phonemes generatable by its PoA.

The thirty alphabets are divided into two types as 18 consonants and 12 vowels. Consonants (mey/physique) are the covered physical organs of the PoA while vowels (uyr/soul) are the covered places of articulation. The vowel is believed to agitate the consonant to live for a suitable matrai/time-interval and vowel can also spring to live on its own to a required matrai. However, the consonant in general is thought incapable of springing to live on its own; hence in Tamil, consonant conjuncts are thought scientifically as non-existent. However, Tamil defines near-voiceless/kuRRiyal vowels as enabling factor for consonant-conjuncts.

Unicode is yet to encode **kuTTiyal markers** and **matrai/timing markers** as defined in Tamil Grammar. In day-to-day usage, Tamil is believed to contain countless number of phonemes because of the scalable nature of each PoA. A request for the use of **diacritic-markers** to define phonemes into sub-spectrum ranges is yet to be made, probably by INFITT. This will be useful for activities such as pronunciation dictionaries and speech recognition interpreters.

Unicode encodes thirty alphabets and one marker (aytham) as the basic Tamil character set. This is called linear encoding. The philosophy of 30 characters is in sync with Tamil grammar. Additionally a few Tamil-Granta characters are also encoded within the Tamil Unicode code range. Though it is in day-to-day usage, Tamil-Granta however does break the principle that alphabets in Tamil represent scalable PoA and not phonemes.

In theory the processing of Tamil data is achievable at linear Plain-1 level. This would imply, for example sorting of Tamil text is processed at 16bit Plain-1 without consideration for any complex processing of 32bit and beyond. In theory, for Tamil, only display and printing should be processed using complex rendering at 32bit and beyond.

However, due to present state of operating system design shortcomings and shortcomings within Unicode (additions) definitions, complex processing is also required for a very few instances of processing in Tamil software, in addition to print and display processing. A fix for these shortcomings cannot be expected in the immediate future.

Therefore, when architecting software solutions, one need to decide if linear processing requirements and complex processing requirements can be modularised so that **about 99%** of the development tasks can be simplified to 16bit processing. Some of the shortcomings in Unicode definitions can be listed as "X as the only one complex conjunct in Tamil", " duplicate au-marker", and clearance for the use of

duplicate “combu-markers”, which cause unnecessary software development initiatives, when Tamil could be developed simply at 16bit level. Inherent “a” is currently not encoded for any of the Indic languages.

For now, developers can assign their own code point to inherent “a” so that software routines for sorting Tamil can be made virtually a simple task. All of the shift and pull operations required to sort in complex requirement, because of the lack of inherent “a” code point, will become a simple task once a code point is assigned for inherent “a”.

There are written materials that talk about the discussions took place in the ancient time about the pros and cons of the use of inherent “a” and visible “combing “a”. The debate still goes on and that requires the facility to express the various theories on the pros and cons. It is therefore necessary to **encode inherent “a”** with an alternative visible “a” form and its usage be permitted for research and software processing purposes.

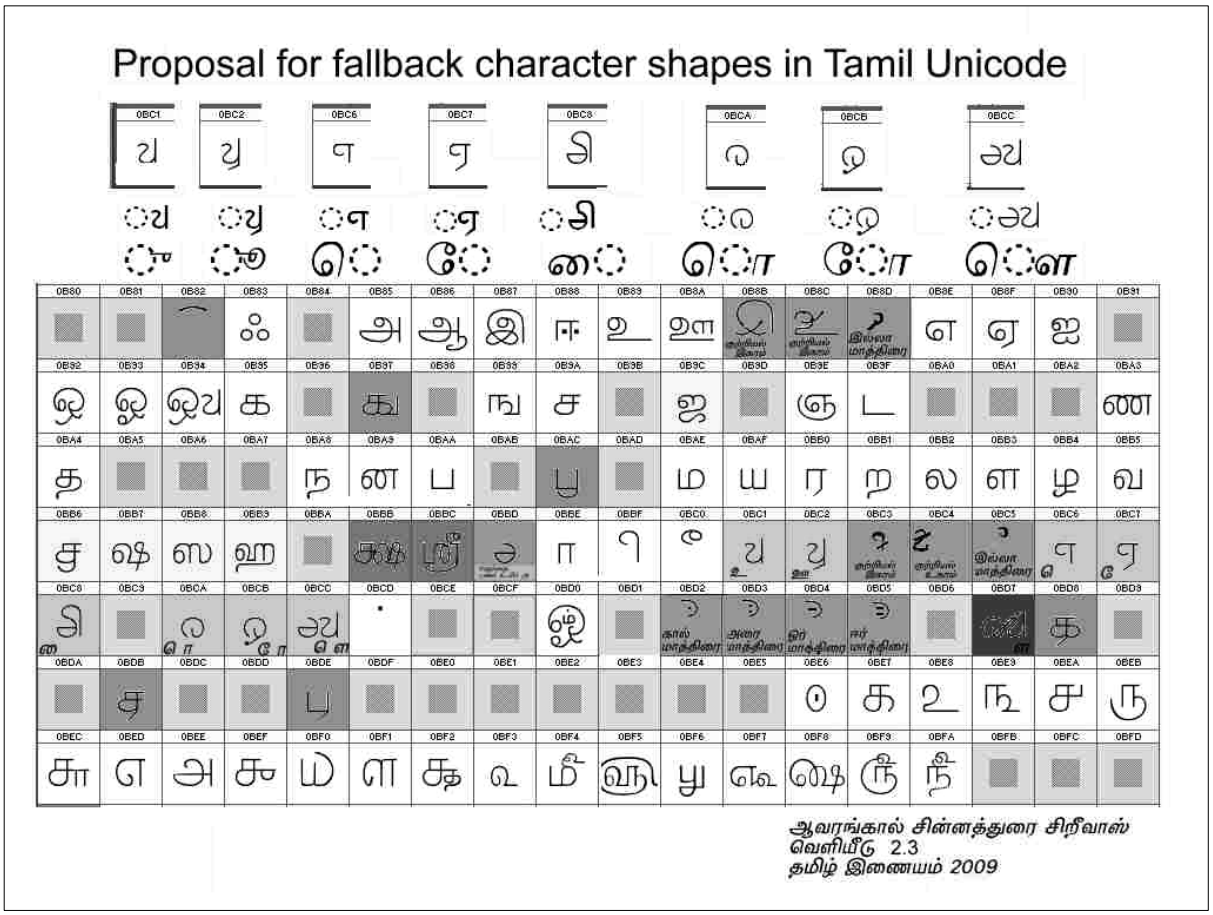
Software developers need to understand that in Tamil Unicode there are no “combukaL (ு, ூ)”; there is no “sangkili kombu (ை)”; there are no “uhara-mey nor uuhaara-mey (கு, கூ, ... னு, னூ)”. The uyr-meykaL that appear on print and on screen are not really there. They are only software illusions. Once a software developer understands this phenomenon, the development cycle would become a walk through exercise. However, erroneously encoded secondary “au marker U+0bd7” in Unicode is real and this can cause havoc if its behaviour is not understood properly. Note that the primary “au marker U+0bcc” behaves normally like any other combining vowels and it is fully fit for purpose.

As depicted in the image below, all the real combining vowel markers only combine with consonants from the right. This is in line with Tamil Grammar, while the contemporary written Tamil combining vowels misleadingly combines with consonants from left or right or top or bottom or left & right and even manufactures new shapes (u-haram, uu-haaram). For software development purposes misleading appearances can be discarded and grammatical definitions of alphabet can be utilised.

To graphically represent the linear nature of combining vowels in Unicode/Tolkappiyam the letters’ shapes “ா ி ீ ு ூ ெ ே ை ொ ோ ௌ” are recommended. All of these shapes combine with consonants, visibly on the right side, as logically coded in Unicode/Tolkappiyam. When there is a lack of complex rendering OS interface, Tamil need to use these linear forms of the combining vowels. Simple electronic devices would always lack complex rendering facility. Displaying the deformed combukaL in these instances would be an utter degrading experience.

Some examples of utter degrading Unicode displays are கெ=க◌ி, கொ=க◌ொ, கே=க◌ே, கோ=க◌ோ, கை=க◌ை. As the simple electronic devices are always going to be in existence, Unicode standard should allow the **Fllback characters** of combining vowels as ◌ா ◌ி ◌ீ ◌ு ◌ூ ◌ெ ◌ே ◌ை ◌ொ ◌ோ ◌ௌ and not as degrading கெ=க◌ி, கொ=க◌ொ, கே=க◌ே, கோ=க◌ோ, கை=க◌ை.

The picture below provides a conscious guide to what is encoded in Unicode, what need to be encoded in Unicode, what need to be deprecated out of Unicode, what need to be clarified as external to Tamil and also gives a head start for Tamil character reform in line with Tholkappiyam.



A long standing claim by me is that CombukaL in Tamil are a highly illogical, disruptive and retarding influencers that were introduced in recent times by Viramamuni, because of his lack of understanding of Tamil, even though intentions were good. So the combukaL related linear representation can be made part of Tamil character reform. It is scientific, logical and simple if all combining vowels take position on the right side of consonant, naturally. For this reason, proposed linear characters can be classed as potentially an important item in the Tamil character reinstatement/reform agenda. The withering of u-hara meykaL and uu-haara meykaL will also come natural, if Unicode fallback characters are made part of any reform agenda.

Numerous phonemes/sounds exist and are used in day to day language. Each PoA generates a number of phonemes. Diacritic markers are essential to document and communicate the use of these phonemes and also to publish pronunciation dictionaries. New Unicode code point ranges need to be allocated for Tamil diacritics or the feasibility of using the existing diacritic markers in Unicode for Tamil with the defined diacritic glyph shape and positions need to be considered. (Unicode code range U+0300 to U+036f. The PoA/pirappidam "*த*" for example, generates multiple phonemes such as *அத்தகு, அந்த, அதன்*. The phoneme *த̪* is not proved to exist in any other languages. Though some say some use this phoneme *த̪* in the word father rather than the phoneme *த*. The PoA/pirappidam "*அ*" for example, generates multiple phonemes such as *அம்மா, அன்னை* and *வகை, விரை*. The effect of diacritics when combining before or after a Tamil character also need thorough investigation, because

of the nature of disruptive combining vowels in Tamil, joining the consonants in every direction in a non-uniform fashion.

In Unicode, the character Virama for other Indic languages was erroneously assumed to have similar properties to the puLLi character in Tamil. However, there are considerable differences between Virama and puLLi. Because Unicode names the puLLi also as Virama (with recent annotation), there is a danger that developers may assume the general characteristics of Indic Virama as the characteristics of puLLi and indulge in wasted development activities. Similarly Aytham in Tamil is wrongly named and defined (with recent annotation) as Visarga. Visarga has totally unrelated properties to Aytham; hence the properties of Aytham are wrongly defined in Unicode. Again, it is the responsibility of developers not to indulge in wasted efforts by the lead-believe, that the Visarga and Aytham as having the same/similar properties. The Aytham in Tamil act as a vibrationary modulator for glotalising and other purposes of speech. Example formation of Aytham are "ஃக, ஃஃத, பஃ, பஃஃ, கஃஃப" and the recent usage of "ஃப" to clearly identify the phoneme "f=ஹ ". For example, the single character KHA in Devanagari translates as KAH=கஃ=க்ஃ. Matrai in Tamil grammar defines the timing mechanism involved with generating spoken sounds, while Matra in Unicode is used to denote the combining vowels (puNar Uyr/Pin uyr ezuththukkaL ா ி ீ ு ூ ெ ே ை ொ ோ ௌ).

References

1. Tamil Unicode Chart: http://www.unicode.org/charts/PDF/U0B80.pdf
2. Combining Diacritics U+0300 to U+036f: http://www.unicode.org/charts/PDF/U0300.pdf
3. Tamil grammar "Tolkappiyam"

Building Tamil Unicode Fonts for Mac OS X

Muthu Nedumaran
(Muthu at Murasu dot Com)

Introduction

The Tamil script, like all other Indic scripts, is a syllabic script. It has 12 independent vowels (உயிர் எழுத்துகள்), 18 consonants (மெய் எழுத்துகள்), Aytham (ஆய்த எழுத்து) and compound forms (உயிர்மெய் எழுத்துகள்) that represent combinations of a consonant and a vowel[1]. In addition, there are grantha letters that are used to write words of non-Tamil origin.

The Unicode Standard (Unicode) encodes the following groups of characters[2]:

a) Letters: Independent vowels, consonants and aytham
b) Dependant vowel signs
c) Tamil numerals
d) Various signs and symbols

Having these encoded characters alone in a Tamil Unicode font (Tamil font) is not sufficient to render a readable Tamil text. The compound forms are also required.

The compound forms are not encoded with single (atomic) code points. Thus, a Tamil Unicode font will contain glyphs that do not have a character code. In other words, there will be more glyphs in the font than there are Tamil characters encoded in Unicode.

In addition to the glyphs, the font should contain rules that dictate the formation of compound forms from the respective consonant-vowel pairs. These rules are called *shaping rules*.

In a Tamil font designed for Microsoft Windows platforms, the shaping rules are defined in OpenType (OT) tables. In Mac OS X, these rules are defined with AAT tables.

Windows XP and later versions of the Windows platform includes a Tamil font called Latha. Mac OS X has one called InaiMathi since version 10.4 (Tiger).

This paper presents the steps required to build a Tamil font for Mac OS X.

Prerequisites

In the interest of space and time, this paper assumes that the reader is familiar with the following:

1. How Unicode defines characters? (http://unicode.org)
2. Difference between a character and a Glyph
3. Code-point order vs presentation order
4. Using the Terminal application in Mac OS X

Glyphs in a typical Tamil font

The figure below shows the glyphs in a typical Tamil font.

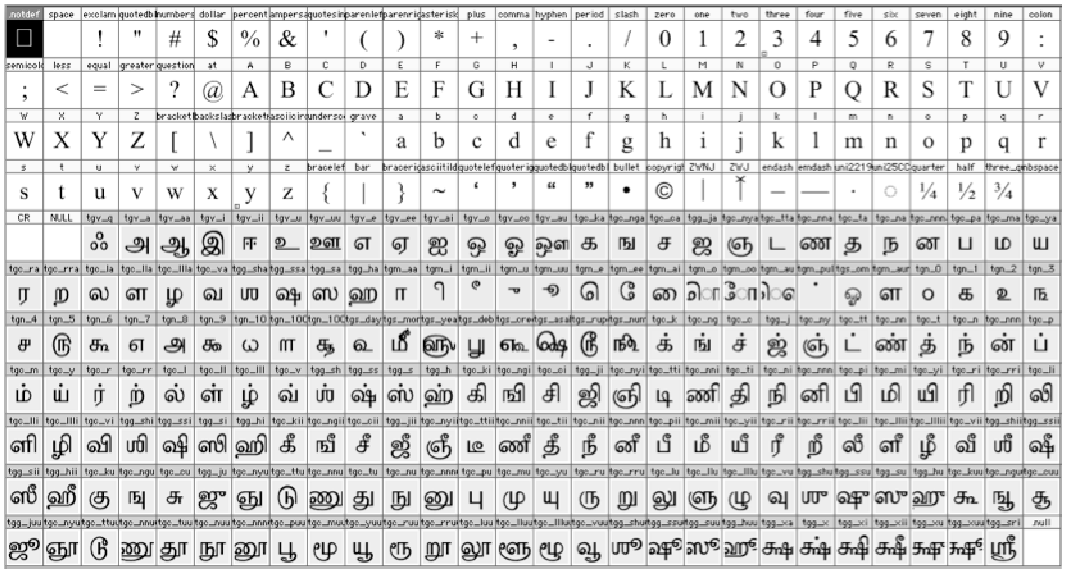


Figure 1: Glyphs in a Tamil font.

The choice of glyphs may vary from font to font. Some may have the pure consonants (base+pulli) pre-composed as with the font above. Others may just have one pulli glyph and use kerning tables to position it above the base glyphs. The font above has all possible combinations pre-composed for simplicity.

Adding shaping rules

Once the required glyphs are drawn, the only remaining step is to add the shaping rules.

Two shaping frameworks are popular: OpenType (OT) used in Windows (and some Linux platforms) and Apple Advanced Typography (AAT) used in Mac OS X.

Both these frameworks differ in design philosophies. OT attempts to do some of the common processing in an external engine called Uniscribe. Uniscribe eliminates the need to define glyph reordering and two part vowel handling in a Tamil font. While this appears to make things easier for the developer, it does limit flexibility. Since most font developers use a common template for shaping rules, across all their fonts, the real benefit Uniscribe provides at the expense of complexity and performance is hard to realise.

AAT on the other hand, provides complete freedom to the developer. There is no script processor. Therefore, all of the shaping rules are defined by the developer and are included in the font. Once a working font has been built, the same rules can be applied to all other fonts by simply compiling the definitions into the font. AAT is a simple and an elegant framework with less overheads in rendering.

Shaping rules with AAT

The rules are defined in a text file, called the Morph Input File (MIF) and compiled into the font file using Apple's font tools.

The command to add the rules from a MIF file is:

ftxenhancer –m <MIF filename> <TTF filename>

For example:

ftxenhancer –m my_tamil.mif my_font.ttf

Creating a MIF file

A MIF file can be created with any text editor. The rules are defined in tables arranged in the order they should be executed. For example, the letter ழ is not assigned a code-point in Unicode. Since this letter has compound forms when combined with vowel signs, it may be best to define the shaping rule for this before hand.

Shaping rules are defined using glyph names and not character codes. While it is desirable to use Adobe Standard names for glyphs, the common practice is to use friendly names or adopt a naming convention that clearly describes how the glyphs are formed. There are no rules for defining glyph names. It is entirely up to the developer.

The convention used for the font in Figure 1 is as below:

Consonants: tgc_<unicode name>. Example: tgc_ka, tgc_mi, tgc_ttoo etc

Vowels : tgv_<unicode name>. Example: tgv_a, tgv_au etc

Grantha: tgg_<unicode name>. Example: tgg_sha, tgg_juu, tgg_sri etc

Vowel Signs: tgm_<unicode name>. Example: tgm_a, tgm_au etc

The following are the shaping rules that are needed for the font in figure 1.

- Substitute BASE+I, BASE+II, BASE+U, BASE+UU and BASE+PULLI with combinations with their respective compound forms. The feature used for this purpose is called ***Ligature Substitution***.
- Rearrange E, EE and AI vowel signs so that they appear before the BASE glyph. The feature used for this purpose is ***Rearrangement***.
- Place the left and right marks of O, OO and AU vowel signs on either side of the BASE glyph. The feature used for this purpose is ***Insertion***.

The sections below describe each of the features.

5.1 Ligature Substitution

Ligature substitution is done in two steps.

First ழ ligature is formed by substituting the characters that make up this glyph: KA + PULLI + SSA (க + புள்ளி + ஷ). This is to ensure that this glyph is already available when vowel combinations are substituted. Likewise SRI can be substituted for SHA + PULLI + RA + VOWEL_SIGN_II (+ புள்ளி + ர + ீ).

Second, all compound forms for I, II, U, UU and PULLI signs are substituted.

Examples of these tables are given below:

```
Type                LigatureList
Name                Ligatures
Namecode            1
Setting             Required Ligatures
Settingcode         0
Default             yes
Orientation         HV
Forward             yes
Exclusive           no

List
    tgg_xa      tgc_ka tgm_pulli tgg_ssa
    tgg_sri     tgg_sha tgm_pulli tgc_ra tgm_ii
```

Figure 2: First table in the Tamil MIF file

```
Type                LigatureList
Name                Ligatures
Namecode            1
Setting             Required Ligatures
Settingcode         0
Default             yes
Orientation         HV
Forward             yes
Exclusive           no

List
    // --- substitutions for glyphs with I & II marks
    tgc_ki      tgc_ka tgm_i
    tgc_ngi     tgc_nga tgm_i
    ....
    tgg_hi      tgg_ha tgm_ii
    tgg_xi      tgg_xa tgm_ii
    // --- substitutions for glyphs with U & UU marks
    tgc_ku      tgc_ka tgm_u
    tgc_ngu     tgc_nga tgm_u
    ....
    tgg_xuu     tgg_xa tgm_uu
    // --- substitutions for pulli
    tgc_k       tgc_ka tgm_pulli
    tgc_ng      tgc_nga tgm_pulli
    ....
    tgg_x       tgg_x tgm_pulli
```

Figure 3: Ligature substitution for compound forms. This table follows the earlier table so that tgg_xa is already available from the first substitution

5.2 Rearrangement

In a string of Tamil text, vowel signs are stored after the base character. This is the same with any Indic script. For example, the word மலைநாடே is stored in memory as below:

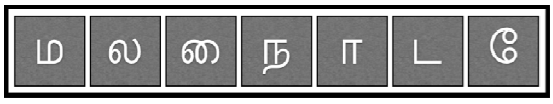


Figure 4: Memory representation of the word மலைநாடே.

When this word is rendered, the AI and EE vowel signs need to be re-ordered so that they appear before the base glyph. In Windows OpenType, this process is not necessary, as the Uniscribe engine will do the reordering internally. In Mac OS X, this rule needs to be defined. It can be easily done with the Rearrangement feature in AAT.

Rearrangement and Insertion features use state tables to mark the positions and perform a defined action when desired glyphs are seen together4. Unlike rearrangements in other Indic scripts where conjuncts and consonant signs may be involved, Tamil rearrangement is a simple act of swapping the positions of the base glyph and vowel sign.

Figure 5 shows the state table required for this feature. When a base glyph, defined in the *Cons* group, is seen, it is marked and the next glyph is examined. If the next glyph is a member of the *RVowel* group, the base and vowel sign are swapped. Since reordering is a required feature, and involves all of the base glyphs in Tamil, this table can be used in any Tamil font created for Mac OS X.

```
Type            Rearrangement
Name            Rearrangement
Namecode        8
Setting         Do Rearrangement
Settingcode     1
Default         yes
Orientation     HV
Forward         yes
Exclusive       no

Cons    tgc_ka tgc_nga tgc_ca tgc_nya tgc_tta tgc_nna
+   tgc_ta tgc_na tgc_pa tgc_ma tgc_ya tgc_ra tgc_la
+   tgc_va tgc_llla tgc_lla tgc_rra tgc_nnna tgg_ja
+   tgg_ssa tgg_sa tgg_ha tgg_sha tgg_xa

RVowel  tgm_e tgm_ee tgm_ai

            EOT OOB DEL EOL Cons    RVowel
StartText   1   1   1   1   2       1
StartLine   1   1   1   1   2       1
SawCons     1   1   1   1   2       3


    GoTo        MarkFirst?  MarkLast?  Advance?  DoThis
1   StartText   no          no         yes       none
2   SawCons     yes         no         yes       none
3   StartText   no          yes        yes       xD->Dx
```

Figure 5: Rearrangement Table

5.3 Insertion

Insertion involves base glyphs with vowel signs O, OO and AU. The word கோ is represented in memory as shown below:

Figure 6: The word கோ in memory

Although three glyphs are needed to present the compound form, there are only two characters in memory. A glyph needs to be inserted somewhere in order to get the required three.

There are many techniques to do this. One of them is to draw the vowel sign OO in the font as just the *kaal* and insert vowel sign EE before KA. This will provide the desired effect. However, the glyph for vowel sign OO becomes misleading.

A more efficient technique can be as described in Figure 7.

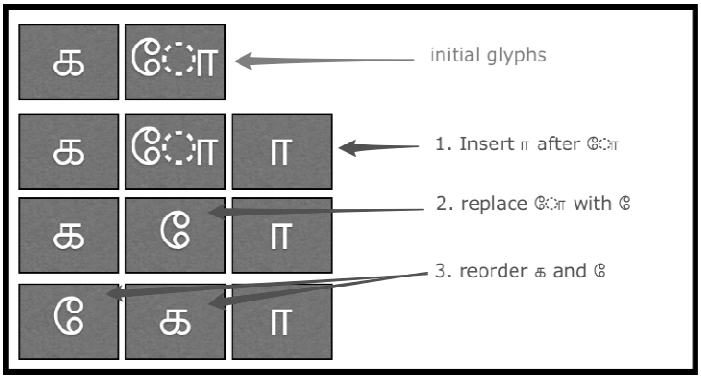


Figure 7: Technique to perform insertion without changing the shape of vowel sign OO.

The steps can be performed with the following MIF file entries:

Step 1: Insertion. This is done with state tables:

```
Type            Insertion
Name            NULL
Namecode        16000
Setting         NULL
Settingcode     0
Default         yes
Orientation     HV
Forward         yes
Exclusive       no

okaram      tgm_o
ookaaram    tgm_oo
aukaaram    tgm_au

                EOT OOB DEL EOL okaram ookaaram aukaaram
StartText       1   1   1   1   2      3        4
StartLine       1   1   1   1   2      3        4

    GoTo            Mark?   Advance?    InsertMark  InsertCurrent
1   StartText       no      yes         none        none
2   StartText       yes     yes         none        oinsert
3   StartText       yes     yes         none        ooinsert
4   StartText       yes     yes         none        auinsert

oinsert
    IsKashidaLike   yes
    InsertBefore    no
    Glyphs          tgm_aa

ooinsert
    IsKashidaLike   yes
    InsertBefore    no
    Glyphs          tgm_aa

auinsert
    IsKashidaLike   yes
    InsertBefore    no
    Glyphs          tgm_aumark
```

Figure 8: State table to perform insertion of right side glyph for a two part vowel sign. Vowel sign AA is inserted for O and OO. AU Length mark is inserted for AU.

Step 2: Replacing O, OO and AU vowel sign glyphs with their respective left side signs. This can be done with substitution as shown below:

```
Type  Noncontextual
Name  Right Side Vowel Signs
Namecode 16000
Setting Right Side Vowel Signs
Settingcode 16000
Default yes
Orientation HV
Forward yes
Exclusive no

tgm_o   tgm_e
tgm_oo  tgm_ee
tgm_au  tgm_e
```

Figure 9: Substituting O, OO and AU vowel signs with their respective right side signs.

Step 3: Since this is the same as rearrangement discussed in 5.2, Step 1 and 2 can be done just before the reordering process in the MIF file. By doing so, rearrangements for step 3 can be done along with the rearrangements for E, EE and AI.

Putting it all together

The completed MIF file will have tables in the following order:

- Substitutions for ஶ and ◌ி. (*Fig 2*)
- Substitutions for I, II, U, UU and Pulli forms (*Fig 3*)
- Inserting right side vowel sign for O, OO and AU (*Fig 8*)
- Substituting O, OO and AU with their respective left side signs (*Fig 9*)
- Rearranging E, EE and AI vowel signs with their base glyphs (*Fig 5*)

Once the MIF file is created, it can then be added to the font with ftxenhancer as described in section 4.0.

References

1. http://en.wikipedia.org/wiki/Tamil_script
2. http://www.unicode.org/charts/PDF/U0B80.pdf
3. http://developer.apple.com/fonts
4. Refer to Apple font tools documentation for details.

Tamil Encoding, Keyboard Layout and Collation Sequence

Standard for ICT Sri Lanka

Balachandran G.
ICT Tamil Consultant – ICTA, Sri Lanka
balag.lk@gmail.com

Abstract: One of the need of localisation is developing standards for operating environments. When the standardisation process is through, it will be easy for developers and vendors to build applications and products for the target environment. The SLS 1326 : 2008 standard for Tamil ICT, was approved by the Sectoral Committee on Information Technology and was authorised for adoption and publication as a Sri Lanka Standard by the Council of the Sri Lanka Standards Institution. This standard defines mainly three(3) standards for Tamil ICT; viz character encoding, keyboard layout and collation sequence. **Character encoding** defines codes for the vowels, consonants, āytam, vowel modifiers, numerals, and symbols in the Tamil language. Some formations of the language are not represented by individual codes, but are generally constructed as a sequence of one or more consonants followed by a vowel modifier which forms a syllable. Standard also provides a **keyboard layout**, which in turn is based on the "Renganathan" typewriter keyboard. Key sequences are defined on the principle "type as you write". Each symbol is typed in the order it is written in, which may be different from the encoding sequence or the display order. In **collation sequence** standard, an effort has been made to preserve the alphabetical order of the Tamil language to a great extent.

Keywords: Tamil ICT, Sri Lanka, Tamil, Keyboard, Character Encoding, Collation sequence.

Introduction

Tamil (தமிழ்) is a *Dravidian* language and it has a literary tradition of over two thousand years. *Tolkāppiyam* by *Tolkāppiyar* is the earliest grammatical treatise now extant in Tamil. Although the exact date is still unascertained, there are strong arguments that the Tamil script first appeared in the early centuries of the Common Era. Tamil was accorded "classical language status" by the Union Government of India, and was the first Indic language to have been accorded such status.

We are currently at the beginning of a new era of computing - one where our people do not have to use a foreign language to benefit from information technology. Computers, phones and the Internet now work in our own language, i.e. Tamil. Initially, local language efforts in Sri Lanka were not directed towards Tamil, as it was expected that this work will be carried out in India, and Tamil Nadu in particular. However, it was observed that Indian national-level initiatives and Tamil Nadu initiatives diverged. Also we realised that the use of Tamil in Sri Lanka often diverged considerably from that in Tamil Nadu, and that independent standards and initiatives are needed in Sri Lanka. The SLS 1326 : 2008 [5] standard for Tamil ICT, was approved by the Sectoral Committee on Information Technology and was authorised for adoption and publication as a Sri Lanka Standard by the Council of the Sri Lanka Standards Institution. This standard defines mainly three(3) standards for Tamil ICT; viz character encoding, keyboard layout and collation sequence.

Tamil Encoding

Although Tamil script system is generally called an alphabet it is in fact an abugida system. An abugida has the segmental writing system in which each vowel-consonant letter represents a pure-consonant accompanied by a specific vowel; the vowels are indicated by modification of the consonant sign, either by means of diacritics or through a change in the form of the consonant. The contemporary Tamil script contains following elements in its systems, as written or printed (Sub total 326).

உயிர் எழுத்துக்கள் (uyir eluttukkal) – vowel letters	12
தமிழ் மெய் எழுத்துக்கள் (mey eluttukkal) – Tamil pure-consonant letters	18
கிரந்த மெய் எழுத்துக்கள் (Grantha mey eluttukkal) – Grantha pure-consonant letters	6
தமிழ் உயிர்-மெய் எழுத்துக்கள் (Tamil uyir- mey eluttukkal) – Tamil vowel-consonant syllables	216
கிரந்த உயிர்-மெய் எழுத்துக்கள் (Grantha uyir-mey eluttukkal) – Grantha vowel-consonant syllables	72
ஆய்த எழுத்து (āyta eluttu) – A special letter – ஃ	1
கூட்டு எழுத்து (kottu eluttu) – Conjunct syllable - ஸ்ரீ	1

Table 1 – Tamil script elements

Moreover, since Tamil is one of oldest languages it has its own numeral representations and symbols in the writing system. Character encoding defines codes for the vowels, consonants, āytam, vowel modifiers, numerals, and symbols in the Tamil language. Some formations of the language are not represented by individual codes, but are generally constructed as a sequence of one or more consonants followed by a vowel modifier which forms a syllable.

This standard simply adopts the Unicode standard (version 5.1) and provides a coding of Tamil for use in computer and communication media. This standard character code encodes the characters of the Tamil language within 128 code positions of the 16-bit Basic Multilingual Plane (BMP) of **ISO/IEC 10646: 2003.** In addition to storage, retrieval and machine to machine communication in Tamil, it also includes provisions to co-exist with other languages as specified in **ISO/IEC 10646: 2003**. In particular, English and Sinhala texts may be intermixed with Tamil. This code set is able to represent contemporary and historical Tamil writings.

Even though it is an adoption of Unicode version 5.1, it does includes following rules;

- The Anusvara and AU length mark are encoded in 0B82 and 0BD7 respectively to map along with the other Indic scripts. However, the Anusvara is not used in contemporary Tamil and AU length mark should not be used to form any Tamil text. Although the vowel ஔ can be represented in ஒ (0B92) + ◌ௗ (0BD7), this form of sequence shall not be used. One shall use the single code point 0B94 for ஔ.
- The representation of a syllable such as கௌ by a consonant character (க) followed by more than one vowel signs (◌ெ + ◌ௗ) are permitted in Unicode, but is discouraged in this standard. One

vowel sign must be used to form the syllables.
e.g. க + ொ, க + ோ, etc

- The Tamil OM may get decomposed to "ஓம்" and the canonical decomposition is 0BD0 = 0B93 0BAE 0BCD. On the other hand, the inverse canonical composition is not always true.
- When the sequence of 0B95 0BCD 0BB7 is specified by default, க்ஷ will be formed by the Unicode engine. In case க்ஷ is needed, the following sequence should be specified with the zero-width non-joiner (ZWNJ) 0B95 0BCD 200C 0BB7.
- ஸ்ரீ may also be formed from the ஸ் letter as follows:
  ஸ்ரீ = 0BB8 0BCD 0BB0 0BC0 (ஸ ◌் ர ீ – Not allowed)
  However, only the following sequence shall be used to form ஸ்ரீ
  ஸ்ரீ = 0BB6 0BCD 0BB0 0BC0 (□ ◌் ர ீ)

Keyboard Layout

Tamil Typewriter Layout

Ramalingam Muttiah of Jaffna origin is ascribed with designing and implementing the Tamil Typewriter [1]. He designed a typewriter which uses 72 keys to type all Tamil letters. His design was based on the frequency of occurrence of each Tamil letter. The க, ப, த, ம, ன, ள, ய and ட vowel-consonants are the most frequent, and were placed on the *home line* of the typewriter, as shown in Figure 1.

Figure 1 - The Tamil Typewriter keyboard

Most Tamil keyboard layouts are based on the typewriter or Remmington scheme. In Sri Lanka this layout is called the Renganathan layout. Mr. Vasu Renganathan designed a layout with some changes during 2004 and made it available for Tamil diaspora [7], and this might have had an impact on the name "Renganathan Layout". Over the years a number of variations of this layout have appeared. The Bamini Tamil font is very popular in Sri Lanka, and the keyboard layout based on this font is widely used. This layout has slight differences from the typewriter / Renganathan layout. The Thibus and Helawadana layouts are also in this category, but have minor differences between each-other and with the typewriter layout.

Inscript Layout

Inscript includes a layout for Tamil [2]. In this layout, vowels and vowel modifiers are on the left-hand side of the keyboard and the consonants are on the right-hand side. A notable feature of this layout is that some Tamil consonants appear on multiple keys, as other Indian scripts contain multiple letters

corresponding to each Tamil consonant.

Romanised Layouts

Romanised Tamil keyboards map Tamil letters to analogous English letters. Such keyboards are preferred by those who work mainly in English and use Tamil occasionally.

The Tamil99 Layout

The Tamil99 keyboard too is a *consonant-vowel* (often called *phonetic* in Tamil) layout. A key feature of this keyboard is that each vowel shares a key with its corresponding vowel modifier, and the keyboard driver produces the vowel at the beginning of a word, and a modifier after a consonant, following Tamil grammar. Another significant feature is that all Tamil (as opposed to Grantha) letters are on unshifted keys, allowing rapid and easy entry. Tamil99 also has features such as *auto-pulli*, where typing the same consonant twice automatically adds a pulli to the to the first letter. Senthilnathan says: "No language keyboard in the world is as simple as this! To represent 26 characters of English, you need 26 lower case and 26 upper case keys. But to represent 247 Tamil letters, you need only 31 keys! How laudable!" [3].

Sri Lanka Standardisation Work

The Tamil working Group of the Information and Communication Technology Agency of Sri Lanka (ICTA) studied the above keyboard layouts in 2003, and recommended the adoption of the Tamil99 layout. In addition to its technical superiority, a main reason for this choice was its adoption by the government of Tamil Nadu.

The Tamil99 keyboard was endorsed by the user community and successfully tested at a pilot govt. site. However it was not accepted by the public. The main reason for its non-acceptance in govt. offices is that experienced typists found it very difficult to adjust not only to a different keyboard layout, but also a completely different keying-in methodology.

At a discussion held in 2006, the reasons given by typists for not using the Tamil99 keyboard were:

1 Text is typed differently from how it is written.
2 The key placements are totally different form the typewriter layout. Although it may be more efficient, the new scheme is unfamiliar and thus hard to use.
3 The lack of vowel symbols printed con the keyboard is dis-concerting.

In opinion, new users (e.g. in schools) would have adopted the Tamil99 keyboard had sufficient awareness, training and support been available. However, although physical Tamil keyboards were sent to schools, most teachers and students were unaware how to use them, and thus did not do so. In view of the above, the ICTA standardized Tamil keyboard based on the Renganathan / Typewriter layout and keying-in sequence.

A team comprising Tamil scholars, IT experts, typists and keyboard operators was formed to develop the layout. The terms of reference of the committee were to:

- be close to the Renganathan / Bamini layout
- be uniform and logical and
- be compatible with the English keyboard.

Over 10 different variations of the typewriter layout were studied and the keys common to all of them identified. for symbols which varied among the layouts, the positions in Bamini and Helawadana

were preferred, as these were the most popular layouts. The common punctuation keys such as 'comma', 'full stop' and 'question mark' as well as the numbers and symbols in the first row were placed on the same keys as in the English keyboard.

Thereafter uniformity was maintained as far as possible. All 18 Tamil consonants were placed on unshifted keys, and all Grantha letters on shifted keys. Although the typewriter keyboard has separate keys (Figure 2) for each consonant in conjunction with the *u* and *uu* modifiers (due to variation in their shape), in the new layout, they are produced by typing the consonant followed by a common key and the correct glyph is produced by the font. The standard layout is shown in Figure 3.

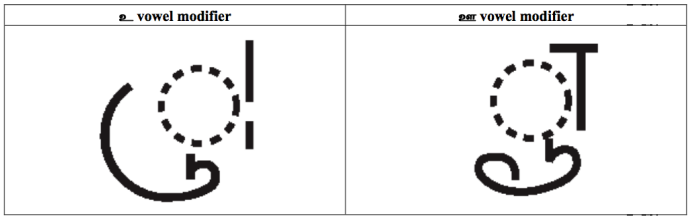


Figure 2 – Shapes for u and uu modifiers

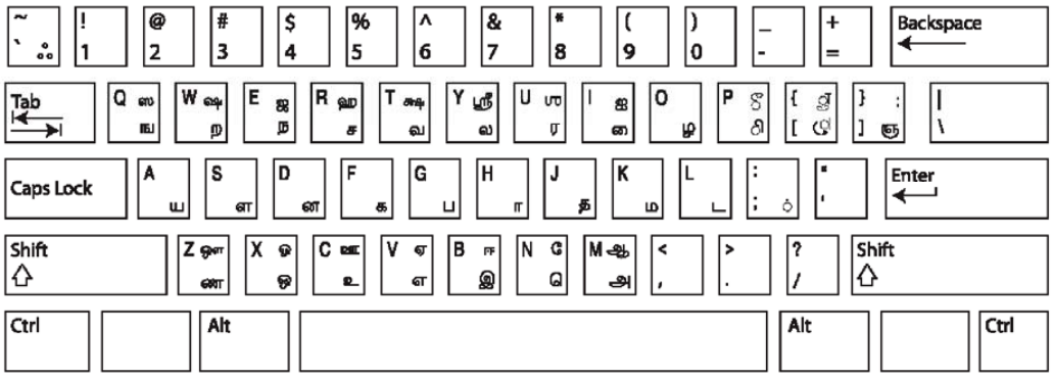


Figure 3 - The Sri Lanka Tamil Keyboard Layout (2008)

Collation Sequence

Tamil collation has similarities with other Indic languages, which follows the Sanskrit collation order. During 16th to early 20th century most of the Tamil dictionaries followed the Sanskrit collation sequence which includes the Grantha letters (especially ஜ - JA) in between the Tamil letters. However the majority of the dictionaries and scholars follows the unique collation sequence which is collating the Grantha letters after all Tamil letters and which been mostly accepted as the standard among Tamil community.

A number of different collation sequences have been used by different authors. The ICTA appointed Mr. G. Balachandran to study these sequences, and to recommend a standard for use in Sri Lanka [6]. The recommendation was to use the following collation order for Tamil: the vowels first, then the Tamil consonants, followed by the Grantha consonants, and then the ஃப (fa) sequence (Figure 4). The symbols and Tamil numerals will come last in the collation order. This recommendation was accepted by the LLWG and the SLSI, and published as the Sri Lanka Standard Tamil Collation Sequence, SLS1326:2008 Part 1 [4].

Figure 4 – Collation Sequence.

அ, ஆ, இ, ஈ, உ, ஊ எ, ஏ, ஐ, ஒ, ஓ, ஔ

∴

க், க, கா, கி, கீ, கு. கூ. கெ. கே, கை, கொ, கோ, கௌ
ங், ங, ஙா, ஙி, ஙீ, ஙு, ஙூ, ஙெ, ஙே, ஙை, ஙொ, ஙோ, ஙௌ
ச், ச, சா, சி, சீ, சு, சூ, செ, சே, சை, சொ, சோ, சௌ
ஞ், ஞ, ஞா, ஞி, ஞீ, ஞு, ஞூ, ஞெ, ஞே, ஞை, ஞொ, ஞோ, ஞௌ

ன், ன, னா, னி, னீ, னு ,னூ, னெ, னே, னை, னொ, னோ, னௌ

ஜ், ஜ, ஜா, ஜி, ஜீ, ஜு, ஜூ, ஜெ, ஜே, ஜை, ஜொ, ஜோ, ஜௌ
ஶ், ஶ, ஶா, ஶி, ஶீ, ஶு, ஶூ, ஶெ, ஶே, ஶை, ஶொ, ஶோ, ஶௌ
ஷ், ஷ, ஷா, ஷி, ஷீ, ஷு, ஷூ, ஷெ, ஷே, ஷை, ஷொ, ஷோ, ஷௌ
ஸ், ஸ, ஸா, ஸி, ஸீ, ஸு, ஸூ, ஸெ, ஸே, ஸை, ஸொ, ஸோ, ஸௌ
ஹ், ஹ, ஹா, ஹி, ஹீ, ஹு, ஹூ, ஹெ, ஹே, ஹை, ஹொ, ஹோ, ஹௌ
க்ஷ், க்ஷ, க்ஷா, க்ஷி, க்ஷீ, க்ஷு, க்ஷூ, க்ஷெ, க்ஷே, க்ஷை, க்ஷொ, க்ஷோ, க்ஷௌ

∴ப், ∴ப, ∴பா, ∴பி, ∴பீ, ∴பு, ∴பூ, ∴பெ, ∴பே, ∴பை, ∴பொ, ∴போ, ∴பௌ

ஔ, ஒ, ஓ, ஒ பூ, ஒ, ஒ, ஒ, ஒ

0, க, உ, ங, ச, ஞ, கூ, எ, அ, க, ஔ, ா, க

Acknowledgements: The work of the ICTA, SLSI, and the members of the various committees which produced the above standards is gratefully acknowledged. The assistance of the UCSC and University of Moratuwa, and especially Prof. Gihan V. Dias, Aruni Goonetilleke and Anura Tissera was instrumental in this work.

References

1 Visagaperumal Vasanthan, "One Hundred Tamils of the 20th Century" on Tamilnation Website. [Online]. Available: http://www.tamilnation.org/hundredtamils/muttiah.htm, 2007.
2 Dept. of IT, Govt. of India, Inscript Keyboard, [Online]. Available: http://tdil.mit.gov.in/keyoverlay.htm
3 C.S. Senthilnathan, (), "New Tamil Font Encoding & Keyboard Standards" on Tamilnation Website [Online], 2007. Available: http://www.tamilnation.org/digital/senthilnathan.htm
4 Sri Lanka Standards Institute, Sri Lanka Standard SLS 1326: Part 1: 2008 - Tamil Character Code for Information Interchange Part 1 – Collation Sequence, SLSI, 2008.
5 Gihan Dias and Aruni Goonetilleke, "Recent Developments in Local Language Computing in Sri Lanka", 27th National Information Technology Conference, 9th – 10th september 2009, Colombo, Sri Lanka
6 Renganathan, Vasu. on thetamillanguage.com Website. [Online]. Available: http://www.thetamillanguage.com/tamil_typewriter.zip, 2004.

Natural Language Processing:

OCR, Text to Speech,

Machine Translation etc.

An Intelligent System for Picture Based Tamil Sentences Generation

Dr.T.Mala, Dr.T.V.Geetha
Department of Computer Science and Engineering
College of Engineering, Guindy, Anna University, Chennai

Abstract: The existing picture dictionaries available electronically are static in nature. The user selects a picture and the contents related to the picture will be retrieved from the database and will be displayed on the screen. The drawbacks of the existing system are many. To mention a few, it is not intelligent, category of the picture is pre-defined, it is not dynamic, new sentences based on pictures cannot be generated and the semantic relationship between the pictures is not maintained. The proposed system aims at developing an intelligent system to generate the Tamil sentences automatically. Domain related pictures are provided. The user has to select more than one picture, based on the selection of the pictures; the semantic relationship between the pictures will be extracted from the semi-automatic domain ontology. Picture words and the semantically related words are sent to the sentence structure framework to obtain the syntactic representation of the Tamil sentence. Then the suffixes are added to the words to generate syntactically, semantically and morphologically correct sentences in Tamil.

Introduction

In the proposed picture based sentence generation system, computer programs are made to produce high-quality natural language text from computer internal representations of information. The system generates automatically a meaningful, grammatical and well formed sentence(s) about the set of pictures on which the user points out. The sentence(s) are generated in Tamil. The main part of the entire system is the domain ontology construction. The corpus is given as the input for the automatic ontology construction subsystem. Automatic domain ontology construction involves two modules. They are ontological word selection and semantic relationship identification between the ontological words. Using the terms extracted from the above modules and the sentence structure ontology the sentence structure is identified and finally suffixes are added to the words to generate syntactically, semantically and morphologically correct Tamil sentences. The entire paper is organized as follows. Section 2 discusses the literature survey in the areas of ontology construction and natural language generation, section 3 gives the overall system architecture, section 4 talks on automatic ontology construction, section 5 gives details related to sentence generation and section 6 gives the conclusion and future works.

Literature Survey

Ontology construction itself is a big challenge. An extension of hierarchical-valued concept context is adopted to elicit concepts from original component descriptions to construct ontological hierarchy for functional concepts [1]. Another method constructs ontology semi automatically by delimiting

documents for learning in the domain of pharmacy involving four steps [2]. Ontology can be constructed automatically from a set of text documents. If documents are similar to each other in content they will be associated with the same concept in ontology. Semantic relationship for ontology construction can be identified from the keywords which are extracted by the clumping properties of content – bearing words [3].

Natural language generation systems are to be studied for automatically generating the sentences. The more "principled" approaches to NLG often make use of a division of the problem into stages such as content determination, sentence planning and surface realization [4][5]. Some of the issues in automatic sentence generation for picture dictionary creation are given below:

- Automatic construction of ontology without any manual intervention.
- What words and syntactic constructions will be used for describing the content?
- How is it all combined into a sentence that is syntactically and morphologically correct?

The proposed system is constructed by tackling the above issues and the architecture of the overall system constructed is explained in next section.

System Architecture

The figure 3.1 represents the entire architecture of an intelligent system for picture based Tamil sentences generation.

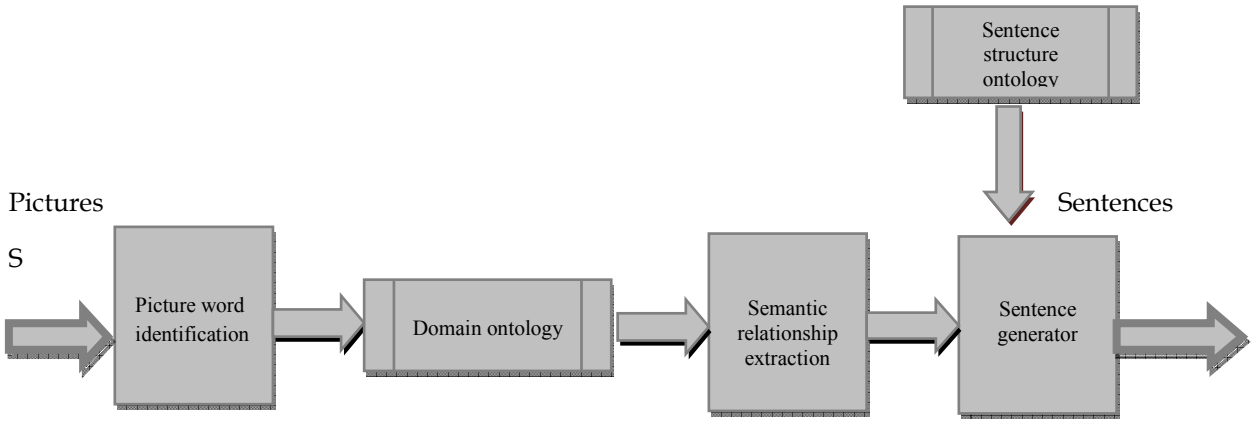


Figure 3.1 Overall system design

Users are intended to select pictures. Based on the selection of the pictures the corresponding picture words are identified. Since the preferred domain is animal domain, the input is restricted to the same. Picture words are searched in the animal domain ontology, which is already constructed semi-automatically as an offline process. The path is traversed from the bottom of the animal domain ontology to the top of it. The words thus extracted from the animal domain ontology are sent to the sentence structure framework to obtain the syntactic representation of the sentence. Finally suffixes are added to the words to generate morphologically correct sentences.

Ontology Construction

Tamil corpus forms the input to the system. Each and every Tamil text document is word segmented and morphologically analyzed to find out the parts of speech, by a tool called Atcharam (Tamil

morphological analyzer). Then, the nouns are extracted and the noun list is confined by TF-IDF (Term frequency-Inverse document frequency) technique. Similarly semantically related words are extracted by the probabilistic framework and thus content bearing words are identified, from which the verbs are extracted to act as the semantically related terms. Using ontological node term and semantically related term, a domain ontology is constructed semi-automatically for Tamil text documents.

Semantic Relationship Extraction

Content bearing terms are identified using probabilistic framework from the text document corpus. Knowing the tendency of important terms to cluster serially, could be useful for extracting the semantic relationship of noun terms. Three measures are used to calculate the content bearing strength of a term and are given as term condensation over textual units, term distribution over textual units and linear clustering. The tamil text documents are given as the input, to tag each and every ontological node term and semantically related term with their corresponding start and end tag. From the tagged tamil text documents, the connection between two ontological node with its semantically related term is identified.

Automatic Sentence Generation

To automatically generate the sentence the sentence structure frame work is to be defined first and then suffixes are added to the terms to generate syntactically and semantically correct sentences. Therefore the next stage is to construct the sentence structure frame work. There is no standard sentence structure for tamil. The following grammar rules were framed and based on the rule sentence structures are obtained [6].

1. NC --- > adj N / N / ADJC N / NNC
2. VC --- > adv V/ adv rpl / ADVC V / vpl / V
3. NNC --- > S con
4. ADJC --- > NC VC
5. ADVC --- > (NC)* vpl
6. S --- > (NC)* (VC)

Addition of suffixes was done using if – then rules. The next section gives the performance evaluation of the developed system and conclusion.

Evaluation And Conclusion

Extraction of ontological terms and semantic relationships are evaluated using that of experts and both gave an accuracy of about 80%. The accuracy of the generation of Tamil sentences is calculated by simple string accuracy formula which is defined in the below equation.

Simple String Accuracy = (1 – (I+D+S)/R)

where, I is the number of insertions, D is the number of deletions, S is the number of substitutions and R is the total number of tokens in the target string. It is inferred that as the total number of tokens gets increased, the accuracy gradually decreases and then gets increased. It proves that the system is efficient, since it can handle any number of tokens without affecting the accuracy. Thus the system generates syntactically, semantically and morphologically correct sentences. It also proves that the system is efficient, by comparing the results of the system with the results of an expert and also by

calculating string accuracy. In future, the project work can be extended to access the picture directly without using any picture word identification. This picture input can be chosen from the pictures provided in the user interface or the input can be given directly by the user. Also, a speech engine can be incorporated to convert the text to speech. Additionally, the user interface can be modified in such a way that it is also applicable for the physically challenged users.

References

Xin Peng, Wenyun Zhao, "*An Incremental and FCA-based Ontology Construction Method for Semantics-based Component Retrieval*", IEEE Seventh International Conference on Quality Software (QSIC), 2007.

1. Mu-hee song, Soo yeon lim, Ki-jun son and sang joo lee , " *Domain ontology construction based on semantic relation information of terminology*", IEEE industrial electronics society, November 2-6 2004.
2. Bookstein. A, Klein S.T, and Raita.T, "*Clumping properties of content bearing words*", IEEE proceedings of the sixth international conference on machine learning and cybernetics Hong kong, 19-22 August 2007
3. Auxilio Medina, Alberto Chavez-Aragon, "*Construction, Implementation and Maintenance of Ontologies of Records*", Proceedings of the Fourth Latin American Web Congress (LA-WEB'06), 2006.
4. Ehud Reiter, "*Building Natural Language Generation Systems*", 1996
5. Saravanan, K., Ranjani parthasarathi, Geetha.T.V., "*Syantactic Parser for tamil*", Tamil internet, 2003.

A HMM Based Online Tamil Word Recognizer

Rituraj Kunwar, Shashi Kiran, Suresh Sundaram, A G Ramakrishnan
Medical Intelligence and Language Engineering Laboratory
Indian Institute of Science, Bangalore -560012, India.

Introduction

In Online handwriting recognition, a machine recognizes, as a user writes on a pressure sensitive screen with a stylus. The stylus captures information about the position of the pen tip as a sequence of points in time. The sequence of point between a PEN DOWN and PEN UP signal defines a stroke. This spatio-temporal information of the character being traced is the only input available to the online recognition system. Also given a character, one can capture the different writing styles using the information from the stylus. Tamil is a popular South Indian language for a significant population in countries such as Singapore, Malaysia and Sri Lanka besides India. There are 313 characters in Tamil alphabet. Of these, there are 12 pure vowels and 23 pure consonants (including the 5 consonants derived from Sanskrit (Grantham consonants)), the remaining being consonant vowel combinations, wherein the vowels modify the consonants and 2 special characters ஃ and ஸ்ரீ. It has been found that to represent all the possible 313 characters, a set of 155 symbols is sufficient. This paper deals with the problem of recognizing online handwritten Tamil words with a Hidden Markov Model framework. Given a Tamil word, we first run a segmentation algorithm to identify the individual symbols. A Tamil symbol may be written with different number of strokes. The extracted symbols are subjected to the following preprocessing modules: smoothing to remove noise, resampling to a fixed number of points for speed normalization and size normalization. A set of seven features are derived at each sample point of the preprocessed Tamil symbol. These features are then fed to the Hidden Markov Model classifier for recognition of the Tamil symbol. Based on Unicode generation rules derived from the language, stroke groups are generated from the Tamil symbols. A Tamil word corresponds to a set of stroke groups. Fig 1 gives a snapshot of a handwritten Tamil word கவசம், collected with a TABLET PC, together with the recognized output using the Hidden Markov Models as the classifier.

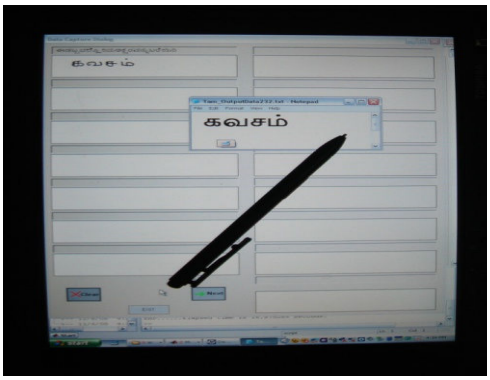


Fig 1. *Data Collection Device with recognized output.*

Segmentation

Word level data is to be segmented to character level as the modeling of the data is done at the character level. Then the recognition is performed at the character level and the results are concatenated to form the words. Segmentation of the Tamil words into characters is simple when

compared to the English cursive handwriting. In Tamil script, two strokes of the same character either overlap or touch at some point. The vowel or consonant modifiers are written as separate symbols and their models are built separately.

For each current stroke, the next stroke is taken and checked whether the x co-ordinate of its starting point is less than the maximum x coordinate of the previous (i.e. current) stroke. Saying in other words, we see if the next stroke overlaps with the current stroke. If there is an overlap (within a threshold empirically set) then the successive stroke is concatenated with the current stroke to form the same symbol (Fig. 2). Other wise the future stroke is considered as the current stroke / new symbol and the same procedure is repeated. However, there are cases wherein the last part of the trace of the successive stroke overlaps with the current stroke, as shown in Fig 3. In such scenarios, one needs to treat these strokes as 2 separate entities. Hence we formulate our condition as follows: If the x co-ordinate of the last point of the current stroke is less than the x max of the previous stroke, we do not concatenate these strokes.

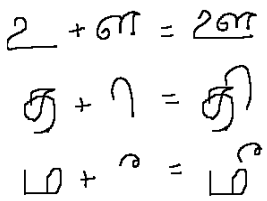


Fig 2. *Simple overlap of the first part of the successive stroke with the current stroke*

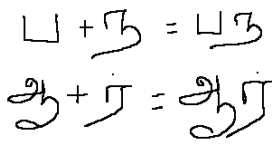


Fig. 3 *Simple overlap of the last part of the successive stroke with the current stroke*

Preprocessing

The raw character, captured from the device, comprising of N points sampled uniformly in time, $\{x_i^{raw}, y_i^{raw}\}_{i=1}^{N}$ is first smoothed using a Gaussian mask, that is applied to the x and y coordinates independently.

$$x_t^{smooth} = \sum_{i=-3\sigma}^{3\sigma} w_i x_{t+i}^{raw}$$

$$y_t^{smooth} = \sum_{i=-3\sigma}^{3\sigma} w_i y_{t+i}^{raw}$$

The weights w_i are given by

$$w_i = \frac{e^{\frac{-i^2}{2\sigma^2}}}{\sum_{i=-3\sigma}^{3\sigma} e^{\frac{-i^2}{2\sigma^2}}}$$

We then normalize each character to a standard size using the transformation.

$$x_i = \frac{x_i^{smooth} - x_{\min}}{x_{\max} - x_{\min}} \qquad y_i = \frac{y_i^{smooth} - y_{\min}}{y_{\max} - y_{\min}}$$

Here x_{min} and x_{max} denote the minimum and maximum x coordinate of the raw character. y_{min} and y_{max} represent the minimum and maximum y coordinate. The characters are r-sampled to fixed number of points (in our work 60) uniformly in space. First, we find the length of each character and the length of all the strokes constituting it. We then assign different number of points to each stroke, depending on the ratio of the length of the stroke to that of the character. The first and last points of the character are retained and the points in between are found which are spaced uniformly. Fig 4(a) (b) depict a snapshot of the raw and preprocessed character ಹ.

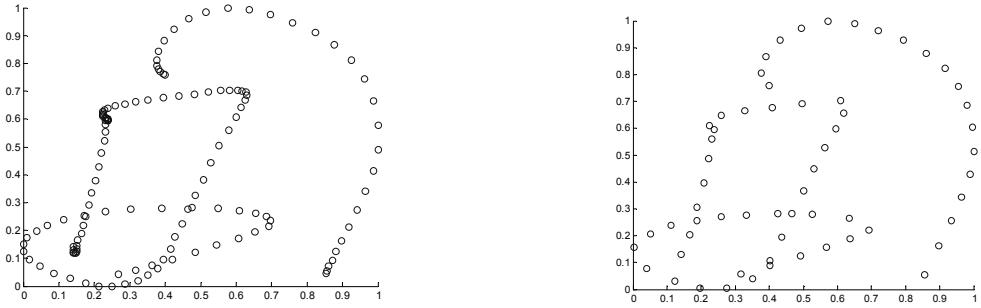


Fig 4 (a): *Raw Character* ***4 (b):*** *Pre processed Character*

Feature Extraction

Each of the segmented characters are subjected to a feature extraction module. The following features are derived at each sample point of the characters.

Normalized x and y coordinates: The x and y coordinates of the normalized sample are taken as two features for the recognition.

Normalized x and y first derivative: The variations in the x and y coordinates are found independent of each other which give the shape features locally at each point. At the current point (x_j, y_j), a window is taken covering the past and the future points and the derivative is calculated using the formulae given below

$$x_j^{'} = \frac{\sum_{i=1}^{2} i(x_{j+i} - x_{j-i})}{2\sum_{i=1}^{2} i^2} \qquad y_j^{'} = \frac{\sum_{i=1}^{2} i(y_{j+i} - y_{j-i})}{2\sum_{i=1}^{2} i^2}$$

The normalized first derivatives in x and y direction at (x_j, y_j) is given by

$$x_{norm}^{'} = \frac{x_j^{'}}{\sqrt{x_j^{'2} + y_j^{'2}}} \qquad y_{norm}^{'} = \frac{y_j^{'}}{\sqrt{x_j^{'2} + y_j^{'2}}}$$

Normalized x and y second derivatives: The second derivatives are computed similar to the first derivatives.

$$y_j^{''} = \frac{\sum_{i=1}^{2} i(y_{j+i}^{'} - y_{j-i}^{'})}{2\sum_{i=1}^{2} i^2}$$

$$x''_j = \frac{\sum_{i=1}^{2} i(x'_{j+i} - x'_{j-i})}{2\sum_{i=1}^{2} i^2}$$

The normalized second derivatives in x and y direction at (x_j, y_j) is given by

$$x''_{norm} = \frac{x''_j}{\sqrt{x''^2_j + y''^2_j}} \qquad y''_{norm} = \frac{y''_j}{\sqrt{x''^2_j + y''^2_j}}$$

Curvature

Curvature at a point on a curve is the inverse of the radius of the osculating circle at that point and can be found using the first and second derivatives as given below.

$$C = \frac{x''_{norm} y'_{norm} - y''_{norm} x'_{norm}}{(x'^2_{norm} + y'^2_{norm})^{3/2}}$$

Hidden Markov Models

The derived features are then fed to the Hidden Markov Model classifier for recognition. More details on the description of HMMs can be found in [1] [2]. Each Tamil symbol is modeled using a separate HMM .Training of the models is performed using the well known Baum Welch Estimation. The Bayesian approach is adopted for recognizing the label for the test symbol. The IWFHR Database has been used for our experiments. This database contains around 345 samples (written on a Tablet PC) for each of the 155 symbols. The resulting system gives an accuracy of 84% at the symbol level.

Acknowledgements: The authors are grateful to Technology Development for Indian Languages (TDIL), Department of Information Technology (DIT), MCIT, Government of India for funding this project and to AVM Matriculation Higher Secondary School, Chennai for contributing their students' time to collect training data for our project.

References

1. L. R. Rabiner and B. H. Juang, ``An introduction to hidden Markov models,'' *IEEE ASSP Mag.*, pp 4--16, Jun. 1986.
2. Duda, Hart, Stork, "Pattern Classification", Second Edition.

Problems related to Eng-Tam Translation

M.B.A.Salai Aaviyamma
Assistant Professor/ CSE, SRM university, Chennai

Dr.K.Kathiravan
Professor and VicePrincipal, Easwari Engg.College, Chennai

Introduction

Countries like India, where many official languages are used, and if there is a commonly used language, like English, then translating from that common language to regional language solves many purposes. Tamil, a South Indian language, not only used in Thamil nadu, a State of India, but is also used as one of the official languages in many countries like Singapore, Malaysia and Sri Lanka.

Stages

There are 6 modules in this project like Tokenization, Parsing, Mapping, Word formation, Sentence structure changing, and Display. The English sentences(to be translated) are separated into words (Tokenization), each word is recognized as affiliations of root words (parsing) i.e. root words and morphemes are identified, root words are mapped into equivalent Tamil words using dictionary (Mapping,). Then according to the parser output, Thamizh words are formed (Word formation), and then structure of sentence is formed according to Thamizh sentence making rules (Sentence structure changing) and the output statement is displayed in Thamizh. **http://wwwlingsoft.fi/cgi-bin/engcg** is used for extracting the parser outputs. Mainly concentrated module of the project is the word formation .

Problems found in the following areas:

1. when tense markers are added
2. when case markers are added
3. when proper nouns are translated

Problems in forming Verb Phrases

Traditionally, a Tamil word is divided into a maximum of six parts, known as pakuthy (prime-stem), sandhi (junction), .viha:ram (variation), idainilai (middle part), sa:riyai (enunciater) and vikuthy (terminator) in that order.

(prime-stem)	(junction)	(Tense marker) middle part	(enunciater)	(terminator)
Pagudhi	Sandhi	Idainilai	Sariyai	Vigudhi

The sixth part is Vigaaram. This is the trouble making part of translation.

For example, if the root word is a finite verb, then changes of penultinating characters are of many type.

First case

Changes can be introduced, when root words are joined with tense markers or additional suffixes. For many root words, verb formed are different.

For example, "HE ACTED",

"nadi" + th + th + aan --- nadiththaan

நடி + த் + த் + ஆன் -- நடித்தான்

" He walked"

"nada" + th(ndth) + th + aan --- nadandthaan

நட + த்(ந்) + த் + ஆன் -- நடந்தான்

To overcome this problem, Dr. Crowl[2] and M.Raagava iyangar[3], in their books

(Thamizhp peragaraathi and Vinaith thiribu Vilakkam), they divided the entire Thamizh verb family into 12 groups as

செய், ஆள், கொல், அறி, அஞ்சு, நகு, உண், தின், கேள், பார், நட

according to the last characters of root words, the tense markers they accept etc. Even though they are grouped, some root words, having same last character but grouped under different tables, makes the rule based translation, a problem.

For example,

"cel" செல் - is grouped under "kol" table, group 3

"kal" கல் - is grouped under "kal" table, group 10

Even though both are having the same last character "l" ல், but when added with past tense markers, they are turned as

"cendraan" - சென்றான் and

"kattrraan" - கற்றான்

Second case

The same root word is kept under 2 or more table and it takes different tense markers under each circumstances.

For example,

"migu" - மிகு

under "ari" table, it is transformed as " migundthaan" மிகுந்தான்

under "nagu" table, it is transformed as "mikkaan". மிக்கான்

So, to inform the system, under which table, that root word is classified is so difficult.

Hence , it is a problem to form the verb phrase for these root words.

Third case

The same root word, even though kept under a same table, because of its different meaning, it is transformed differently, which creates problem to decide how to transform it. For example,

"madi" - மடி under "ari" table,

when having meaning as 'die', transformed as "madindhaan"- மடிந்தான்

when having meaning as 'fold' , transformed as "madiththaan"- மடித்தான்

For this problem, a solution is obtained as, for verbs denoting self deeds, it will be transformed with "thth". And for denoting other's deeds, it will be "ndhth"

But there is a problem on this solution, for some roots .

For example,

"vadi" - வடி

(kanneer) vadiththaan – கண்ணீர் வடித்தான்

(azudhu) vadindhaan - அழுது வடிந்தான்

Both are denoting self deeds but are transformed in both the ways.

But, as a different case,

root word "pidi" - பிடி is transformed as

("pidiththaan") - பிடித்தான், in both the cases.

Problems with case markers

Eight cases are there. They are Nominative, Accusative, Dative, Benefactive, Instrumental, Sociative, locative and Ablative.

Some prepositions are marked as case markers. But, these case markers give different meanings in different places. So, translation becomes difficult.

For example,

(i) " HE ATE WITH THE SPOON"

Here , 'with' comes as instrumental case marker.

(2) "HE ATE WITH HIS FRIEND"

Here, the same preposition 'with' comes as Sociative case marker.

More than this, in the parser output[5], for both the sentences, the subject and object are no where mentioned as either instrumental or sociative case markers. But mentioned as nominative in both the cases . And the word 'with' is simply mentioned as preposition .

Problems with Gender suffixes and nouns:

In English, generally, the names of male persons , are spelled excluding the last letter. For example,

'Rama went to Srilanka', instead of 'Raman went to Srilanka'**.**

Now, there is a problem, whether to translate this as

இராமன் இலங்கைக்கு சென்றான்?

(or)

இராமா இலங்கைக்கு சென்றாள்?

he ate with the spoon.

"<he>"
"he" <NonMod> PRON PERS MASC **NOM** SG3 SUBJ @SUBJ
"<ate>"
"eat" <SVO> <SV> V PAST VFIN @+FMAINV
"<with>"
"with" **PREP** @ADVL
"<the>"
"the" <Def> DET CENTRAL ART SG/PL @DN>
"<spoon>"
"spoon" N **NOM** SG @<P
"<$.>"

he ate with his friend.

"<he>"
"he" <NonMod> PRON PERS MASC NOM SG3 SUBJ @SUBJ
"<ate>"
"eat" <SVO> <SV> V PAST VFIN @+FMAINV
"<with>"
"with" **PREP** @ADVL
"<his>"
"he" PRON PERS MASC GEN SG3 @GN>
"<friend>"
"friend" N **NOM** SG @<P
"<$.>"

Therefore, by rule based method alone, translation from English to Tamil cannot be done. So, we have to train the system accordingly and then only we can translate the sentences.

Conclusion

In this paper, the problems created by case markers and problems created when forming the Thamizh words by adding tense, gender, plural suffixes with root words are dealt in detail. And it is more important to solve these problems, since translation from 'English to Thamizh' is a very important and timely needed task for Tamilnadu state government, and countries like Singapore, Malaysia and Sri Lanka, where Tamil is accepted as one of the official language, in order to improve the communication and education.

References

1. A grammar of modern Tamil - Thomas Lehmann (Pondicherry University)
2. தமிழ்ப் பேரகராதி - Dr. Crowl
3. வினைத் திரபு விளக்கம் - M. Raagava Iyangar,1958
4. Hidden problems and challenges in Tamil computing
   i. -S.Srinivasan, A James (Tamil Virtual University)
   ii. -S.AnanthaKrishnan, K.R.S.Narayanan (IARC Kalpakkam)
5. **http://wwwlingsoft.fi/cgi-bin/engcg**

Tamil-English Cross Lingual Information Retrieval System for Agriculture Society

D. Thenmozhi and C. Aravindan
Department of Computer Science & Engineering
SSN College of Engineering, Chennai, India
{theni_d, aravindanc}@ssn.edu.in

Abstract: Cross Lingual Information Retrieval (CLIR) system helps the users to pose the query in one language and retrieve the documents in another language. We developed a CLIR system in Agriculture domain for the Farmers of Tamil Nadu which helps them to specify their information need in Tamil and to retrieve the documents in English. In this paper, we address the issue of translating the given query in Tamil to English using Machine Translation approach. It uses a Morphological Analyzer to obtain the root terms of source query. We developed language resources like Bi-lingual Dictionary and Named Entity Recognizer using which the query is translated to English. Local word reordering is performed according to Subject-Verb-Object pattern in order to preserve the relative dependency among the words. Word sense disambiguation is done that identifies the correct sense of an ambiguous word that is being used in a query. The system exhibits a dynamic learning approach wherein any new word that is encountered in the translation process could be updated to the bilingual dictionary. The translated query is given to an existing search engine like Alta Vista, Google, etc. This Machine Translation approach retrieves the pages with Mean Average Precision of 95%. The recall value is also considerably improved.

Introduction

The World Wide Web (WWW), a rich source of information is growing at an enormous rate. According to Online Computer Library Center, English is still the dominant language in the web that contributes most of the content [10]. However, global internet usage statistics reveal that the number of non-English internet users is steadily on the rise, but all of them are not able to express their basic needs in English. Tamil users who are not able to express their needs in English are also growing in the Internet. They generally search for the information using the Tamil search engines. But the content provided by these search engines is less in number [13]. Making the huge repository of information on the web, which is available in English, accessible to non-English internet users has become an important challenge in recent times. When the non-English users want to access the existing search engines, most of the time they arrive at improper formulation of English queries.

Cross-Lingual Information Retrieval (CLIR) systems aim to solve the above problem by allowing the users to pose the query in their own (source) language which is different from the language of the documents that are searched. This enables users to express their information need in their native language while the CLIR system takes care of matching it appropriately with the relevant documents in the target language.

CLIR focuses on the cross-language issues from the Information Retrieval perspective rather than Machine Translation perspective [12]. The basic idea in Machine Translation (MT) is to replace each term in the query with an appropriate term or a set of terms from the lexicon syntactically. If the query is translated based on MT approach, the search will give better result. For example, a Tamil query "*udal nalaththirrku ettra payirkal*" translated to English query "*body health suitable for crops*" in a word by word approach will give an average result. Whereas the machine translation approach translates the query to "*crops suitable for body health*" which gives better result.

We propose a CLIR system using Machine Translation approach in Agricultural domain for Tamil Farmers. The system retrieves relevant documents from an English corpus in response to a query expressed in Tamil language. Here, the query given in Tamil language is translated syntactically and semantically to English (not word by word translation/transliteration) for Information Retrieval process.

Section 2 briefly describes the various works done related to Cross Lingual Information Retrieval systems. Section 3 explain the various phases that are involved in translating the given Tamil query to English using MT approach in Agriculture domain. Section 4 elaborates the various experiments conducted to analyze the performance namely (i) comparison of word by word translation with machine translation, (ii)comparison of Tamil Search Engine with CLIR system and (iii)comparison of irrelevant query formed by non-English users with query translated by CLIR system.

Literature Survey

Cross Lingual Information Retrieval Systems for Indian Languages

Several organizations in India are working on the CLIR system for Indian Languages [13]. Jadavpur University has developed a Bengali, Hindi and Telugu to English CLIR system as part of the ad-hoc bilingual task [15]. IIT Bombay has developed Hindi-English and Marathi-English CLIR systems [10]. IIIT, Hyderabad has developed a Hindi and Telugu to English CLIR system [12]. IIT kharagpur has developed a CLIR system for two most widely spoken Indian languages, Hindi and Bengali [4]. All these works uses bilingual dictionaries. Microsoft Research India has also work on Hindi to English cross-lingual System [8] in which they used a word alignment table that was learnt by a Statistical Machine Translation (SMT) system trained on aligned parallel sentences. These organizations have experimented their results on English corpus of LA Times 2002. AU-KBC had developed Tamil-English Cross Lingual Information Retrieval Track [11] for news articles taken from "The Telegraph", English news magazine in India. All these organizations have developed their CLIR systems using word by word translation approach in news domain.

Machine Translation Systems

Statistical machine translation (SMT) is an approach to MT that is characterized by the use of machine learning methods. A complete survey and methodologies to build SMT systems is found in [1]. Tamil-English [5] and English-Tamil [2] statistical machine translation system are developed by constructing parallel corpus.

Applications of Agriculture

Many researches are working on the development of applications in the domain of Agriculture. Food and Agricultural Organization of United Nation has developed an Agricultural Ontology – AGROVAC [7] that provides different concepts and their relations for agricultural domain in different

languages of European and Asian languages including Hindi. Knowledge elicitation methods for multiple experts in domain of Agriculture are developed as part of the expert system [3].

We have developed a Cross Lingual Information Retrieval System for Tamil language using MT approach in Agriculture domain.

System Architecture

The proposed CLIR system uses a number of phases to translate the given Tamil query in Agriculture domain to an English query using MT approach. This is illustrated in the figure 3.1.

Morphological Analysis

Morphological Analyzer accepts the input query string and performs a database lookup operation to check whether the given query is directly present in the bilingual dictionary. If present, the translated query is returned. Otherwise split the query into the individual constituent words. By applying morphological rules for handling plurals, case suffices, oblique, etc., the root words are obtained.

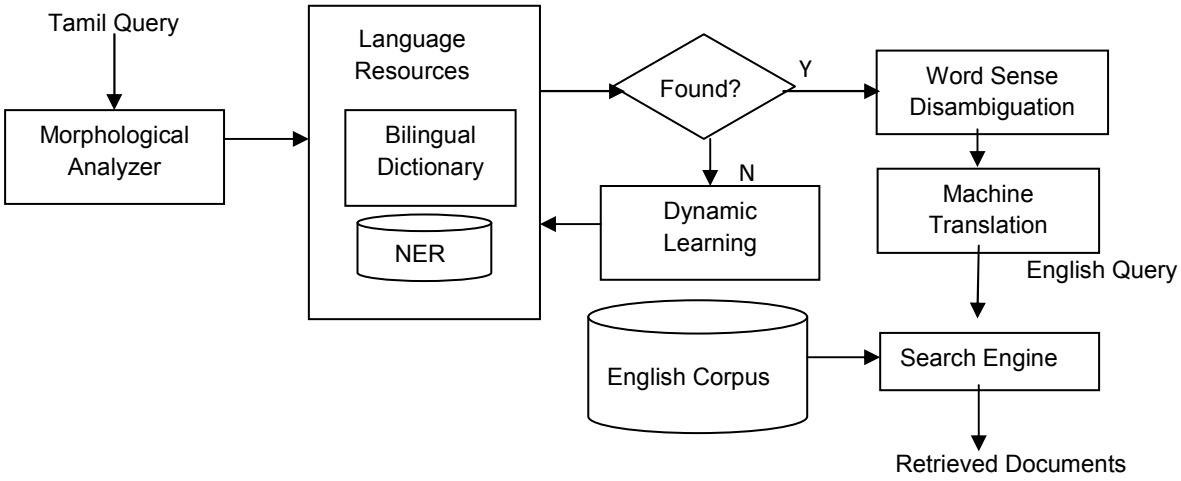


Figure 3.1. System Architecture

Dictionary Lookup

We have developed a Tamil-English bilingual dictionary of size 5.08MB that contains most the words related to agricultural domain. The dictionary had to be built from the scratch as no resource is available for this domain. After each intermediary step in the Morphological Analyzer, the extracted word is mapped with the bilingual dictionary to check whether it is a root word. If it is available, meaning of the word is returned. If not, the word is then passed on to the subsequent stages in the Morphological Analyzer. At the final stage of the Morphological Analyzer, if the word returned is a root word that is available in the bilingual dictionary, then its meaning in the target language is returned. Otherwise the word is processed so as to bring it to a form that is available in the dictionary and relevant to the context. For example, for the given word "*veelaan*" meaning agriculture, the root word available in the dictionary is "*veelaanmai*". The closest match for "*veelaan*" is identified as "*veelaanmai*" and the meaning is returned.

The system exhibits a dynamic learning approach wherein any new word that is encountered in the translation process could be updated to the bilingual dictionary by allowing the user dynamically to insert it into the dictionary along with its corresponding English meaning.

Machine Translation

Tamil is a subject-object-verb (SOV) language. SOV is the type of language in which the subject, object and verb appear in that order. Subject-verb-object (SVO) is a sentence structure where the subject comes first, the verb second and the object third. English is one such language. Tamil to English translation involves classifying the individual translated words into subject, verb and object and placing them in correct ordering. The individual words are processed and identified as to whether they belong to noun or verb and the classification is performed. The words are then arranged according to the SVO pattern to obtain the translated query in English. In order to perform the translation part of speech (pos) tagging should be done for all the words in the dictionary. A local word reordering is performed based on POS tagging to obtain SVO patter of English query [9].

Word Sense Disambiguation

A complete survey of Word Sense Disambiguation is found in [14]. This phase uses the word-net, a variation of Lesk algorithm [6] to retrieve the possible senses of a word. For each sense of a given word, it is compared with all possible senses of the surrounding words in the given query. The count of number of words common between the sense descriptions is calculated and assigned as the score for the particular sense of the word. The sense that has the highest score is declared the most appropriate one for the target word in the given context. For example, for the query *"aarukalil ulla miin vakaikal"*, the word "*aaru*" is ambiguous having two different meanings, "*The digit six*" and "*River*". The second sense of the word obtains the highest score when compared with the senses of the other words in the query. Thus the correct sense of the word in the given query is "*river*". Hence the query is translated to "*Fish type present in river*".

Experimental Results

We have developed a small GUI using which the users can enter their query in Tamil and are translated to English using the CLIR system. The translated query is given to an existing search engines like Alta Vista, Google, etc and the pages are retrieved in English. Various experiments have been done to compare the performance of the developed system with an existing system.

Precision comparison between Word By Word Translation (WBWT) and MT

To determine the relevance of each retrieved page, a four-point scale was used which enabled us to calculate precision. A page representing full text of research paper, seminar/conference proceedings or a patent is given a score of three and its abstract is given a score of two. A page corresponding to a book or a database is given a score of one. A page representing other than the above (i.e. company web pages, dictionaries, encyclopedia, organization, etc.) is given a score of zero. A page occurring more than once under different URL is assigned a score of zero.

The machine translated queries retrieves documents whose precision is greater than the precision of the documents retrieved using the Word by Word translation Technique which is illustrated in the following table.

Tamil Query	Translated query		Precision(%)	
	WBW trans	Machine Trans	WBWT	MT
Nerppayir saakupatikku ettra urangkal	Paddy crop cultivation for Suitable pesticide	Pesticide suitable for paddy crop cultivation	72	97
Utal nalaththirrku Ettra payirkal	Body health suitable for crops	Crops suitable for body health	69	96
Mann thottarpaana itarpaatukal	Soil related to hurdle	Hurdle related to soil	70	89
Velan thurayil ulla tharpothaya valarssikal	agricultural department present in current development	Current development present in agricultural department	83	91

Performance comparison between a Tamil search engine and the CLIR System

The non-English (Tamil) users who do not know how to give query in English generally use the Tamil Search Engines. We experimented by giving query in Tamil to Webulagam search and observed that the recall value was very less and the precision was also very low due to the lack of content availability with Tamil Search Engines. We obtained a result with good precision and recall when the same query was given to our CLIR system.

Search System	Webulagam Search	CLIR Search
Query	பயிர் பாதுகாப்பு	Crop protection
No. of. documents retrieved	57	1,40,000
Precision(%)	23	97

Search Result and Precision for an Improper query formed by non-English user and Correct query formed using MT

When the non-English(Tamil) users try to formulate their queries in English, most of the time they arrive at improper queries. We have experimented with some improper queries given to an existing search engines and the performance of the search result is low when compared to the query that was translated by the CLIR system.

Search request	Irrelevant query in English	translated to English using CLIR
Munthiri valarkka ettra mann	Cashew grow soil	Soil suitable for cashew growth
No: of documents retreived	14,700	65,700
precision	44	82

Conclusion

The CLIR System helps the Farmers of Tamil Nadu, India to pose their information need in Tamil and to retrieve the documents from a large corpus in English language. The system focuses on the Machine Translation technique rather than the word by word translation and gives better result. The CLIR systems generally display the search result in English. It is appropriate, if the results are displayed in their own language for the users who do not know how to give query in English. This system can be further extended to Rank the pages and provide a summary (in English) of top pages, translate the summary to Tamil or provide an answer to the query in Tamil (like an expert system).

Acknowledgement: We wish to thank C.Karthika and M.Nandhini for their valuable contributions in collecting data related to Agricultural domain, developing the bilingual dictionary and implementing the code for CLIR system. We also thank our management for their continuous motivation and support.

References

1. Adam Lopez, "Statistical Machine Translation", ACM Computing Surveys, Vol. 40, No. 3, 2008.
2. Amrita Vishwa Vidyapeetham, "valluvan-English to Tamil Statistical Machine Translation", Center for Excellence in Computational Engineering and Networking (CEN), 2005.
3. Bertrand Legar, Oliver Naud, "Experimenting Statecharts for Multiple Experts Knowledge Elicitation in Agriculture, An International Journal on Expert Systems with Apllications, 2009.
4. Debasis Mandal, Sandipan Dandapat, Mayank Gupta, Pratyush Banerjee, Sudeshna Sarkar, "Bengali and Hindi to English Cross-language Text Retrieval under Limited Resources", in the working notes of CLEF 2007
5. Fedric C.Gey, "Prospects for Machine Translation of the Tamil Language", in the proceedings of Tamil Internet 2002, California, USA
6. http://en.wikipedia.org/wiki/Lesk_algorithm
7. http://www.fao.org
8. Jagadeesh Jagarlamudi and A Kumaran, "Cross-Lingual Information Retrieval System for Indian Languages", in the working notes of CLEF 2007
9. Maja Popović and Hermann Ney. "POS-based Word Reorderings for Statistical Machine Translation". 5th International Conference on Language Resources and Evaluation (LREC), pages 1278-1283, Genoa, Italy, May 2006
10. Manoj Kumar Chinnakotla, Sagar Ranadive, Pushpak Bhattacharyya and Om P. Damani, "Hindi and Marathi to English Cross Language Information Retrieval at CLEF 2007"
11. Pattabhi R. K. Rao and Sobha L, "AU-KBC FIRE2008 Submission - Cross Lingual Information Retrieval Track: Tamil-English", First Workshop of the Forum for Information Retrieval Evaluation (FIRE), Kolkata. pp 1-5, 2008
12. Prasad Pingali and Vasudeva Varma, "IIIT Hyd at CLEF 2007-Adhoc Indian Language CLIR task"
13. Prasenjit Majumder, Mandar Mitra Swapan parui and Pushpak Bhattacharyya, "Initiative for Indian Language IR Evaluation", Invited paper in EVIA 2007 Online Proceedings.
14. Roberto Navigli, "Word Sense Disambiguation: A Survey", ACM Computing Surveys, Vol. 41, No. 2, Article 10, February 2009.
15. Sivaji Bandyopathyay, Tapabrata Mondel, Sudip Kumar Naskar, Asif Ekbai, Rejwanuj Haque, Sinivasa Rao Godavarthy, "Bengali, Hindi and Telugu to English ad-hoc Bilingual task at CLEF 2007", in the proceedings of Cross Lingual Evaluation Forum(CLEF) in 2007.

Tamil in Mobile Phones and Handhelds

அறிவியல் மற்றும் தற்காலத் தேவைகளுக்கேற்ற தமிழெழுத்துச் சீர்திருத்தமும் அவற்றைக் கணினி மற்றும் கைபேசி செயல்பாட்டிற்குப் பயன்படுத்தும் முறையும்

த .ஞான பாரதி

மத்திய தோல் ஆராய்ச்சி நிலையம், சென்னை

Abstract: Expansion of wisdom and introduction of various science, communication and technology along with arrival of various cultures from different parts of the world has to be accepted and adopted as per the needs for the development of a society. As the Tamil people accepted many such changes, Tamil language also in need of some changes to adopt the new environment. The article emphasizes the need for improvement in the letters of the language for acceptance various science and technology and for its own developments. The new letters introduced are ங் ,ச் ,ட் ,ப் ,த் ,வ் ,ஜ் for an***g****a*m, ***s****a*ngam, pa***d****a*m, aim***b****a*dhu, san***dh****a*m, u***v****(****w****)****a*mai and zigz***a***g, respectively. The Tamil letter)ஃப் (used for sound 'F' is modified for a new and better one as 'ஃ'. The new letters introduced are similar to the existing Tamil letters thus easily be accepted without any difficulties by the people who can read and write Tamil. The article also gives a new keyboard arrangements system for computers and mobile phones for Tamil letters that includes the newly introduced letters. These changes will revolutionize usage of Tamil in its advancement in every field on its acceptance.

அறிமுகம்

உலகில் இயற்கையின் அங்கங்களும் நிகழ்வுகளும் புதிதாக இன்றும் கண்டறியப்படுகின்றன. மேலும் புதிது புதிதாக பல கருவிகளும் செயல்களும் உருவாக்கப்படுகின்றன அல்லது மேம்படுத்தப்படுகின்றன இவற்றை பெயரிட்டு அழைக்கின்றோம். இவ்வழக்கம் பல்வேறு பகுதிகளிலும் நிகழ்வதால் அவை வெவ்வேறு பெயர்களில் அழைக்கப்படுகின்றன. மேலே குறிப்பிட்டவற்றை பிறர் மூலம் அறியும் பொழுது, ஒரு சமுதாயம் அதில் தமக்கு தேவையானதை தம்மொழிக்கேற்ற ஒரு புதுப்பெயரிலோ அல்லது தமக்கு தெரிவித்தவர் உரைக்கும் பெயரிலோ அடையாளப்படுத்தி ஏற்றுக்கொள்கிறது. சமூகங்களுக்குள் ஏற்பட்ட தொடர்புகளினால் வகைப்படுத்துதல் மற்றும் பெயரிடுதலில் வர்த்தகம் மற்றும் அறிவியல் சார்ந்தவை ஒருமுகப்படுத்தப்பட்டுள்ளன. இப்பெயர்கள் பல்வேறு பகுதிகளில் இருந்தும் அறிமுகமாகுதலால், முதலில் உலகிற்கு வெளிபடுத்தியவர் குறிப்பிடும் பெயரே பெரும்பாலும் நிலைக்கிறது.

எனவே பல திக்குகளிலிருந்து வரும் புதியனவற்றை ஏற்கும் வகையில் மொழியின் எழுத்துக்கள் அமைந்திருக்க வேண்டும் அல்லது அமைக்கப்பட வேண்டும். தமிழில் தற்போது உள்ள எழுத்துக்களால் இவ்வாறு ஏற்றுக்கொண்ட சொற்களைத் தெளிவாக எழுத இயலாத நிலை உள்ளது. தமிழ் சொற்களைப் போல பிறசொற்களையும் தவறின்றி எழுத மற்றும் அவற்றை தெளிவாகப் படித்துணரக்கூடிய நிலையில் மொழி இருக்க வேண்டுமாதலால் எழுத்துச் சீர்த்திருத்தம் அவசியமாகிறது.

எழுத்துச் சீர்திருத்தத்தின் அவசியம்

எழுத்தொலியும் இலக்கணமும் இணைந்து தமிழ் சொற்களை உருவாக்குகின்றன. ஒவ்வொரு மொழியிலும் அவற்றிக்கு மட்டுமே உரிய சில ஒலிகளும் இருக்குமென்றாலும் பொதுவான ஒலிகள் அனைத்து மொழிகளிலும் காணப்படுகின்றன. அவ்வாறு தன்னிடம் இல்லாத ஒலிகளையும் ஏற்கக்கூடிய நிலையில் மொழியின் எழுத்துக்கள் அமைந்திருப்பது மொழியின் வளர்ச்சிக்கும் அதனால் அச்சமூகத்தின் மேம்பாட்டிற்கும் அவசியமாகும்.

தமிழ், பல்லாண்டுகளாக தனித்திருந்தது. கடந்த சில நூற்றாண்டுகளாக அயல் நாடுகளின் அறிவியல் மற்றும் தொழில்நுட்பத்தை ஏற்றுக்கொண்டதோடு மட்டுமல்லால் பிற மொழிகளின் தாக்கத்தையும் தமிழ் எதிர்கொள்ள வேண்டியிருந்தது. ஆனால், அவ்வாறு ஏற்றுக்கொண்ட சொற்களில் பல, தமிழின், தமிழரின் ஆளுமையால், இம்மொழியின் இலக்கண மற்றும் வழக்குமுறைக்கேற்ப தமிழாக்கப்பட்டன. எ.கா. லக்.ஷ்மன் – இலக்குவணன், டேவிட் – தாவீது, அலாடின் – அலாவுதீன், ஃப்ரான்ஸ் – பிரான்சு, காஃபி – காப்பி, எஞ்சின் – இயந்திரம். அரை நூற்றாண்டாக பல்வேறு காரணங்களால் தனித்துவத்தை ஆளுமையை தளர்த்தத் தொடங்கியது எ.கா. மனோகரன் – மனோகர். நாளடைவில், ஏற்றுக் கொண்ட சொற்களின் மீது எந்த ஆளுமையையும் செலுத்த தயங்கியது. எ.கா. சங்கரன் – சங்கர் – ஷங்கர். இன்று அறிவியல், கலை, தகவல், வணிகம், தொடர்பு மற்றும் தொழில்நுட்பம் என பல்வேறு துறைகளிலும் தமிழர் ஈடுபட்டிருப்பதாலும், மேலும் பல காரணங்களினாலும் இலக்கண நெறியைத் தளர்த்த வேண்டிய நிலை ஏற்பட்டுள்ளது. இதனால் அறிவியல், வர்த்தகம், தொழில்நுட்பம், கலை, இடங்களின் பெயர்கள் மற்றும் பலவற்றை அதே நிலையில், அதே பெயரில், எவ்வித மாற்றமுமின்றி ஏற்றுக்கொள்ளத் தொடங்கியிருக்கிறது. இத்தளர்வால், தமிழ் எழுத்துக்களைக் கொண்டு நாம் ஏற்றுக்கொண்ட பல்வேறு சொற்களைத் தெளிவாக எழுத, எழுதியவற்றை சரியான முறையில் படிக்க இயலாத நிலை/தடங்கல் ஏற்பட்டிருக்கிறது. எனவே, தமிழ் சொற்களைப் போல பிறசொற்களையும் தவறின்றி எழுத, அவற்றை தெளிவாகப் படித்துணரக்கூடிய நிலையில் மொழி இருக்க வேண்டும் என்பதால் தமிழில் எழுத்துச் சீர்த்திருத்தம் அவசியமாகிறது.

சீர்திருத்த முறை

தமிழர் பல்லாண்டுகளாக பயன்படுத்தும் ஒலிகளில், குறிப்பிட்டவற்றிக்கு தமிழ் எழுத்துக்களை வரையறுத்து, பிற ஒலிகளுக்கும், மேலும் தற்காலத்தில் பரவலாக ஏற்றுக்கொண்டிருக்கும் தமிழல்லாத ஒலிகளுக்கும், புதிய எழுத்துக்கள் உருவாக்கப்பட்டுள்ளன. தமிழில் உருவாக்கப்படும் புதிய எழுத்துக்களும் மாற்றங்களும் பின்வரும் முறைகளின் அடிப்படையில் அமைக்கப்பட்டுள்ளன:

1. தமிழ் எழுத்துக்களுடன் ஒத்திருக்கும் வரிவடிவங்களாக இருத்தல். (எ.கா. பெரும்பாலான தமிழ் எழுத்துக்களுக்கு 90˚ திருப்பங்களும், பல எழுத்துக்களின் மேலிருக்கும் கோடு வலதுபுறம் நீண்டும் இருக்கும்)
2. தமிழெழுத்துக்களை அறிந்தோர் உணருமாறு இருத்தல்.
3. மாற்றங்களை ஏற்றும்/தவிர்த்தும் படிக்க கூடியதாக இருத்தல்.
4. நடைமுறைப்படுத்துவதில் சிக்கல்கள் குறைந்திருத்தல்.

புதிய எழுத்துக்கள்

நாம் பரவலாக பயன்படுத்தப்படும் பொதுவான ஒலிகளைப் பிரித்துணரும் வகையில் இச்சீர்திருத்தம் வடிவமைக்கப்பட்டுள்ளது. கீழ்கண்ட முறைகளின் அடிப்படையில் புதிய எழுத்துக்கள் உருவாக்கப்பட்டுள்ளன:

1. தமிழில் உள்ள ஒலிகளுக்கு தனித்தனி எழுத்துக்களை வரையறுத்து தேவையானவற்றிக்கு புதிய எழுத்துக்களை உருவாக்குதல்

2. தமிழில் (எழுத்துக்களுக்கு) இல்லாத, ஆனால் தற்காலத்தில் பயன்படுத்தப்படும், ஒலிகளுக்குப் புதிய எழுத்துக்களை ஏற்படுத்துதல்

தமிழில் உள்ள ஒலிகளுக்கான வரிவடிவங்கள்

தமிழ்ச் சொற்களில் உள்ள ஒலிகளை தனித்தனியாக வகைப்படுத்தி, தனிமைப்படுத்தவேண்டிய ஒலிகளுக்கு புதிய எழுத்துக்களைத் தோற்றுவிப்பதின் மூலமே பெரும்பாலான அறிவியல் சொற்களையும் அவற்றின் ஒலிகளையும் தமிழில் பிழையின்றி கொண்டுவர முடியும்.

ககரத்தை ஒத்துவரும் ஒலிகள்: கல்வி, அங்கம் என்ற இரு சொற்களில், கல்வி என்பதில் ககரம் வல்லொலியாகவும் அங்கம் என்பதில் மெல்லொலியாகவும் ஒலிக்கிறது. வல்லொலியை 'க' என்றும் மெல்லொலியை 'ங' என்றும் பிரிப்பதால் ஒலிக்கேற்ப தனித்தனி எழுத்துக்கள் அமைகின்றன.

சகரத்தை ஒத்துவரும் ஒலிகள்: பாய்ச்சல், சங்கு என்ற வார்த்தைகளில் உள்ள சகரம், பாய்ச்சல் என்பதில் வல்லொலியாகவும், சங்கம் என்பதில் மெல்லொலியாகவும் ஒலிக்கிறது. மெல்லொலியாக வரும் 'ச', 'ஞ' வாக மாறி தனியெழுத்தாகிறது. ஆனால், தற்பொழுது 'ஸ' என்ற வரிவடிவமும் பயனில் உள்ளது. இவ்வெழுத்துக்களின் பயன்பாட்டிலுள்ள பிரச்சினைகளும் அதற்கான நடைமுறை தீர்வுகளும் பின்னர் விவரிக்கப்பட்டுள்ளது.

டகரத்தை ஒத்துவரும் ஒலிகள்: தமிழில் டகரம் வல்லொலியாக சட்டம் என்ற சொல்லிலும், மெல்லொலியாக படம் என்ற சொல்லிலும் வருகிறது. இங்கு படம் என்ற சொல்லில் உள்ள 'ட', 'ண' வாக மாறி மெல்லொலிக்கு தனியெழுத்து உருவாகிறது.

தகரத்தை ஒத்துவரும் ஒலிகள்: தண்ணீர், சந்தம் என்ற இரு சொற்களில் தண்ணீர் என்பதில் தகரம் வல்லொலியாகவும், சந்தம் என்பதில் மெல்லொலியாகவும் ஒலிக்கிறது. இங்கு மெல்லொலியாக வரும் 'த', 'ந' வாக மாறி அதன் ஒலிக்கேற்ப தனியெழுத்தாகிறது.

பகரத்தை ஒத்துவரும் ஒலிகள்: பசி என்ற சொல்லில் பகரம் வல்லொலியாகவும், ஐம்பது என்ற சொல்லில் மெல்லொலியாகவும் வருகிறது. ஐம்பதில் வரும் 'ப', 'ம' வாக மாறி மெல்லொலிக்கு தனியெழுத்தாக அமைகிறது.

வகரத்தை ஒத்துவரும் ஒலிகள்: அவன், உவமை என்ற சொற்களிலுள்ள வகரம் முதல் வார்த்தையில் மேற்பல் கீழுதட்டை தொட்டு விலகும் போதும், இரண்டாம் வார்த்தையில் இரு உதடுகளும் குவிந்து விரியும் போதும் உருவாகிறது. எனவே இரண்டாவது வார்த்தையில் வருமாறு அமையும் வகரம் 'வ' வாக மாறி தனியெழுத்தாகிறது. இவ்வாறு 'வ'வின் நடுவில் வைக்கப்படும் புள்ளியால் உருவாகும் இவ்வெழுத்து, எந்தவொரு உயிர்மெய்யாகும்போதும் எழுத்தின் வடிவை மாற்றாமலும் தனித்து தெரியும்படியும் அமைந்திருக்கும். இதன்படி வியன்னா, வாஷிங்டன் என்பனவற்றை வியன்னா, வாஷிங்டன் என ஒலிக்கேற்ப பிரித்துணரலாம்.

தமிழில் அல்லாத ஒலிகளுக்கான வரிவடிவங்கள்

- ஹ, ஜ, ஷ போன்ற வரிவடிவங்கள் குறிப்பிட்ட ஒலிகளுக்காக பல்லாண்டுகளாக பயன்படுகின்றன. தமிழ் எழுத்துக்களைப்போல் இல்லாமலிருந்தும் தொடர்ந்து பல ஆண்டுகளாக பயன்படுவதால், அதே நிலையில் ஏற்றுக்கொள்வதே சிறந்தது.

- இன்ஃபோசிஸ், ஃபாரன்ஹீட், ஃப்ரான்ஸ், ஃபின்லாந்து என்ற சொற்களில் வரும் ஒரு ஒலி, ஃ, ப என்ற இரு வேறுபட்ட ஒலிகளைக் கொண்ட எழுத்துக்களைச் சேர்த்தமைத்து ஓரெழுத்தாக தமிழில் எழுதப்படுகின்றது. இதை எளிமையான ஒரெழுத்தாக 'ஹ' என்று எழுதுவது மேன்மையைத்தரும். இவ்வாறு 'ப'வின் நடுவில் வைக்கப்படும் புள்ளியால் உருவாகும்

இவ்வெழுத்து, எந்தவொரு உயிர்மெய்யாகும்போதும் இடர்படாமலும், எழுத்தின் வடிவை மாற்றாமலும் மேலும் தனித்து உணரும்படியும் அமைந்திருக்கும். எனவே இச்சொற்கள் இனி , இன்ஃபோசிஸ், ப்ராண்ஹீட், ப்ரான்ஸ், ப்ின்லாந்து என்றமையும்.

- Zambia, Zen, zip, benzene என்ற வார்த்தைகளில் வரும் ஒரு ஒலியை தமிழில் சரியாக எழுத தற்போது எழுத்துக்கள் இல்லையென்பதால், அவ்வெழுத்துக்குறிய ஒலியை, பொதுவாக, 'ஜ' என்ற எழுத்தை பயன்படுத்தி ஜாம்பியா, ஜென், ஜிப், பென்ஜின் என்று எழுதுகின்றோம். இவ்வெழுத்து பல இடங்களில் ஒலியை மட்டுமல்லாமல் விளக்கத்தையும் மாற்றக்கூடியது. எனவே இவ்வேறுபாட்டை வெளிப்படுத்தும் விதமாக இங்கு 'ஜ' விற்கு பதிலாக 'ஐ' என்ற புதிய எழுத்தைப் பயன்படுத்த, சொல்லும் எழுத்தும் தெளிவாகும். இதன்படி அவ்வொலியுடைய சொற்கள் இனி ஐாம்பியா, ஐன், ஐப், பென்ஐின் என்று எழுதப்படும்.

திருத்தம்/பிரச்சினைகளும் தீர்வுகளும்

- ஙகரம் உயிர்மெய் எழுத்தாக அமைந்து, அங்**ஙனம்**, ஆங்**ஙனம்**, இங்**ஙனம்**, எங்**ஙனம்** என்ற சொற்களிலுள்ள 'ங' வைத் தவிர, வேறு எந்தவொரு சொல்லையும் ஏற்படுத்தாததால், ஆயுத எழுத்தைப் போன்று 'ங்' என்ற சொல் தனித்து ஓரெயெழுத்தாக இயங்கும். மேலும் இச்சொற்கள் தற்காலத்தில் பெரிதும் பயன்படுத்துவதில்லை. இதனால் மேற்குறிப்பிட்ட சொற்களைப் பயன்படுத்த வேண்டிய நிலை ஏற்பட்டால் அச்சொற்களில் உள்ள ஙகரம் இனி அங்கனம், ஆங்கனம், இங்கனம் எங்கனம் என ககரமாக அமையும்.
- ிவுடன் உகரமும் ஊகாரமும் சேர்ந்து உயிர்மெய்யாகும்போது உருவாகும் எழுத்துக்கள் ரு, ரூ என்ற எழுத்துக்களைப்போல் அமையுமாதலால், அவ்விரு உயிர்மெய் எழுத்துக்களையும் ரு, ரூ என்று இரு நேர்க்கோடுகளை மேலே இணைக்காமல் இடையில் இணைத்து எழுதுதல் வேண்டும்.
- 'ஸ' என்ற எழுத்து பல்லாண்டுகளாகப் பயன்பாட்டில் இருந்தும் எல்லா நிலைகளிலும் பயன்படுத்தப்படுவதில்லை. எடுத்துகாட்டாக, சங்கம், சட்டசபை, சேவல் என்ற சொற்களை ஸங்கம், ஸட்டஸபை, ஸேவல் என்று எழுதுவது கிடையாது. இவற்றை சங்கம், சட்டசபை, சேவல் என்று எழுதுவதே அதனினும் சிறப்பானதாக இருக்கும். அதே நேரத்தில், சகரம் மெய்யெழுத்தாக வரும்போது இவ்வெழுத்து நன்கு பழக்கப்படும் வரை குழப்பம் ஏற்படும். எ.கா. விஸ்வபாரதி, ஸ்காட்லாந்து என்பனவற்றை விச்வபாரதி, ச்காட்லாந்து என்று எழுதினால் தடுமாற்றம் ஏற்படக்கூடும். எனவே, மெய்யெழுத்தாக வரும்போது 'ஸ்' என்றும் உயிர்மெய்யாக வரும்போது 'ச' என்றும் ,சில காலங்களுக்கு, எழுதுதல் வேண்டும். எ.கா. சரஸ்வதி.

தூய, பயனிலுள்ள மற்றும் சீர்திருத்த முறையில் உருவான தமிழ் எழுத்துக்களின் அட்டவணை கீழே காணலாம்:

	தூய தமிழ்	**பயனிலுள்ளவை**	**புதிய எழுத்துக்களுடன்**
உயிர் எழுத்து	12	12	12
மெய் எழுத்து	18	23 [18 + ஸ், ஜ், ஷ், ஹ், ஃப்]	28 [17 (ங் தவிர) + ஜ், ஷ், ஹ், க், ச், ட், ப், த், ப், வ், ஐ]
உயிர்மெய்	12 X 18 = 216	12 XX 23 = 276	12 XX 28 = 336
தனியெழுத்து	1 [ஃ]	1 [ஃ]	3 [ஃ, ங், ஸ்]
மொத்தம்	**247**	**312**	**379**

ஔகாரம் பழந்தமிழில் இல்லை. இப்புதிய எழுத்துச் சீர்திருத்தத்தின் மூலமும் ஔகார எழுத்துக்களின் ஒலியை சிக்கலின்றி வெளிப்படுத்தலாம். இவ்வெழுத்து 'ஃப'வைப் போன்று இரு வேறுபட்ட ஒலிகளுக்கான எழுத்துக்களைக் (கெ, ள) கொண்டு எழுதப்படுகிறது. எனவே ஆயுத எழுத்தைப்போல 'ஔ' வைத் தனி எழுத்தாக்கலாம். ஆனாலும், கௌதமன், கௌரி, கௌதாரி, சௌக்கியம், சௌதி அரேபியா, பௌத்தம், பௌர்ணமி, ஔவை போன்ற ஔகாரத்தைப் பயன்படுத்தும் சொற்கள் இன்றும் பயனில் உள்ளன. எனவே, ஔகாரத்தை தனி எழுத்தாக்கலாமா வேண்டாமா என்பது தீர விவாதிக்கப்பட வேண்டியதாகும். அவ்வாறு ஏற்றுக்கொண்டால் உயிர்மெய் எழுத்துக்கள் 308 ஆகவும் மொத்த எழுத்துக்கள் 351 ஆகவும் அமையும்.

எழுதும் முறையின் அமைப்பு

இப்புதிய எழுத்துக்களை அச்சில் எவ்விடத்திலும் சீரான முறையில் எழுதமுடியுமென்றாலும், கையில் சரியான முறையில் எழுதுவதில் சிரமம் இருக்குமாதலால், கூ, சூ, தூ, டூ, பூ என்ற எழுத்துக்களை கூ, சூ, டூ, தூ, பூ என்று எழுதலாம்.

கணினி தட்டெச்சு, கைபேசி தட்டெச்சு மற்றுமுள்ள மின்னணு தட்டெச்சுக்ளுக்கேற்றபடி தமிழ் எழுத்துக்களை சில கோர்வைகளில் ஒருங்கிணைக்க வேண்டும். இவை தட்டச்சு இயந்திரங்களுக்கு பயன்படாது. இவ்வாறு அமைத்த சில மாதிரிகள் கீழே கொடுக்கப்பட்டுள்ளன. இவற்றை செயலாக்க மற்றும் நடைமுறைப்படுத்த மென்பொருள் உருவாக்குதல் அவசியம்.

கணினியின் தட்டெச்சு மாதிரி

தற்பொழுது பொதுவாக காணப்படும் ஆங்கில தட்டெச்சை அடிப்படையாகக் கொண்டு இந்த மாதிரி வடிவமைக்கப்பட்டுள்ளது. தமிழில் உள்ள அனைத்து எழுத்துக்களையும் கணினியின் 27 பட்டன்களைக் (';' வும் 26 ஆங்கில எழுத்துக்களும்) கொண்டு எழுதலாம்.

ஒரெழுத்தாக வருபவை (12): இவ்வெழுத்துக்கள் தட்டெச்சின் ஒரு பட்டனை தட்ட வெளிப்படும். ஒருமுறைக்கும் மேல் தட்ட அதே எழுத்து திரும்பவரும்.

க்	கீ	ஹ்ற	ச்	ஜ்	ப்	ஃப்	ஃபு	ம்	ய்	வ்	வ்ஃ
K K	G G	H	C C	JJ	PP	BB	FF	M M	YY	V V	W W

குறில் நெடிலாக வருபவை (5): இவ்வெழுத்துக்கள் ஒருமுறை தட்ட உயிர் குறிலாகவும் இருமுறை சேர்த்துத்தட்ட உயிர்நெடிலாகவும் வரும்.

அ	இ	உ	எ	ஒ
ஆ	ஈ	ஊ	ஏ	ஓ
A	EE	UU	;	OO

இருவேறு எழுத்துக்கள் (10): மேல் வரிசையிலுள்ளவை அதற்குரிய பட்டனைத் தட்டவும், கீழ் வரிசையிலுள்ளவை கூடுதலாக H பட்டனையோ அல்லது SHIFTSHIFT SHIFT பட்டனையோ சேர்க்க வெளிப்படும். சில காலங்களுக்கு ஸ்/ஸ் என்ற எழுத்துக்களில் மெய் 'ஸ்' ஆகவும் உயிர்மெய் 'ஸ்' ஆகவும் வரும்.

ஐ	ங்	ஸ்/ஸ்	ட்	ஃட்	ண்	ன்	ர்	ல்	ழ்
ஃஃ	ஞ்	ஷ்	த்	தூ	ஔ	ந்	ற்	ள்	ழ்
I	Q	S	T	D	X	N	R	L	Z

கைபேசி தட்டெச்சு மாதிரி

கைபேசியின் பயன்பாட்டில் பல்வேறு முறைகளில் சொற்களை அமைக்க முடியும். இங்கு குறிப்பிட்டுள்ள மாதிரியில் உயிரெழுத்துக்களுக்கு முதன்மை கொடுக்கப்பட்டுள்ளது. இதனால், முதலில் உயிர்க்குறியும் உயிர்நெடியும் வெளிப்பட்டு பின்னர் மெயெழுத்துக்கள் தோன்றும். உயிரும் மெய்யும் இணைய உயிர்மெய் உருவாகும். கீழ்க்கண்ட அட்டவணையில் ஒன்று முதல் ஒன்பது வரையிலான எண்களுக்கு மாதிரி முறை ஒன்று கொடுக்கப்பட்டுள்ளது.

1	2	3	4	5	6	7	8	9
அ	ச்	இ	உ	எ	ஐ	ஒ	வ்	ம்
ஆ	ச்/ஸ்	ஈ	ஊ	ஏ	ல்	ஓ	வ்	ய்
க்	ஷ்	ட்	த்	ப்	ள்	ந்	ர்	ங்
க்	ஜ்	ட்	த்	ப்	ழ்	ன்	ற்	ஞ்
ஹ்	ஸ்	-	-	ப்	-	ண்	ஃ	ஔ

முடிவுரை

இச்சீர்திருத்தத்தில் அறிமுகப்படுத்திய எழுத்துக்களையும் தமிழில் ஏற்றுக்கொண்டால், சிறந்த முறையில் கணினி மற்றும் கைபேசி பயன்பாட்டில் தமிழை செயலாக்க முடியும். மேலும்:

- இச்சீர்திருத்தம் தமிழ் அல்லாத அல்லது தமிழ் இலக்கண நெறிகட்க்கு உட்படாத சொற்களுக்காக உருவாக்கப்பட்டதாகும்
- மருத்துவம், பொறியியல், வர்த்தகம், தொழில்நுட்பம், அறிவியல், கலை மற்றும் பல துறைகளிலுள்ள மிகப் பெரும்பாலான ஒலிகளை தமிழில் தெளிவாக எழுதவும் படிக்கவும் முடியும்.
- இச்சீர்திருத்தத்திலுள்ள மாற்றங்களை ஏற்றும்/தவிர்த்தும் படிக்கலாம்.
- அறிவியல் மற்றும் பிற சொற்களுக்கு அடுத்து அடைகுறிப்பினுள் ஆங்கிலத்தில் எழுதுவதைத் தவிர்க்கலாம்.
- படிக்க மட்டுமே தெரிந்தவரும் சொல்லின் தனித்தனி எழுத்துக்களைச் சேர்த்து உச்சரித்தாலே வார்த்தை தெளிவாக உருவாகும் - தமிழுக்குரிய சிறப்புநிலை மேன்மையடைகிறது.

Tamil on Mobile Devices

Challenges and Opportunities

Muthu Nedumaran
(Muthu at Murasu dot Com)

Introduction

When the Compounded Annual Growth Rate (CAGR) for mobile penetration is compared to that of the Internet, almost every country in the world records a higher number for the former. Mobile is certainly growing faster. The number of mobile phone users worldwide has exceeded 4 billion in the year 2009 and is expected to touch 5 billion in 2010[1]. It is no surprise why this industry is getting so much attention and drawing new and exciting innovations, let alone new players.

In recent years, the landscape of the mobile industry has seen dramatic changes in two main areas: *characteristics of devices* and *types of mobile content*. These two has opened the doors for numerous new types of content and services that were never possible before. In addition, there is a third factor that is driving the growth of mobile content: *push from mobile operators*. In developed countries, where mobile penetration has exceeded 100%, operators need to find innovative ways to increase their ARPU (Average Revenue Per User). Traditional revenue from voice and SMS alone does not promise them a healthy growth rate. Thus the increased focus on content and VAS (Value Added Services)

The demand for Tamil content on mobile platforms has not been as dramatic in comparison. It certainly does not correspond to the volume of Tamil content available on the Internet. We can attribute this to two main reasons: (a) lack of native Tamil support on mobile devices in general and (b) unlike the Internet, mobile content is not free.

Murasu Systems Sdn Bhd (Malaysia) developed the world's first Tamil SMS application in 2003[2] and launched it as a live service together with Mediacorp Radio's Oli96.8FM (Singapore) in 2005[3]. Since then the product, called Sellinam, has evolved both horizontally to support more devices and vertically to support more services and content. This paper provides a summary of experiences gained in developing a mobile content management and distribution framework over the years.

Distribution of Mobile Content

Broadly, distribution of mobile content involves three components: Content Management Platform (CMP), Over-the-air channel (OTA) and the terminal (Handset). Let's look at each of these in the light of Tamil content.

Content Management Platform

The CMP contains the database, connectivity to operator networks and the glue code to bind the two. If the content that is managed does not involve subscriptions and delivered only via Data networks (3G/EDGE/GPRS/WAP), operator connectivity may not even be required. However, if subscriber status is important, operator support may be necessary in order to obtain subscriber information as well as device details.

Obtaining device information can be critical for certain types of content. In particular those that need to be formatted for specific screen sizes or device capabilities.

It is for these reasons that Sellinam is offered through mobile operators. With operator integration, content can be pushed via SMS and MMS in addition to IP based data networks. When content is requested via Data networks, the operators pass through the device information as well as the subscriber's mobile number (MSISDN). This enables management of content for the subscriber at the CMP end. For example, content that has already been read need not be presented again to the same subscriber.

Tamil content in the CMP can be in stored Unicode. Other data formats can be considered when the content is pushed over the air into custom applications.

OTA Delivery

In a typical GSM network content can be delivered to mobile devices over two possible channels: SMS or Data. MMS, which is a relatively recent messaging service utilizes both where the alert is pushed via SMS and the content is pulled via Data.

a. SMS Channel

The SMS service was designed exactly for what its name indicates: short-messages. SMS message can contain 160 characters in a segment when sent as a text message or 140 characters when sent as a binary message. The size of a character in a text message is 7bits. In other words, it can only contain 128 possible characters and these are defined in the GSM 03.38 standards document.

The GSM 03.38 Default Character Set

Dec		0	16	32	48	64	80	96	112
	Hex	0	10	20	30	40	50	60	70
0	0	@	Δ	SP	0	¡	P		p
1	1	£	_	!	1	A	Q	a	q
2	2	$	Φ	"	2	B	R	b	r
3	3	¥	Γ	#	3	C	S	c	s
4	4	è	Λ	¤	4	D	T	d	t
5	5	é	Ω	%	5	E	U	e	u
6	6	ù	Π	&	6	F	V	f	v
7	7	ì	Ψ	'	7	G	W	g	w
8	8	ò	Σ	(	8	H	X	h	x
9	9	Ç	Θ	)	9	I	Y	i	y
10	A	LF	Ξ	*	:	J	Z	j	z
11	B	Ø	<ESC>	+	;	K	Ä	k	ä
12	C	ø	Æ	,	<	L	Ö	l	ö
13	D	CR	æ	-	=	M	Ñ	m	ñ
14	E	Å		.	>	N	Ü	n	ü
15	F	å	É	/	?	O	§	o	à

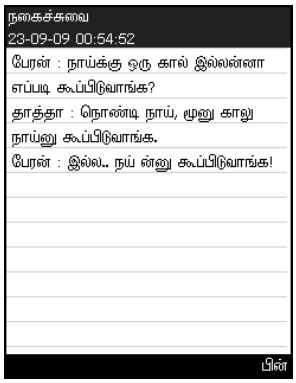
Additional characters in ASCII that are not in the Default Character Set can be sent via escapes. These characters are: { } [ ] \ | ` ~ and the Euro sign□.

If the text message contains characters that are not defined above, the message is typically converted to binary format by the handset before taking it over the air. In a binary format message, each character is 8bits wide. Unicode messages are sent as binary messages. The text in a binary message

can contain standard Unicode characters and in the case with Tamil, each Tamil-Unicode-Character will occupy 2 bytes.

Because of the limitations in the number of characters that can be put into a single SMS message, long messages are split into multiple segments and sent as multiple SMS messages. Each segment will contain header information for the receiving handset to concatenate the individual segments back together.

Words in an SMS message are typically short. If we estimate an average Tamil word to contain 7 Unicode characters (as in வணக்கம்) we will need 14 bytes. In a 140-byte segment, we can fit 8-10 words. This should be sufficient for a simple and short Tamil message. Longer messages can be concatenated into multiple segments.



Content delivery over SMS: While we can live with the GSM standard for peer-to-peer SMS messaging in Tamil, it may not serve well for SMS content delivery. A reasonable SMS content in Tamil will not fit into a single segment. A simple joke as shown in the screen capture requires 300 bytes, which means it requires 3 SMS message segments to deliver it to the handset. Pushing this content from the CMP over the operator network can be costly.

This is where operator support comes in handy. A mobile operator can choose to absorb the cost of delivering this content to the handset and offer it as a bundled subscription service.

Sellinam OTA format: Sellinam is a custom mobile application for messaging and content in Tamil. It serves to compose, deliver and receive content in Tamil over both SMS and Data networks. Since all of the communication happens between Sellinam and the CMP, a proprietary format was designed for push content where more characters can be packed into a single segment. The format was inspired by the way GSM standards pack text messages as 7-bits to realise 160 characters in a single segment (Google for SMS PDU format specification for details).

b. Data channel

Data channel is similar to what we typically see in IP (Internet Protocol) networks. The difference is that the data is transported over mobile networks such as 3G, Edge, GPRS or WAP.

For mobile content offered to custom applications like Sellinam, the Data channel provides tremendous opportunities.

However, there are challenges. Unlike the SMS channel, which simply works without any form of setup of configuration at the user end, Data networks require device support, Data-plan subscriptions and coverage areas. The coverage area for Data may be not as wide as SMS. These are improving with technologies like over-the-air device detection, configuration messages pushed over SMS and bundled plans offered by operators.

Introduction of newer handset models, which are becoming more and more data centric, is helping with the promotion of data usage.

A key difference in content delivery in Data networks versus SMS is that the subscriber will be able to 'pull' content on-demand instead of waiting for a push.

c. Sellinam Content Delivery

Sellinam, as it is offered today in Malaysia and Singapore, pulls content via Data network. In Chennai, content is pushed to the handsets as SMS messages via the operator's SMS-C.

Mobile Devices and Tamil Support

For the purpose of discussion, we can classify mobile devices into three categories: low-end, smart-phones and web-phones.

Low-end phones are those manufactured primarily for voice and SMS. Data support is extremely limited, if at all included. As such, the only reliable communication channel to deliver content to these devices is SMS. Tamil content can only be delivered to handsets that have built-in support for Tamil.

In recent years, manufacturers like Nokia and Samsung have been adding Indian language support to their low-end handsets. However, the number of models that are supported is minimal and the usage is limited to user interface and messaging. Mid-range devices incorporate basic PDA features and reasonable Internet access. These devices incorporate a browser that can browse content formatted for mobile screens. Symbian is the operating system that is dominant in this category. The latest trend in mobile devices is the Web phone. Apple's iPhone was a phenomenal invention in this space where it made Internet on mobile easier than before. Google's Android, RIM's BlackBerry and Microsoft's WindowsMobile are other players who are providing technologies for Web phones.

Tamil support through client applications: Indian language support is generally unavailable on high-end devices. The features that are put into these are driven by market demand. Unlike China, English is predominantly the language for communications in India. As such, there is no real need for Indian languages to be resident in smart-phones for it to be successful in India. This is certainly changing in the Web-phone space where users will like to access Indian/Tamil content that is readily available on the Internet. However, with a bit of hard work, it is possible to realise Tamil on smart phones today. Sellinam did this by building a text input and presentation engine built from ground-up in Java ME, which is predominantly available on almost all mid-range phones and most Web-phones. Sellinam was recently ported to the iPhone, making it the first Tamil application on this platform. All of the content provisioned to Sellinam applications, both JavaME and iPhone, are served by the same Content Management Platform; using the same data formats. It is rather tedious to build all of the client features for every platform. As such, a native Windows Mobile version and a native BlackBerry version are still incomplete. However, since both these platforms support Java as well, Sellinam runs reasonably well on the virtual machine in these handsets. It is our hope that the device manufacturers will add support for Indian languages in general and Tamil in particular some day soon. We need to generate sufficient demand in the market for this to become a reality.

References

1 http://www.eito.com/pressinformation_20090807.htm
2 Maxis Communications Bhd (Malaysia) 2005 Annual Report
3 http://murasu.com/mobile
4 http://en.wikipedia.org/wiki/GSM_03.38

TAMIL E-TEXTS, CORPORA AND DIGITIZATION OF ANCIENT TAMIL TEXTS

Tamil Corpus Generation and Text Analysis

M. Ganesan
Annamalai University

Introduction

Language can be studied from the point of view of language structure and language use. Study of language structure is called structural or formal linguistic study. Study of language use is called as functional linguistic study. Languages are described qualitatively in terms of grammatical units like nouns, verbs, noun phrases, verb phrases, subject, object, agent, goal, etc. to explain the structure of the language. Most of the grammars take a sentence as the minimum unit for the description of the structure. The structure has been studied from different view points and many linguistic theories have emerged to account the syntactic pattern of the sentence.

Languages can also be studied quantitatively in terms of frequency, place of occurrence, pattern of occurrence, etc. of various linguistic units in a text. The quantitative study basically needs a large quantum data and a mechanism to browse them fast. Now the computer technology facilitates to store and study a huge texts to the tune of hundreds of millions of words in few seconds or minutes. A new method of language study called corpus linguistics has emerged in recent years. Corpus is a large collection of written or spoken texts available in machine readable form accumulated in scientific way to represent a particular variety or use of a language. It serves as an authentic data for linguistic and other related studies. The size, text type, organization, accessing method, etc. are some of the basic features of a corpus which have to be carefully decided while generating a corpus. There are different types, which are again determined by the purpose for which the corpus is built.

In this article I share my experience in the generation of CIIL corpus and explain the scheme of POS tagging using morphological analysis. I also discuss the various tools that I have developed to analyze Tamil texts (Corpus Analysis Tools for Tamil) and their use in different applications.

Tamil Corpus Generation

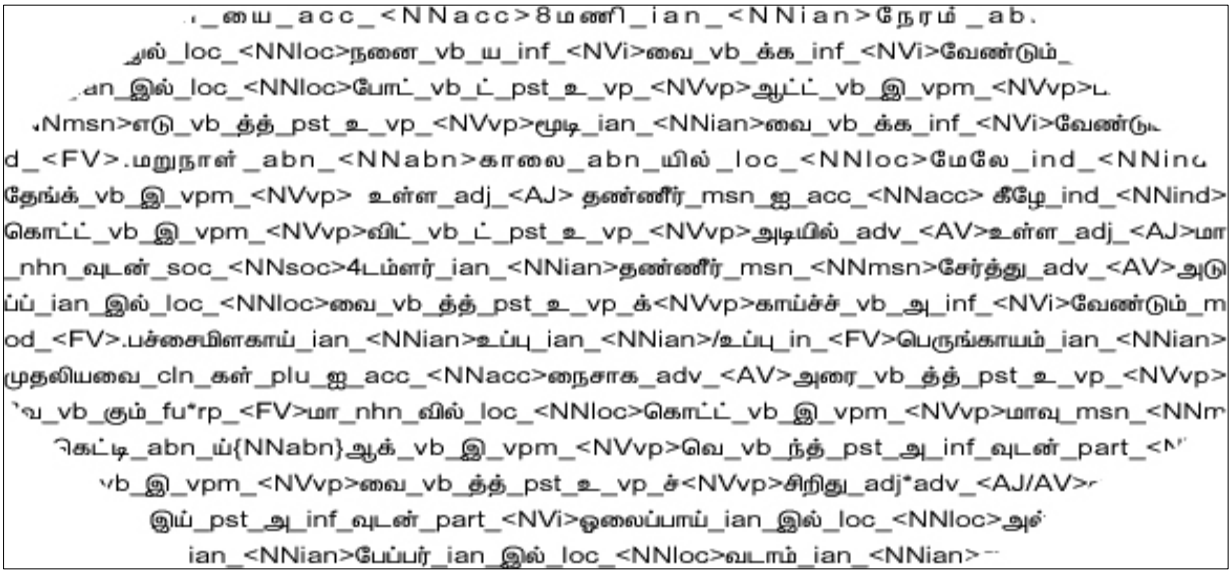
The first corpus for modern written Tamil was built in the Central Institute of Indian Languages (CIIL), Mysore in 1987. The CIIL in collaboration with the Tokyo University of Foreign Studies, Japan built a corpus on Tamil textbooks. In 1991 under the scheme Technological Development for Indian Languages(TDIL), the Department of Electronics, Govt. of India launched a project called Development of Corpora of Texts of Indian Languages. The CIIL was entrusted to build Corpora for the four major Dravidian languages, where the present author worked as an investigator. Texts printed during the period 1981 to 1990 were selected to represent the modern Tamil. They are collected from 6 major categories, viz. Aesthetics (Literature and Fine arts), Social Sciences, Natural, Physical and Professional Sciences, Commerce, Administration and Technology, and Translated Texts. They are further classified into 76 minor categories, to cover various domains of language use. The size of the Tamil corpus is 3.6 million words. The major objectives of the project was to build corpora of not less than 3 million words, and to develop software for grammatical tagging (at word level), KWIC Concordance, and for corpus management. In 1993 a spoken Tamil corpus (1 lakh words) was generated on the transcribed spoken data collected by Eric Pederson, Netherlands. The Mozhi Truest, Chennai has build a corpus of around 3 million word for modern written Tamil. The CIIL is currently augmenting the corpora of Tamil texts and creating corpus for spoken Tamil.

Corpus Organization

The data for the CIIL corpus are collected from books, textbooks, magazines, newspapers and Government documents in order to represent the contemporary Tamil. The collected data are organized with the following information: 1) major category, 2) sub category, 3) title of the text, 4) author name, 5) source 6) publishers, 7)year of publication, and 8) page numbers. These information help the user to retrieve any data selectively from the corpus. Further the organization of data is needed for any addition or deletion of data from the corpus. A software called "corpus manager" does these jobs.

POS tagger

The collection of texts, called Raw Corpus can be provided with many additional information, particularly grammatical ones at different levels, viz. phonological, morphological, syntactic, semantic and discourse level. It is called annotating or tagging the corpus. The POS (Parts of speech) tagging is a popular and common type of annotation successfully implemented on a number of corpora in English and other European languages. The tagging can be achieved in the following four ways: 1) rule- based tagging, 2) statistics –based tagging, 3) pattern – based tagging, and 4) manual tagging. The first three are automatic tagging (with manual post-editing) and the last one, manual tagging is slow, labour intensive and liable to error and inconsistency(Leech, 1992:131). The present author has developed an automatic tagger called "Morph and POS tagger for Tamil"(Ganesan, 2007) which tags at morph and word level. At present the tagset has 82 tags at morph level and 22 at word level. A sample tagged text is given below.



Syntactic Tagger

A software for tagging at phrase and clause level has been developed for Tamil by the present author. The texts tagged at morpheme and word level will be the input for the syntactic tagging. Tamil is an agglutinative language; therefore its morphology is complex. But, its morphotactics is very tight. Therefore identifying the morphs is not very difficult. But, syntax of Tamil is very loose; because it is comparatively a free word ordered language. It is very difficult to identify the phrase boundary.

Tools for Text Analysis

The potentiality of corpus in language studies, both theoretical and applied, are enormous. Language description, testing of grammatical theories, natural language processing(NLP), language teaching, dictionary making, translation (both human and machine), style analysis, etc. are the major areas where corpus can throw lot of insights. To bring out all those information that are needed for these applications, a number of tools have to be developed. Frequency count (letter, syllable, word frequency), searching (for particular pattern, in particular context, at different levels), sorting (forward and reverse), indexing, concordance, KWIC, tag search, word list, lemma extraction, type/token ratio, etc. are some of the tools which can be used to analyze the corpus and to get a variety of quantitative information. Corpus Analysis Tools for Tamil (CATT) (Ganesan, 2007), a software provides a number facilities to extract different quantitative information from the corpora. Using the quantitative information language can be described from a new angle.

Frequency Count

The frequency of occurrence of a letter, syllable, morph, word, or a phrase, in a text can be counted using a software. For language like Tamil word frequency can be studied only after removing all the affixes from the stem. A Lemma Extractor (Ganesan, 2007) does this job and provides the list of roots in alphabetic order with frequency. For example, if one wants to study the words which are used in the primary school textbooks, using Lemma Extractor he can get them in no time. Such information facilitate to know what are the words introduced at what level, whether all the words that are intended to teach at primary level are there in the textbooks, etc. Similarly phrases and sentence types can be studied. This kind of study is practically not possible without proper computational tools. One can also study the frequency of different word forms. For example, in Tamil the verb forms Infinitive, Verbal participle and Relative participle are more frequent than the finite or imperative forms.

Verb	Total	Inf.	V.P.	R.P
col	428	56	29	33
kuuRu	1109	83	52	100
viLakku	197	34	19	11

These frequencies are from the text of 3lakh words. It clearly shows that while teaching Tamil these forms Infinitive, Verbal participle and Relative participle must be given priority and more attention than the finite forms. Another observation is that among the words *col* and *kuuRu* 'to say' the word *kuuRu* which is more literary, occurred more frequently than the word *col*. In another study the frequency of occurrence of letters are made from the data of one lakh words.

dot (on the consonants)		16.85%
ka		07.94%
vowel *u* after consonants		07.38%
ta		06.73%
vowel *i* after consonants	06.55%	etc.

Such study helps in designing the Keyboard layout for Tamil. The Keyboard based on the frequency study will be more scientific and faster to use.

KWIC Concordance

KWIC concordance is a list consists of a keyword in the middle and the contexts of 4 or 5 words on either side of the keyword. A sample is given below. KWIC concordance can be extracted from corpus for any word, part of a word, suffix, infix, prefix or even a phrase. Sorting can be done on the keyword or on the previous word or following word. It is more useful in identifying

the different meanings of a polysemous word, lexical association, collocation properties of lexical items, etc. In dictionary compilation the KWIC concordance facilitate the lexicographer to find out the various meanings of a word, subject area, registers, idiomatic usages, etc.

Word List

This tool makes a word list for all the words in a selected texts or a corpus. The words will be sorted and presented with frequency. Type / token ratio will also be given for the total word. Words with a frequency or more than particular frequency or less than a frequency can be listed separately. All the outcomes can be stored as a separate file. It helps to find out the various words found in a text with their frequency.

Word-part Search

Like the word list, here the words are sorted from the end of the word. All the same suffixes come together and therefore it helps to study the inflectional and derivational properties of different words and affixes. Using this tool a reverse dictionary for Tamil can made.

Multi-Conditional Search

A string with multi condition can be searched with this tool. If a string is given as an infix for search, the other conditions like prefixes and suffixes can also be given. All the words in the corpus satisfying all the conditions will alone be listed. For example, all the Finite verb with present tense marker can be extracted from the corpus.

Structure Search

Here the data must be a POS tagged texts. Pattern in terms of word-tags can searched. All the sentences matching the pattern will be extracted and listed. There are two options: pattern as a sentence and pattern available anywhere in a sentence. It helps to find out the usage of various phrases, clauses, particular pattern of word association, etc.

Tag Search

This tool also works on a tagged corpus. Words inflexed / derived to particular forms can extracted using this tool. First Word level tag information must be supplied, then morph level tag information in the next window. The tags may be one or more than one. There are three options: 1) include 2) exclude and 3) stem extraction. The first option lists all the words matching the word level tag and morph level tag. The second extracts all the words matching the word level tag, but excluding the morph level tag. The third option removes all the affixes and lists the stem portions alone in sorted order. It is useful to list for example, all the verb forms conjugated to particular forms or other than a particular forms.

Conclusion

The Quantitative Analysis is getting importance on par with Qualitative analysis. Language use is given priority for the description of the language. Various quantitative information extracted from the corpus provide new insights on language structure and are useful for textbook preparation, dictionary compilation, machine learning, Machine Translations, etc. The size of corpus at present available for Tamil is very small. Sometimes, many words used in day-to-day context have not attested in the corpus. Therefore there is a need to increase the size of the corpus to a minimum of 200 million words.

References

1. Ekka. Francis, B. D. Jayaram and Ganesan "Final Report Development of Corpora of Texts of Indian Languages" in Machine Readable Form, Part II (Tamil, Telugu, Kannada, Malayalam) Mysore, CIIL, 1995.
2. Ganesan, M. "A Scheme for Grammatical Tagging of Corpora in Indian Languages" in **Technology and Languages,** (Ed) BB Rajapurohit, Mysore, CIIL, 1994.
3. Ganesan, M. 'Corpus Analysis Tools for Tamil (CATT) (Software) Annamalai University, Annamalai Nagar, 2007.
4. Ganesan, M. 'Morph and POS Tagger for Tamil' (Software) Annamalai University, Annamalai Nagar, 2007.
5. Leech, Geoffery and Steven Fligelstone "Computers and Corpus Analysis" in **Computers and Written Texts** (Ed) Christopher S. Buller, Oxford: Basil Blackwell Ltd, 1992.

6. Leech Geoffery, "Corpora Annotation Schemes" in **Literary and Linguistic** Computing, Vol 8 No4 1993.
7. Shanno, C and Weaver, W. **The Mathemetical Theory of Communication,** UIP: Illinois, 1949.
8. Stubbs. Micheal, **Tect and Corpus Analysis** Oxford: Blackwell Publishers Ltd, 1996.

OMNIS/2 Integrating Libraries with Digital Multimedia Database

C.Radha
IT Department, Pre-Final Year Student,
Vivekanandha College of Engineering for Women,
Tamilnadu,India.
E-mail:c.radha007@gmail.com

Abstract: Nowadays more complementary information is stored in the electronic media. There is an increasing demand for the integration of traditional digital library systems and multimedia systems. An advanced Meta system or retrieval systems are needed for these integrations of traditional library systems. For that the OMNIS/2 system is presented in this paper which enhances existing digital library system by additional storing and indexing of user-defined multimedia documents, automatic and personal linking concepts, annotations, filtering and personalization. The OMNIS/2 system forms the multimedia storage layer, linking layer and personalization layer.OMNIS/2 is part of the Global Inventory Project of the G7 countries. This general approach ensures the integration and transparent combination of digital library systems. Users will be able to use the personalization feature to create their own view on the documents and to "work" with digital library systems by themselves. Most of the digital library systems are mere retrieval systems that can be enriched to interactive multimedia DL-systems and are combined into one virtual personal digital library. The Meta system OMNIS/2 which provides the environment in which the anchor and linking concept is successfully used in conjunction with the object oriented document model.

Introduction

Many digital library systems exist which store a variety of information. This information is usually available and stored in many different media types. The system as a whole emerged into established tools and the users in a simplified view received systems with powerful retrieval capabilities, but still miss features that would improve their ability to work with documents in digital library systems as it is common with books printed on paper (i.e. adding references, marking pages and annotating text) [1]. Some of these ideas were approached by some systems over the last decade, but a cross platform solution was always out of scope. This led us to the development of the OMNIS/2 system [2] which is a Meta system for various existing digital libraries.

The OMNIS/2 system can provide online access to historical and cultural documents whose existence is endangered due to physical decay. The major areas which offer digital libraries great exploitation are: Information retrieval, multimedia database, data mining, data warehouse, on-line information repositories, image processing, hypertext, World Wide Web and wide area information services (WAIS) [fox] [3].

The following are some of the advantages of digital libraries:

- Users can access the information everywhere
- Reduction of bureaucracy by access to the information
- The information is not necessarily located in same place
- Understanding the catalog structure is not necessary
- Cross references to other documents speed up the work of users
- Full text search
- Protection of the information source
- Wide exploration and exploitation of the information

The OMNIS philosophy is based on the "document" which is the unit for the archiving process and retrieval results. Each OMNIS document represents e.g. a catalog entry and contains information in three different sorts of attributes.

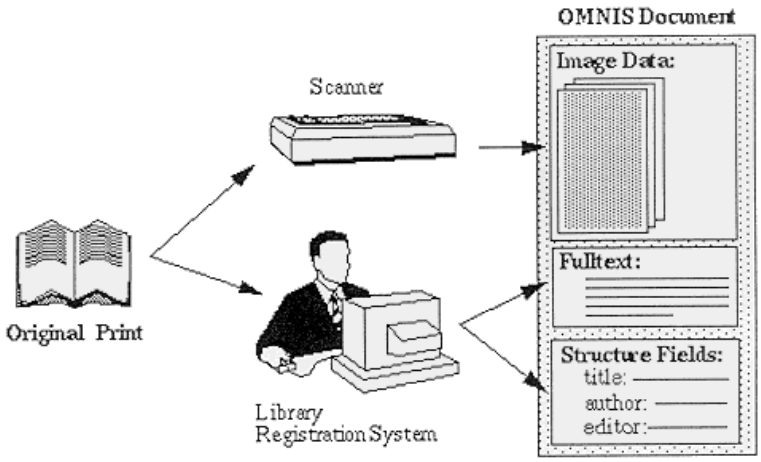


Figure 1.1.OMNIS/2 Document

- **Structure Fields** describe the catalog entry in a relational way. Fields like author, title, etc., provide structured information as known from traditional literature retrieval systems. Structure fields are the basis for relational queries and may be accessed by the user. Only a small part of each catalog entry is stored as OMNIS structure fields in Myriad databases.
- **Full text** contains the whole catalog entry as a text body. It is the basis for comfortable full text queries and may be accessed by retrieving users. A document's full text attribute is an unstructured sequence of words stored in Myriad databases and represents a superset of the document's structure fields.
- **Image Data** in some pixel format may be attached to each document. These images are stored as BLOBs (Binary Large Objects) [7] in TransBase databases and can be shown to the user.

OMNIS System Components

OMNIS is intended for retrieval and archiving from multiple remote locations. The atomic unit for the archiving and retrieval process is the "document" which may include several catalog records and provides information in different forms: as attributes, as full text form and as image data.

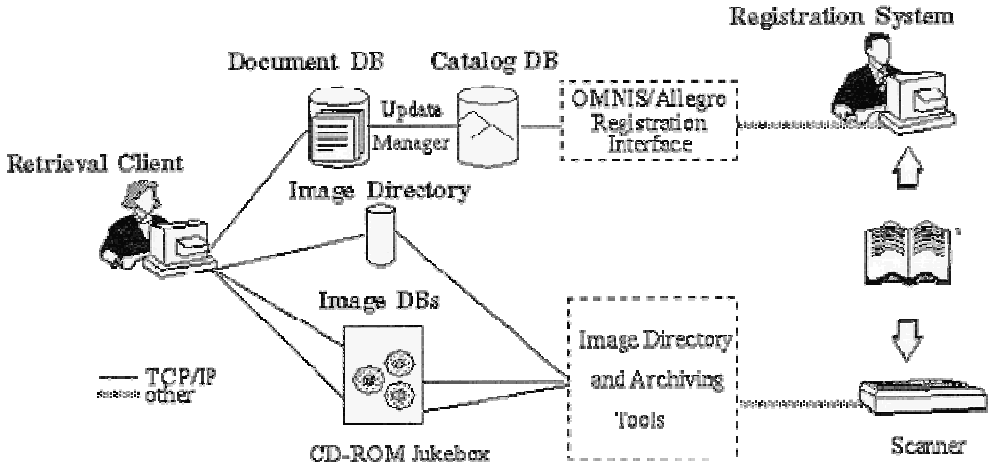


Figure 2.OMNIS System Components

The OMNIS system components that are catalog and document management, distributed images servers and retrieval interface are described below.

Catalog and Document Management

The catalog server deals with organizing, storing and providing textual catalog entries. Six registration centers spread all over Germany (Munich, Berlin, Wolfenbuttel, Dresden, Gotha, Halle) are participating via the Internet.

Distributed Image Servers

The distributed image databases allow decentralized image management [4]. Images are scanned with 1-bit color depth (black-white) and resolutions of 300 dpi, compressed with loss free TIFF G4.

Retrieval Interface

High-speed network transfer allows quick and easy retrieval, especially for the display of pixel images, via the WWW. Digitized key-pages may be requested and displayed at the clients' desktop in a few seconds. The display function allows a variety of image operations, e.g. selecting a portion of image for display, scaling of images, etc. Full text retrieval allows easy and comfortable on-line search.

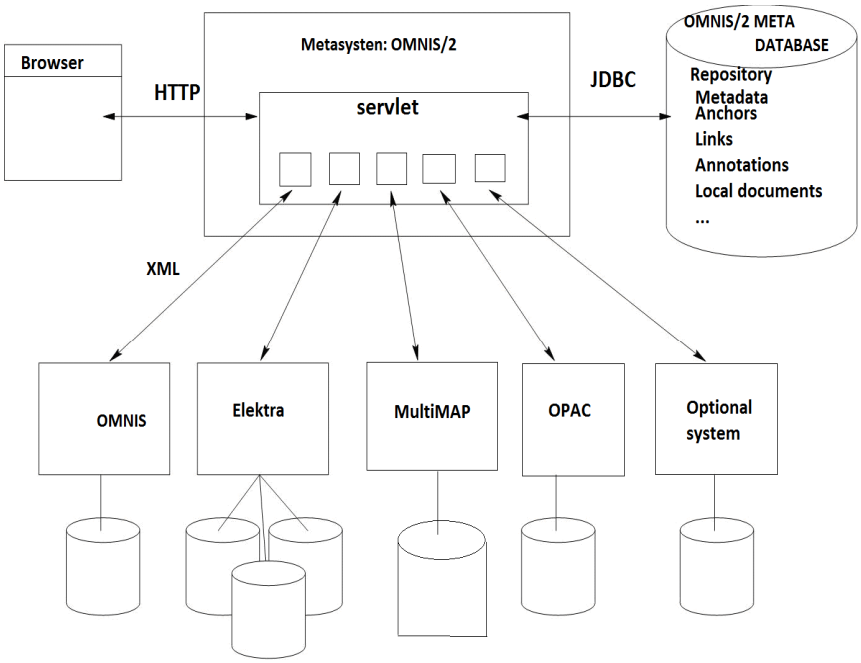
OMNIS System Architecture

OMNIS/2 is an integration of the digital library system OMNIS and the multimedia database system MultiMAP. The goal is to create a stand-alone, interactive digital multimedia library system. The full text retrieval capabilities of OMNIS and the storage capabilities of multimedia documents in MultiMAP including parts of the database scheme are incorporated into the OMNIS/2 system, which enables the user to interactively create, store and search for multimedia documents in digital libraries.

The architecture of OMNIS/2 is shown in Figure1.3. The system is modeled as a three-tier architecture where the databases are separated from the web server in a layer of its own. There is no difference in the handling of local documents and the handling of results from connected external systems. This enables OMNIS/2 [5] to search various other systems and to automatically link all documents, to annotate them, to extend them with multimedia components and to personalize them. The original documents themselves remain in the original database systems and are never modified. They are

represented in the OMNIS/2 database simply by their address and Meta data. The linking, including the anchor positions, is stored in OMNIS/2 exclusively and is included dynamically into the retrieved documents at run-time. In the same way documents can be annotated with user-related, group-related or general annotations. To create user-defined \multimedia documents or to enhance existing ones, OMNIS/2 is equipped with an easy to use authoring tool. The ability to integrate various other systems gives OMNIS/2 the characteristics of a Meta system. It is also possible to look at OMNIS/2 as a stand alone system since it offers features to create, store and search for its own multimedia documents.

Figure 1.3 Architecture of OMNIS/2



Conclusion

The digital libraries provide many advantages to the information infrastructure. But there are still many issues to be addressed, such as migration, intellectual property rights, etc. To ensure the longevity of digital collections and to save them for the future, continual maintenance will be required, i.e. integration of new media, new formats, and migration of data and so on. It is not possible to create a user defined link in any document of a digital library system to another document in another digital library although the source document and also it is not possible for users to work with the libraries as they are retrieval systems only. The OMNIS/2 system enables a user to make use of various information sources, i.e. digital library systems, and which combines them into one virtual personal digital library.

Reference

1. BayerR., The digital library system OMNIS /Myriad, Proc.18thAustralasianComputer Science Conference (ACSC'95), Glenelg, South Australia, Feb. 1995.
2. Baldonado M., Chang C. K., Gravano L., Paepcke A., The Stanford Digital Library Metadata Architecture, Int. Journal on Digital Libraries, 1(2), 1997, pp. 108-121.
3. Daniel,R.,Lagoze, C,Payette,S.D.,A Metadata Architecture for Digital Libraries, Proc. of ADL'98,

Santa Barbara, CA, IEEE Computer Society, 1998, pp. 276-288.
4. DorrM., HaddoutiH., Wiesener S., The German National Bibliography 1601-1700: Digital Images in a Cooperative Cataloging Project, Proc. of ADL'97, Washington DC, IEEE Computer Society, 1997, pp. 50-55.
5. Endres A., Fuhr N., The MeDoc Digital Library Operates as a Network of Distributed Servers, Comm. of the ACM, 41(4), 1998.
6. Federated Repositories of Scientific Literature, University of Illinois at UrbanaChampaign, http://dli.grainger.uiuc.edu
7. Frew J. et al., The Alexandri a Digital Library Architecture, Proc. of ECDL' 98, Crete, Greece, Springer Verlag, Sept. 1998, pp. 61-73.
8. Halasz F., Schwartz M., The Dexter Hypertext Reference Model, Comm. of the ACM, 37(2), Feb. 1994, pp. 30-39.

மின்பதிப்பு - பல்லூடக தகவல் தரவுக் களஞ்சியம், ஓலைச்சுவடிகளின் பேரட்டவணை உருவாக்கல்

சுபாஷினி டிரெம்மல்
துணைத் தலைவர், தமிழ் மரபு அறக்கட்டளை (http://www.tamilheritage.org)
Technical Consultant, Hewlett Packard Germany.
Email: ksubashini@gmail.com

இலத்திரன் வடிவில் ஒரு மொழியின் தொண்மங்களை (intellectual property) மின்னாக்கம் செய்ய முனையும் போது தொழில் நுட்ப அடிப்படைகளைக் கருத்தில் கொள்ள வேண்டிய அவசியம் உள்ளது. இணையத்தில் பதிப்பிக்கப்படும் மின்னூல்களின், மின்பதிப்பாக்கம் எப்படி இருக்க வேண்டும், அதன் கோட்பாடுகள், தொழில் நுட்பத் தரம், பதிப்பிக்கும் முறை, மின்னூல்களின் மின்பதிப்புக்களைச் சேமிக்கும் முறைகள் என பல்வேறு விஷயங்களை மின் பதிப்பு செய்ய விரும்புபவர்கள் எதிர்நோக்க வேண்டியுள்ளது. தமிழ் மரபு அறக்கட்டளை என்னும் தன்னார்வ தொண்டுழிய நிறுவனம் 2001ம் ஆண்டு தொடங்கப்பட்டது. ஓலைச்சுவடிகள் மற்றும் மறு பதிப்பு காணாத பழம் நூல்களின் மின்பதிப்பு நடவடிக்கைகளை மையமாகக் கொண்டு தொடங்கப்பட்ட இந்த முயற்சி பின்னர் படிப்படியாக வளர்ந்து ஓலைச் சுவடி மட்டுமல்லாது, கல்வெட்டுக்கள், வாய்மொழி இலக்கியங்களின் மின் சேகரிப்பு, வரலாற்றுச் செய்திகளின் தொகுப்பு, மின் செய்திகள் தொகுப்பு என விரிவடைந்துள்ளது.

சமீபத்தைய கணினி சார் தொழில்நுட்ப வளர்ச்சி தந்திருக்கும் வாய்ப்புகள்,,வழக்கில் குறைந்து வரும் தமிழ் மரபுச் செல்வங்களைப் பாதுகாப்பதோடு மட்டுமல்லாமல், அவற்றை இலகுவாகப் பகிர்ந்துகொள்ளவும் வழிவகுக்கிறது. இந்தக் கணினித் தொழில்நுட்பங்கள் தமிழ் மரபுச் செல்வங்களை, ஒலி, ஒளி, எழுத்து வடிவம் என பல்வேறு வழிகளில் அவற்றை இலக்கப்பதிவாக்க உதவுகின்றன. அத்தோடு மட்டுமல்லாமல் அவற்றை நாம் இன்றுள்ள இணைய வசதிகள் துணை கொண்டு இலகுவாக உலகின் பல மூலைகளில் உள்ள தமிழ் ஆர்வலர்களோடு பகிர்ந்துகொள்ளவும் வாய்ப்பளிக்கிறது. தமிழ் மரபு அறக்கட்டளை முக்கியமாக இப்பணியில் ஈடுபட்டு வருவதோடு தமிழின் மரபைப் பாதுகாக்கும் ஆர்வலர்களை இணைக்கும் பாலமாகவும் அமைந்துள்ளது.

கடந்த சில ஆண்டுகளில் உலகெங்கிலும் உள்ள தமிழ் ஆர்வலர்கள் மத்தியில் தமிழ் பழம் நூல்கள் மற்றும் ஓலைச்சுவடிகளைப் பாதுகாக்க வேண்டும் என்ற விழிப்புணர்ச்சி ஏற்படுத்தி வந்துள்ளது தமிழ் மரபு அறக்கட்டளை. அதே வேளை, மின்னாக்கப் பணிகளில் ஈடுபட விரும்புபவர்களின் தேவைகளுக்காக மின்னாக்கம் தொடர்பான கலந்துரையாடல்கள், செய்திப் பகிர்வுகள், பேட்டிகள், என்ற ரீதியில் தொழில் நுட்ப விஷயங்களிலும் தமிழ் மரபு அறக்கட்டளைத் தொடர்ந்து ஈடுபட்டு வருகின்றது.

தமிழ் நூல்கள் மின்பதிப்பாக்கம்

பல அரிய தமிழ் நூல்கள் ஒரு முறை சுவடி நூல்களிலிருந்து அச்சுப் பதிப்பாகப் பதிப்பிக்கப்பட்ட பின்னர் மறுபதிப்பிற்கு வருவதில்லை. இப்படி பல நூல்கள் நாள் செல்லச் செல்ல மறக்கப்பட்ட ஒன்றாகிப் போவதோடு தாட்கள் கிழிந்து, மக்கிப் போய் அழிந்தும் விடுகின்றன. இவை மின்னாக்கம் செய்யப்படும் போது இவ்வகையில் அழிவதிலிருந்து பாதுகாக்கப்படுவதோடு நமது பயன்பாட்டிற்கும் கிடைக்கின்றது. தமிழ் மரபு அறக்கட்டளை இவ்வகை நூல்களை அதன் அசல் வடிவம் மாறாத வகையில் வருடி(scanner) மற்றும் புகைப்படக் கருவிகளைப் பயன்படுத்தி மின்பதிவுகளை உருவாக்கி வருகின்றது. அவ்வகையில் 19ம் நூற்றாண்டு நூல்கள், 20ம் நூற்றாண்டின் ஆரம்ப கால நூல்கள் என குறிப்பிடத்தக்க பல நூல்கள் மின்பதிப்பாக்கம் செய்யப்பட்டு நமது வலைப்பக்கத்தில் சேர்க்கப்பட்டுள்ளன.

ஓலைச் சுவடிகள் மின்பதிப்பாக்கம்

தமிழ் அச்சுப் பதிப்புக்கள் தோன்றுவதற்கு முன்னர் பனை ஓலைகளில் நூல்கள் பதிப்பிக்கப்பட்டன. இவ்வகை சுவடி நூல்கள் அனைத்தையும் அச்சு வடிவில் பதிப்பிக்க பல தமிழறிஞர்கள் கடந்த இரண்டு நூற்றாண்டுகளில் பெருமளவில் தொடர்ந்து முயன்று அதில் ஓரளவு வெற்றியும் கண்டுள்ளனர். ஆனால் இதனை வைத்து அனைத்து தமிழ் சுவடி நூல்களும் அச்சு வடிவம் பெற்று வெளி வந்துள்ளன என்று கூறி விட முடியாது. ஆக மின்பதிப்பாக்க முயற்சிகள் தொடர்ந்து மேற்கொள்ளப்படும் போது அழியக் கூடிய நிலையிலுள்ள பல அறிய நூல்களை நாம் பாதுகாத்து நமது அடுத்த சந்ததியினருக்கு அதனை வழங்குவதற்கான சாத்தியங்கள் உள்ளன.

தமிழகம் மற்றும் அயல்நாடுகளில் உள்ள பல்வேறு நூலகங்களில் ஓலைச்சுவடிகள் பாதுகாக்கப்பட்டு படிப்படியாக நூல் வடிவம் பெற்று வந்தாலும் அனைத்து ஓலைச் சுவடிகளும் முழுமையாக அச்சு வடிவம் பெறாத நிலையே உள்ளது. ஆக சுற்றுச் சூழல், சுகாதார பிரச்சனைகள், முறையான சுவடி நூல் பாதுகாப்பு பற்றிய தொழில் நுட்ப திறன் இல்லாமை போன்ற காரணங்களினால் தொடர்ந்து பல சுவடி நூல்கள் அதன் சுவடு தெரியாமல் அழிந்துள்ளன. இதனைப் போக்க இவ்விஷயம் தீவிர கவனத்திற்குட்படுத்தப்பட்டு சுவடி நூல்கள் அச்சுப் பதிப்பாக்கமும் மின்பதிப்பாக்கமும் பெற வேண்டிய அவசியம் நிச்சயமாக உள்ளது.

தமிழகத்திற்கு அயல்நாட்டவரின் வருகையால் ஏற்பட்ட பல்வேறு தாக்கங்களில் பனை ஓலைச் சுவடிகளிலிருந்து நூல்களைத் தற்கால தமிழுக்குக் காகித வடிவில் அச்சுப் பதிப்பாக மாற்றம் செய்ய ஏற்பட்ட முயற்சிகளும் அடங்கும். தமிழ் நாட்டில் அச்சில் வெளிவந்த முதல் தமிழ் நூல் திருக்குறள் என்பதும் 1835ம் ஆண்டு வரை நூல்கள் அச்சிடுவதற்கு அரசின் தடை இருந்ததும், 1835க்குப் பின்னர் இந்தத் தடை நீக்கப்பட்டதும் குறிப்பிடத்தக்க விஷயங்கள். அரசாங்கத்தின் இந்தத் தடை இருந்த போதிலும் தமிழரறிஞர்கள் பலரது தீவிர உழைப்பின் காரணமாக பல சுவடி நூல்கள் அச்சில் வெளிவந்தன. (**சுவடிப்பதிப்பியல்**, *டாக்டர்.வே.மாதவன்*)

இதில் குறிப்பிடத்தக்கவையாக திருநெல்வேலி அம்பலவாண கவிராயரின் முயற்சியில் 1812ல் முதன் முதலில் அச்சில் வெளிவந்த திருக்குறள் மூலபாடம், நாலடியார் மூலபாடம் ஆகியவற்றினைக் கூறலாம். இப்படி சுவடி நூல்களைப் பனை ஓலையிலிருந்து அச்சு வடிவத்திற்கு கொண்டு வர உதவியவர்களில், குறிப்பிடத்தக்கவர்கள், தணிகைமணியார், டி.கே.சிதம்பரநாத முதலியார், பேராசிரியர் உ.வே.சாமிநாதையர், திருவாரூர் வி.கல்யாணசுந்தரனார், மகாவித்துவான் மு.இராகவையங்கார், யாழ்ப்பாணம் வை.தாமோதரம் பிள்ளை, திரிகோணமலை த.கனகசுந்தரம் பிள்ளை, யாழ்ப்பாணம் சபாபதி நாவலர், யாழ்ப்பாணம் அம்பலவாண பண்டிதர், தாண்டவராய முதலியார், புதுவை நயனப்ப முதலியார், அ.முத்துசாமிப்பிள்ளை, யாழ்ப்பாண நல்லூர் ஆறுமுக நாவலர், மகாவித்துவான் மீனாட்சி சுந்தரம் பிள்ளை, வடலூர் இராமலிங்க அடிகள், தாண்டவராய முதலியார்,, பேராசிரியர் எஸ் வையாபுரிப்பிள்ளை போன்றோர்.

இவர்கள் மட்டுமன்றி மேலும் பல பதிப்பாசிரியர்களின் துணையினால் தான் கடந்த நூற்றாண்டுகளில் எழுதப்பட்ட நூல்களில் சில இன்றளவும் நமக்கு வாசிக்கக் கிடைக்கின்றன. பதிப்பாசிரியர்களின் விபரங்கள் அடங்கிய பட்டியலை த.ம.அறக்கட்டளையின் கீழ்க்காணும் வலைப்பக்கத்தில் காணலாம். (http://www.tamilheritage.org/manulogy/pubs/pubs.html)

தமிழ் மரபு அறக்கட்டளை தனியார் பாதுகாப்பிலுள்ள, இதுவரை அச்சில் வெளிவராத பனை ஒலை நூல்களை மின்பதிப்பு செய்து அவற்றைப் பாதுகாக்க வேண்டுமென்பதில் மிகுந்த அக்கறை கொண்டுள்ளது. அதன் அடிப்படையில் ஓலைச் சுவடி மின்பதிப்பாக்க முறை என்பது ஐந்து படி நிலைகளில் பிரிக்கப்பட்டு செயல்முறையாக்கம் பெற வேண்டும் என்பத பரிந்துரைக்கின்றது.

1. ஓலைச் சுவடிகளைப் பெறுவது மற்றும் சேகரிப்பது.
2. சுவடி நூல்களை இனம் பிரிப்பது. மருத்துவம், ஸ்தல புராணம், கதைகள், செய்திகள், ஆவணங்கள், தொழில் நுட்பம், ஓவியம், சமய நூல்கள், சங்க இலக்கியங்கள் என வகை பிரிப்பது.
3. மூன்றாவது நிலையாக சுவடி நூல்களை மின்பதிப்பு செய்வது. வருடியால் மின்பதிப்பாக்கம் செய்து தகுந்த கணினி தொழில் நுட்பத்தின் அடிப்படையில் நூல்களின் பக்கங்களைக் கோப்புக்களாகச் சேகரித்து மின்னூலாக பிரசுரிப்பது.
4. இந்த மின்பதிப்பு செய்யப்பட்ட ஓலைகளை, ஓலைச் சுவடி நூல்களை வாசிக்கத் தெரிந்த அறிஞர்களைக் கொண்டு வாசிக்க வைத்து, அதனை ஒலிப்பதிவு செய்வது, மற்றும் எழுத்துக்களை அதாவது முழு நூலையும் தற்கால தமிழுக்கு மாற்றம் செய்வது.
5. இந்த ஒலிப் பதிவுகளையும், தட்டச்சு செய்யப்பட்ட நூலையும் மின்பதிப்பாக்கம் செய்து நிரந்தரப்படுத்துவது, வாசிப்புக்குட்படுத்துவது.

அச்சு வடிவ பழம் நூல்களின் மின்பதிப்பு

பனை ஓலை நூல்கள் அச்சுப்பதிப்புக்கு வந்தாலும் முழுமையாக அந்நூல் பாதுகாக்கப்பட்டு விட்டது என்று கூறுவதற்கில்லை. ஓலையிலிருந்து அச்சு வடிவத்தில் பதிப்பு கண்ட பல நூல்கள் மறு பதிப்புக் காணாத நிலையும் உள்ளது. பல நூல்கள் பாதுகாப்பு இன்மையால் தாட்கள் மடிந்து, இக்கிப் போய் நொறுங்கி உடைந்து விடும் நிலையும் உள்ளது. ஆக ஒரு நூல் அச்சுப் பதிப்பினை அடைந்தாலும் அது மின்பதிப்பாக்க முறையில் பதிப்பிக்கப்படும் போது அதனைத் தற்கால தொழில் நுட்பத்தைக் கொண்டு எளிய முறையில் பாதுகாக்க முடிகின்றது. ஆகவே அச்சு வடிவ நூல்களையும் பாதுகாக்க வேண்டிய நிலை இருப்பதால் மின்பதிப்பாக்கம் இதற்கு ஒரு சிறந்த வழியாக அமைகின்றது.

வாய்மொழி இலக்கியங்கள், கலைகள், மின்பதிப்பு

தமிழர் கலை கலாச்சார பண்பாட்டு வாழ்க்கை முறை சம்பந்தப்பட்ட பல்வேறு தகவல்கள் இன்றளவும் கூட முழுமையாக பதிப்பிக்கப்படவில்லை என்பதை மறுக்கமுடியாது. கிராமங்களில் முன்னர் வழக்கிலிருந்த விஷயங்கள் பல படிப்படியாக, நகர்ப்புற வாழ்க்கை மற்றும் வாழ்க்கை முறையில் ஏற்பட்டுள்ள மாற்றங்களின் காரணத்தால் கொஞ்சம் கொஞ்சமாக வழக்கு ஒழிந்து வருகின்றன. இது முற்றிலும் தடுக்கப்பட முடியாத ஒன்று. ஆனால் இந்த வாய்மொழி இலக்கியங்களை நமது பண்பாட்டின் வரலாற்று ஆவணங்களாகப் பதிப்பிக்க கூடிய தொழில் நுட்பம் இன்று கிடைத்துள்ள வேளையில் இதனை நாம் இத்தொழில் நுட்பம் கொண்டு பாதுகாக்க முடியும்.

உதாரணமாக வாய்மொழி இலக்கியங்களான தாலாட்டுப் பாட்டு, ஒப்பாரிப் பாட்டு, கும்மி, அம்மானை போன்றவை, மற்றும் கிராமத்து சட்ட திட்டங்கள், பஞ்சாயத்து, பாரம்பரிய உணவுகள், பாட்டி வைத்தியம், மூலிகைகளின் சிறப்புக்கள், பாண்டி, சில்லாட்டம், அம்மானை, பள்ளாங்குளி போன்ற கிராமிய விளையாட்டுகள், நாடகக் கூத்துக் கலைகளான கரகாட்டம், ஒயிலாட்டம், பறையாட்டம், சிலா ஆட்டம் போன்றவை அயல் நாடுகளுக்குக் குடி பெயர்ந்து விட்ட தமிழர்களுக்கு மட்டுமின்றி தமிழகத்திலேயே கூட வழக்கில் இல்லாத கலாச்சார பிம்பங்களாக உருவாகி உள்ளன. இவ்வகை வாய்மொழி இலக்கியங்களை, அதன் நுணுக்கங்களை விளக்கும் தகவல்களை ஒலிப்பதிவுகளாகச் சேகரித்து வைக்க வேண்டிய அவசியம் உள்ளது. இதனை கடுத்தில் கொண்டு, செய்தித் தகவல்கள், ஒருவரது பல்வேறு காலகட்டங்களில் நிகழ்ந்த வரலாற்று விஷயங்களின் தகவல்கள், தனிப்பட்ட சிந்தனையின் வெளிப்பாடுகள் கலை வடிவங்கள் போன்றவற்றை வாய்மொழி பதிவுகளின் தொகுப்பாக (Oral history archive) உருவாக்குவதில் தமிழ் மரபு அறக்கட்டளைத் தொடர்ந்து ஈடுபட்டு வருவதோடு இது தொடர்பான தொழில்நுட்ப விஷயங்களை அலசுவதிலும் மின்தமிழ் மற்றும் இ-சுவடி மின்னரங்கங்களின் வழியாக கலந்திரையாடல்களையும் நிகழ்த்தி வருகின்றது.

ஒரு இனத்தின் முக்கிய அம்சங்களாகத் திகழ்பவை அந்த இனத்தின் பண்பாட்டு விழுமியங்கள். வரலாற்று விஷயங்கள் முறையாக பதிவு செய்யப்பட்டு, தகுந்த முறையில் தொகுத்து பராமரித்து வைக்கப்படும் போதே அந்த இனத்தின் சிறப்புகள் பாதுகாக்கப்படுகின்றன. ஆக, தமிழ் இனத்தின் சிறப்புகளை, சிந்தனை ஆழத்தைப், பண்பாட்டுக் கூறுகளை, கலைகளின் பெருமிதத்தை நாம் சுலபமாகக் கருதி விட்டு விடாமல், அதன் சுவடுகள் மறக்கப்படும் முன்னர் அவற்றைத் தக்க முறையில் பாதுகாத்து வைக்க வேண்டும் என்பது தமிழ் மரபு அறக்கட்டளையின் முக்கிய நோக்கங்களில் ஒன்று. இந்தக் கலைகள் வருங்காலங்களில் வழக்கில் இல்லாமல் போனாலும் இவை இருந்ததற்கான அடையாளங்களாவது இவ்வகை மின்பதிப்புகளின் வழி பாதுகாக்கப்படுவதற்கான சாத்தியங்கள் உண்டு.

மின்பதிப்பாக்க வழி முறைகள், தொழில் நுட்பக் கூறுகள்

எந்த ஒரு மின்னூல் மின்பதிப்பாக்க நடவடிக்கைக்கும் அடிப்படையாக அமைவது அதன் தொயழில்நுட்ப அடிப்படைக் கூறுகள். மின்பதிப்பு செய்யப்படும் ஒவ்வொரு பக்கமும் துல்லியமாகவும் தெளிவாகவும் வாசிப்புக்கு உகந்ததாகவும் இருக்க வேண்டியது மின்பதிப்பாக்கத்தின் அடிப்படை தேவைகளில் மிக முக்கியமான ஒன்று. மின்னூல்கள் பதிப்பாக்கம் என்பது தற்சமயம் வருடி(scanner) பயன்படுத்தி மின்பதிப்பாக்கம் செய்வது அல்லது புகைப்படக் கருவி பயன்படுத்தி மின்பதிப்பு செய்வது என்ற இரண்டு நிலைகளில் உருவாக்கப்படக் கூடிய சாத்தியம் உள்ளது.

ஒரு நூல் வருடியில் வைத்து பக்கங்கள் திருப்பப்பட்டு மின்பதிப்பு செய்யும் போது சில வேலைகளில் பக்கங்கள் உடைந்து விடும் நிலை உண்டு. பழம் நூல்கள் மின்பதிப்பு செய்யும் போது நூல்களின் பக்கங்கள் உடைந்து விடுவதிலிருந்து தடுக்க புகைப்படக் கருவி பயன்படுத்தி மின்பதிப்பு செய்வது தற்போது வழக்கத்தில் உள்ள சிறந்த முறை என்று கூறலாம். பனை ஓலைச் சுவடி மின்பதிப்பிற்கு வருடி பயன்படுத்தி மின்பதிப்பு செய்வது எளிமையானதாகவும் அதே சமயம் உகந்ததாகவும் அமைகின்றது.

பக்கங்களின் தன்மைகள்

மின் நூல்கள். மின்பதிப்பு செய்யபடும் போது பொதுவாக ஒரு மின்பதிப்பு வடிவம் jpg, tiff, png, gif வடிவங்களில் உருவாக்கப்பட்டே சேமிக்கப்படுகின்றன. உதாரணமாக வருடி பயன்படுத்தி உருவாக்கப்படும் மின் நூல்களின் பக்கங்கள் பெரும்பாலும் பட வடிவ கோப்புக்களுக்குப் புகழ் பெற்ற jpg வடிவத்தில் தான் உருவாக்கப்படுகின்றன. jpg முறை புகைப்படங்களில் அதிலும் தத்ரூபமான படங்களில் சிறிய மாறுபாடுகளையும் உள்வாங்கி அதன் அசல் வடிவத்தை மின்பதிப்பு வடிவங்களில் கொண்டு வருவதில் சிறந்த தன்மையைக் கொண்டது. புகைப்படக் கருவிகளில் எடுக்கப்படும் மின்பதிவுகள் இந்த compression வடிவத்தை பெரும்பாலும் அடிப்படையாகக் கொண்டவை.

அடிப்படையில் jpg பதிவுகள் மின்பதிவின் போது தகவல்களை இழக்கக் கூடிய தன்மை கொண்டவை (lossy compression method). இதனால் jpg வடிவ compression மிகத்துல்லியமான பதிவுகளுக்கு, உதாரணமாக வரைபடங்கள், சுவடிகளில் உள்ள வரி வடிவங்கள் போன்றவற்றிற்கு உகந்தவையல்ல. மின்னாக்கம் செய்யப்படுகின்ற பல பழம் நூல்கள் பல வேளைகளில் மிகச் சிதைந்து சில பகுதிகளில் எழுத்துக்கள் தெளிவில்லாத நிலையிலும் கூட மின்பதிப்பு செய்ய வேண்டிய நிலை உள்ளது. ஆக இவ்வகை நூல்களைக் கருத்தில் கொண்டால் வரி வடிவ மின்பதிப்புக்கு உகந்த compression முறையைப் பயன்படுத்த வேண்டிய அவசியம் உள்ளது. tiff மற்றும் png முறைகள் இவ்வகை தேவைகளுக்கு மிக உகந்தவை. பொதுவாகவே நூல்கள் மின்பதிப்பாக்கம் என்பது வரிவடிவ அசல் பக்கத்தின் ஒரு மறுபதிப்பு. இதற்கு மிகத் தகுந்த முறை tiff compression ஆகும். அதிலும் குறிப்பாக OCR (optical character recognition) கொண்டு இந்த நூல்கள் பின்னர் வாசிப்பில் உட்படுத்த tiff compression பயன்படுத்தி சேமிக்கப்பட்ட கோப்புக்களே சிறந்தவையாக அமைகின்றன.

மின்பதிப்பாக்கத்தில் சந்திக்கக்கூடிய தொழில்நுட்ப பிரச்சனைகளைக் கையாள பொதுவான கையேடுகள் உருவாக்கப்பட வேண்டிய அவசியம் உள்ளது. கணினி கொள்ளளவு விலை அதிகம்

இல்லையென்ற நிலை இப்போது இருந்தாலும் மின்னூல்கள் தரவிறக்கத்திற்காக தேவைப்படும் நேரம் அதிகமாக இருப்பது ஒரு வகை பிரச்சனையாக உள்ளது. இதனை மேலும் செம்மைப் படுத்த தகுந்த தொழில்நுட்ப கையேடுகள் இருப்பது அவசியமாகின்றது. அதே போல ஒலி ஒளி வடிவ பதிவுகளை இணையத்திலிருந்து துரிதமாக தரவிறக்கம் செய்வதற்கும் கோப்புக்களின் அளவில் கவனம் செலுத்த வேண்டிய நிலை உள்ளது. இதனைக் கையாள, கையேடுகள் உருவாக்கப்படுதல் அவசியமாகின்றது. இதற்கு pdf மற்றும் djVu தொழிநுட்பங்கள் சிறந்தவையாக இருக்கின்றன.

மின்பதிப்பு செய்ய விரும்புபவர்களுக்குப் பெரும்பாலும் தொழில்நுட்ப அடிப்படை விஷயங்களில் பல்வேறு ஐயங்கள் இருப்பது இயல்பு. இதனை நிவர்த்திக்க எளிய கையேடுகளைத் தமிழ் மரபு அறக்கட்டளை உருவாக்கி வெளியிட்டுள்ளது. அந்த வகையில் இதுவரை அடிப்படை மின்பதிப்பாக்கம், FTP செய்வது எப்படி எனும் கையேடு மற்றும் ஒலிப்பதிவு கட்டுரை, மின் நூல்களைப் புகைப்படக் கருவி கொண்டு உருவாக்கும் விதம், ஒலிப்பதிவுகள் மற்றும் குரல் அரட்டை ஒலிப்பதிவு விளக்கக் கட்டுரைகள், தமிழில் மின்னஞ்சல் செய்யும் முறை அடிப்படை கையேடு, மின்னூலாக்கக் கையேடு போன்றவை வெளியிடப்பட்டு மின் நூலாக்கத்திலும் மின்பதிவுகளிலும் ஆர்வமுள்ளோர் பயன்படுத்தும் வகையில் இலவசமாக வழங்கப்பட்டுள்ளன.

தமிழ் மரபு அறக்கட்டளையின் அடுத்த கட்ட முயற்சிகள்.

ஒருங்கிணைக்கப்பட்ட மின்னூல்களுக்கான அட்டவணை மற்றும் ஓலைச் சுவடிகளுக்கான பேரட்டவணை

மின் நூல்களை வெளியிடும் தன்னார்வக் குழுவினர் பலர் தற்போது தமிழ் மின்னூல்களை உருவாக்குவதில் ஈடுபாட்டுடன் இயங்கி வருகின்றனர். பல்வேறு குழுவினர் 'தமிழ் நூல்கள் மின்பதிப்பாக்கம்' என செயல்படத் தொடங்கும் போது ஒருவர் மின்பதிப்பு செய்த அதே நூலை வேறொருவர் செய்வதற்கான நிலை ஏற்பட வாய்ப்புண்டு. ஆக தமிழ் மரபு அறக்கட்டளையின் வெளியீடுகள் குறித்த ஒரு தகவல் வங்கி உருவாக்கப்பட வேண்டியது அவசியம் என்பதால் இந்த முயற்சி தொடங்கப்பட்டு தேடு இயந்திரம் மற்றும் வரிசைப்படுத்தும் சாத்தியத்துடன் கூடிய பட்டியல் ஒன்றும் உருவாக்கப்பட்டுள்ளது. அதே வேளை இணையத்தில் உள்ள அனைத்து மின்னூல்களுக்குமான ஒருங்கிணைக்கப்பட்ட ஒரு அட்டவணை உருவாக்கப்பட வேண்டிய தேவை இருப்பதால் இவ்வகை முயற்சிகளுக்கும் தமிழ் மரபு அறக்கட்டளை துணை நிற்கின்றது.

பணை ஓலைச் சுவடிகள் தமிழகத்திலும் அயல் நாடுகளிலுமுள்ள பல்வேறு நூலகங்களிலும் உள்ளன. இவற்றைப் வாசிப்புக்கு உட்படுத்தும் அட்டவணை தயாரிக்கப்பட வேண்டிய அவசியம் உள்ளது. தொடர்ந்து பல நிறுவனங்கள் சுவடிகளின் தகவல்களைச் சேகரிக்கும் முயற்சிகளில் இயங்கி வந்தாலும் இதுவரை சேகரத்திலுள்ள அனைத்து சுவடிகளுக்குமான முழுமையான ஒரு பட்டியல் இல்லாதது கவனத்தில் கொள்ளப்பட வேண்டிய ஒன்று. அத்தோடு இதுவரை கண்டெடுக்கப்பட்டு தொகுக்கப்பட்டு பராமரிக்கப்பட்டு வரும் ஓலைச்சுவடிகளுக்கான ஒரு பேரட்டவணையும் இணையத்தில் ஒருங்குறியில் தேடுதல் (digital searchable catalogue) வசதியுடன் உருவாக்கப்பட வேண்டிய அவசியம் உள்ளது. இவ்வகையிலான முயற்சிகளுக்கான ஆரம்பப்பணிகளை தமிழ் மரபு அறக்கட்டளை இப்போது தொடங்கியுள்ளது.

Tamil Inscriptions and on line search

Database Compilation, Grammatical analysis, Lexicon and Translation

Appasamy Murugaiyan
Ecole Pratique des Hautes Etudes –Section des Sciences historiques et philologiques, Paris.
a.murugaiyan@wanadoo.fr

Abstract: This is a description of a database tool for the analysis of a small linguistic corpus. It aims at representing syntactic and semantic information at both clause and argument levels in their sociolinguistic components. This paper summarizes the categories that are encoded, and describes how they are encoded. Few examples to show their usage are also provided here as an illustration.

Introduction

The use of large computerized bodies of text for linguistic analysis has emerged in recent decades as one of the most significant fields in the study of language. This project is the result of our experience for more than a decade of teaching and conducting research in Tamil epigraphy. The major aim of this project is to make the Tamil inscription database accessible to people and researchers in a variety of fields (e.g. linguistics, anthropology, sociology, archaeology, folklore, history, art history, religion etc.), both native and non-native speakers of the Tamil language. A corpus based linguistic analysis lays emphasis on the importance of empirical data. Empirical data alone would enable researchers to make objective statements, rather than those that are subjective based upon one's own perception of language and society.

The structure of the epigraphic text is complex and the order of constituents is motivated by pragmatics (Information structure) rather than by syntax. Tamil inscriptions consist of a large number of technical vocabularies, some are native Tamil lexemes, some are coined in Tamil, some are loan words from Indo-Aryan, and yet some others are compounds of both Tamil and Indo-Aryan lexemes. All of these features of epigraphic texts constitute a major challenge for any body working with Tamil inscriptions. In order to overcome these complexities, we felt the need for a specific tool –grammatical and lexical- for reading and analyzing Tamil inscriptional texts.

Aim

The aim of this project includes: (1) to archive Tamil epigraphic texts, (2) to present the data in a queryable format in order to extract linguistic information, (3) to compile a dictionary with an electronic interface, based on this queryable format, (4) to write a historical grammar of the Tamil language based on the database and 5) to contribute to the field of Dravidian historical linguistics.

Linguistic analysis, as broadly conceived, includes not only phonological, morphological, and grammatical information, but also includes sociolinguistic components as well. It aims above all to examine the usage and change of language in its real context. This linguistic database includes several features: phonological (texts in transliteration based on the Madras University Tamil lexicon), different scripts (vaṭṭeḻuttu, grantha, and tamiḻ), morphosyntactic features (derivational suffixation,

cliticization, development of postpositions, process of grammaticalization etc.,), and etymological features (for the analysis of patronyms and toponyms).

Small Linguistic Corpus

A corpus of written texts can be analysed for different purposes including linguistic, lexicographic, rhetoric and stylistics for instance. Our approach is linguistic and seeks to focus the analysis of our corpus in the social and cultural contexts in which these inscriptions were written and read. This project at this stage is concerned about a small linguistic corpus as opposed to any corpora containing several millions of words (for example, The British National Corpus -10 million words, CIIL Tamil database around 3 Million words and the Cre-A data bank www.crea.in contains 2.5 million words and will be the largest Tamil data bank very shortly). Except these two databases, Tamil does not have any other database to our knowledge.

The total number of Tamil inscriptions discovered to date counts around thirty thousand. Such a large number of inscriptional texts cannot be handled in a single attempt at a stretch mainly because of lack of technical and human resources. The present database "Kalveṭṭu" is conceived in several phases in chronological order. In the first phase, the database includes texts from 0450 A.D. to 0799 A.D. The Tamil-Brāhmī inscriptions are not included in the present project. The first phase of the database comprises mainly Pallava inscriptions, Copper plates and Hero Stone inscriptions (naṭu kal). In our database, incomplete or defective inscriptions and clauses (phrases) are not included. In case of some doubtful inscriptions, we took maximum precaution to check with, whenever possible, original estampage and with inscriptions insitu. We thus have avoided all errors as much as possible.

Tamil Inscriptions

Since the last few decades, the importance of Tamil inscriptions as source material has been recognized, at least to some extent, by researchers in humanities and social sciences. But comprehension, interpretation and information retrieval of Tamil epigraphic texts, no doubt, constitute a major challenge. It is evident from our own experience that a better reading and unambigous interpretation of Tamil inscriptions is possible only through a multidisciplinary approach. Rather, Tamil inscriptions provide source material for constituting a monumental multi disciplinary database.

Language structure

There is a striking difference (in word order) between Modern Tamil and inscriptional Tamil (IT). These deviations offer valuable clues on the historical changes that the grammar of Tamil went through. The purpose of this paper is to bring out the characteristic features of the underlying linguistic system of the inscriptional Tamil. The IT is characterised by an alternative coding system which has resulted from distribution of grammatical patterns motivated by semantic and pragmatic (information structure) parameters

The field of historical syntax can be divided into two parts: the study of the grammars of languages of the past and the study of changes in grammar as attested in the historical record. The first subfield is best considered as a branch of comparative syntax which tries to reconstruct, through textual evidence, the grammars of languages that lack living native speakers. The second subfield studies the problem of the diachronic instability of syntax and the transition between grammars. For this reason, we have chosen to focus on the diachronic aspect of historical syntax in our database.

Scripts

Four different scripts were used in Tamil inscriptions: tamiḻ-brāhmi, vaṭṭeḻuttu, tamiḻ and grantha. We are only concerned with the last three scripts besides the historical, social and cultural contexts of their usage. Broadly speaking, use of vaṭṭeḻuttu and tamiḻ scripts can be explained by historical and political factors. But the use of grantha script implies complex sociolinguistic factors (language contact, bilingualism, cultural and political hegemony and so).

DATABASE - "Kalveṭṭu"

The construction of our extensive database consists of the following four stages: 1) selection and documentation of epigraphic texts, 2) identification of a meaningful sequence of units (minimal clause structures), 3) linguistic, and lexical analysis of clauses (morphology, syntax, borrowings), and 4) translation. Due to the complex structure of epigraphic texts, we have opted for a methodology based on semantics and pragmatics instead of a traditional –subject-object-predicate- approach. In our approach, the text is divided into minimal meaningful units (minimal linguistic structure or "clauses"). Each such minimal unit is subject to a multilevel analysis: morphology, morphosyntax, syntax and semantics.

The database consists of four major components: 1) general information, which is composed of 21 subunits, 2) Tamil inscription in transliteration, 3) translation and 4) linguistic analysis. The fourth part, which is linguistic analysis, is the essential part of the "Kalveṭṭu database". It consists of a list of "words" (any meaningful unit whether segmentable or not). Each "word" is associated with a canonical form (unsegmentable unit- morpheme or lexical unit) and a set of "grammatical categories" (specifying tense, mood, number, gender or other linguistic functions). In our database, we have listed so far 200 different types and subtypes of "grammatical categories". This list is subject to modification depending on a combination of semantic and morphosyntactic functions. A fine-grained analysis of inscriptions necessitates such a vast sub categorisation of different constituents of the identified minimal meaningful units. This will allow the users to view the detailed grammatical and lexical analysis of each minimal meaningful unit identified in our database.

The database can be used for various purposes and to handle many issues, most of them remain unanswered in the study Tamil inscriptions. Some of them may be listed below. To retrieve information about the type and structure of a particular inscription: Whether it is a hero stone, copper plate or temple inscription; whether the inscription contains invocation and meykkīrtti or not; the meykkīrtti is in Sanskrit or in Tamil; where the inscription is situated in the temple physically– on the wall of the main shrine or on the wall of the secondary shrines and finally to calculate the correlation among all of these information.

The usage of grantha script, as mentioned above, is a complex phenomenon. For each identified item, among other information, the script used is also mentioned (vaṭṭeḻuttu, tamiḻ and grantha). This database may shed light on the social and cultural significance of grantha scripts in Tamil inscriptions: whether grantha was used only for Indo-Aryan loan words; whether all Indo-Aryan loan words were marked systematically in grantha, whether there were variation in the usage of grantha between different dynasties on the one hand and on the other hand whether there were variations from one region to another; whether all personal names (kings, king's consorts, chieftains, Brahmans, royal officers, artisans and cultivators etc.) were marked in grantha and is there any social and cultural significances in this pattern. The same analysis is also possible for toponyms.

This database is created principally as a tool for linguistic analysis of inscriptions. Every searchable (identified) item 'word' in this corpus is tagged by its formal category (e.g. proper noun, place noun, post-position, case marker, auxiliary verb etc.) so that users can list all occurrences of specific searchable (or identified) unit in the database.

For instance, a searchable unit like "paṭṭāṉ" can be retrieved in two distinct ways: 1) a list of all occurrences irrespective of its different grammatical functions as finite verb or as participial noun, and 2) it is also possible to retrieve the unit "paṭṭāṉ" distinctly either as finite verb or as a participial noun.

e.g. ēṛaṉ eṛintu paṭṭāṉ (finite verb) "Eran was dead wounded"

akkantai kōṭaṉ toṛu viṭuvittup paṭṭāṉ kal (participial noun) "Akkandai Kotan is killed while liberating the cow herds and this is his memorial stone".

This database provides for each identified "phrase" the order of constituents. This constituent order pattern reveals without doubt that the {SOV} order considered as basic is only a possible word order among other common word order as noticed in Tamil inscriptions.

It is possible to list all case forms and different case functions among the searchable units. This function allows the users to find the functional and semantic variations or evolution for each case morpheme both diachronically and regionally. It is possible to examine the different case functions substituted by the oblique form (genitive or locative) and their morphosyntactic contexts.

Using this database - which is currently used with a Windows application - one can identify all of the main clauses to determine how each clause is constructed. One can also identify all the alternate structures for each clause identified and study the underlying linguistic structure.

Whenever it is difficult to identify the functional category of any unit, a question marker is used to indicate such problems. The final stage of this project will be creation of user-interface to search this database online over the internet.

பல்லூடகவழிப் பழந்தமிழ் இலக்கியக் கருத்தாடல் இலக்கியம் கற்றல் கற்பித்தல் அடிப்படையில்

முனைவர் வா. மு. சே. முத்துராமலிங்க ஆண்டவர்
தமிழ் இணைப் பேராசிரியர்
மு.வ. முதுகலை உயராய்வு மையம்
தமிழ்த் துறை, பச்சையப்பன் கல்லூரி, சென்னை- 30
Email: sethuandu@yahoo.co.in

அயலக ஆய்வறிஞர்களால் ஏற்கனவே செம்மொழியாக ஏற்புப் பெற்றிருந்த தமிழ், நீண்ட போராட்டங்களுக்குப் பிறகு 2004ஆம் ஆண்டு இந்தியப் பேரரசால் முறையாக அறிவிக்கப்பட்ட இந்திய மொழி தமிழ். இந்தப் பின்னணியில் பழந்தமிழ் இலக்கியங்களைப் பற்றித் தெரிந்துகொள்ள வேண்டும் என்ற ஆர்வம் அயல்நாட்டு, தாய்நாட்டு அறிஞர்களுக்கும் ஆர்வலகளுக்கும் ஏற்பட்டுள்ளது. இன்றைய தகவல் தொழில்நுட்ப வளர்ச்சியினை முழுமையாகப் பயன்படுத்தி எவ்வாறு பழந்தமிழ் இலக்கியங்களை பல்லூடகவழி அறிவு நிலையிலும் அறிந்துகொள்ளும் நிலையிலும் பரப்புவதற்குரிய கணினிவழித் திட்டமிடுவது இவ்வாய்வுக் கட்டுரையின் நோக்கம்.

பல்லூடகவழிப் பழந்தமிழ் இலக்கியங்களைக் கருத்தாடல் வழி எவ்வாறு கற்பது, கற்பிப்பது, அதற்கான வழிமுறைகள் என்ன எப்படிச் செயலாற்றுவது?, இதுவரை கணினி சார்ந்து நிகழ்ந்த இலக்கிய, இலக்கணக் கருத்தாக்கங்களை எவ்வாறு வளர்த்தெடுப்பது?, வளப்படுத்துவது போன்ற செய்திகள் இக்கட்டுரையில் ஆய்வுக்கு உட்படுத்தப்படுகின்றன.

முதல் முயற்சியாகப் பல்லூடகவழி சங்கத் தமிழுக்கான குறுவட்டினை உருவாக்கியுள்ளேன். இக்குறுவட்டில் உரையாடல் அடிப்படையில் எல்லா வயதினரும் புரிந்துகொள்ளும் வகையில் சங்க இலக்கியங்களை உருவாக்கியுள்ளோம். அவற்றைப் பின்வரும் மூன்று நிலைகளில் வரையறுக்கலாம்.

1. ஆர்வமூட்டல்

2. அறிவூட்டல்

3. திறந்தநிலை இலக்கியம் கற்பித்தல்

இந்த அடிப்படையில் உருவாக்கப்பட்டுள்ள குறுவட்டைக் கணினி அறிஞர்களின் மேலான கருத்துகளுக்காக முன்வைக்கிறோம்.

தகவல் தொழில்நுட்பத்தில் ஏற்பட்ட கணினி நுட்பத்தின் ஊடாக வளர்ச்சி அனைத்துத் துறைகளிலும் புதிய மாற்றத்தினை ஏற்படுத்தியது. மொழி / இலக்கியம் கற்பித்தலிலும் மாற்றம் ஏற்பட்டு புதிய அணுகுமுறைகளும் வழிமுறைகளும் கண்டறியப்பட்டுள்ளன.

இன்றைய கணினித் தமிழ் வளர்ச்சிக்கு 'உத்தமம்' போன்ற தன்னார்வ நிறுவனங்கள் ஆற்றிய / ஆற்றும் தொண்டு மகத்தானது. தமிழ் இணைய மாநாடுகள் வழி மொழி / இலக்கிய / நுட்ப அறிஞர்களையும், கணினி - நுட்ப அறிஞர்களையும் ஒன்றிணைத்து ஐரோப்பிய மண்ணில், வரலாற்றுச் சிறப்பு மிக்க மாநாட்டில் பல்லூடக வழியாகப் பழந்தமிழ் இலக்கியக் கருத்தாடல் என்ற தலைப்பில் ஆய்வுச் செய்திகளை முன்வைக்கிறேன்.

மதுரைத் திட்டம், பழந்தமிழ் நூல்களை மின்பதிப்பாக்கம் செய்து மாபெரும் பணியாற்றுகிறது. தமிழ் மரபு அறக்கட்டளை, இலக்கியங்களைப் பாதுகாக்க வேண்டும் என்ற அடிப்படையில் கன்னித் தமிழைக் கணினித் தமிழாக இணைத்துப் பணியாற்றுகிறது.

மொழி தொடர்பான ஆய்வுகள், ஒலி, ஒலியன், உருபன், தொடர், பொருள் என விரிவடைந்து வருகின்றன. மேலும் இது இலக்கியம் சார்ந்து விரிவடைகிற போது, கரு, நடை, கருத்தாடல் அமைகிறது.

மொழி ஒருவரையொருவர் விளங்கிக்கொள்ளும் ஆற்றல் மிக்க சமூகப் பிணைப்புக்கு இன்றியமையாததாக அமைகிறது. கருத்துத் தொடர்பிலே ஈடுபடுவர்கள், ஒரு மனிதக் குழுவையோ, சமூகத்தையோ, ஒரு பண்பாட்டையோ சார்ந்தவராகக் குறிப்பிடலாம். அவர்கள் யாவரும் தம்முடைய மொழி, பழக்கங்கள், வழக்கங்கள் ஆகியன பற்றிய விதிமுறைகளைப் பின்பற்றி நடப்பவர்கள். அவ்விதிமுறைகளை அவர்களே உண்டாக்கினார்கள். அவர்களே அனுசரித்து நடக்க வேண்டியவர்களாய் உள்ளனர்.

எனவே மொழி வழியாக நடத்தப்படும் கருத்துத் தொடர்பும் சமூக வயப்பட்டது எனப் பெறப்படும் (அ.சண்முகதாஸ், 2006)

என்ற அறிஞரின் கூற்றுப்படி மொழியின் கருத்துத் தொடர்புக்குச் சமூகமே அடித்தளமாக விளங்குகிறது. சமூகத்தில் ஏற்படும் மாற்றங்கள் எல்லாம் மொழியாலும் ஏற்படும். பழந்தமிழ் இலக்கியங்களின் ஊடாகக் கற்றுத் தரும்பொழுது சமூகம் சார்ந்த விடயங்களைக் கணக்கில் எடுத்துக்கொள்வது இன்றியமையாதது.

தொடர்புபடுத்துகின்ற செய்தி, கருத்து, குறியீடு என்ற மூன்றின்வழி கருத்துத் தொடர்பு நிகழ்த்தலாம். பழந்தமிழ் இலக்கணமான தொல்காப்பியத்திலேயே கருத்துத் தொடர்புக்கான கூறுகள் இடம் பெற்றிருக்கின்றன.

செப்பும் வினவும் வழால் ஓம்பல்

(தொல்.சொல்.இளம்பூரணம்)

எழுத்து, சொல், பொருள், உரையாடல், கருத்துத் தொடர்புக்கான கூறுகளைச் சேனாவரையரும் இளம்பூரணரும் குறித்துள்ளனர்.

ஒவ்வொரு ஊடகம் சார்ந்து கருத்தாடல் முறைளும் வேறுபடும். வானொலிக்கான கருத்தாடல் நிகழ்த்தும் முறை வேறு.

தொலைக்காட்சி போன்ற காட்சி ஊடகத்தில் நேயர் கருத்துச் சொல்பவரின் முகத்தை நேரில் பார்க்கிறார். ஒரு திரைப்படப் பாடலை வானொலி மூலம் கேட்பதற்கும் தொலைக்காட்சியில் பார்ப்பதற்கும் உள்ள வேறுபாட்டினை நாம் இதனோடு தொடர்புபடுத்திப் பார்க்கலாம்.

வானொலிக் கருத்தாடலில் செவிப்புலன் சார்ந்து கருத்து அரங்கேறுகிறது. ஆனால், தொலைக்காட்சி போன்ற காட்சி ஊடகத்தில் கட் - செவி - புலன் வழியாகக் கருத்தாடல் அரங்கேறுகிறது. பனுவலைக் கருத்தாடல் வழி புரிந்துகொள்ளலாம் என்பதை இதன்வழி நாம் புரிந்துகொள்ளலாம். இதன் அடிப்படையில் இரண்டு வகைகளாகப் பிரிக்கலாம்.

1. மொழி சார் கூறுகள் (verbal)

2. மொழி சாராத கூறுகள (non-verbal)

வானொலி மூலமாக ஒலி வடிவில் சங்கத் தமிழ்க் கருத்தாடல் நிகழ்த்தினால், எந்தப் பகுதியில் வானொலி உரை நிகழ்த்துகிறோமோ, அந்தப் பகுதி மொழியின் வட்டாரத் தன்மையினை நிகழ்த்துபவர் அறிந்திருக்க வேண்டும்.

கேட்கிற வாசகர்களின் மொழியினை நிகழ்த்துநர் பயன்படுத்துவதா அல்லது தகுதமிழ் நிகழ்த்துவதா அல்லது இரண்டும் இணைந்து நிகழ்த்துவதா என்ற சிக்கல் வரும். அதனை மனத்தில் கொண்டு உரையினை நிகழ்த்த வேண்டும். கேட்பவருக்குப் பொருள் எளிதில் விளங்கிக்கொள்ளும் வண்ணம் உரை அமைதல் வேண்டும்.

மேலும் வானொலிக் கருத்தாடலுக்கு மொழிசார் கூறுகள் நிறைய இடம் பெற்றிருக்க வேண்டும்.

குரலிலே ஏற்ற இறக்கம், அழுத்தம், பாடலை மேற்கோள் காட்டுதல், குழப்பமில்லாமல் சொல்லுதல் போன்ற அடிப்படையில் வானொலிக் கருத்தாடலைச் சிறப்பாக நிகழ்த்த முடியும்.

திருக்குறளைக் கற்பிக்கிறோம் என்றால், திருக்குறளைப் பற்றித் தெளிவாகப் புரிந்துகொள்ளுதல், நேர ஒழுங்கு, உரையாற்றுதல் குறித்து முன்கூட்டியே திட்டமிடல், கருத்தாடல் நிகழ்த்துவதற்கு

அடிப்படையாகும். மாணவர்களின் வயது, தகுதி, கற்றல் திறன் அடிப்படையில் கருத்தாடல் நிகழ்த்தலாம்.

உரையாடல் முடிவில், தொடர் நிகழ்ச்சியா, தனி நிகழ்ச்சியா என்பது அறிந்து, அதற்கான கருத்தாடல் குறிகள் இடம்பெற வேண்டும்.

தொலைக்காட்சி ஊடகம் போன்றவற்றில் கருத்தாடல் நிகழும்பொழுது, மொழிசாராக் கூறுகளும் மொழிசார் கூறுகளும் இரண்டின் அடிப்படையில் அமைய வேண்டும்.

1. முகபாவனைகள், 2. செய்கைகள், 3. ஒப்பனைகள் போன்றவற்றில் கூடுதல் கவனம் செலுத்தினால் காட்சி ஊடாகக் கருத்தாடல் சிறப்பாக அமையும். குறிப்பாக, உடல்சார் மொழி மிக முக்கியம்.

பல்லூடக வழி பழந்தமிழ் இலக்கியங்களைக் கருத்தாடல் முறையில் அணுகும்போது பல சிக்கல்கள் ஏற்படும். அச்சிக்கல்களுக்குத் தீர்வு காண்பதாக நம் முயற்சிகள் அமைய வேண்டும்.

குறிப்பாக, சங்கத் தமிழை இன்றைய தமிழாக மொழி மாற்றம் செய்ய வேண்டும். நிகழ்த்துநருக்கும் - கேட்பவருக்கும் சமகாலத் தன்மை / தொடர்பு அமைய வேண்டும்.

குறுவட்டில் பல்லூடக இலக்கியக் கருத்தாடல் நிகழ்த்தும் பொழுது, கண்ணுக்குத் தெரியாத மக்கள் பலரை அது சென்றடையப் போகிறது என்ற நோக்கில் திட்டமிட்டுக் கருத்தாடல் நிகழ்த்த வேண்டும்.

ஆசிரியர் இல்லாத வகுப்பறை அமையலாம். மாணவர்களுக்கு வினா நிரல் தந்து, கற்றல் பணியினை மேம்படுத்தலாம். அடுத்த நிலையில் இணையவழி பழந்தமிழ் இலக்கியங்களை எவ்வாறு கற்கலாம், கற்பிக்கலாம் என முயற்சி செய்ய வேண்டும்.

பல்லூடக வழி பழந்தமிழ் இலக்கியங்களைக் கற்பிக்கும் பொழுது ஏற்படும் சிக்கல்களை அறிந்தால்தான் தீர்வுகளுக்கு நாம் திட்டமிட முடியும்.

பனுவலை ஒலி வழியாகக் கற்பிக்கும்பொழுது ஏற்படும் சிக்கல்கள்

பனுவலைக் காட்சி ஊடக வழி கற்பிக்கும்பொழுது ஏற்படும் சிக்கல்கள்

இலக்கியங்களைக் குறுந்தகட்டில் பதிந்து கற்பிக்கும்பொழுது எழும் சிக்கல்கள் இணையவழி கற்பிக்கும்பொழுது ஏற்படும் சிக்கல்கள்

மேற்கண்ட முறைகளில் சிக்கல்கள் இருப்பதைக் கண்டறிய வேண்டும். அதன் அடிப்படையில் கருத்தாடல் அமைந்தால், கற்றல் சிறப்பாக அமையும்.

குறுவட்டில் பல்லூடகக் கருத்தாடல் நிகழ்த்தும்பொழுது, கண்ணுக்குத் தெரியாத நேயர்கள் பலரை இது சென்றடைகிறது என்ற நோக்கத்தோடு நம் உரை அமைய வேண்டும்.

ஆசிரியர் இல்லாத வகுப்பறையில் குறுவட்டின் வழி மாணவர்கள், இலக்கியங்களை எவ்வாறு கற்றுக்கொள்கின்றனர் என ஒரு வினா நிரல் உருவாக்கி, அவர்களைப் பதில் அளிக்கச் செய்து, கருத்தாடல் மூலம் கற்பிக்கும் திறனை மேம்படுத்தலாம்.

தமிழர்கள் மட்டுமில்லாமல், தமிழ் மேல் ஆர்வமுள்ள பிற மொழியினரும் பழந்தமிழ் இலக்கியங்களைக் கற்றுக்கொள்வதற்குக் கணினி வழியில் தமிழ் தயாராக இருக்கிறது. நாம் தயாராக இருக்கிறோமா என்பதே கேள்வி.

துணை நின்ற நூல்கள்

1. மொழியும் பிற துறைகளும், அ.சண்முகதாஸ், 2006
2. மொழியும் அதிகாரமும், எல்.இராமமூர்த்தி, 2005
3. கருத்தாடல் ஆய்வு, ந.இளங்கோ, 1996
4. தொல்காப்பியம் சங்கத் தமிழ் புதிய பார்வை, அண்ணாமலைப் பல்கலைக்கழகம், 2006
5. தொல்காப்பியம், சொல், தமிழண்ணல், 1996

COMPUTATIONAL LINGUISTICS

தமிழ் வினை வடிவங்கள்: கணினிப் பகுப்பாய்வு

Computer Analysis of Tamil Verb Forms

முனைவர் ப. டேவிட் பிரபாகர்
இணைப் பேராசிரியர் - தமிழ்
சென்னைக் கிறித்தவக் கல்லூரி, சென்னை-600 059, இந்தியா
Email: tamilprofessor@gmail.com

Abstract: The aim of computational morphology is to take a string of characters as input and deliver an analysis as output. The input string can be analyzed for underlying morphemes and syntactic interpretation.

As the 'verb' is a dynamic part of a sentence, the present study focuses on both simple (finite and non-finite) and complex verb forms in Modern Tamil. Algorithms have been developed for analysis of all the verb forms, that can be adopted for various Tamil related NLP tasks.

Seven rule sets are developed in the form of flow charts, which can analyze any verb form from right to left, to carve out morphemes and from what is left of the verb, reconstruct the verb root. This work also gives morpho- syntactic interpretation for the given structure. The details regarding data representation, methods and modules are given in the full paper.

மொழியின் பண்புகளைக் கண்டறிய கணினியைப் பயன்படுத்தும் நோக்கிலும் கணினிக்கு மொழி அறிவைப் புகட்டும் நோக்கிலும் ஆய்வுகள் மேற்கொள்ளப்பட்டு வருகின்றன. இத்தகைய ஆய்வு 'இயற்கை மொழியாய்வு' எனப்படுகிறது. கணினிக்குச் செயற்கை நுண்ணறிவாற்றலை வழங்கும் முயற்சியின் ஒரு பகுதியாகவும் இது விளங்குகிறது. இதற்கு, இயற்கை மொழியின் அமைப்பு மற்றும் பயன்பாடு பற்றிய இலக்கண நியதிகளை கணினியின் ஏற்புக்குத்தக விதிமுறையாக்கமாக மாற்றியமைக்க வேண்டும். இதனால், இயற்கை மொழிகளிலேயே கணினியோடு உறவாடவும், தேவையான தகவல்களைத் தேவையான வடிவத்தில் பெறவும் வழி ஏற்படும்; மொழி சார்ந்த பல்வேறு பணிகளுக்குக் கணினியைத் துணையாகக் கொள்ளவும் இயலும்.

இலக்கும் நோக்கும்

இக்கட்டுரை, தமிழில் உள்ள வினை வடிவங்களைப் பகுத்து, அவற்றில் உள்ள வினையடிகளையும், அவற்றோடு ஒட்டுமுறையில் இணையும் கூறுகளையும் அடையாளம் காண, கணினியின் ஏற்புத்தன்மைக்கு இயைந்த வழிமுறை வரைவுகளை வழங்குகிறது. இதனைக் காட்சிப்படுத்தவும் முற்படுகிறது. திணை, பால், எண், இட இயைபுகளைப் புலப்படுத்தும் விகுதிகளைத் தமிழ் வினைச்சொற்கள் ஏற்பதால், சொற்றொடர் பற்றிய அமைப்பு ஆய்விலும் வினை வடிவங்கள் முக்கியப் பங்காற்றுகின்றன. பில்மோரின் வேற்றுமை இலக்கணக் கோட்பாட்டில்,வினையும் வினையோடு வேற்றுமை உறவு கொண்டுள்ள பெயர்களும் முதன்மை உறுப்புகளாகக் கொள்ளப்படுகின்றன. வினைச்சொல் இன்றியமையாத சொல் வகையாக விளங்குவதோடு, சொற்றொடர் பகுப்பாய்விற்கும் அடிப்படையாக விளங்குவதால் வினை வடிவங்களை முதன்மைப்படுத்தி இக்கட்டுரை அமைகிறது.

இக்கட்டுரை, தற்காலத் தமிழில் காணப்படும் எல்லா வகையான வினை வடிவங்களையும் விவரிக்கிறது. வழக்கொழிந்த வடிவங்கள் தவிர்க்கப்பட்டுள்ளன. வண்ணனை மொழியியலாரின் கொள்கையையொட்டி, சந்தியும் சாரியையும் முறையே இடைநிலையுடனும் விகுதிகளுடனும் இணைந்த நிலையில் பகுப்பு மேற்கொள்ளப்படுகிறது. வலமிருந்து இடமாக மொழிக்கூறுகளைப் பகுத்துக் காணும் வண்ணம் வழிமுறைகளும், எஞ்சிய வினைப்பகுதியிலிருந்து வினையடியினை ஆக்கிக் கொள்ளும் விதித்தொகுப்புகளும் உருவாக்கப்பட்டுள்ளன. பகுப்பை மட்டுமே இவ்வாய்வு இலக்காகக் கொள்கிறது; தோற்றுவிப்புக்கான வழியமைப்புகள் உருவாக்கப்படவில்லை.

ஆய்வு முறை

இவ்வாய்வில், வினையடி, பாலிட விகுதிகள், கால இடைநிலைகள், எதிர்மறை உருபுகள், எச்ச விகுதிகள், பிற ஒட்டுருபுகள் ஆகியன தனித்தனிப் பட்டியலாக்கிக் கொள்ளப்பட்டுள்ளன. பகுப்பால் பயனற்ற தனி, வேறு, சரி, பொது ஆகிய வினைகளைக் கணினி முதலில் தேடும். எதிர்மறை ஒட்டை ஏற்காது வினைமுற்று வடிவில் வரும் உண்டு, அல்ல, இல்லை முதலிய வினைகள் தனிப்பட்டியலாக்கப்பட்டுள்ளன. அல், உள், இல், உடை, உரி முதலிய வினைகள் பொதுப்பட்டியலிலும் சேர்க்கப்பட்டுள்ளன. குறை வினையடிகள் குறிப்பிட்ட பாலிட விகுதிகளை மட்டும் ஏற்பதால், குறிப்பிட்ட பாலிட விகுதிகள் கண்டறியப்படும்போது மட்டும் குறை வினையடிப் பட்டியல் நோக்கப்படுகிறது.

தனிவினைப் பகுப்பு

தனி வினை வடிவங்களின் முதற்பகுதி வினையடியாக அமைகிறது. விகுதி எதுவும் பெறாத நிலையில் இது ஏவல் வினை எனப்படுகிறது. வினையமைப்பில் மூன்று பகுதிகள் இடம் பெற்றிருந்தால், முதற்பகுதி வினையடியாகவும் நடுப்பகுதி கால இடைநிலைகளையோ எதிர்மறை

உருபுகளையோ கொண்டிருக்கிறது. இறுதிப் பகுதியில் பாலிட விகுதிகளோ, எச்சம், முற்று, எதிர்மறை, தொழிற்பெயர் ஆகியவற்றை உணர்த்தும் விகுதிகளோ அமைகின்றன.

வலமிருந்து இடமாகப் பகுத்துக் கொள்ளப்படும் இவ்வாய்வில், பாலிட விகுதிகள் தவிர்த்த பிற விகுதிகள் முதலிலும் அடுத்து பாலிட விகுதிகளும், அதனைத் தொடர்ந்து கட்டுண்ட மாற்றுப்பெயர் வடிவங்களும் தொகுத்துக் கொள்ளப்படுகின்றன.

விகுதிகள், ஒட்டுகள் படிப்படியாகப் பகுக்கப்பட்டு எஞ்சிய வினைப் பகுதியிலிருந்து விதிகள் மூலம் வினையடி பெறப்பட்டு, வினைப்பட்டியலில் ஒப்பு நோக்கப்படுகிறது. முறை சாரா வினையடிகளும் விதிகளின் மூலமே பெறப்படுகின்றன.

வினையடிகளைப் பெற உதவும் ஏழு விதித்தொகுப்புகள் இவ்வாய்வில் உருவாக்கப்பட்டுள்ளன. பகுக்கப்படும் ஒட்டு அல்லது விகுதிக்கேற்ப இவ் விதித்தொகுப்புகளில் ஒன்று செயல்படும். வினையடியும் பகுக்கப்பட்ட ஒட்டு அல்லது விகுதியும் கொண்டுள்ள அமைப்பையொட்டி, அதற்கு இலக்கண வரைவு வழங்கப்படுகிறது. ஒன்றுக்கு மேற்பட்ட இலக்கண வரைவுகளுக்கு இடமளிக்கும் ஒத்த அமைப்புகளுக்குத் தொடரியல் அடிப்படையில் தீர்வு காணப்பட்டுள்ளது.

வினை வடிவங்களின் அமைப்பும் இலக்கண வரையறையும்

'Ø' இறுதி

வினையடி + Ø	ஏவல்

'அ' இறுதி

வினையடி + அ	செயவென் எச்சம்
வினையடி + இறப். உருபு+அ	இறந்தகாலப் பெயரெச்சம்

வினையடி + நிகழ். உருபு+அ	நிகழ்காலப் பெயரெச்சம்
வினையடி + எதிர். உருபு+அ	எதிர்மறைப் பெயரெச்சம்
வினையடி + க/க்க	வியங்கோள்

'ஆ' இறுதி

வினையடி + ஆ	ஈறுகெட்ட எதிர்மறைப் பெயரெச்சம்/பலவின் வினைமுற்று

'இ' இறுதி

வினையடி + இ	வினையெச்சம்

'உ' இறுதி

வினையடி + இறப். உருபு+உ	வினையெச்சம்
வினையடி + ஆத்+அது	எதிர்மறைத் தொழிற்பெயர்/ எதிர்மறை வி.அ பெயர்
வினையடி + ஆ+து	எதிர்மறை வினையெச்சம் / ஒன்றன்பால் எதிர்மறை வி.மு

'ஐ' இறுதி

வினையடி + இறப். உருபு+அமை	தொழிற்பெயர் (வந்தமை)
வினையடி + நிகழ். உருபு+அமை	தொழிற்பெயர் (வருகின்றமை)
வினையடி + இறப். உருபு+அமை	எதிர்மறைத் தொழிற்பெயர் (வராதமை)

'ம்' இறுதி

வினையடி + ம்	ஏவல்(மரியாதை ஒருமை)/ எதிர்கால வினைமுற்று/ எதிர்காலப்பெயரெச்சம்
வினையடி + உம்	ஏவல்(மரியாதை ஒருமை/பன்மை)
வா/தா +உம்	ஏவல்(மரியாதை ஒருமை/பன்மை)
வினையடி + அவும்	பொது ஏவல்/ எச்சம்
வினையடி + (க்)கும்	எதிர்கால வினைமுற்று/ எதிர்காலப்பெயரெச்சம்
வினையடி + அட்டும்	வினைமுற்று(இசைவு)
வினையடி + ஆத் +ஏயும்	எதிர்மறை ஏவல்
வினையடி + இறப். உருபு+அதும்	வினைமுற்று(தொடர் நிகழ்வு)
வினையடி + அலாம்	வினைமுற்று(சாத்தியம்)

'ய்' இறுதி

ஆ/போ +ய்	வினையெச்சம்

'ல்' இறுதி

வினையடி + இறப். உருபு+ஆல்	நிபந்தனை எச்சம்
வினையடி + ஆ+மல்	எதிர்மறை வினையெச்சம்
வினையடி + -(க்)கையில்	உடன் நிகழ்வு எச்சம்
வினையடி + -அல்/ (த்)தல்	தொழிற்பெயர்

'ள்' இறுதி

வினையடி + -(உ)ங்கள் ஏவல்(மரியாதை ஒருமை/பன்மை)

பாலிட விகுதிகளும் கட்டுண்ட மாற்றுப்பெயர்களும் இறுதியில் அமைதல்

வினையடி + இறப். உருபு+பாலிடவிகுதி இறந்தகால வினைமுற்று
வினையடி + நிகழ். உருபு+ பாலிடவிகுதி நிகழ்கால வினைமுற்று
வினையடி + எதிர். உருபு+ பாலிடவிகுதி எதிர்கால வினைமுற்று
(உயர்திணை)
வினையடி + பாலிடவிகுதி (உயர்திணை) எதிர்மறை வினைமுற்று
வினையடி + ஆத்+ பாலிடவிகுதி(ஆய் ஒழிந்த..மு.வி)எதிர்மறை ஏவல்
வினையடி + காலஉருபு+அவன்/அவள்/அவர்/அவை வினையாலணையும் பெயர்
வினையடி + ஆத்+அவன்/அவள்/அவர்/அவை வினையாலணையும் பெயர்
(எதிர்மறை)

கூட்டுவினைப் பகுப்பு

கணினிய நோக்கில் துணைவினைகளின் பட்டியல் சிறியதாக அமைவதால், ஒவ்வொரு துணைவினையையும் எடுத்துக்கொண்டு அதன் வருகை, ஏற்கும் அலகுகள், கட்டுப்பாடுகள் முதலிய குறிப்புகளைக்கொண்ட விவரத்தொகுப்பு உருவாக்கப்பட்டுள்ளது. கலப்பு வினையடிகள் நேரடியாக வினைப்பட்டியலில் சேர்க்கப்பட்டுள்ளன.

கூட்டுவினை அமைப்பில் வலப்புறத்தில் அமைந்துள்ள வினை முதல் வினையாகக் கொள்ளப்படுகிறது. தனிவினைப் பகுப்பிற்கான விதித்தொகுப்புகளைப் பயன்படுத்தி முதல் வினையைப் பகுக்க இயலும். முதல் வினை பகுத்துக்கொள்ளப்பட்ட பின்பு, வினைகளுக்கிடையே அமையும் ஒற்று, உடம்படுமெய், இடைச்சொற்கள் ஆகியவற்றைப் பகுத்துக் கொண்டு, அடுத்தடுத்த வினைகளை கண்டறியும் பொது வழியமைப்பும் உருவாக்கப்பட்டுள்ளது. வினைகள் இடைவெளியின்றி அமைந்திருக்கக் கூடும் இந்நிலையில், எஞ்சிய வினைப் பகுதியிலிருந்து ஒவ்வொரு எழுத்தாக எடுத்துக்கொண்டு வினைப்பட்டியலில் தேட வேண்டும். ஓரெழுத்து வினையாக இருப்பின் அது நெடிலாகவே அமைகிறது. ஒன்றிலிருந்து நான்கெழுத்தை மேல் எல்லையாகக் கொண்டே தமிழில் வினைச் சொற்கள் அமைகின்றன. அ, ஆ, இ, ஈ, உ, ஐ, ஓ, ண், ய், ர், ல், ள், ழ் ஆகிய எழுத்துகளையே தமிழ் வினையடிகள் இறுதியாகக் கொள்கின்றன. இந்நிலையில் மட்டும் நேரடியாக வினைப்பட்டியலில் தேட வேண்டும்.பிற ஈறுகளைக் கொண்டிருப்பின் பொருத்தமான விதித்தொகுப்பைப் பயன்படுத்தி தமிழில் வினையடியைக் காண வேண்டும். அசை அமைப்பைக் கொண்டும் படிப்படியாக வினையடிகளைப் பிரித்தெடுக்கும் வழிமுறைகளும் மேற்கொள்ளப்பட்டுள்ளன.

Dravidian Wordnet

Dr.S.Rajendran
Professor & Head, Department of Linguistics,
Tamil University, Thanjavur 613010,
Email: raj_ushush@yahoo.com

Abstract: Languages work as an integrated and interconnected system. Basically it connects the world with the human brain. The conceptualization of the world by perception and naming results in building the basic units of language, the words. This becomes the input into the brain from which the output of communication is made. Human beings more or less look at the world with similar experience and make abstractions about the world by conceptualization so that it is possible for them to exchange their ideas without difficulty. Language works at different levels between sounds, words and sentences. As sounds do not have meaning on its own human beings depend on words as basic units of communication. We visualize, conceptualize and name things based on our conceptual percepts. While doing so we make classifications of the objects of the world depending on their physical and telic properties. The psychology behind the abstraction leads to typology which in turn turned into ontology. That is the reason the base of WordNet's is ontology.

WordNet is a nonlinear lexical structure based on semantic features of individual words. It allows linking one word to another in a more meaningful way than in a conventional dictionary or thesaurus, since every word shares one or more of semantic features with the other in a language.

The primary use of wordNet is making use of it as a multilingual information accessing system through lexical items. The WordNet by its nature turns to be an ideal lexical accessing system as it links concepts with another concept by multifarious meaning relations. The wordNet is a lexical data base which has many practical utilities. One of them is to use of wordNet for translation. WordNet not only links one concept with another concept through semantic relations, but also captures the contextual meaning variations of a particular word i.e. the polysemy of a word. The wordNet has amble scope to link meaning with a context there by capturing the different meanings of a word contextually.

Development of Dravidian WordNet is an on-going project activity shared by Tamil University at Thanjavur, Amrita Vishwa Vidyapeetham at Coimbatore, and Dravidian University at Kuppam. The main objectives of the project are:

- Developing an extensive and high quality multilingual semantic lexical database for Dravidian languages
- Developing language-independent set of semantic concepts linking the language networks together
- Standardizing semantic classification of information for all Dravidian languages and providing resources for development of applications

International Status

WordNet was originally conceived and developed as a lexical database for English on the basis of psycholinguistic properties. The major lexical categories like nouns, verbs, adjective and adverbs are

organized in terms of sets of synonyms (synsets), each representing a lexical concept. A synset is a set of synonyms (word forms that have to the same or similar meaning) and two words are said to be synonymous if their mutual substitution does not alter the truth-value of a given sentence in which they occur, in a given context. For example, {car; auto; automobile; machine; motorcar} form a synset because they can be used to refer to the same concept. These synsets are interconnected by certain relations - lexical relations such as synonymy, antonymy, and semantic relations such as hyponymy (between specific and more general concepts) and meronymy (between parts and wholes).

The success of the English WordNet has paved way for the emergence of several projects with the aim of constructing WordNets in various languages and developing multilingual WordNets. EuroWordNet, a conglomeration of WordNets in European languages is an important project that has come up with a multilingual WordNet.

Dravidian WordNet

Dravidian WordNet is a natural chunk in the Indo-WordNet. It is only ideal that we should have Dravidian WordNet before we develop a larger Indo-WordNet, because the genetic relationship among the Dravidian languages can be maximally exploited in a more natural way. It allows, for example, a search tool to infer other terms, from the terms provided by the user and coming up with the most optimal search for retrieving information.

Among Indian languages, Dravidian languages such as Tamil, Telugu, Kannada and Malayalam share a number of lexicalized concepts in terms of morphology and semantics besides others as in typological and culture-specific features. Building a common WordNet for this family of languages makes it easier the task of developing an IndoWordNet.

The WordNet is a natural answer in machine translation systems. It has the potential to interpret source language words and come up with lexical equivalents in the target language in a more natural way as a bilingual does. This is particularly useful for users working in a second language who may not have appropriate knowledge of vocabulary. The network will also be used as a basic resource for supporting other applications. The semantic knowledge embodied in the network makes it suitable as a component in expert systems, question-answer systems, language learning systems and automatic summarizers. Dravidian WordNet with explicitly stated semantic features will become an essential lexical resource to be used in all practical NLP applications. It will give a boost for NLP research in India.

The resources produced by Dravidian WordNet will have a wide range of users who are interested in language learning, language generation, machine aided translation, language understanding, information retrieval, electronic publishing and the production of WordNets in other languages. The end users of the resources will be all those people who utilize the applications that incorporate Dravidian WordNet resources.

Design and Implementation

The design and implementaion of the Dravidian WordNet will be based on EuroWordNet as explained by Pike Vossen (1998). Designing and implementation of word net are the two major tasks assigned to computer scientists. To achieve them they have to work in collaboration with lexicographers and linguists. Once the lexicographers complete their work of collecting the data required for building the word net, the job will be handed over to computer scientists. The semantic

information collected on lexical items from the basic building blocks for the computer scientist to construct word net. As the words and meanings are related to one another and mapped as such in word net, it is but natural that the word net gives the impression of an on-line thesaurus. The word net automatically inherits the all the powers of a thesaurus. It also resembles an on-line dictionary as it provides meanings for lexical items. Being superior to these two tools, word net provides much more information that has been loaded in an on-line thesaurus as well as in an on-line dictionary.

Design of Individual WordNets

The task of developing the on-line database of a language can be conveniently divided into two interdependent tasks (Beckwith, Miller and Tengi, 1993). These tasks bear a vague similarity to the traditional tasks of writing and printing a dictionary:

- To write the source files that contain the basic lexical data - the contents of those files are the lexical substance of WordNet.
- To create a set of computer programs that would accept the source files and do all the work leading ultimately to the generation of a display for the user.

In line with English WordNet, the WordNet system can be divided into four parts based on the specific tasks assigned to them:

- Lexical resource system
- Compiler system
- Storage system
- Retrieval system

Expand/Merge approach

The WordNet database will be built (as much as possible) from available existing resources and databases with semantic information developed in various projects. This will be not only more cost-effective given the limited time and budget of the project, but also will make it possible to combine information from independently created WordNets. Two models will be involved in the built up.

Merge Model: the selection will be done in a local resource and the synsets and their language-internal relations will be first developed separately, after which the equivalence relations to Tamil WordNet will be generated.

Expand Model: the selection will be done in Tamil WordNet and the Tamil WordNet synsets will be translated (using bilingual dictionaries) into equivalent synsets in the other language. The WordNet relations will be later on adopted across languages.

The Merge Model will result in a WordNet that will be independent of Tamil WordNet, possibly maintaining the language-specific properties. The Expand model will result in a WordNet that is very close to Tamil WordNet but which is also biased by it. What approach should be followed also depends on the quality of the available resources.

After the first production phase the results will be converted to the Dravidian WordNet import format and loaded into the common database. At that point, various consistency checks will be carried out, both formally and conceptually. By using the specific options in the database, it is then possible to further inspect and compare the data, to restructure relations where necessary and to measure the

overlap in the fragments developed at the separate sites.

Design of the multilingual database

The design of the Dravidian WordNet-database will be first of all based on the ontological structure of Tamil WordNet which in turn is based on a thesaurus for Tamil prepared by Rajendran (2001). Tamil wordNet relies on extensive preliminary investigations of the vocabulary of Tamil (Rajendran, 1976-2003) based on the componential analysis of meaning (Nida, 1975a & 1975b) and structural semantics (Lyons, 1977). Portions of this work have been compiled into a Tamil thesaurus (Rajendran, 2001). The Tamil thesaurus in electronic form represents the Ontological Structure of Tamil (shortly OST) vocabulary giving scope to take care of any kind of semantic/lexical relations that hold between lexical items.

The notion of a synset and the main semantic relations will be taken over in Dravidian WordNet. However, some specific changes will be made to the design of the database, which are mainly motivated by the following objectives:

- to create a multilingual database;
- to maintain language-specific relations in the WordNets;
- to achieve maximal compatibility across the different resources;
- to build the WordNets relatively independently (re)-using existing resources;

The most important difference of Dravidian WordNet with respect to a language specific WordNet is its multilinguality, which however also raises some fundamental questions with respect to the status of the monolingual information in the WordNets. In principle, multilinguality will be achieved by adding an equivalence relation for each synset in a language to the closest synset in Tamil WordNet. Synsets linked to the same Tamil WordNet synset will be supposed to be equivalent or close in meaning and can then be compared. However, we have to take into consideration the differences across the WordNets. If 'equivalent' words are related in different ways in the different resources, we have to make a decision about the legitimacy of these differences.

In Dravidian WordNet, we will take the position that it must be possible to reflect such differences in lexical semantic relations. The WordNets are seen as linguistic ontologies rather than ontologies for making inferences only. In an inference-based ontology it may be the case that a particular level or structuring is required to achieve a better control or performance, or a more compact and coherent structure. For this purpose it may be necessary to introduce artificial levels for concepts which are not lexicalized in a language or it may be necessary to neglect levels that are lexicalized but not relevant for the purpose of the ontology. A linguistic ontology, on the other hand, exactly reflects the lexicalization and the relations between the words in a language. It is a "WordNet" in the true sense of the word and therefore captures valuable information about conceptualizations that are lexicalized in a language: what is the available fund of words and expressions in a language. In addition to the theoretical motivation there is also a practical motivation for considering the WordNets as autonomous networks. To be more cost-effective, they will be derived (as far as possible) from existing resources, databases and tools. Each sites therefore will have different starting points for building their local WordNet, making it necessary to allow for a maximum of flexibility in producing the WordNets and structures.

The Database Modules

To be able to maintain the language-specific structures and to allow for the separate development of independent resources, we will make a distinction between the language-specific modules and a separate language-independent module. Each language module represents an autonomous and unique language-specific system of language-internal relations between synsets. Equivalence relations between the synsets in different languages and Tamil WordNet will be made explicit in the so-called Inter-Lingual-Index (ILI). Each synset in the monolingual WordNets will have at least one equivalence relation with a record in this ILI, either directly or indirectly via other related synsets. Language-specific synsets linked to the same ILI-record should thus be equivalent across the languages.

The ILI will be an unstructured list of meanings, mainly taken from Tamil WordNet, where each ILI-record consists of a synset, an Tamil gloss specifying the meaning and a reference to its source. The only purpose of the ILI is to mediate between the synsets of the language-specific WordNets. No relations are therefore maintained between the ILI-records as such. The development of a complete language-neutral ontology is considered to be too complex and time-consuming given the limitations of the project. As an unstructured list, there is no need to discuss changes or updates to the index from a many-to-many perspective. Note that it will nevertheless be possible to indirectly see a structuring of a set of ILI-records by viewing the language-internal relations of the language-specific concepts that are related to the set of ILI-records. Since Dravidian WordNet will be linked to the index in the same way as any of the other WordNets, it is still possible to recover the original internal organization of the synsets in terms of the semantic relations in WordNet. Once the WordNets are properly linked to the ILI, the Tamil WordNet database will make it possible to compare WordNet fragments via ILI and to track down differences in lexicalization and in the language-internal relations. In this view, the ILI-records will be represented by a Tamil gloss. Below a synset-ILI pair, the language-internal relations can be expanded, as is done for the hyperonyms. The target of each relation will again be represented as a synset with the nearest ILI-equivalent (if present).

Conclusion

The theme of lexical semantics, computational lexicography, and computational semantics are altering rapidly. The availability of machine-readable resources and newly developed tools for analyzing and manipulating lexical entries make it possible to build a massive word net for a language. In present state of affairs it is quite feasible to build Dravidian WordNet. Building of a word net for Dravidian languages is an immediate requirement in the context of information technology equipped with internet in which the web sites in Dravidian languages are getting added up day by day.

References

1 Rajendran, S. 2001. *taRkaalat tamizc coRkaLanjciyam* [Thesaurus for Modern Tamil]. Thanjavur: Tamil University.

2 Rajendran, S. 2002. 'Preliminaries to the preparation of a Word Net for Tamil.' Language in India 2:1, www.languageinindia.com

3 Rajendran, S., S. Arulmozi, B. Kumara Shanmugam, S. Baskaran, and S. Thiagarajan. 2002. "Tamil WordNet." Proceedings of the First International Global WordNet Conference. Mysore: CIIL, 271-274.

4 Vossen P. (eds.) 1998. EuroWordNet: A Multilingual Database with Lexical Semantic Networks. Dordrecht: Kluwer Academic Publishers.

Unsupervised Approach to Tamil Morpheme Segmentation

K.Rajan (Muthiah Polytechnic College, Annamalainagar)
Dr.V.Ramalingam (Department of Computer Science & Engineering)
Dr.M.Ganesan (CAS in Linguistics, Annamalai University, Annamalainagar)

Abstract: This paper presents an unsupervised learning of Tamil morphology from untagged text corpus. An unsupervised approach to the segmentation of morphemes is attractive for highly inflective languages. Morpheme segmentation is the task of segmenting a word into morphemes such as prefixes, stems and suffixes. The unsupervised learning approach requires no or minimal linguistic knowledge. Many such algorithms have been applied and tested on English, Finnish, Turkish and other European languages. This is the first such kind of work being carried out for Tamil. The objective of this work is not the realisation of a complete morphological analysis, but the production of the list of morphemes for the language. This paper discusses about Letter Successor Varieties, N-gram based approach for morpheme identification and the application of an iterative process for the segmentation. This method is trained on a list of words collected from the CIIL Tamil corpus.

Keywords: Unsupervised morpheme segmentation, Tamil morphology, machine learning

Introduction

Word segmentation is an important problem in many natural language processing tasks such as speech recognition where there is no explicit word boundary information within a continuous speech utterance, or in interpreting written languages such as Chinese, Japanese and Thai where words are not delimited by space. In other languages, words are a combination of smaller meaning bearing units referred to as morphemes. The act of separating a word into its morphemes is called morphological analysis and/or morpheme segmentation. The morpheme segmentation algorithm attempts to find morpheme boundaries within word forms. The identified morphemes can be used to produce clustering of word forms of the same lemma with a quite high precision. These morphemes are classified into prefixes, stems and suffixes. Morphemes are used to identify words, which are semantically similar, and improve the performance of the systems in document retrieval and speech recognition. Commonly, algorithms designed for word segmentation utilize very little prior knowledge or assumptions about the syntax of the language. Instead prior knowledge about typical word length may be applied, and small seed lexicons are sometimes used for bootstrapping. The segmentation algorithms try to identify character sequences that are likely words, without the context of the words. Several approaches, based on machine learning, aiming at word segmentation and morphology have been published recently.

Machine Learning

NLP applications typically rely on large databases of linguistic knowledge. The manual design of such resources is labor-intensive and requires considerable effort by linguistic experts. To reduce the amount of manual work, machine learning can be utilized. 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. There are three major types of machine learning: *supervised, unsupervised and reinforcement learning*. In supervised learning, there is a "teacher" that provides the learner with a set of input-output pairs. In unsupervised learning, there is no teacher, providing desired answers, but since the data are not entirely random, there are statistical regularities that can be captured and that can be applied in new cases. Reinforcement learning corresponds to something between supervised and unsupervised approaches. It differs from supervised learning in the sense that explicit input-output pairs are not available. An agent explores environment and is able to take actions. Depending on the outcome of the series of actions taken, the agent is rewarded or penalized (Mathias Cruetz, 2007).

Researchers in the area of data compression, dictionary construction, and information retrieval have all contributed to the literature on automatic morphological analysis. Work in automatic morphological analysis can be divided into four major approaches (John Goldsmith, 2001). The first approach proposes to identify morpheme boundaries first, and thus indirectly to identify morphemes on the basis of the degree of predictability of the n+1 st letter given the first n letters. This was first proposed by Zellig Harris (1951) and further developed by Hafer and Weiss (1974). The second approach seeks to identify bigrams(and trigrams) that have high likelihood of being morpheme internal. The third approach focuses on the discovery of patterns of phonological relationships between pairs of related words. The fourth approach is top-down and seeks an analysis that is globally most concise (i.e MDL , Minimum Description Length approach).

Unsupervised learning of morphology aims to model one or more of three properties of written morphology: segmentation, clustering around a common stem, and generation of new word forms with productive affixes (Taesun Moon, 2009). For unsupervised learning algorithm, the sole input is the corpus, but no dictionary and no morphological rules are needed. The goal of the algorithm is to provide the correct analysis of words into component pieces. Recently, a number of approaches to unsupervised morphological segmentation have been proposed for English and other European languages.

In this paper we have tried morpheme segmentation using Letter Successor Variety on N-grams of words and two iterative approaches for segmentation.

Letter Successor Variety (LSV)

The idea of LSV is to count the amount of different letters encountered after (or before) a part of a word and to compare it to the counts before and after that position. Morpheme boundaries are likely to occur at sudden peaks or increase of that value (Harris 1955). When the successor varieties for a given word have been derived, the information is used to segment the word. Hafer and Weiss (1974) discuss four methods of performing this:

1. The cutoff method – in which a threshold is selected for boundary detection

2. Peak and plateau method-in which a segment break is made after a character whose successor variety exceeds that of its neighbors.

3. Complete word method – in which a break is made after a segment if the segment is a complete word.

4. Entropy method – based on the entropy value calculated for each letter in the word.

Our algorithm is based on the peak and plateau model. If Sn is the successor count of n th character in a word, a word is segmented if Sn forms a local peak or a plateau of the count vector.

Successive Split Method

Tamil morphology is concatenative and agglutinative in nature. By analysing the corpus we can see that if one word is a substring of another word they are morphologically related. The words in the corpus are collected and ordered first. The morphologically related words appear adjacent to each other. This list is processed by the algorithm in a recursive manner. Initially all the words are treated as stems.

In the first pass, these stems are split into new stems and suffixes based on the similarity of the characters. They are split at the position where the two words differ. The right substring is stored in a suffix list and the left substring is kept as a stem. In the second pass, the same procedure is followed, and the suffixes are stored in a separate suffix list. We tested our algorithm using four iterations. For this algorithm we represent all words in vowel and consonant form.

In another variant of this method, we use some seed words. For the given seed word, first we collect all the suffixes and store them in a suffix file. Then, for each suffix of the file, we collect the stems which have these suffixes, in a new stem file. We use a threshold value for the minimum length of the stem as 2. In this way, both the suffix and stem files are augmented iteratively. During this process we mark the words which are used from the corpus. This process is stopped when there are no new stems are collected. After completing this process, the words which are unmarked are also added to the stem file. These unmarked words represent function words or uninflected word forms.

Conclusion

In this paper we have discussed unsupervised machine learning approaches for Tamil morpheme segmentation. These methods are based on the existing approaches which are tested on English and European languages by various researchers. Our objective of this work is to study the performance of these methods on Tamil corpus. The results are encouraging and the performances are comparable with that of the rule based approaches.

References

1. Creutz. M and Lagus.K. 2002. Unsupervised discovery of morphemes. In ACL ’02 workshop on Morphological and phonological learning-Volume 6, pages 21–30.
2. M. Creutz and K. Lagus. 2007. Unsupervised models for morpheme segmentation and morphology learning. ACM Trans. Speech Lang. Process., 4(1):3.
3. Harris, Z. (1951) Structural Linguistics. University of Chicago Press.
4. Hafer M.A. and Weiss.S.F. 1974. Word Segmentation by Letter Successor Varieties. Information Storage and Retrieval, 10:371–385.
5. John Goldsmith. 2001. Unsupervised learning of morphology of a natural language. Computer Linguistics, 27:153-198.
6. Kazakov D. 1997, Unsupervised Learning of naïve morphology with genetic algorithms. In W. Daelemans, et al., eds., ECML/Mlnet Workshop on Empirical Learning of Natural Language Processing Tasks, Prague, pp. 105-111.

7. Patrick Schone and Daniel Jurafsky. 2000. Knowledge-free induction of morphology using latent semantic analysis. In Proceedings of CoNLL-2000 and LLL-2000, 67-72, Lis-bon, Portugal.
8. P. Schone and D. Jurafsky. 2001. Knowledge-free induction of inflectional morphologies. In NAACL '01, pages 1–9.
9. Sylvain Neuvel and Sean A.Fulop. 2002. Unsupervised Learning of Morphology without morphemes. Proceedings of the 6th workshop of ACL Special Interest Group in Computational Phonology (SIGPHON). 31-40. Philadelphia.
10. Taesun Moon, Katrin Erk, and Jason Baldridge. 2009. Unsupervised morphological segmentation and clustering with document boundaries. Proceedings of the Conference on Empirical methods in NLP. pp 668-677. Singapore.

தொல்காப்பியத்தின் எழுத்ததிகாரத்துக்கான இடம் சாரா இலக்கணம்

இல. பாலசுந்தரராமன், ஈசுவர் சிரீதரன்
Email: sundarbecse@yahoo.com, ishwarsridharan@yahoo.com

முன்னுரை

பின்னாளில் இடம் சாரா இலக்கணம் (*context-free grammar*) என்று அறியப்பட்ட முறையை நோம் சாம்சுக்கி 1956-ல் சொற்றொடர் அமைப்பு இலக்கணம் *(phrase structure grammar)* என்ற பெயரில் இயல் மொழிகளின் இலக்கணங்களைக் குறிக்கும் நோக்கில் அறிமுகம் செய்தார். சுருங்குறித்தொடர்களைக் கொண்டு இயற்றப்படும் சீருற்ற இலக்கணங்களைக் *(regular grammars)* காட்டிலும் இடம் சாரா இலக்கணங்கள் பகர்திறன் மிகுதியாகக் கொண்டவை என்று அவர் நிறுவினார். ஆனால் ஆங்கிலம் உட்பட எந்த ஒரு இயல் மொழியின் இலக்கணத்தையாவது முழுமையாக இடம் சாரா இலக்கணத்தைக் கொண்டு வரையறுக்க இயலுமா என அவரால் உறுதிபடக் காட்ட முடியவில்லை. இன்றைய ஆய்வர்கள் இயல்மொழி இலக்கணங்கள் இடம் சாராதவை அல்ல என்றே இணங்கியுள்ளனர். அதே வேளையில் இயல்மொழிகளின் பெரும்பகுதி இடம்சாரா இலக்கணம் கொண்டது என்று காட்டியுள்ளனர். இது போன்ற பகுதிகளுக்கான இடம் சாரா இலக்கணங்களாக எழுதியுள்ளனர்.

இவ்வாய்வுரை தொல்காப்பியத்தின் எழுத்ததிகாரத்துக்கான இடம் சாரா இலக்கணம் இயற்றுதற்கான ஒரு முயற்சியை விவரிக்கிறது. செம்மொழி இலக்கணத்தை தொல்காப்பியம் வரையறை செய்யும் விதமும் முறையும் இப்பணிக்கு வெகு இணக்கமாக அமைந்துள்ளது. தவிர யாப்பருங்கலக்காரிகையை அடிப்படையாகக் கொண்டு வெண்பாக்களை அலகிடும் இடம் சாரா இலக்கண அடிப்படை பகுப்பாய்வி இம்முயற்சிக்கான வாய்ப்பை வலுப்படுத்துகிறது.

எழுத்ததிகாரப் பகுதிகளுக்கான இடம் சாரா இலக்கணம்

பின்வருபவை எழுத்ததிகாரத்தின் சில பகுதிகளுக்கான இடம் சாரா இலக்கணத்தைக் காட்டுகின்றன.

1.முதலெழுத்து -

சொல்லின் முதலெழுத்தாக வரக்கூடியது

<முதலெழுத்து> -> <உயிர் எழுத்து>
<முதலெழுத்து> -> {க, த, ந, ப, ம உயிர்மெய்கள் அனைத்தும்}
<முதலெழுத்து> -> {சகர உயிர்மெய்கள் (அகர, ஐகார, ஔகாரம் நீங்கலாக)}
<முதலெழுத்து> -> {வகர உயிர்மெய்கள் (உகர, ஊகார, ஒகர, ஓகாரம் நீங்கலாக}
<முதலெழுத்து> -> {ஞகர உயிர்மெய்கள் ஆகார, எகர, ஒகரம் மட்டும்}
<முதலெழுத்து> -> {யகர உயிர்மெய்கள் 'யா' மட்டும்}

குறிப்புகள்:

1. சுருக்கம் வேண்டி பேக்கசு-நார் முறைக் குறியீட்டிலிருந்து விலகி பொதுவாகப் புரிந்து கொள்ளத்தக்க அணிக் குறிகளையும் பயன்படுத்தியுள்ளோம்.
2. இப்பகுதி முழுமையையும் சீருறு இலக்கணமாகவே எழுதி விடலாமென்றாலும் இடம் சாரா இலக்கண உருவகங்களைக் கொண்டு எழுதியுள்ளோம். இதன்வழி மேலும் உயர்நிலை இலக்கண நெறிகளை இவற்றைக் கொண்டு விரித்தெழுத ஏதுவாகிறது.

2. ஈரொற்று உடனிலை

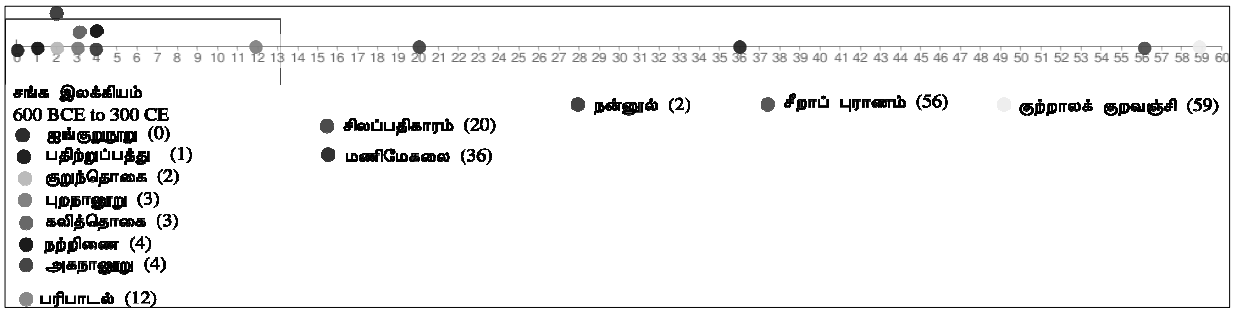
<ஈரொற்று உடனிலை> -> <முன் ஒற்று அசை><பின் ஒற்று>
<முன் ஒற்று அசை> -> <அசை> <உயிரெழுத்து><ய்>
<முன் ஒற்று அசை> -> <அசை> <உயிர்மெய்யெழுத்து><ய்>
<முன் ஒற்று அசை> -> <உயிரெழுத்து><ய்>
<முன் ஒற்று அசை> -> <உயிர்மெய்யெழுத்து><ய்>
<முன் ஒற்று அசை> -> <நெடில்><ர், ழ்>
<முன் ஒற்று அசை> -> <அசை><குறில்><ர், ழ்>
<அசை> -> <அசை><அசை>
<அசை> -> <அசை><ஒற்று>
<அசை> -> <குறில்>
<அசை> -> <நெடில்>
<அசை> -> <குறில்><குறில்>
<அசை> -> <குறில்><நெடில்>
<பின் ஒற்று> -> <க், ச், த், ப், ங், ஞ், ந், ம்>
<மெய்யொலிக் கூட்டம்> -> ன்ம்$

குறிப்புகள்:

1. அசைகளின் தொடரையும் சுருக்கம் வேண்டி அசையெனக் குறித்துள்ளோம்.
2. தொடர் முடிதலை $ குறி கொண்டு சொல்லியுள்ளோம்.
3. குறில், நெடில், உயிரெழுத்து, உயிர்மெய்யெழுத்து, ஒற்று போன்றவற்றை விரித்து எழுதவில்லை.

பயன்கள்

இயல்மொழிகளுக்கான இடம் சாரா இலக்கணங்களை பகுப்பாய்விகள் எழுதுவதற்கும், எழுத்துணரிகளில் மயக்கம் களைதற்கும், உரைச்செயலிகளில் பிழைதிருத்தம் செய்யும் கருவிகளுக்கும், இன்னும் பலவற்றிற்கும் பயன்படுத்தியுள்ளனர். இவை தவிர, இக்கட்டுரையில் பழந்தமிழ் இலக்கியங்களைக் கால வரிசைப் படுத்துவதற்கும் கூட இவற்றைப் பயன்படுத்தலாம் எனக் காட்டியுள்ளோம். எளிமை பொருட்டு சகர, சைகார, சௌகார உயிர்மெய்கள் முதலெழுத்தாக வாரா என தொல்காப்பியம் கூறும் நெறியை எடுத்துக் கொண்டுள்ளோம். பின்னாளில் தமிழில் சேர்ந்த சொற்களில் இந்நெறிமுறை முழுவதுமாகப் பின்பற்றப்படவில்லை. வெவ்வேறு இலக்கியங்களில் இந்நெறிப்பிறழ்வுகளின் எண்ணிக்கையை ஒரு எண்வரிசைக் கோட்டில் குறித்துப் பார்த்ததில் பின்வரும் நிலையைக் காண முடிகிறது. இவற்றில் சில தொகுதிகள் இருப்பதைக் காண முடிகிறது.



இவ்வாய்வில் கோப்புகளில் உள்ளபடியே இருந்த சீர்களை எடுத்துக் கொண்டதால் இடையில் வரும் முதலெழுத்துக்கள் விடுபட்டுள்ளன என்றாலும் இது ஓரளவு பயன் தரக்கூடும்.

தற்போதைய நிலையும் இனி வருவதும்

தற்போதைய நிலையில் எழுத்ததிகாரத்தின் குறிப்பிடத்தக்க அளவு நெறிகளை இடம் சாரா இலக்கணமாக எழுதியுள்ளோம். புணர்ச்சியில் சாரியை பெற்று வருவதை எப்படிக் குறிப்பது என எண்ணிப் பார்த்து வருகிறோம். தேவை ஏற்பட்டால் இடம் சாரா இலக்கணத்தைக் காட்டிலும் கூடுதல் பகர்வுத்திறன் கொண்ட இலக்கண முறைகளில் எழுதத் திட்டமிட்டுள்ளோம். நிறைவடையாத நிலையில் இவ்விலக்கணம் பயன்படுமா என்ற கேள்வி எழக்கூடும். பகுதி இலக்கணமும், பகுதி உரைப்புள்ளியியல் உதவியுடனும் இயங்கும் பகுப்பாய்விகளை உருவாக்க முடியுமென ஆய்வுகள் காட்டியுள்ளன. சொற்றொடர் அமைப்பு கலப்புமுறை மொழி மாதிரிகளைக் கொண்டு பகுப்பாயுதல், இடம்சாரா இலக்கணத்தையும் கிளவி எண்ணிக்கைப் புள்ளிகளையும் கொண்டு பகுப்பாயுதல், சொற்றொடர் எண்ணிக்கையையும் இடம்சாரா நெறிகளின் நிகழ்தகவுகளைக் கொண்டுப் பகுப்பாயுதல் எனப் பல முறைகளில் இவ்விலக்கணம் பயன்படும். பகுதி பகுதியாக பகுப்பாய்வது கணினியின் நினைவகத் தேவையின் அளவைக் குறைப்பதும், கையாளுவதையும் விரிவுபடுத்துவதையும் எளிமைப்படுத்துவதும் அறியப்பட்டுள்ளது.

இயல்மொழிகளுக்கான பிற இலக்கணங்கள்

பாணினியின் வடமொழி இலக்கணத்தின் பல பகுதிகளை இடம் சாரா இலக்கணங்களாகவும் பிற கட்டுக்கோப்பான இலக்கண முறைகளிலும் எழுதியுள்ளனர். உருபன்களைப் பிரித்துணரும் மென்பொருட்கள் ஒலியியல் மற்றும் உருபனியல் நெறிகளைக் கொண்டு உருவாக்கப்பட்டுள்ளன. இதைத் தவிர, ஈடான அணிகளினால் ஆன இலக்கணம் ஒன்றும் தமிழ்ச் சொற்றொடர்களுக்கென எழுதப்பட்டுள்ளது. வெண்பா இலக்கணத்துக்கான இடம்சாரா இலக்கணமும், அதையொட்டிய அலகீட்டு மென்பொருளும் உள்ளன.

முடிவுரை

தொல்காப்பியத்தின் எழுத்ததிகாரத்துக்கான இடம்சாரா இலக்கணம் எழுதும் முயற்சியை இக்கட்டுரை எடுத்துரைக்கிறது. நிறைவடையாத இலக்கணங்களைக்கூட வழக்கமான மொழியியல் பயன்பாட்டுக்கு எப்படிக் கொண்டு வருவது என்பதற்கான பல எடுத்துக்காட்டுக்கள் சுட்டப்பட்டுள்ளன. வழக்கமான பயன்களைத் தவிர மொழியின் படிவளர்ச்சியை அறியவும், இலக்கியங்களைக் காலக்கோட்டில் குறிப்பதற்கும்கூட இதைப் பயன்படுத்தலாமெனக் காட்டப்பட்டுள்ளது.

மேற்கோள்கள்

1. Chomsky, Noam (1956). "Three models for the description of language". IRE Transactions on Information Theory (2): 113–124.
2. Shieber, Stuart (1985). "Evidence against the context-freeness of natural language". Linguistics and Philosophy 8: 333–343. doi:10.1007/BF00630917.
3. Pullum, Geoffrey K.; Gerald Gazdar (1982). "Natural languages and context-free languages". Linguistics and Philosophy 4: 471–504. doi:10.1007/BF00360802
4. L, BalaSundaraRaman; Ishwar.S, Sanjeeth Kumar Ravindranath (2003-08-22). "Context Free Grammar for Natural Language Constructs - An implementation for Venpa Class of Tamil Poetry". Proceedings of Tamil Internet, Chennai, 2003. International Forum for Information Technology in Tamil. pp. 128-136.
5. M.G.J. van den Brand, M.P.A. Sellink, and C. Verhoef (2000). "Generation of Components for Software Renovation Factories from Context-free Grammars". Science of Computer Programming

36: 209-266.
6. "VisaiNeri". SourceForge. http://sourceforge.net/projects/visaineri/. Retrieved on 2009-08-15.
7. M. Selvam, N. AM, and R. Thangarajan, "Structural Parsing of Natural Language Text in Tamil Using Phrase Structure Hybrid Language Model," International Journal of Computer, Information, and Systems Science, and Engineering 2: 4.
8. E. Charniak, "Statistical parsing with a context-free grammar and word statistics," in Proceedings of the National Conference on Artificial Intelligence, 1997, 598–603.
9. D. Linares, J. M Bened'ı, and J. A Sánchez, "A hybrid language model based on a combination of n-grams and stochastic context-free grammars," ACM Transactions on Asian Language Information Processing (TALIP) 3, no. 2 (2004): 113–127.
10. H. Meng et al., "GLR parsing with multiple grammars for natural language queries," ACM Transactions on Asian Language Information Processing (TALIP) 1, no. 2 (2002): 123–144.
11. M. D Hyman, "From Paninian Sandhi to Finite State Calculus," Sanskrit CL 2008 (2007): 253–265.
12. P. M Scharf, "Modeling Paninian Grammar," in Proceedings of First International Symposium on Sanskrit Computational Linguistics, 77.
13. V. Ranganathan, "Development of Morphological Tagger for Tamil," in (presented at the Tamil Internet 2001, INFITT, 2001), http://www.infitt.org/ti2001/papers/vasur.pdf.
14. Arulmozhi, P., Sobha, L, Kumara Shanmugam. B. (2004) "Part of Speech Tagger for Tamil" In the Proceedings of Symposium on Indian Morphology, Phonology and Language Engineering, IIT Khadagpur, pp. 55-57, India.

MORPHOLOGICAL TAGGER

Amrita Morph Analyzer and Generator For Tamil

A Rule Based Approach

Dr.A.G. Menon, Amrita and Leiden (Netherland)
S. Saravanan, R. Loganathan and Dr. K. Soman (Amrita University, Coimbatore, India)

Email: menon.govindankutty@gmail.com / sarwanster@gmail.com
loganathn@gmail.com kp_soman@amrita.edu

The Context

From 2006 CEN (Centre of Excellence for Engineering and Networking) of the AMRITA University, Ettimadai, Coimbatore under the guidance of Prof. K. Soman, is engaged in research and development in the field of Natural Language Processing (NLP). It is a young and dynamic university. AMRITA is part of a consortium of six IITs and two IIITs and CDAC, Pune, which are involved in the research and development of tools for the translation of English to Indian Languages, which is funded by DIT. A new project on Machine Translation is started recently (May 2009) with the funding of the MHRD for developing linguistic resources and machine translation tools. AMRITA is also developing its own engine for Machine Translation. The present Amrita Morph Analyzer and Generator (AMAG) for Tamil is an independent project carried out in CEN.

The Need

More than a dozen Tamil Morphological Analyzers and Generators are announced through the Internet and websites of many renowned institutions. The only DEMO version available is ATCHARAM displayed on the website of the IT Ministry, Resource Centre for Indian Language Technological Solutions – Tamil. However, none is available for our research and development from the open source. This deplorable situation has compelled us to build our own MAG for developing a system for the MT and other NLP applications.

Morphology for Computer

Morphology deals, primarily, with the structure of words. Morphological analysis detects, identifies and describes the meaningful constituent morphs in a word, which function as building blocks of a word. The densely agglutinative Dravidian languages such as Tamil, Malayalam, Telugu and Kannada display a unique structural formation of words by the addition of suffixes representing various senses or grammatical categories, after the roots or stems. The senses such as person, number, gender and case are linked to a Noun stem in an orderly formation.

Verbal categories such as transitive, causative, tense and person, number and gender are added to a verbal root or stem. The morphs representing these categories have their own slots behind the roots or stems. The highly complicated nominal and verbal morphology do not stand alone. It regulates the direct syntactic agreement between the subject and the predicate. Another important aspect of the addition of morphs is the change which often takes place in the space between these morphs and within a stem.

A Morphological Analyzer and Generator (MAG) should take care of these changes while assigning a suitable morph to the correct slot to generate a word. The combination of sense and form in a morph and the possibility to identify the governing rules are the incentives to attempt to build an engine

which can automatically analyse and generate the same processes taking place in the brain of a native speaker.

Challenges in Building a Morph Analyzer and Generator For Tamil

The slots behind the root/stem can be filled by many morphs. The rules governing the order of the morphs in a word and the selection of the correct morph for the correct slot should be formulated for analysis and synthesis. The inflections and derivations are not the same for all the nouns and verbs. The biggest challenge is the grouping of nouns and verbs in such a way that the members of the same group have similar inflections and derivations. Otherwise one has to make rules for each noun and verb, which is not feasible. The most difficult slot in a verb is the one which follows the verb root/stem. This position is occupied by the suffixes belonging to the category transitive. The elusive behaviour of these suffixes poses many problems, and most of the earlier Morphological Analyzers did not handle this problem adequately. Our system, as mentioned earlier, works on rules and these rules are capable of solving this clumsiness in an elegant manner.

Many changes take place at the boundaries of morphs and words. Identifying the rules that govern these changes is a challenge because dissimilar changes take place in similar contexts. In such cases it is necessary to look into the phonological as well as morphological factors which induce such changes.

The system we design involves building an exhaustive lexicon for noun, verb and other categories. The performance is directly related to this exhaustiveness. It is a laborious task.

Structure of a Dravidian Verb

<table>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td rowspan="3">Root/Stem</td><td rowspan="2">Intransitive</td><td>Personal object base</td><td rowspan="2">tense/mode</td><td rowspan="3">Personal endings (person, number and gender)</td></tr>
<tr><td>plural action base</td></tr>
<tr><td>Transitive</td><td>motion base</td><td>negative</td></tr>
</table>

The third position is not relevant for Tamil verbs.

The structure of a Tamil verb is given below:

The finite verb: Root/Stem + Transitive + Causative + Tense / Negative + Empty + PNG
Clitics can be added after the Person Number and Gender (PNG) marker.
Non -Finite Verb:
Root/Stem + Transitive + Causative + Negative + Infinitive / Conditional infinitive suffix
Root/Stem + Transitive + Causative + Tense / Negative + Relative Participle / Verbal Participle / Conditional Verbal Participle
Root/Stem + Transitive + Causative + Negative + Verbal Noun suffix

The above descriptions mention only the slots on the right side of the root/stem. Apart from this, many non-finite verbs occur on the left side of the root/stem and form complex verb structures such as main + auxiliary verbs.

Predictability of the Verb Suffixes

One of the challenges is the predictability of the suffixes which fill the three slots after the verb root/stem: transitive, causative and tense. There are two types of verbs: verbs which have intransitive and transitive contrast such as *tāḻntāṉ* as against *tāḻttiṉāṉ* and such as *cirittāṉ* without such contrast. We can divide the verbs broadly into three groups on the basis of the past tense suffixes *–nt-* , *-iṉ-* and *–t-*. They can be further divided into eight groups taking into account the first three positions after the root/stem. The fillers of each position are determined by the verb root/stem. As far as the non-finite forms are concerned, the predictability of the verbal noun suffixes is an important task of MAG.

Structure of a Noun

Stem
Stem + Formative / Oblique suffix
Stem + Formative / Oblique suffix + Case marker
Stem + Formative / Oblique suffix + Empty suffix + Case marker
Stem + Formative suffix + Plural + Case marker
Stem + Formative /Oblique suffix + Pronominal suffix

The closing slot can be followed by a clitic such as *–um* or interrogative morph such as *–ā*, *-ē* or *-ō*.

Handling the Noun Suffixes

The suffixes occupying the slots after the stem do not vary according to the stem. There is no direct relationship between them unlike the verb stems. However, the noun stems themselves vary before they take suffixes. This phenomenon is limited to a small number of groups of nouns. We have mentioned above that some of the stems take an oblique suffix before the addition of case markers. Since they are identifiable on the basis of their endings or specific phonological features of the stems, it is also easier to make rules for the changes which take place within the stem. For example, forms like *maram* 'tree' become *maratt-* and *nāṭu* becomes *nāṭṭ-*. The first and second person pronouns have also two different forms along with the third person neuter plural pronoun. However, when nouns like *kal* are preceded by another noun, problems arise in handling the *sandhi* rules because these rules are based on the phonemic makeup of the final noun alone. Creating rules governing the distribution of case suffixes is an important step towards the building of a MAG. The noun morphology is relatively less complex than the verb morphology.

Technology

Finite State Transducer (FST) is used for morphological analyzer and generator. We have used AT &T Finite State Machine to build this tool. FST maps between two sets of symbols. It can be used as a transducer that accepts the input string if it is in the language and generates another string on its output. The system is based on lexicon and orthographic rules from a two level morphological system. For the Morphological generator, if the string which has the root word and its morphemic information is accepted by the automaton, then it generates the corresponding root word and morpheme units in the first level (Fig 1). The output of the first level becomes the input of the second level where the orthographic rules are handled (Fig 3), and if it gets accepted then it generates the inflected word.

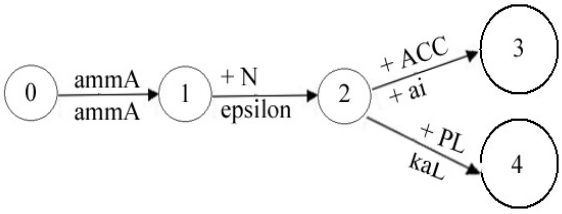


Fig 1: Morphotactics Rule

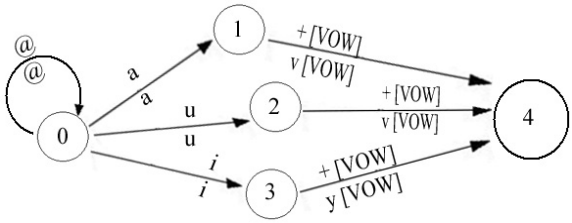


Fig 2: Sandhi Rule

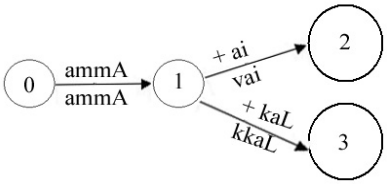


Fig 3: Application of Sandhi Rule

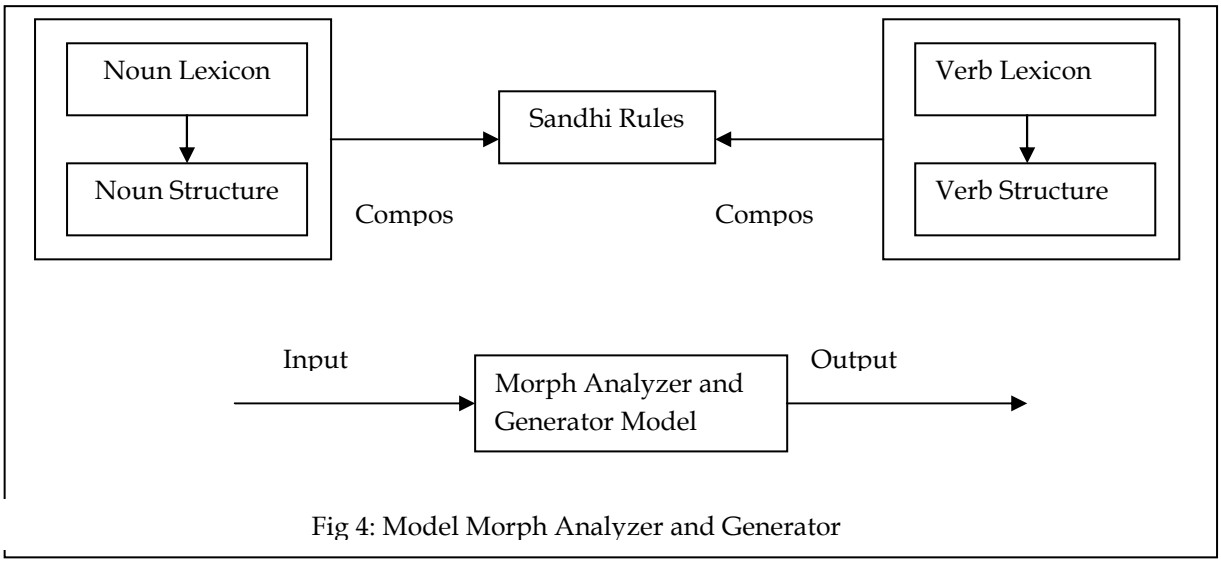


Fig 4: Model Morph Analyzer and Generator

Conclusion

At present we started with a list of fifty thousand nouns, around three thousand verbs and a relatively smaller list of adjectives. Our MAG is capable of analysing and generating more grammatical categories than ATCHARAM. In the future we are planning to expand our lexicons for more exhaustiveness. The uniqueness of our MAG is its capacity to generate and analyse transitive, causative and tense forms apart from the passive constructions, auxiliaries and verbal nouns. A demo version of AMAG will be soon uploaded for testing.

Reference

1. Anandan, P., Ranjani Parthasarathy & Geetha, T.V., 2001. "Morphological Generator for Tamil", Tamil Internet 2001 Conference, Kuala Lumpur, Malaysia.
2. Anandan, P., Ranjani Parthasarathy & Geetha, T.V., 2001. "Morphological Analyser for Tamil", *ICON 2002*, RCILTS-Tamil, Anna University, India.

3. Beesley, Kenneth R., 1996. "Arabic Finite-State Morphological Analysis and generation", *Proceedings of the 16th International Conference on Computational Linguistics*, Vol. 1.Copenhagen, Denmark. pp. 89-94.
4. Beesley, Kenneth R. & Karttunen, Lauri, 2003. *Finite State Morphology*, Stanford, CA: CSLI Publications.
5. Koskenniemi, Kimmo., 1984. "General Computational Model for Word-Form Recognition and Production", *COLING* 84. pp. 178-181.
6. Lakshmana Pandian, S and T.V. Geetha, 2008. "Morpheme based Language Model for Parts-of-Speech Tagging", *POLIBITS – Research Journal on Computer Science and Computer Engineering with applications,* Volume 38 (July-December 2008), Mexico. pp. 19-25.
7. Menon, A.G., 1976. "Tamil Verb Classification", *Actes du XXIXe Congres international des Orientalistes* 1973. Inde Ancienne, Vol. II.3. Etudes Dravidiennes. Paris: L'Asiatheque pp. 136-148.
8. Menon, A.G., 1988. "Tamil Verb Stem Formation", *International Journal of Dravidian Linguistics*, Vol. 27, part 1. Trivandrum: International Association of Dravidian Linguistics. pp. 13-40.
9. Menon, A.G. & Schokker, G.H., 1990. "Linguistic Convergence: the Tamil-Hindi auxiliaries", *Bulletin of the School of Oriental and African Studies*. London: University of London. pp. 266-282.
10. Renganathan, Vasu, 2001. "Development of Morphological Tagger for Tamil", *Tamil Internet 2001 Conference*, Kuala Lumpur, Malaysia (26-28 August 2001).

A Novel Approach to Morphological Analysis for Tamil Language

Anand kumar M, Dhanalakshmi V, Soman K P
Amrita Vishwa Vidyapeetham, Ettimadai, Coimbatore, Tamilnadu, India .

Rajendran S
Tamil University, Thanjavur, Tamilnadu, India .

Email: m_anandkumar, v_dhanalakshmi, kp_soman@ettimadai.amrita.edu, raj_ushus@yahoo.com

Abstract: This paper presents the morphological analysis for complex agglutinative Tamil language using machine learning approach. Morphological analysis is concerned with retrieving the structure, syntactic rules, morphological properties and the meaning of a morphologically complex word. The morphological structure of an agglutinative language is unique and capturing its complexity in a machine analyzable and generatable format is a challenging job. Generally rule based approach is used in building morphological analyzer. In rule based approach what works in the forward direction may not work in the backward direction. The Novel approach to morphological analyzer is based on sequence labeling and training by kernel methods. It captures the non-linear relationships and various morphological features of Tamil language in a better and simpler way. The efficiency of our system is compared with the existing morphological analyzers which are available in net. Regarding the accuracy our system significantly outperforms the existing morphological analyzer and achieves a very competitive accuracy of 95.65% for Tamil language.

Introduction

Morphological analysis is the process of segmenting words into morphemes and analyzing the word formation. It is a primary step for various types of text analysis of any language. Morphological analyzers are used in search engines for retrieving the documents from the keyword (Daelemans Walter et al., 2004). Morphological analyzer increases the recall of search engines. It is also used in speech synthesizer, speech recognizer, lemmatization, noun decompounding, spell and grammar checker and machine translation.

Tamil language is morphologically rich and agglutinative. Each root word is affixed with several morphemes to generate word forms. Generally, Tamil language is postpositionally inflected to the root word. Computationally each root word can take a few thousand inflected word forms, out of which only a few hundred will exist in a typical corpus. For the purpose of analysis of such inflectionally rich languages, the root and the morphemes of each word have to be identified. Generally rule based approaches are used for building morphological analyzer (Rajendran.S et al., 2001). We have implemented a novel method for the morphological analysis of the Tamil language using machine learning approach.

Challenges in Morphological Analyzer for Tamil

The morphological structure of Tamil is quite complex since it inflect to person, gender, and number markings and also combines with auxiliaries that indicate aspect, mood, causation, attitude etc in verb. Noun inflects with plural, oblique, case, postpositions and clitics suffixes. For the purpose of analysis of such inflectionally rich languages, the root and the morphemes of each word have to be identified. 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 sandi rules which account for the shape changes due to inflection. Classification of Tamil verbs based on tense inflections is evolved. The inflection includes finite, infinite, adjectival, adverbial and conditional forms of verbs (Rjendran.S et al., 2001). For the computational need we have made thirty two paradigms of verb.

Compared to verb morphological analysis noun morphological analysis is less challenging. Noun can occur separately or with plural, oblique, case, postpositions and clitics suffixes. We have classified seventeen noun paradigms to resolve the challenges in noun morphological analysis. Based on the paradigm we classified the root words into its group. We have prepared the corpus with all morphological feature information. So the machine by itself captures all morphological rules. Finally the morphological analysis is redefined as a classification task which is solved by using sequence labeling approaches.

Creating Data for supervised Learning

Nowadays machine learning approaches are directly applied to all the natural language processing tasks machine learning approaches can be supervised or unsupervised. In supervised learning set of input and output examples are used for training. So, data creation plays the key role in supervised machine learning approaches.

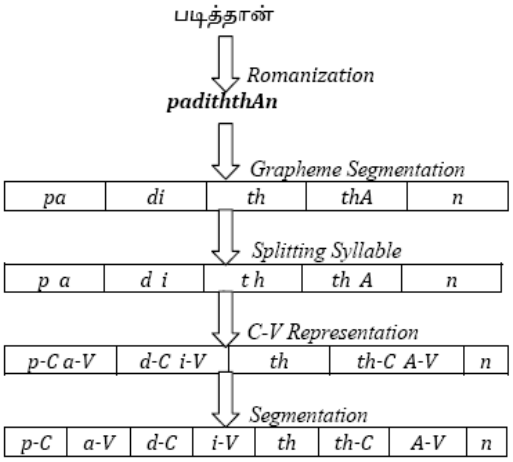


Figure 1: Preprocessing steps

The first step involved in the corpora development for morphological analyzer is classifying paradigms for verbs and nouns. The classification of Tamil verbs and nouns are based on tense markers and case markers respectively. Each paradigm root word will inflect with the same set of

inflections. The second step is to collect the word list for each noun and verb paradigm. Figure 1 explains the preprocessing steps involved in the development of morphological corpus.

Morphological corpus which is used for machine learning is developed by following steps:

Romanization: The set of most commonly used noun and verb forms are generated manually for input structure and similarly the output structure is developed. These data are converted to Romanized forms using the Unicode to Roman mapping file.

Segmentation: After Romanization each and every word in the corpora is segmented based on the Tamil grapheme and additionally each syllable in the corresponding word is further segmented into consonants and vowels. In segmented syllable append "–C" and "–V" to the consonant and vowel respectively. We name it as C-V representation i.e. Consonant – Vowel representation. Morpheme boundaries are indicated by "*" symbol in output data.

Alignment and mapping: The segmented words are aligned vertically as segments using the gap between them. And the input segments are consequently mapped with output segments. Sample data format is given in the fig.2 .First one represents the input data and the latter one represents output data."*" indicates the morpheme boundaries.

Input	p-C a-V d-C i-V th th-C A-V n
Output	p a d i* th th* A n*

Figure 2: Sample Data Format

Mismatching: It is the key problem in mapping between the input and output data. Mismatching occurs in two cases i.e., either the input units are larger or smaller than that of the output units. This problem is solved by inserting null symbol "$" or combining two units based on the morph-syntactic rules to the output data. And the input segments are mapped with output segments. After mapping machine learning tool is used for training the data.

Case 1:

Input Sequence:

P-C | a-V | d-C | i-V | k | k-C | a-V | y-C | i-V | y-C |a-V | l-C | u-V | m (14 segments)
Mismatched Labels:
p | a | d | i* | k | k | a* | i | y | a | l* | u | m* (13 segments)
Corrected labels:
p | a | d | i* | k | k | a* | $ | i | y | a | l* | u | m* (14 segments)

In case 1 input sequence is having more number of segments than the output sequence. For the Tamil verb padikkayiyalum is having 14 segments in input sequence but in output it has only 13 segments. The second occurrence of "y" in the input sequence becomes null due to the morphosyntactic rule. So there is no segment to map with "y". For this reason, in training data "y" is mapped with "$" symbol ("$" indicates null).Now the input and the output segments are equalized.

Case 2:

Input Sequence:

O | d-C | i-V | n-C | A-V | n (6 segments)

Mismatched Labels:

O | d | u* | i | n* | A | n (7 segments)

Corrected labels:

O | du* | i | n* | A | n (6 segments)

In case 2 the input sequence is having less number of segments than the output sequence. Tamil verb OdinAn is having 6 segments in input sequence but output has 7 segments. Due to morphosyntactic change the segment "d-C" in the input sequence is mapped to two segments "d" &"'u*" in output sequence. For this reason, in training "d-C" is mapped with "du*".Now the input and the output segments are equalized and thus the problem of sequence mismatching is solved.

Implementation of Morphological Analyzer

Using machine learning approach the morphological analyzer for Tamil is developed. We have developed separate engines for noun and verb. Morphological analyzer is redefined as a classification task using the machine learning approaches. Three phases are involved in our morphological analyzer.

- Preprocessing
- Segmentation of morphemes
- Identifying morpheme

Fig.3 gives an outlook of the morphological analyzer system. In this machine learning approach two training models are created for morphological analyzer. These two models are represented as model-I and model-II. First model is trained using the sequence of input characters and their corresponding output labels. This trained model-I is used for finding the morpheme boundaries. Second model is trained using sequence of morphemes and their grammatical categories. This trained model-II is used for assigning grammatical classes to each morpheme.

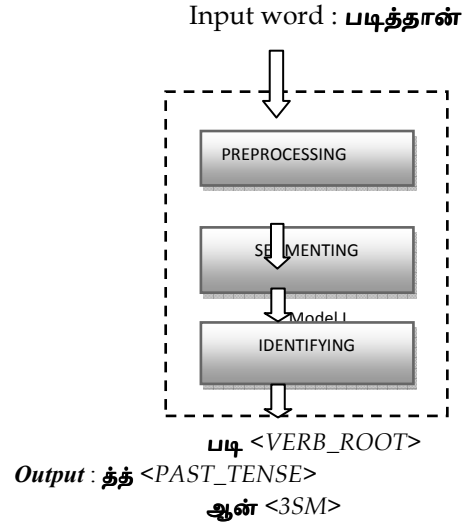


Figure 3: Schematic Representation

Preprocessing: The surface form is converted into sequence of units which is given as the input for the

morphological analyzer tool. The input word is converted into input segments and syllables are identified. Using C-V representation consonants and vowels are represented to the syllable.

Segmentation of Morphemes: Preprocessed words are segmented into morpheme according to the morpheme boundary. The input sequence is given to the trained model-I. The trained model predicts each label to the input segments.

Identifying Morpheme: The Segmented morphemes are given to the trained model-II. It predicts grammatical categories to the segmented morphemes. We have trained the system to give multiple outputs to handle the compound words.

Morphological Analyzer Using Machine Learning

In morphological analysis a complex word form is transformed into root and suffixes. Generally rule based approaches are used for morphological analysis which are based on a set of rules and dictionary that contains root and morphemes. For example a complex word is given as an input to the morphological analyzer and if the corresponding morphemes are missing in the dictionary then the rule based system fails (Daelemans Walter et al., 2004). In rule based approaches every rule is depends on the previous rule. So if one rule fails, it will affect the entire rule that follows. In machine learning all the rules including complex spelling rules are also handled by the classification task. Machine learning approaches don't require any hand coded morphological rules (Daelemans Walter et al., 2004). It needs only corpora with linguistical information. These morphological or linguistical rules are automatically extracted from the annotated corpora. Here input is a word and output is root and inflections. Input word is denoted as 'W' root and inflections are denoted by 'R' and 'I' respectively.

[W]Noun/Verb = [R] Noun/Verb + [I] Noun/Verb

Machine learning using SVM

Support vector approaches have been around since the mid 1990s, initially as a binary classification technique, with later extensions to regression and multi-class classification. Here Morphological problem is converted into classification problem. These classifications can be done through supervised machine learning approach.

Support Vector Machine is a new approach to supervised pattern classification which has been successfully applied to a wide range of classification problems. SVM is based on strong mathematical foundations and results in simple yet very powerful algorithms. SVMs are learning systems that use a hypothesis space of linear functions in a high dimensional feature space, trained with a learning algorithm from optimization theory that implements a learning bias derived from statistical learning theory.

SVMTool

SVMTool is an open source generator of sequential taggers based on Support Vector Machine. Generally SVMTool is developed for POS tagging but here this tool is used in morphological analyzer for classification task. The SVMTool software package consists of three main components, namely the model learner (SVMTlearn), the tagger (SVMTagger) and the evaluator (SVMTeval). SVM models (weight vectors and biases) are learned from a training corpus using the SVMTlearn component (Jes´us Gim´enez and Llu´ıs M`arquez, 2006). Different models are learned for the different strategies. Finally, given a correctly annotated corpus, and the corresponding SVMTool predicted annotation, the SVMTeval component displays tagging results. SVMTeval evaluates the performance in terms of

accuracy.

System Evaluation

Efficiency of the system is compared. The morphological analyzer system for verb and noun are trained with 130,000 and 70,000 words respectively. This system is also tested with 40,000 verb forms and 30,000 nouns from an Amrita POS Tagged corpus (Dhanalakshmi.V et al., 2008). The SVM based machine learning tool affords better results compare to MBT and CRF++. Training time is very less in MBT compare to SVM and CRF++. But in testing SVM holds good. The outputs which are incorrect are noticed and its corresponding input words are added in the training file and trained again. This increases the efficiency of our system. This is the main advantage of using machine learning approach to rule based approach.

Conclusion

This paper has described the morphological analyzer based on the new and state of the art machine learning approaches. We have demonstrated a new methodology adopted for the preparation of the data which was used for the machine learning approaches. We have not used any morpheme dictionary but from the training model our system has identified the morpheme boundaries. The accuracy obtained from the different machine learning tools shows that SVM based machine learning tool gives better result than other machine learning tools. A GUI to enhance the user friendliness of the morphological analyzer engine was also developed using Java Net Beans. We are currently implementing the same methodology for the other Dravidian languages like Malayalam, Telugu, and Kannada. Preliminary experimentation gave promising results. We are confident that the proposed method is general enough to be applied for any languages.

References

1. Anandan. P, Ranjani Parthasarathy, Geetha T.V.2002. Morphological Analyzer for Tamil, ICON 2002, RCILTS-Tamil, Anna University, India.
2. Daelemans Walter, G. Booij, Ch. Lehmann, and J. Mugdan (eds.)2004 , Morphology. A Handbook on Inflection and Word Formation, Berlin and New York: Walter De Gruyter, 1893-1900
3. Dhanalakshmi V, Anandkumar M, Vijaya M.S, Loganathan R, Soman K.P, Rajendran S,2008, Tamil Part-of-Speech tagger based on SVMTool, Proceedings of the COLIPS International Conference on Asian Language Processing 2008 (IALP), Chiang Mai, Thailand. 2008: 59-64.
4. Jes´us Gim´enez and Llu´ıs M`arquez,2006 SVMTool:Technical manual v1.3, August 2006.
5. John Goldsmith. 2001. Unsupervised Learning of the Morphology of a Natural Language. Computational Linguistics, 27(2):153–198.
6. Rajendran, S., Arulmozi, S., Ramesh Kumar, Viswanathan, S. 2001. Computational morphology of verbal complex. Paper read in Conference at Dravidan University, Kuppam, December 26-29, 2001.

POS Tagger and Chunker for Tamil Language

Dhanalakshmi V, Anand kumar M, Soman K P
Amrita Vishwa Vidyapeetham, Ettimadai, Coimbatore, Tamilnadu, India.

Rajendran S
Tamil University, Thanjavur, Tamilnadu, India.
Email: v_dhanalakshmi, m_anandkumar, kp_soman@ettimadai.amrita.edu, raj_ushus@yahoo.com

Abstract: This paper presents the Part Of Speech tagger and Chunker for Tamil using Machine learning techniques. Part Of Speech tagging and chunking are the fundamental processing steps for any language processing task. Part of speech (POS) tagging is the process of labeling automatic annotation of syntactic categories for each word in a corpus. Chunking is the task of identifying and segmenting the text into syntactically correlated word groups. These are done by the machine learning techniques, where the linguistical knowledge is automatically extracted from the annotated corpus. We have developed our own tagset for annotating the corpus, which is used for training and testing the POS tagger generator and the chunker. The present tagset consists of thirty-two tags for POS and nine tags for chunking. A corpus size of two hundred and twenty five thousand words was used for training and testing the accuracy of the POS tagger and Chunker. We found that SVM based machine learning tool affords the most encouraging result for Tamil POS tagger (95.64%) and chunker (95.82%).

Introduction

Part of speech (POS) tagging and chunking are well studied problems in the field of Natural Language Processing (NLP). Different approaches have already been tried to automate the task of POS tagging and chunking for English and other languages. The basic processing step consists of assigning POS tags to every token in the text. A subsequent step after POS tagging focuses on the identification of basic structural relations between groups of words in a sentence. This recognition is usually referred to as chunking. It is essential for many NLP tasks such as structure identification, information extraction, parsing and phrase based machine translation system. Chunker divides a sentence into its major-non-overlapping phrases and attaches a label to each chunk. Chunking falls between tagging and parsing. The structure of individual chunks is fairly easy to describe, while relations between chunks are harder and more dependent on individual lexical properties. The capability for a computer to automatically POS tag and chunk a sentence is very essential for further analysis in many approaches to the field of NLP. Many of the machine learning techniques and algorithms are used in this task. Our POS tagger and chunker based on machine learning techniques using SVM are trained and tested with the tagged corpus of size about two lakh and twenty five thousand words.

POS Tagging in Tamil

The Part of speech (POS) tagging is the process of labeling a part of speech or other lexical class marker to each and every word in a sentence. It is similar to the process of tokenization for computer languages. POS tagging is considered as an important process in speech recognition, natural language parsing, information retrieval and machine translation. Tamil being a Dravidian language has a very rich morphological structure which is agglutinative. Tamil words are made up of lexical roots

followed by one or more affixes. So tagging a word in a language like Tamil is very complex. The main challenges in Tamil POS tagging are solving the complexity and ambiguity of words [Dhanalakshmi V et al., 2009].

Various methodologies have been developed for POS Tagging in different languages. In case of Tamil language a rule-based POS tagger for Tamil was developed and tested [Arulmozhi et al., 2004]. This system gives only the major tags and the sub tags are overlooked while evaluation. A hybrid POS tagger for Tamil using HMM technique and a rule based system was also developed [Arulmozhi P and Sobha L, 2006]. Our POS tagger is based on machine learning techniques using SVM. We tagged our raw corpus of size about two hundred and twenty five thousand words using our Amrita tag set and then trained our corpus with the machine learning based SVMTool by tuning the parameters and feature patterns based on Tamil language. A raw corpus was tested using SVMTool and obtained an overall accuracy of 95.64%.

Customized POS Tagset

Many tagsets are already in existence for Tamil (AUKBC, Vasuranganathan tagset, CIIL Tagset for Tamil, etc). However, we encountered the following problems with these tagsets:

1. For each word, the grammatical categories as well as grammatical features are considered. Hence we need to split each and every inflected word in the corpus, which makes the tagging process very complex.
2. The number of tags is very large. This leads to increased complexity during POS tagging which in turn reduces the tagging accuracy.

For simple POS level, we wanted a tagset which has just the grammatical categories excluding grammatical features. Since the grammatical features can be obtained from the morphological analyzer. We needed a tagset with minimum tags without compromising on tagging efficiency. Hence we decided to create our own tagset for Tamil following the guidelines as mentioned in AnnCorra, Annotating Corpora Guidelines for POS and Chunk Annotation for Indian Languages [Akshar Bharati et al., 2006]. Our customized tagset uses only 32 tags. We do not consider the inflections or the grammatical features of the words. We use compound tag for compound nouns (NNC) and compound proper nouns (NNPC). We consider the tag VBG for verbal nouns and participle nouns. The tagset is shown in the figure below:

S.No	POS	Description		S.no	POS	Description
1	<NN>	NOUN		17	<VINT>	VERB INFINITE
2	<NNC>	COMPOUND NOUN		18	<CNJ>	CONJUNCTION
3	<NNP>	PROPER NOUN		19	<CVB>	CONDITIONAL VERB
4	<NNPC>	COMPOUND PROPER NOUN		20	<QW>	QUESTION WORD
5	<ORD>	ORDINALS		21	<COM>	COMPLEMENTIZER
6	<CRD>	CARDINALS		22	<NNQ>	QUANTITY NOUN
7	<PRP>	PRONOUN		23	<QTF>	QUANTIFIERS
8	<PRIN>	PRONOUN INTROGATIVE		24	<PPO>	POSTPOSITIONS
9	<PRID>	PRONOUN INDEFINITE		25	<DET>	DETERMINERS
10	<ADJ>	ADJECTIVE		26	<INT>	INTENSIFIER
11	<ADV>	ADVERB		27	<ECH>	ECHO WORDS
12	<VNAJ>	VERB NON FINITE ADJECTIVE		28	<EMP>	EMPHASIS
13	<VNAV>	VERB NON FINITE ADVERB		29	<COMM>	COMMA
14	<VBG>	VERBAL GERUND		30	<DOT>	DOT
15	<VF>	VERB FINITE		31	<QM>	QUESTION MARKS
16	<VAX>	VERB AUXILARY		32	<RDW>	REDUPLICATION WORDS

Figure 1. Amrita POS Tagset

Chunking in Tamil

A typical chunk consists of a single content word surrounded by a constellation of function words [S.Abney, 1991]. Chunks are normally taken to be a non recursive correlated group of words. Tamil being an agglutinative language have a complex morphological and syntactical structure. It is a relatively free word order language but in the phrasal and clausal construction it behaves like a fixed word order language. So the process of chunking in Tamil is less complex compared to the process of POS tagging. Various methodologies have been developed for chunking in different languages. In Tamil language TBL was used for text chunking [Sobha L et al., 2006]. vaanavil of RCILTS identifies the syntactic constituents of a Tamil sentence. Our Chunker is based on machine learning techniques (YamCha) using SVM.

Customized Chunk Tagset

We followed the guidelines mentioned in AnnCorra, while creating our tagset for chunking. Our Amrita chunking tagset contains nine tags. The tagset is described below:

Noun Chunks will be given the tag NP. It includes non-recursive noun phrases and postpositional phrases. The head of a noun chunk would be a noun. Noun qualifiers like adjective, quantifiers, determiners will form the left side boundary for a noun chunk and the head noun will mark the right side boundary for it. Examples for NP chunk are given below.

[அந்த <DET> (B-NP) அழகான <ADJ> (I-NP) பெண் <NN> (I-NP)] NP

An adjectival chunk is tagged as AJP. This chunk will consist of all adjectival chunks including the predicative adjectives. However, adjectives appearing before a noun will be grouped together with the noun chunk.

[திரைப்படம் <NN> (B-AJP) சார்ந்த <ADJ> (I-AJP)] AJP

Adverbial chunk <AVP> is tagged accordance with the tags used for POS tagging.

[அருகே <ADV> (B-AVP)]AVP

Conjunctions are the words used to join individual words, phrases, and independent clauses. It is labeled as CJP.

[ஆனால் <CNJ>(B-CJP)] CJP

Complimentizer are the words equivalent to the term subordinating conjunction in traditional grammer. For example, the word that is generally called a Complimentizer in English. In Tamil, enru and its variations falls into this category. Complimentizer is tagged in accordance with the tages used for POS tagging. It is tagged as COMP.

[என்று <COM> (B-COMP)] COMP

Verb chunks are mainly classified into Verb finite chunk and verb non-finite chunk. Verb finite chunk includes main verb and its auxiliaries. It is tagged as VFP. Examples for verb –finite chunk are given below.

[உள்ளது<VF> (B-VFP)] VFP

Non-finite verb comprise all the non-finite form of verbs. In Tamil we have four non-finite forms i.e., relative participle, adverbial participle, conditional and infinitive verb. It is tagged as VNP. Examples for verb non-finite chunk are given below.

[வெளிவந்த (VNAJ) (B-VNP)] VNP செய்திக் <NNC> < B-NP> குறிப்பு <I –NP> <NNC>

[விரைந்து <VNAV>(B-VNP)] VNP முடித்தான் <VF>

Gerundial forms are repersented by a seperate chunk. It is tagged as VGP. Example for gerundial chunk is given below.

தொழிற்சாலை <NN> [அமைப்பதில் <VBG>(B-VGP)] VGP தாமதம் <NN>

Symbols like .(Dot) and ? (question mark) are tagged as <O> . , (Comma) is tagged with the preceeding tag.

Corpus Development

POS tagged corpus containing two lakh and twenty five thousand words was prepared by collecting corpora from Dinamani newspaper, yahoo Tamil news, online Tamil short stories, etc. Dhanalakshmi.V et al., 2008. This POS tagged corpus is used for chunking corpus development. Our customized tagset was used to tag the POS tagging and chunking corpus. The tagged corpus is given for training using the machine learning tools. After training, the untagged corpus is tagged by tagger generator. The output of tagger generator is manually corrected to increase the corpus size.

Training data format: The training data should be in a particular format. The training data must consist of multiple tokens, these token are nothing but words, and a sequence of token becomes a sentence. Each token should be represented in one line, with the columns separated by white space. Many numbers of columns can be used, but the columns are fixed through all tokens. There should be some kinds of 'semantics' among the columns, i.e. first column is a 'word', second column is 'pos tag', and third column is 'chunk tag' and so on. The last column represents the answer tag which is going to be trained by SVM based Tools. We have fixed three column formats. Following is a sample of the training data.

வளாகத் <NNC> <B-NP>

தேர்வில் <NNC> <I-NP>

வேலைவாய்ப்பு <NN> <B-NP>

பெற்ற <VNAJ> <B-VNP>

மாணவர்களின் <NN> <B-NP>

பட்டியல் <NN> <I-NP>

வெளியிடும் <VNAJ> <B-VNP>

விழா <NN> <B-NP>

திங்கள்கிழமை <NNP> <B-NP>

நடைபெற்றது <VF> <B-VFP>

. <DOT> <O>

SVM based Tools for Tamil POS Tagger and chunker

The SVMTool is a simple, flexible, and effective generator of sequential taggers based on Support Vector Machines and how it is being applied to the problem of part-of-speech tagging. This SVM-based tagger is robust and flexible for feature modeling (including lexicalization), trains efficiently, and is able to tag thousands of words per second. YamCha(Yet Another Multipurpose Chunk Annotator by Taku Kudo) is a generic, customizable, and open source text chunker. Yamcha is using a state-of-the-art machine learning algorithm called Support Vector Machines (SVMs), introduced by Vapnik.

Support Vector Machine

SVM is a machine learning algorithm for binary classification, which has been successfully applied to a number of practical problems, including NLP. Tagging a word in context is a multi-class classification problem. Since SVMs in general are binary classifiers, a binarization of the problem must be performed initially before applying them. Here a simple one-per- class binarization is applied, i.e., a SVM is trained for every POS tag in order to distinguish between examples of this class and all the rest. When tagging a word, the most possible tag according to the predictions of all binary SVMs is selected.

SVMTool for Tamil POS Tagger

The SVMTool software package consists of three main components, namely the model learner (SVMTlearn), the tagger (SVMTagger) and the evaluator (SVMTeval).

SVM model is learned from a training corpus using the SVMTlearn component. Different models are learned for the different tagging strategies. During tagging time, the SVMTagger component is used to choose the tagging strategy that is most suitable for the purpose of the tagging. Finally, when we give a correctly tagged corpus and the corresponding SVMTool predicted annotation, the SVMTeval component displays tagging results and reports. Tagged corpus is used for training a set of SVM classifiers. This is done using SVMlight, an implementation of Vapnik's SVMs in C, developed by Thorsten Joachims.

Yamcha for Tamil Chunker

YamCha is an open source text chunker and so called Support Vector machines (SVMs). SVMs are binary classifiers and thus must be extended to multiclass classifiers to classify three cases for NP chunking with (I, O, B). By mapping the n-dimensional input space into high dimensional feature space in which a linear classifier is then typically constructed. This approach is used for chunking, YamCha is used to perform the initial tagging, basic features in Yamcha are used, later all possible POS tag for the words in the corpus are added. This information is added to the training corpus and then it is trained using SVM thereby predicting the chunk boundary names using Yamcha, Finally the chunk labels and the chunk boundary names are merged to obtain the chunk tag.

Conclusion

This paper has described the POS tagger and Chunker for Tamil using Machine learning approach. For the POS tagging and chunking we have used a corpus of size 2, 25,000 words. The corpus is divided into training set (1, 65,000 words) and test set (60,000 words). Machine learning tools like SVMTool and Yamcha are trained and tested for the same corpus. We have found that automatic POS

tagging and chunking done by SVM based Machine learning tools gives better result. A GUI to enhance the user friendliness of the tool was also developed.

References

1. Akshar Bharati, Rajeev Sangal, Dipti Misra Sharma and Lakshmi Bai. 2006. AnnCorra:Annotating Corpora Guidelines for POS and Chunk Annotation for Indian Languages, Technical Report, Language Technologies Research Centre IIIT, Hyderabad.
2. Arulmozhi P, Sobha L, 2006. A Hybrid POS Tagger for a Relatively Free Word Order Language. In proceedings of MSPIL-2006, Indian Institute of Technology, Bombay.
3. Dhanalakshmi V, Anandkumar M, Shivapratap G, Soman, K P, Rajendran S. May 2009. Tamil POS Tagging using Linear Programming, In International Journal of Recent Trends in Engineering, 1(2):166-169.
4. Gim´enez, J and L M`arquez, 2003. Fast and Accurate Part of- Speech Tagging: The SVM Approach Revisited, in Proceedings of the Fourth RANLP.
5. Sobha L,Vijay Sundar Ram R. 2006. Noun Phrase Chunking in Tamil, In proceeding of the MSPIL-06, Indian Institute of Technology, Bombay.pp-194-198.
6. Taku Kudo, Yuji Matsumoto. 2001.YamCha: Yet Another Multipurpose Chunk Annotator http://chasen.org/~taku/software/YamCha/.

ELECTRONIC DICTIONARIES

AND GLOSSARY OF TECHNICAL TERMS

தகவல் தொழில்நுட்ப கலைச்சொற்களை வளப்படுத்துங்கள்

மா. ஆண்டோ பீட்டர்

நிறுவனர், சாப்ஃட்வியூ கம்ப்யூட்டர்ஸ்
email: softviewindia@gmail.com / http://www.softview.in
தலைவர், கணித்தமிழ்ச் சங்கம் <http://www.kanithamizh.in>
117, நெல்சன் மாணிக்கம் ரோடு, சென்னை - 600029, தொலைபேசி : +91-44-42113535

தகவல் தொழில்நுட்பத்தின் அசுர வளர்ச்சி அனைத்து துறையையும் அசர வைத்துள்ளது. கடந்த பத்து ஆண்டுகளில் பத்துகோடி தமிழனிடமும் தகவல்தொழில்நுட்பம் பல மாற்றங்களை ஏற்படுத்தியுள்ளது. பொறியியல், மருத்துவம் ஆகியன நம்மிடையே பெரும்தாக்கத்தை ஏற்படுத்தியிருந்தாலும், தமிழன் தன் வசப்படுத்திய ஒரே புரட்சி தகவல் தொழில்நுட்பம் தான். அதற்கேற்ப தகவல் தொழில்நுட்ப கலைச்சொற்களின் புழக்கம் மற்றும் பயன்பாடு அதிகரித்தாலே, சர்வதேச அளவில் தமிழின் பயன்பாடு மிக அதிக அளவில் உயரும். தமிழில் தகவல் தொழில்நுட்ப கலைச்சொற்களை வளர்த்திட்டால் உலக அளவில் தமிழன் மத்தியில் ஒருங்கிணைப்பும், பயன்பாடும் உயரும். இவற்றிற்கு அடிப்படையாக நாம் எதிர்நோக்க வேண்டியவை என்னென்ன என்று ஆய்வோமா ?

தொடக்கப்பள்ளி அளவிலேயே பாடங்களின் வாயிலாக கலைச் சொற்களை வளர்த்தல்.

ஒரு மருத்துவரின் குறிப்பும், மருந்தும் பிறருக்கு புரியாததன் காரணம் அவற்றிற்கான தமிழ் சொற்கள் இல்லாமையே ஆகும். தகவல் தொழில்நுட்பத்தை பொறுத்த வரை ஆயிரக்கணக்கில் தமிழ்மொழி சொற்கள் புழக்கத்தில் உள்ளது. குறிப்பிட்ட வட்டத்திற்குள்ளேயே புழக்கத்திலுள்ள இச்சொற்கள் வெளிக்கொண்டு வரவேண்டும். பள்ளிப்பருவத்திலேயே மாணவர்களுக்கு கலைச் சொற்களை பொருள்பட போதித்தால் நம் மொழி, மேலும் வளர்ச்சி பெறும். தொடக்கப்பள்ளி அளவிலேயே படத்துடன், பொருள்பட கணினி பயன்பாட்டை மாணவர்களுக்கு விளக்கி கற்பித்தால், மாணவ பருவத்திலேயே தமிழ் சொற்கள் ஆழமாக குழந்தைகள் மனதில் பதியும்.

உலக குழுமங்களுடன் புதிய கலைச் சொற்களை உருவாக்க விவாதித்தல்.

தமிழ் தொழில்நுட்ப அமைப்புகளான உத்தமம், கணித்தமிழ்ச்சங்கம் மற்றும் பல்கலைக்கழகங்கள், அரசு அமைப்புகள், தமிழ் அமைப்புகள் மற்றும் புதிய கலைச்சொற்களை உருவாக்க பாடுபடும் மையங்கள் என அனைத்தையும் இணைக்க பாடுபடுதல் அவசியமாகும். இவ்வமைப்புகள் மூலமாக உருவாக்கப்படவுள்ள சொற்களை உருவாக்கிட இணையம் வாயிலாகவோ அல்லது நேரிடையாகவோ விழைய வேண்டும். ஆழமாக விவாதித்தால் பொருள்பட அழகிய தகவல் தொழில்நுட்ப கலைச்சொற்களையும் உருவாக்க முடியும்.

உலக குழுமங்களுடன் கலைச் சொற்களை உருவாக்கி தேவையான விதிகளை புதுப்பித்தல்

கலைச்சொற்களை பல துறைகளுக்கும் தொகுக்க மற்றும் உருவாக்க, தமிழக அரசு முனைவர்.வா.செ.குழந்தைசாமி தலைமையில் அறிஞர்கள் குழுவை நியமித்தது. இக்குழுவில் கலைச்சொற்களை உருவாக்க மற்றும் தொகுக்க பொதுவான விதிகளை, 2000 ஆம் ஆண்டு முனைவர்.வா.செ.குழந்தைசாமி தலைமையில் புதிய விதிகளை உருவாக்கியது. இவை தகவல்தொழில்நுட்பம் மட்டுமின்றி பொதுவாக 14 துறைகளுக்காக உருவாக்கப்பட்டது. இவற்றின் விதிகள் பலருக்கும் பரப்பப்படுதல் வேண்டும். மேலும் காலமாற்றத்திற்கேற்ப விதிகளில் மாற்றத்தை காண ஆய்வுகள் நடத்தப்படவேண்டும். கலைச் சொற்கள் உருவாக்கம் குறித்த விதிகள் அனைத்து தொழில்நுட்ப வல்லுநர்களும், ஆசிரியர் பெருமக்களும், அரசு அதிகாரிகளும் அறியும் வகை செய்தல் வேண்டும்.

தமிழ்வழி தகவல்தொழில்நுட்பத்துறை கலைச் சொற்களையும், படைப்புகளையும் தொகுத்தல்.

தமிழில் வெளியான அனைத்து தகவல் தொழில்நுட்ப செயல்பாடுகளையும், தொகுப்புகளையும், ஆய்வுகளையும், சொற்களையும் தொகுத்தல் வேண்டும். அண்ணாப்பல்கலைக்கழக வளர்தமிழ் மன்ற தகவல்தொழில்நுட்ப கையேடு, 2000 ஆம்ஆண்டின் தமிழக அரசின் தகவல் தொழில்நுட்ப தொகுப்பு, இலங்கை தகவல் தொழில்நுட்ப அகரமுதலி மற்றும் பிற பல்கலைக்கழகங்கழகங்களும், பதிப்பகங்களும் வெளியிட்டுள்ள அனைத்து கலைச்சொல் படைப்புகளை தொகுத்தல் வேண்டும். இப்பணியின் மூலமாக புதிய சொற்களின் வரவு, வேறுபாடு, சிறப்பு ஆகியவற்றை ஆராயலாம்.

தமிழ்வழி மற்றும் பிறமொழி அனைத்து கலைச்சொல் படைப்புகளையும் தொகுத்தல்.

இந்திய மற்றும் பிற உலக மொழிகளில் வெளியான அனைத்து தகவல் தொழில்நுட்ப செயல்பாடுகளையும், தொகுப்புகளையும், ஆய்வுகளையும், சொற்களையும் தொகுத்தல் வேண்டும். பன்னாட்டு மையங்கள், அரசு நிறுவனங்கள், தகவல் தொழில்நுட்ப நிறுவனங்கள், பல்கலைக்கழகங்கழகங்கள் மற்றும் பதிப்பகங்கள் வெளியிட்டுள்ள அனைத்து கலைச்சொல் படைப்புகளை தொகுத்தல் வேண்டும். இப்பணியின் மூலமாக புதிய சொற்களின் வரவு, வேறுபாடு, சிறப்பு ஆகியவற்றை ஆராயலாம். தகவல் தொழில்நுட்ப கலைச்சொற்களுக்கு அப்பாற்பட்டு அனைத்து துறையையும் தொகுத்தால் தொலைநோக்கு பார்வையுடன் நாம் பணியாற்றலாம். அனைத்து பிறமொழி கலைச்சொல் படைப்புகளையும் தொகுத்தால் அதன் வளர்ச்சிகளையும், மாற்றங்களையும் அறிய முடியும்.

கலைச்சொற்கள் படைக்கும் இணையங்களை தொகுத்து பாமரருக்கும் அளித்தல்.

தமிழில் செயல்படும் இணையங்களின் அனைத்து தகவல் தொழில்நுட்ப தொகுப்புகளையும், ஆய்வுகளையும், சொற்களையும் தொகுத்தல் வேண்டும். www.infitt.org, www.tcwords.com, www.tamilvu.org, www.bhashaindia.com, www.microsoft.com மற்றும் பிற பல்கலைக்கழகங்கழகங்கள் வெளியிட்டுள்ள அனைத்து இணையம் சார்ந்த கலைச்சொல் படைப்புகளை தொகுத்தல் வேண்டும். இப்பணியின் மூலமாக இணையம் வாயிலாகவே புதிய சொற்களின் வரவு, வேறுபாடு, சிறப்பு ஆகியவற்றை ஆராயலாம்.

சர்வதேச மென்பொருள் மற்றும் வன்பொருள் நிறுவனங்களை கலைச்சொற்களை பயன்படுத்தக்கோரி வலியுறுத்தல்.

மைக்ரோசாப்ட், ஐபிஎம் மற்றும் லைனக்ஸ் குழுக்கள் பிரத்யேக கலைச்சொல் குழுக்களையும், பிரிவுகளையும் இயக்கி வருகின்றன. உலகின் அனைத்து மென்பொருள், வன்பொருள் மற்றும் தொழில்நுட்ப கையேடுகளில் தமிழ் கலைச்சொல் வளத்தை அறியச்செய்தல் வேண்டும். மென்பொருள், வன்பொருள் மற்றும் கையேடுகளில் தமிழ் கலைச்சொற்களை பயன்படுத்த வலியுறுத்த வேண்டும். அவர்களின் பயன்பாட்டு தேவை, பயன்பாட்டில் குறை, புதிய தேவைகளுக்கான உதவிக்குழுக்களை அமைத்தல் வேண்டும்.

அரசுகளின் பயன்பாட்டில் கலைச்சொற்களை வளர்த்தல்

அரசாங்கத்தின் பயன்பாட்டில் தமிழ்மொழி அங்கீகரிக்கப்பட்ட மொழியாக உள்ள நாடுகளில் தமிழ் தகவல் தொழில்நுட்ப கலைச்சொற்களுக்கு முக்கியத்துவம் அளிக்கவேண்டும். தமிழ்வளர்ச்சித்துறை, கல்வித்துறை மற்றும் தகவல் தொழில்நுட்பத்துறை ஆகியவற்றில் ஒருங்கிணைத்து கலைச்சொற்களின் பயன்பாட்டை வளர்த்தல் வேண்டும்.

"தகவல் தொழில்நுட்ப கலைச்சொற்கள் வளர்ச்சி பெற்றாலே வருங்காலத்தில் தமிழ் அறிவியல் மொழியாக போற்றப்படும்".

Electronic Dictionary for Sangam Literature

சங்க இலக்கியத்திற்கான மின் அகராதி

Dr.K.Umaraj
Fellow, Central Institute of Classical Tamil, Chennai
e_mail: k_umaraj@yahoo.com

Abstract: In recent years, a tremendous development has been achieved in the field of Educational Technology. As a result, Dictionaries, Thesauruses and Encyclopedias have been developed electronically for the benefit of e_learners and researchers of Tamil.

To define the term Electronic dictionary, it is a machine readable dictionary, which provides search facilities to identify meaning and grammatical information of a particular word in a particular context. Many a number of Electronic dictionaries have been published for Tamil by different Research Institutes and Commercial organizations as well. A few may be mentioned as Lexicon of Madras University, Tamil-Tamil dictionary of Prof.M.Shanmugom Pillai and PALS dictionary of Palaniappa Brothers. But an Electronic dictionary exclusively for Sangam literature is yet to be developed.While developing one such dictionary, a number of problems arised in that venture. Those problems are discussed in detail in this paper.

அறிமுகம்

அண்மைக் காலத்தில் தகவல் தொழில் நுட்பவியல் வளர்ச்சியின் விளைவாகக் கல்வி தொழில் நுட்பவியலும்)Educational Technology) பெரிய மாற்றமும் வளர்ச்சியும் ஏற்பட்டுள்ளன. குறிப்பாகத் தமிழாய்வுக்குத் தேவையான கருவி நூற்களான அகராதிகள்,பேரகராதிகள்,சொற்களஞ்சியங்கள், கலைக்களஞ்சியங்கள் கணினி உதவி கொண்டு உருவாக்குவதில் பல்வேறு முயற்சிகள் மேற்கொள்ளப்பட்டு வெற்றி பெற்றுள்ளன .அவற்றுள் தமிழ் – தமிழ் – ஆங்கில அகராதியாகிய சென்னைப் பல்கலைக்கழகத்தின் பேரகராதி)Lexicon), பழனியப்பா பிரதர்ஸ் நிறுவனத்தாரின் ஆங்கில – தமிழ் அகராதி, பேராசிரியர் மு .சண்முகம் பிள்ளை அவர்களின் தமிழ் – தமிழ் அகராதி ஆகியவை குறிப்பிடத்தக்க அகராதிகளாகும் .இக்கட்டுரையில் சங்க இலக்கியத்திற்கான மின் அகராதியை உருவாக்கும் போது ஏற்படும் மொழியியல்ச் சிக்கல்கள் விவாதிக்கப்பட்டுள்ளன .

இலக்கண அடைவு உருவாக்கும்போது ஏற்படும் சிக்கல்கள்

தொல்காப்பியரும் நன்னூலாரும் சொற்களைப் பெயர், வினை, இடை, உரி என நான்காகப் பகுக்கின்றனர் .இந்த நான்கு வகைச் சொற்களை ஆஷர் ஆறு வகையாகவும் தாமஸ் லேமன் எட்டு வகையாகவும் கோதண்டராமன் பத்து வகையாகவும் வகைப்படுத்தி உள்ளனர் .இவ்வாறு சொற்களை பலவாறாக வகைப்படுத்துவதால் ஒரே பொருளைக் குறிக்கும் சொல்லுக்கு வேறு வேறு இலக்கணக் குறிப்புகள் தரப்பட்டுகின்றன .எடுத்துக்காட்டாக **‘அஃதான்று’** என்ற சொல்லுக்குத் தமிழ்ப் பேரகராதியில் **‘வினையடை’** என்றும் வரலாற்று முறை தமிழ் இலக்கியப் பேரகராதியில் **‘வினை’** என்றும் குறிக்கப்பட்டுள்ளது .

பொது	**பேரகராதி (36-1924)**	**பெருஞ்சொல்லகராதி (1998)**	**வரலாற்று முறை தமிழ் இலக்கியப் பேரகராதி (2004)**
அ	பெயர்	பெயர்	பெயர்
அ	இடை	இடை	இடை
அக்கால்		வினையடை	
அஃகிய			வினை
அஃகிய நுட்பம்			பெயர்
அஃகியோன்			வினை
அஃது	சுட்டு	சுட்டு	பெயர்
அஃதே		இடை	இடை
அஃதான்று	**வினையடை**	**வினையடை**	**வினை**

பொருள் அடைவு உருவாக்கும்போது ஏற்படும் சிக்கல்கள்

ஒரு சொல் எத்தனை இடங்களில் வருகிறதோ அத்தனை இடங்களையும் தொடர்களுடன் எடுத்து அவற்றை ஆராய்ந்து பொருள் மாறுபடும் இடங்களும் தொடர்களும் குறித்துக்கொள்ளப்பட வேண்டும்.

சான்றாக

அடர்

என்னும் சொல் 13இடங்களில் வருகிறது 11 .இடங்களில் **"தகடு"** என்னும் பொருளிலும் 2இடங்களில் **"அடர்ந்த"** என்னும் பொருளிலும் வருகிறது .இவை இரண்டு பொருளுக்கும் ஒவ்வொரு மேற்கோள் தொடர் கொடுக்க வேண்டும் .

அடர் .1**தகடு :"நுண் உருக்கு உற்ற விளங்கு அடர்ப் பாண்டில்" (மலை(4-**

.2**அடர்ந்த :"கடல் படை குளிப்ப மண்டி அடர் புகர்" (புற (12-6**

ஆனால் 'அக' என்ற சொல் '**பெயரடையாக'** வரும்போது ஒரு பொருளையும், **'வேற்றுமை உருபாக'** வரும்போது வேறொரு பொருளையும் தருகிறது .**வேற்றுமை உருபாக'** வந்து தருகின்ற பொருள் சங்க இலக்கியத் தொடர் எண்ணுடன் அகராதிகளில் தரப்படவில்லை.

ஆகவே, சங்க இலக்கிய அகராதி தயரிக்கும் போது ஒரு சொல் அது இடம் பெறுகின்ற அனைத்து த்தொடர்களையும்)key word in Context) இடத்திற்கேற்ப ஆராய்ந்து பொருள்களை)meanings) அகராதியில் பதிவுசெய்ய வேண்டும்.

கணிணியியலில் நேர்பெயர்ப்புச் சொற்களும், ஒலிபெயர்ப்புச் சொற்களும்

இலக்குவனார் திருவள்ளுவன்

Email: thiru2050@gmail.com

அறிவியல் துறைகளைப் புரிய வைப்பதற்கும் அறிந்து கொள்வதற்கும் கையாளப்படும் கலைச்சொற்கள் தன்-விளக்கமாயும் எளிமையாயும் அமைய வேண்டும். அவ்வாறு இல்லாச் சூழலில், தவறாகப் புரிந்து கொள்ளவோ விளங்காமல் குழப்பம் அடையவோ வாய்ப்புகள் ஏற்படுகின்றன. எனவே, விரைந்து வளரும் கணிணியியலில் துறை வளர்ச்சிக்கேற்ற கலைச்சொல் பெருக்கமும் அமைய வேண்டும். இத்தகைய கலைச் சொல்பெருக்கத்திற்குத் தடையாக இருப்பது சொல்லைப் புரிந்து கொண்டு படைக்காமல், 'சொல்லுக்குச் சொல்' என்ற நேர்முறையில் ஆக்கப்படும் கலைச்சொற்களும் தமிழ்ச் சொற்களைக் கையாளாமல் ஒலிபெயர்ப்புச் சொற்களாக மூலச் சொற்களைக் கையாளலுமாகும். இவற்றை உணர்ந்து, புத்தம்புதுக் கலைச்சொற்களை நாளும் உருவாக்கவும், உருவாக்கப்பட்ட கலைச் சொற்களைப் பயன்படுத்தவும் நாம் முன்வர வேண்டும். கலைச் சொற்கள் சுருங்கியனவாகவும், அவற்றின் அடிப்படையில் மேலும் புதிய கலைச் சொற்களை ஆக்க வாயிலாகவும் அமைய வேண்டும்.

கணிணியியலில் அமையும் கலைச் சொற்களைப் பின்வருமாறு பகுக்கலாம்:

1. பெரும்பான்மையர் தமிழில் கையாளும் சொற்கள்: சான்றாக, பெரும்பான்மையர் 'கோப்பு' என்றே எழுதி வந்தாலும், சிறுபான்மையர் 'Lபைல்' என்றே குறிப்பிடுவது.
2. சிறுபான்மையர் தமிழில் கையாளும் சொற்கள்: சான்றாக 'இண்டர்நெட்' எனப் பெரும்பான்மையராலும், 'இணையம்' எனச் சிறுபான்மையராலும் கையாளப்படுவது.
3. ஆங்கிலத் தலைப்பெழுத்துச் சொற்களையே கையாளுதல். சான்று: RAM, ROM
4. அனைத்துத் தரப்பினரும் ஆங்கிலச் சொற்களையே கையாளுதல். modem - மோடம் எனல். சிலர் ஆங்கிலச்சொற்களையே - ஆங்கில வரிவடிவங்களைக் கொண்டே - தமிழ்க் கட்டுரையில் பயன்படுத்தல்.சான்று: "Syntax error என்பது எளிதான தவறு; Semantic error என்பது கடுமையான தவறாகும்." என ஆங்கிலச் சொற்களை அவ்வாறே கையாண்டுள்ளமை. (இவற்றை, முறையே 'அமைவுத்தவறு', 'பொருள் தவறு' எனக் குறிப்பிட்டிருக்கலாம்.)
5. ஒவ்வொருவர் ஒவ்வொரு வகையாகக் கையாளுதல். சில நேரங்களில் ஒருவரே வெவ்வேறு வகையாகக் கையாளும் நேர்வும் உள்ளது. சான்றாகக், 'கம்ப்யூட்டர்' என்பதற்குக் கணிப்பொறி, கணிணி, கணினி, கணணி, கணிப்பான், கணிப்பொறி இயந்திரம் என வெவ்வேறு வகையாகக் கையாளல். இவற்றைத் தலைப்பில் ஒரு வகையாகவும், உள்ளடக்கங்களில் வேறுவகையாகவும் கையாளுதல். அதேபோல், ஒத்த பொருளுடையதாய் வெவ்வேறு சொற்களைப் பயன்படுத்தல். சான்றாக, home என்பதற்கு, வீடு, முகப்பு, மனை, இல்லம், தலைவாயில் என்பன போன்று பல வகையாகக் கையாளுதல். நடைமுறைக்கு நல்ல சொற்கள் வந்துவிட்டபின்னும் கொச்சையாகக் கையாளுதல்.
6. சுருங்கிய கலைச்சொல்லாக இல்லாமல், விளக்கச் சொற்றொடராகக் கையாளுதல். எ.கா.: debugging aids - பிழை நீக்க உதவும் பொருள்கள் - பிழை நீக்குதவி என்று சொல்லலாம். bootstrap input program - கணணி உயிர்ப்பூட்(டு)டல், உள்ளீட்டு திட்ட நிரல் - தொடக்கத் தரவு நிகழி என்று சொல்லலாம். இருப்பினும் (தமிழில் எழுதினால்தானே சுருக்கம், விரிவு என்றெல்லாம் சூழல்

ஏற்படும்;) பொதுவாக ஆங்கிலச் சொற்களையே கையாளும் போக்கு காணப்படுவதால், இவற்றிற்கான வாய்ப்பு குறைவாகவே காணப்படுகின்றன. ஆனால், தமிழில் அமைந்தனவற்றுள் பல மேலும் எளிமையாகவும் சுருக்கமாகவும் அமைதலே நன்று.

7. பொருள்விளக்கமான கலைச்சொல்லைக் கையாளாமல், நேருக்குநேர் மொழி பெயர்த்துக் கையாளுதல். mouse என்றால் சுண்டெலி என்பது போன்றவை.

8. தவறான சொல்லாக்கத்தைக் கையாளுதல். எ.கா.: barcode – சட்டக் குழூஉக்குறி; bar என்றால் சட்டம் என bar council என்ற முறையில் எண்ணியிருந்தாலும், frame என்று பொருள் கொண்டிருந்தாலும் தவறுதான். (பட்டைக்குறி என்று சொல்லலாமே!)

9. ஒரு சொல்லே வெவ்வேறு பொருள்களில் கையாளப்படுதல். எ.கா.: அடையாளம் அல்லது சின்னம் என்றே symbol, logo, icon ஆகிய சொற்களைக் குறித்தல். தனித் தனிச் சொல்லாக முறையே குறியீடு, முத்திரை, குறியுரு எனலாம்.

10. பிற அறிவியல் துறைச் சொற்களைக் கையாளுதல். எடுத்துக்காட்டாக, எண்- மதிப்புகள் கணக்குத் துறையில் கையாளப்படுகின்றன. இங்கும் கையாளப்படுகின்றன. ஆனால், அங்கும் தமிழ் இல்லை; இங்கும் தமிழ் இல்லை. பிற அறிவியல் துறைகளில் தமிழ்க்கலைச்சொற்களைக் கையாண்டிருந்தால், அவற்றையே இத்துறையிலும் கையாளுவதே ஏற்ற முறையாகும்.

மேற்குறித்த ஒவ்வொரு வகைப்பாட்டிலும், கணினிக் கலைச்சொற்களை ஆராய்தல் இன்றைய அடிப்படைத் தேவையாகும். எனினும் நாம், இங்கு நேர்பெயர்ப்புச் சொற்களையும் ஒலிபெயர்ப்புச் சொற்களையும் பற்றி மட்டும் பார்ப்போம்.

இவற்றுக்கு முன்னதாகச் சொல்லாக்க நெறிமுறைகள் குறித்துக் கருத்தில் கொள்ள வேண்டும். எந்த ஒரு சொல்லுக்கும் தனிப்பட்ட முறையில் பொருள் இல்லை. அச்சொல் பயன்படும் இடத்திற்கேற்பத்தான் அச்சொல்லுக்குப் பொருள் உண்டாகிறது. சொல் என்பது பொருளைச் சுமந்து செல்லும் ஊர்திதான். குறிப்பிட்ட சூழலில் எந்த ஒரு பொருளை வெளிப்படுத்துகிறதோ அதுதான் அந்த இடத்தில் அந்தச் சொல்லின் பொருளாகிறது. அந்தச்சொல்லே வேறு இடத்தில் வேறு பொருளை விளக்கும்பொழுது சொல்லின் பொருள் வேறாகின்றது. நீர், தான் இருக்கும் இடத்திற்கேற்ற வடிவைப் பெறுவதுபோன்று, சொல்லும் இடத்திற்கேற்ற உருவையும் வனப்பையும் பெறுகிறது எனலாம்.

எடுத்துக் காட்டாக soft ware, hard ware என்பவற்றை நேர்பொருளில் மொழி பெயர்த்து மென்பொருள் அல்லது மென்மி, வன்பொருள் அல்லது வன்மி என்று சொல்வது தவறாகும். பயன்பாட்டு அடிப்படையில் கணியம், கருவியம் எனக் குறிப்பிடுவதே முறையாகும். இவை போன்றே சாப்டுகாப்பி (soft copy) என்றும் ஆர்டுகாப்பி (hard copy) என்றும் சொல்லப்படுவன முறையே மென்படி என்றும் வன்படி என்றும் குறிக்கப்படுவதும் தவறாகும். சாப்டுவேர் என்பதைக் கணியம் என்றாலும் சாப்டுகாப்பி என்பதை நாம் கணினியில் காட்சியாகக் காணும் படி என்ற பொருளில் காட்சிப்படி என்றும் அச்சுப்படியாக நாம் எடுக்கும் ஆர்டுகாப்பி என்பதை அச்சுப்படி அல்லது கைப்படி என்றும் சொல்லுவதே முறையாகும். தவறான தமிழ்ச் சொல்லாக்கங்கள் குறித்து எள்ளி நகையாடுவதோடு நிற்காமல், சரியான சொல்லாக்க முயற்சியில் ஈடுபட்டு அவற்றைப் படிப்போரிடையே பரப்ப வேண்டும்.

தமிழ் மொழியில் - உருவாக்கப்படும் 'தொடர் சொற்கள்' உரிய பயன்பாட்டை இழந்து விடுகின்றன. 'பயன்பாடு இல்லாத சொல் இருந்து பயன் என்ன? நாம் கலைச் சொற்களை உருவாக்குவதன் நோக்கம், அவை அயற் சொற்களை யகற்றி அல்லது அயற்சொற்களுக்கு இடந்தராமல் நின்று நிலைத்துப் பொருள் தரவேண்டும் என்பதே! எனவே சுருங்கிய எளிய சொற்களையே கையாளல் வேண்டும். நடைமுறையில் வழங்கப்படும் பழஞ்சொற்களையும் புனையப்படும் புதுச்சொற்களையும் பயன்படுத்தி உயிர்ப்பூட்ட வேண்டும். ஆகவே, கலைச் சொல்லாக்கத்தில் ஈடுபடும் பொழுது, (மூலச்) சொல்லுக்கு நேரான (பெயர்ப்புச்) சொல்லை அமைக்காமல், (மூலப்)பொருளுக்கு ஏற்ற (பெயர்ப்புச்) சொல்லையே ஆக்க

வேண்டும். சொல் செறிவாயும் செவ்விதாயும் இருத்தல் வேண்டும். பண்பாட்டுப் பின்னணியைக் கருத்தில் கொண்டு ஆக்கப்பட வேண்டும். 'குன்றக் கூறல்' முதலான நூற்குற்றங்கள் பத்தும் 'சுருங்கச் சொல்லல்' முதலான நூல் அழகுகள் பத்தும் சொல்லுக்கும் மிகப் பொருந்தும்.

"சொல்லுக சொல்லைப் பிறிதோர்சொல் அச்சொல்லை
வெல்லும் சொல் இன்மை யறிந்து."

எனும் திருக்குறளை நினைந்து தக்க சொல்லைத் தெரிவு செய்ய வேண்டும். அயற்சொல் கலப்பை அறவே நீக்க வேண்டும். உரிய சொல் கண்டறியும் இடைநேரத்தில் அயற்சொல்லைப் பயன்படுத்த வேண்டிய தவிர்க்க இயலா நேர்வுகளில் பெயர்ப்பு மொழியின் வரிவடிவிலேயே எழுத வேண்டும்.

கணிணியியலில் தற்போது நடைமுறையில் உள்ள தவறான சொல்லாக்கங்களையும் (அடைப்பில் உள்ளன) உரிய பொருளில் எவ்வாறு குறிப்பிட வேண்டும் என்பனவற்றையும் பின்வருமாறு அளிக்கின்றேன். விரிவு அல்லது விளக்கம் வேண்டுவோர் முழுமையான கட்டுரையைப் படித்துப் பயனுற வேண்டுகிறேன். (மின்வரியில் தொடர்புகொண்டால் முழுக் கட்டுரை அளிக்கப்படும்) ஒலி பெயர்ப்புச் சொற்களையும் ஆங்கிலத் தலைப் பெழுத்துச் சொற்களையும் கையாளுவதைத் தவிர்க்க இவை துணை புரியும்.

ஒலி பெயர்ப்புச் சொற்கள்

Bit–(பிட்; துண்டு)> ஓர்மி	Garbage–(குப்பை)>தேவையிலி
Byte – எண்மி	Gate(வாயில்)>இடுவாய்
NIBBLE – நான்மம் > நான்மி	Joy stick -(மகிழ் குச்சி)> இயக்கப் பிடி
Bomb - (குண்டு)> அழிப்பி	Language-(மொழி)> விளம்பி
Brush – (புருசு, தூரிகை)> மின்னிகை	Lifeware–(உயிர்ப்பொருள்)> செயலியர்
Bug – (பூச்சி)> பிழை	Motion capture – (நகர்வு/அசைவு, சிறைப்பிடிப்பு)> அசைவு கவர்வு/அசைவுஈர்ப்பு
Debugging–(பூச்சிநீக்கம்)> பிழை நீக்கல்	Menu– (மெனு)> நிரல்
Cache–(மறைவு, விரைவு)> இடைச்சேமம்	Noise–(இரைச்சல், சப்தம்)> குறுக்கீடு
Cell – (செல், சிற்றறை)> அலககம்	Opensystems– (திறந்தவெளி முறைமைகள்)>
Chain–(செயின்,சங்கிலி)> தொடர் வினை > தொடரி	பொதுமை முறைமைகள்
Chip–(செதுக்கல்,சில்லு)> சிமிழ்	Prompt – (காட்டி / தூண்டி)> குறிப்புதவி
Clock Pulses – (கடிகாரத் துடிப்புகள்)> பதிவாரத் துடிப்புகள்	Raw data - (பச்சை விவரம்)> மூல விவரம்
CPU Handshaking - (சி.பி.யு கைகொடுத்தல்)> மை.வ.அ.பரிமாற்றி	Scroll – (அழிப்பான், திரை உருளல்)> சுருணை
Drill and practice – (ஓடிப் பழகு)> பழகிப் பயில்	Slave application – (எடுபிடி பிரயோகம், / அடிமை – ஆணைகேள்)> சார் வினைப்பாடு
Dumb-(ஊமை)>சார் செயலி	Smart peripheral –(கெட்டிக்கார வெளிப்புறப் பிரிவு)> திறன் புறக் கருவி
Dump-(திணி)>சேம மாற்றம்	Intelligent terminal – (கெட்டிக்கார முகப்பு)> திறன் முகப்பு
Error message - (பிழைச் செய்தி)> வழு குறிப்பி	Tree- (மரம்)> கிளைப்பி
Face – (முகம்)> முகப்பு	Forest – (காடு)> கிளைப்பித் திரள்
Folder–(மடக்கி)>அடங்கல்	Virus – (வைரசு , நச்சுயிரி)> சிதைப்பி

தலைப் பெழுத்துச் சொற்கள்

Baud – (போ(ஓ)ட் / பாடு (ஓர் அலகு) / பாட்)> முடுகுமானம்	DBMS– (டிபிஎம்எசு) > விவரணைத் தள மேலாண் முறை; விவமேல்
Biquinary- (பிக்குனரி)>ஈரைமம்	DCE – (டிசிஇ) > விவரணைத் தொகுப்புக் கருவி; விவமேல்
Capstan – (கேப்சுடன்)> சுழற்றி	EBCDIC - (ஈபிசிடிஐசி) > நீட்சி ஓர்மிக் குறியீட்டு பதின்மப் பரிமாற்றக் குறியீடு; நீட்டீடு
Cartridge – (கார்டிரிட்சு)> பொதியுறை	EDP - (ஈடிபி) > மின்னணு விவரணை வனைமம்; மிவிவனை
Diode – (டையோடு)> ஒருமுகி	ENIAC – (ஈனியாக்) > மின்னணு எண் ஒருங்கிக் கணக்கி; ஒருக்கி
Do-loop – (டூ லூப்)> மாறுநிலைத் தொடரி	EOF – (இஒஎஃப்) > கோப்பு முடிவு; கோ.மு.
Dongle – (டாங்கிள்)> தவிரி	EOM – (இஓஎம்) > செய்தி முடிவு; செ.மு.
ESC Key / Escape key – (எஸ்க் கீ)> விடுவிசை	EOT-இஓடி) > மாற்றீட்டு முடிவு; மா.மு.
Fetch – (பெட்ச்)> கொணர்வு	EPROM–(ஈப்ராம்); அழிவகுபடிஏற்றம்; அழிஏற்று (அழியேற்று)
I/O Input / Output – (இன்புட்/அவுட்புட்)> த/பெ; தரவு/பெறவு	Fortan - (போர்ட்ரான்)> விதி மாற்றம்;
Paddle - (பாடில்)> துடுப்பம்	Forth–(போர்த்து)>விரைவு விளம்பி > விரைவி
ABEND – (அபென்ட்)> அல்லியல் முடிவு; அ.மு.	HLL-(எச்எல்எல்) உயர்நிலை விளம்பி; உ.நி.வி
Abnormal termination - (திடீர் நிறுத்தம்)> அல்லியல் நிறுத்தம்; அத்தம்	IC–(ஐசி)> ஒருங்கிணைந்த மின் சுற்று ; ஒருமின்
ADONIS – (அடோனிசு)> இ.த.மு.க.; இணைதகவு	IBM-PC – (ஐபிஎம் பிசி)>பன்னாட்டு வணிகப் பொறியத்
ADP – (ஏடிபி)> த.வி.வ.; தன்வனைவு	தனியர் கணினி; பவபொ தகணி
ALGOL – (அல்காரிதம்)> தீ.வி./தீர்விளம்பி	INFLIBENT– (இன்பிலிபெண்ட்)> கணினித் தகவல் வலையமைப்பு. நூற் தக வலை
APL – (ஏபிஎல்)> க.செ.வி./கணிவிளம்பி; கணிவி	IOCS - (ஐஓசிஎசு)>தரவு பெறவு கட்டுப்பாட்டு முறைமை ; த.பெ.க.மு
ASCII- (அசுகி)> பரியீடு	ISD - (ஐஎசுடி)> அயல் நேரி
Basic Language – (பேசிக் மொழி)>அறிமுறை விளம்பி	ISO - (ஐஎசுஓ)> தரப்படுத்துப் பன்னாட்டு அமைப்பு; த.ப.அ.
BBC MICRO – (பிபிசி மைக்ரோ)> பிரி.ஒலி.கணி	LISP–லிஸ்ப்)>அல்லெண் விவரவனைவு;
BCD – (பிசிடி)> இருமக் குறியீட்டுப் பதின்மம்>	விவனை
இகுமம்	Memory - (நினைவு, நினைவகம்)>ஏற்றம்,
BCPL - (பிசிபிஎல்)> வனைவி	Modem – (மோடம்)> தருவி
BIOS – (பயாசு)> அடிப்படைத் தரவு / பெறவு	MPU– (எம்.பி.யு)> நுண் வனைம

முறை	அலகு>நு.வ.அ.
அ.த.பெ.மு.	On-line- (ஆன்லைன்)>
CAD – (கேட்)> க.உ.வ./கணி வரை	இணைநிலை
CAI – (கேய்)> க.உ.அ./கணிமுறை	Off-line– (ஆப்லைன்)> அணைநிலை
CAM – (கேம்)> க.உ.உ.; கணித்தி	Pixel–(பிக்செல்)> படக்கூறு
CBX - (சி.பி.எக்சு)> கணிக்கட்டுப்பாட்டுக் கிளை நிலையம்; கணிக்கிளையம்	PL/1-(பி.எல்.1)>நிகழி விளம்பி-1
	Program – (அறிவுறுத்தல்) > நிகழி
C-DAC - (சி-டேக்)> கணிமேலம்	RPG - (ஆர்.பி.ஜி)> அறிக்கை நிகழி உருவாக்க விளம்பி; அறிநிவி
CIM – (சி.ஐ.எம்.)> கணிணித் தரவு நுண்படம்;கதநு.	SBC– (எஸ்பிசி) > தனி அட்டைக் கணிப்பொறி; த.அ.க.
COBOL – (கோபால்)> பொது வணிகமுக விளம்பி; வணிவி	SEA-ME-WE: (). தென்கிழக்கு ஆசிய,
CPL - (சிபிஎல்)> தொகுப்பு நிகழி விளம்பி; தொநிவி	மத்தியக் கிழக்கு மேற்கு ஐரோப்பா; தெ.கி.ஆ-ம.கி.மே.ஐ.
CPL - (சிபிஎல்)> வரி ஒன்றிற்கு எழுத்துரு எத்தனை; வ.எ.எ.	WORM – (வோர்ம்)ஒருமுறை எழுது பன்முறை படி; ஒஎபப; எழுபடி
CPM – (சிபிஎம்)> கட்டுப்பாட்டு நிகழி - நுண்வனைமம்; கநிநுவனை.	WYSIWYG – (விசுவிக்)> காண்பதே கிடைக்கும்; கா.கி.

மைக்ரோ, மில்லி, நானோ, பீக்கோ, கிலோ, மெகா, கிகா முதலான அளவைகளெல்லாம் கணிணியியலுக்கு மட்டுமே உரியன அல்ல. கணக்கறிவியலில் இருந்து பயன்படுத்தப்படும் பொதுவான சொற்களே. இவையும் எல்லா இடங்களிலும் தமிழிலேயே குறிக்கப்பட வேண்டும். வேறு வேலை இல்லையா? உலகளாவிய கணக்குச் சொற்களைப் பயன்படுத்த வேண்டியதுதானே! என்பர் சிலர். உண்மையில் பழந்தமிழ்நாட்டிலேயே வேறு எவ்வினத்திலும் இல்லாத அளவு மிகமிகக் கூடுதல் மதிப்பிலும் (அனந்தம் 10 இன் 189 அடுக்கு) மிகமிகக் குறைந்த மதிப்பிலும்(தேர்த்துகள் 1 / 2323824530227200000000) எண்களைப் பயன்படுத்தி வந்தது தமிழினமே! இன்றும் பல மொழியினர் தத்தம் மொழியிலேயே எண்களைக் குறிப்பிட்டு வருகின்றனர். அவ்வாறிருக்க நாம் நாம் பழஞ் செல்வத்தைக் குப்பையில் போடாமல் பேணிக் காத்துப் பயன்படுத்துவதே சரியானதாகும். எனவே, கணிணியில் குறிப்பிடப்படும் எண்மதிப்புகளையும் பின்வருமாறு தமிழிலேயே குறிப்பிடலாம்.

4 இருமம் (4 Bits) – 1 நான்மம் (1 Nibble)
8 இருமம் (8 Bits) – 1 எண்மம் (1 Byte)
1024 எண்மம் (1024 Bytes) – 1 அயிரை எண்மம் (1 Kilo Byte)
1024 அயிரை எண்மம் (1024 K.B) – 1 மா அயிரை எண்மம் (1 Mega Byte)
1024 மா அயிரை எண்மம் (1024 MB) – 1 பேரயிரை எண்மம் (1 Giga Byte)
1024 பேரயிரை எண்மம் (1024 GB) – 1 மாப்பேரயிரை எண்மம் (1 Tera Byte)
1024 மாப்பேரயிரை எண்மம் (1024 TB) – 1சீரயிரை எண்மம்(1 Petta Byte)
1024 சீரயிரை எண்மம் (1024 PB) – 1 மாச்சீரயிரை எண்மம்(1Exa Byte)
1024 மாச்சீரயிரை எண்மம் (1024 EB) – 1செவ்வயிரை எண்மம் (1 Zeetta Byte)
1024 செவ்வயிரை எண்மம் (1024 ZB) – 1 மாச் செவ்வயிரை எண்மம் (1Yotta Byte)

கீழ் வாய் எண்களையும் பழந்தமிழ் முறையில் பின்வருமாறு குறிக்கலாம்.

deci-	10^{-1}	கீழ்ப் பதின்
centi-	10^{-2}	கீழ் நூறன்
milli-	10^{-3}	கீழ் அயிரை
micro-	10^{-6}	கீழ் மீ அயிரை
nano-	10^{-9}	கீழ்ச் சிற்றயிரை
pico-	10^{-12}	கீழ் மீச் சிற்றயிரை
femto-	10^{-15}	கீழ்ச் சீரயிரை
atto-	10^{-18}	கீழ் மீச் சீரயிரை

தமிழ்ப்படைப்புகளில் அயற் சொற்களும் கிரந்த எழுத்து முதலான அயல் எழுத்துகளும் பயன்படுத்தக்கூடா. அனைத்துத் துறைகளிலும் தமிழில் எண்ணுவதற்கு வழி வகுக்கும் வகையில் எளிய இனிய செவ்விய தமிழில் எழுத வேண்டும். அதற்கு கணிணியமைப்புகளும் அறிவியல் துறை யமைப்புகளும் உதவ முன்வரவேண்டும்.

ஆதலின் கலையியல் ஆர்வலர்கள், தமிழ்ப்புலமையர், சொல்லாக்கநெறி ஆட்சியர் ஆகியோர் ஒன்றிணைந்து புத்தம்புதுக் கலைச் சொற்களை உடனே உருவாக்கும் முயற்சியில் ஈடுபட வேண்டும். பழந்தமிழ்ச் சொல் வளத்தை வீணாக்காமல் இம்முயற்சிக்குப் பயன்படுத்திக்கொள்ள வேண்டும். கணிணியியல் கட்டுரையாளர்கள், நூலாசிரியர்கள், இதழாளர்கள், உரையாளர்கள் ஆகியோருக்குச் சொல்லாக்கப் பயிற்சிகள் அளிக்க வேண்டும். தமிழ்க்கலைச் சொற்களைப் பயன்படுத்தும் நூல்களை மட்டுமே பாட நூல்களாக வைக்க வேண்டும்; கலப்பு நடையைக் கைவிட்டு நல்ல தமிழில் எழுதப்படும் நூல்களுக்கு மட்டுமே பரிசுகள் வழங்க வேண்டும். ‘இத்தகைய விதி செய்வோம்! அதை எந்த நாளும் காப்போம்!’ என்னும் உறுதி எடுத்துக் கொள்ள வேண்டும். முயன்றால் முடியாதது என எதுவும் உண்டோ? எனவே, கணிணிச் செல்வத்தைத் தங்குதடையின்றித் தமிழுலகிற்கு அளிப்போம்! சொற்கள் புதியன புனைவோம்! நூல்கள் புதியன படைப்போம்! அன்னைத் தமிழ் நோக்கி நம் துறையறிவுப் பயணம் இருக்க வேண்டும். இந்த வகையில் நாளும் வளர்ந்துவரும் கணிணியியலிலும் நற்றமிழ் முழுமையாய் ஆட்சி செய்ய வேண்டும் என்பதை வலியுறுத்தவே இக்கட்டுரை. தவறான சொல்லாக்கம் என்பது தமிழின் பிழையன்று; தமிழனின் பிழையே என்பதை உணர்த்தித் திருத்தவே இக்கட்டுரை.

அறிவியல் செல்வத்தைச் சேர்ப்போம்! தமிழைச் செழிப்பாக்குவோம்! நாமும் செழிப்பாவோம்!
வாழ்க தமிழாக! வளர்க நலமாக!

Critical editions of Tamil works

Exploratory survey and future perspectives

[Jean-Luc Chevillard, CNRS, University Paris-Diderot Paris 7]
<http://www.linguist.univ-paris-diderot.fr/~chevilla/>

Abstract: this paper will deal with the topic of variation in ancient Tamil texts. It does not offer a technical (software) solution for that problematic question. It is rather the exposition, through several examples, of a recurring feature which requires attention and which is important for proper understanding of the history and transmission of those texts. Logical frameworks, such as the one provided by the Text Encoding Initiative (TEI), exist, but software tools for comfortably dealing with the situation remain a desideratum.

Various possible divisions for a treatise

In this example, we imagine a user who wants to read the *Tolkāppiyam Collatikāram* (=TC) and who has acquired the text under several forms, one of them being the Project Madurai (=PM) file pm0100.pdf available on the internet[3] and the others being several books, referred to as A, B, C, D, E, F and G.[4] He counts the number of sūtra-s *(cūttiram)* and he realizes that the figures obtained differ (see chart 1):

	PM	A	B	C	D	E	F	G
கிளவியாக்கம்	62	62	62	61	62	59	61	61
வேற்றுமையியல்	22	22	17	22	21	21	22	22
வேற்றுமை மயங்கியல்	35	35	35	34	35	33	34	35
விளிமரபு	37	37	37	37	37	36	37	37
பெயரியல்	43	43	43	43	43	41	43	43
வினையியல்	51	51	49	51	51	54	51	51
இடையியல்	48	48	48	48	48	47	48	48
உரியியல்	98	98	99	100	98	100	100	100
எச்சவியல்	67	67	66	67	61	61	67	67

Chart 1

[3]On the on the Project Madurai web site, we would find a text at the following URL:
<HTTP://WWW.PROJECTMADURAI.ORG/PM_ETEXTS/PDF/PM0100.PDF> (**தொல்காப்பியம்**)

[4]A IS THE 1981 NCBH REPRINT OF THE MURRAY RAJAM EDITION OF *TOLKĀPPIYAM*. B, C, D AND E ARE THE VOLUMES IN THE 2003 TAMI□MA□ PATIPPAKAM EDITION CONTAINING THE *COLLATIKĀRAM* WITH THE COMMENTARIES BY I□AMPŪRA□AR, CĒ□ĀVARAIYAR, NACCI□ĀRKKI□IYAR AND TEYVACCILAIYĀR. F IS THE TOLKĀPPIYAM IN THE PULIYŪR KĒCIKAN EDITION (*Ā□ĀM PATIPPU*, 1980) AND G IS THE *TOLKĀPPIYAM ĀRĀYCCI KĀ□□IKAIYURA*I BY PĀVALARĒ□U CA. PĀLACUNTARAM (1988).

HE REALIZES UPON EXAMINING THE TEXT THAT THE REASON FOR THE DIFFERENCE IN THE COUNT OF SUTRAS LIES IN THE FACT THAT SOME TEXTS TAKE AS ONE SUTRA WHAT IS TAKEN AS TWO SUTRAS BY OTHER TEXTS. FOR INSTANCE, THE FOLLOWING CHART SHOWS SOME DIFFERENCE BETWEEN IḷAMPURAṇAR AND CEṉAVARAIYAR:

மூன்றாகுவதே ஒடுவெனப் பெயரிய வேற்றுமைக் கிளவி வினைமுதல் கருவி அனைமுதற் றதுவே அதனின் இயற லதற்றகு கிளவி அதன்வினைப் படுத லதனின் ஆதல் அதனிற் கோடல் அதனொடு மயங்கல் அதனோ டியைந்த ஒருவினைக் கிளவி அதனோ டியைந்த வேறுவினைக் கிளவி அதனோ டியைந்த ஒப்ப லொப்புரை இன்னான் ஏது ஈங்கென வரூஉம் அன்ன பிறவும் அதன்பால என்மனார். (TC73i)	*மூன்றாகுவதே* ஒடுவெனப் பெயரிய வேற்றுமைக் கிளவி வினைமுதல் கருவி அனைமுதற் றதுவே (TC73c) அதனின் இயற லதற்றகு கிளவி அதன்வினைப் படுத லதனின் ஆதல் அதனிற் கோடல் அதனொடு மயங்கல் அதனோ டியைந்த ஒருவினைக் கிளவி அதனோ டியைந்த வேறுவினைக் கிளவி அதனோ டியைந்த ஒப்ப லொப்புரை இன்னான் ஏது ஈங்கென வரூஉம் அன்ன பிறவும் அதன்பால என்மனார்.(TC74c)

Chart 2

Various possible authoritative readings for a treatise

IN THIS EXAMPLE, WE IMAGINE OUR USER MAKING A MORE DETAILED COMPARISON BETWEEN THE VERSIONS OF THE TEXT TRANSMITTED BY THE COMMENTATORS. HE IS COMPARING THIS TIME THE WORDING TRANSMITTED BY IḷAMPURAṇAR AND PERACIRIYAR FOR THE LAST SUTRA INSIDE THE *MARAPIYAL*, WHICH IS THE LAST CHAPTER INSIDE THE *TOLKAPPIYAM PORUḶATIKARAM*. THE DIFFERENCES ARE GIVEN IN CHART 3A, 3B AND 4[5]:

Chart 3a	List A1: TP656i list (the 32 *தந்திரவுத்தி* as per Iḷampūraṇar's reading)
(I1) நுதலியது அறிதல், (I2) அதிகார முறை, (I3) தொகுத்துக் கூறல், (I4) வகுத்து மெய்ந் நிறுத்தல், (I5) மொழிந்த பொருளோ டொன்ற வைத்தல், (I6) மொழியாததனை முட்டின்றி முடித்தல், (I7) வாராததனான் வந்தது முடித்தல், (I8) வந்தது கொண்டு வாராதது முடித்தல், (I9) முந்து மொழிந்ததன் தலைதடுமாற்று, (I10) ஒப்பக் கூறல், (I11) ஒருதலை மொழி, (I12) தன்கோட் கூறல், (I13) உடம்பொடு புணர்த்தல், (I14) பிறன் உடம்பட்டது தான் உடம்படுதல், (I15) இறந்தது காத்தல், (I16) எதிரது போற்றல், (I17) மொழிவாம் என்றல், (I18) கூறிற்று என்றல், (I19) தான் குறியிடுதல், (I20) ஒருதலை அன்மை முடிந்தது காட்டல், (I21) ஆணை கூறல், (I22) பல் பொருட்கு ஏற்பின் நல்லது கோடல், (I23)	

[5]These charts are adapted from charts found pp.82-83 in Chevillard [2009], "The Metagrammatical Vocabulary inside the Lists of 32 Tantrayukti-s and its Adaptation to Tamil", in Wilden, Eva (Ed.), *Between Preservation and Recreation: Proceedings of a workshop in honour of T.V. Gopal Iyer*, Collection Indologie – 109, IFP/EFEO, Pondicherry.

தொகுத்த மொழியான் வகுத்தனர் கோடல், (I24) மறுதலை சிதைத்துத் தன் துணிபு உரைத்தல், (I25) பிறன் கோள் கூறல், (I26) அறியாது உடம்படல், (I27) பொருள் இடையிடுதல், (I28) எதிர் பொருள் உணர்த்தல், (I29) சொல்லின் எச்சம் சொல்லியாங்கு உணர்த்தல், (I30) தந்து புணர்ந்து உரைத்தல், (I31) ஞாபகம் கூறல், (I32) உய்த்துக்கொண்டு உணர்த்தல்.

Chart 3b	List A2: TP665p list (the 32 தந்திரவுத்தி as per Pērāciriyar's reading)
(P1) நுதலியது அறிதல், (P2) அதிகார முறைமை, (P3) தொகுத்துக் கூறல், (P4) வகுத்து மெய்ந் நிறுத்தல், (P5) மொழிந்த பொருளோ டொன்ற வவ்வயின் மொழியாததனை முட்டின்றி முடித்தல், (P6) வாராததனான் வந்தது முடித்தல், (P7) வந்தது கொண்டு வாராதது உணர்த்தல், (P8) முந்து மொழிந்ததன் தலைதடுமாற்று, (P9) ஒப்பக் கூறல், (P10) ஒருதலை மொழிதல், (P11) தன்கோட் கூறல், (P12) முறை பிறழாமை, (P13) பிறன் உடம்பட்டது தான் உடம்படுதல், (P14) இறந்தது காத்தல், (P15) எதிரது போற்றல், (P16) மொழிவாம் என்றல், (P17) கூறிற்று என்றல், (P18) தான் குறியிடுதல், (P19) ஒருதலை அன்மை, (P20) முடிந்தது காட்டல், (P21) ஆணை கூறல், (P22) பல் பொருட்கு ஏற்பின் நல்லது கோடல், (P23) தொகுத்த மொழியான் வகுத்தனர் கோடல், (P24) மறுதலை சிதைத்துத் தன் துணிபு உரைத்தல், (P25) பிறன் கோட் கூறல், (P26) அறியாது உடம்படல், (P27) பொருள் இடையிடுதல், (P28) எதிர் பொருள் உணர்த்தல், (P29) சொல்லின் எச்சஞ் சொல்லியாங் குணர்த்தல், (P30) தந்து புணர்ந்து உரைத்தல், (P31) ஞாபகம் கூறல், (P32) உய்த்துக்கொண்டு உணர்தல்	

As seen by comparing these two charts, each list has 32 items, but only 24 of them are found identical in both lists, although their rank may differ. The discrepancies are the following:

I2 and P2, I8 and P7, I11 and P10, and I32 and P32 differ slightly.

I13 (உடம்பொடு புணர்த்தல்) is unique to List A1

P12 (முறை பிறழாமை) is unique to List A2

I5 (மொழிந்த பொருளோ டொன்ற வைத்தல்) and I6 (மொழியாததனை முட்டின்றி முடித்தல்) are combined into P5 (மொழிந்த பொருளோ டொன்ற வவ்வயின் மொழியாததனை முட்டின்றி முடித்தல்)

I20 (ஒருதலை அன்மை முடிந்தது காட்டல்) is broken into P19 (ஒருதலை அன்மை) and P20 (முடிந்தது காட்டல்)

Combining the two lists alphabetically, we see that a list of 32 tantiravutti-s has become a list of 40 items

1. atikāra muṟai, I2	21. TĀṉ KUṟIYIṭUTAL, I19, P18
2. ATIKĀRA MUṟAIMAI, P2	22. TOKUTTA MOḻIYāṉ VAKUTTAṉAR KōṬAL, I23, P23
3. AṟIYĀTU UṬAMPAṬAL, I26, P26	23. TOKUTTUK KūṟAL, I3, P3
4. ĀṇAI KūṟAL, I21, P21	24. NUTALIYATU AṟITAL, I1, P1
5. IṟANTATU KĀTTAL, I15, P14	25. PAL PORUṬKU ĒṟPIṉ NALLATU KōṬAL, I22, P22
6. UṬAMPOṬU PUṇARTTAL, I13	26. PIṟAṉ UṬAMPAṭṭATU TĀṉ UṬAMPAṬUTAL, I14, P13
7. UYTTUKKOṇṭU UṇARTAL, P32	27. PIṟAṉ KōṬ KūṟAL, I25, P25
8. UYTTUKKOṇṭU UṇARTTAL, I32	28. PORUḷ IṬAIYIṬUTAL, I27, P27
9. ETIR PORUḷ UṇARTTAL, I28, P28	29. MAṟUTALAI CITAITTUT TAṉ TUṇIPU URAITTAL, I24, P24
10. ETIRATU PŌṙṙAL, I16, P15	30. MUṭINTATU KāṭṭAL, P20
11. OPPAK KūṟAL, I10, P9	31. MUNTU MOḻINTATAṉ TALAITAṬUMĀṙṙU, I9, P8
12. ORUTALAI MOḻI, I11	32. MUṟAI PIṟAḺĀMAI, P12
13. ORUTALAI MOḻITAL, P10	33. MOḻINTA PORUḷŌṬU OṉṟA VAITTAL, I5
14. ORUTALAI AṉMAI MUṭINTATU KāṭṭAL, I20	34. MOḻINTA PORUḷŌṬU OṉṟA VAVVAYIṉ MOḻI-

	YĀTATAṉAI MUṯṯIṉṟI MUṭITTAL, P5
15. ORUTALAI AṉMAI, P19	35. MOḻIYĀTATAṉAI MUṯṯIṉṟI MUṭITTAL, I6
16. KŪṟIṟṟU EṉṟAL, I18, P17	36. MOḻIVĀM EṉṟAL, I17, P16
17. COLLIṉ ECCAM COLLIYĀṅKU UṇARTTAL, I29, P29	37. VAKUTTU MEYN NIṟUTTAL, I4, P4
18. ÑĀPAKAM KŪṟAL, I31, P31	38. VANTATU KOṇṭU VĀRĀTATU UṇARTTAL, P7
19. TANTU PUṇARNTU URAITTAL, I30, P30	39. VANTATU KOṇṭU VĀRĀTATU MUṭITTAL, I8
20. TAṉKŌṭ KŪṟAL, I12, P11	40. VĀRĀTATAṉĀṉ VANTATU MUṭITTAL, I7, P6

Chart 4 (Items in Lists A1 AND A2) [I = IḶAMPURAṇAR; P = PERACIRIYAR]

Providing dual texts, with metrical and simplified readings

WE NOW EXAMINE A SITUATION WHERE A TEXT IS AVAILABLE UNDER TWO FORMATS, A TRADITIONAL ONE, WHERE IT APPEARS AS AN INSTANCE OF A SPECIFIC METRICAL FORM, AND A SIMPLIFIED ONE, WHERE THE CIR-S ARE NO LONGER RECOGNIZABLE AND THE SANDHI RULES ARE NOT RESPECTED. GOOD EXAMPLES OF SUCH A SITUATION ARE THE HYMNS TO CIVAṉ WHICH ARE CONTAINED IN THE *TEVARAM*. WE CONSIDER FOR INSTANCE:

செந்நெல் அம் கழனிப் பழனத்து அயலே செழும்
புன்னை வெண் கிழியில் பவளம் புரை பூந்தராய்
துன்னி, நல் இமையோர் முடி தோய் கழலீர்! சொலீர்---
பின்னுசெஞ்சடையில் பிறை பாம்புஉடன் வைத்ததே? (TEV. 2-1_1)

and

செந்நெ லங்கழ னிப்பழ னத்தய லேசெழும்
புன்னை வெண்கிழி யிற்பவ ளம்புரை பூந்தராய்
துன்னி, நல்லிமை யோர்முடி தோய்கழ லீர்சொலீர்
பின்னு செஞ்சடை யிற்பிறை பாம்புடன் வைத்ததே? (Tēv. 2-1_1)

The first presentation is the text as it appears inside the Digital Tēvāram CD[6] and the second presentation is the one found for instance on the Project Madurai site (as PM0157.pdf), where the pattern தேமா கூவிளம் கூவிளம் கூவிளம் கூவிளம் appears on each line[7] and where word boundaries are not respected.

Critical edition of a classical text

WE HAVE DISCUSSED IN PAR. 2 THE EXAMPLE OF A TEXT FOR WHICH THERE ARE WIDELY DIVERGENT READINGS, EACH BASED ON THE AUTHORITY OF AN ANCIENT SCHOLAR. A MORE GENERAL SITUATION IS THE PREPARATION OF A CRITICAL EDITION, DEFINED IN THE FOLLOWING WAY BY GOODALL [2000],[8]

"Since this expression is today so variously understood among Indologists, I must state what I understand by it. A critical edition is an editor's reconstruction of a text as he supposes it to have been at a particular time in its transmission (. . .). Although it is a hypothesis, it is made on the basis of all evidence for the wording of the text that the editor can consult (ideally all surviving evidence) and by

[6]V.M. Subramanya Aiyar, Chevillard, J.-L., S.A.S. Sarma, *Digital Tēvāram. Ka□i□it Tēvāram*, Collection Indologie n° 103, IFP / EFEO, 2007

[7]ACCORDING TO THE 1991 *TĒVĀRAM ĀYVUT TU□AI*, T.V.GOPAL IYER, COLLECTION INDOLOGIE N° 68.3, IFP, PONDICHERRY, THE METER IS *KALINILAITTU□AI*.

[8]"Problems of Name and lineage: relationships between South Indian authors of the Saiva Siddhânta", *Journal of the Royal Asiatic Society*, 10/2, p. 205-216.

an editor who has striven to understand as far as possible the ideas of the author(s) as well as the relationships between the sources that make up that evidence, and it is equipped with an apparatus that reports all of that evidence that is relevant to the constitution of the text (in some cases this means all the evidence). Such editions, as yet all too rare, are invaluable tools for all who are interested—from any perspective—in texts and their transmissions." (pp. 214-215, fn.38, op.cit.)

In the case of Tamil classical texts, a recent example of critical edition is Wilden[2008],[9] which is based on 5 Manuscripts and 5 printed editions of the *Naṟṟiṇai*, and provides all the readings found in those sources.

Existing protocols for dealing with the situation

I could present more examples but those provided here should suffice for the present purpose. In all those cases, it should be clear that the knowledge concerning the various forms has to be preserved and that choosing one form and eliminating the others will impoverish the scholarly community. This raises two different questions: **1.** how to encode a text when it is a dual (or "plural") text? **2.** how to allow the user to manipulate and to read that complex text comfortably? Fortunately, some progress has been made towards answering those questions, thanks to the members of a collective endeavour called the "Text Encoding Initiative"[10]. It is to be desired that the intellectual adventure which was started by the inspired pioneers of the Project Madurai, emulating the Project Gutenberg, will in the future be continued by the younger generations, eager to preserve their rich heritage, in all its complexity, thanks to ever more powerful tools.

[9]Wilden, Eva, 2008, *Na□□i□ai / a critical edition and an annotated translation of the Na□□i□ai,* Ecole française d'Extrême-Orient : Tami□ma□ patippakam, Pondicherry / Chennai (3 vol.)

[10]The goals of that consortium are found on the following web site: <http://www.tei-c.org/index.xml>: "The Text Encoding Initiative (TEI) is a consortium which collectively develops and maintains a standard for the representation of texts in digital form. Its chief deliverable is a set of Guidelines which specify encoding methods for machine-readable texts, chiefly in the humanities, social sciences and linguistics. Since 1994, the TEI Guidelines have been widely used by libraries, museums, publishers, and individual scholars to present texts for online research, teaching, and preservation. [...]"

The English Dictionary of the Tamil Verb

What can it tell us about the structure of Tamil?

Harold F. Schiffman
University of Pennsylvania

The *English Dictionary of the Tamil Verb* was undertaken because of a number of needs that were not being met by existing or previously-extant English-Tamil dictionaries. The main goal of this dictionary is to get an English-knowing user to a Tamil *verb*, irrespective of whether he or she begins with an English verb or some other item, such as an adjective; this is because what may be a verb in Tamil may in fact not be a verb in English, and vice versa. The web and DVD versions of this dictionary are searchable, so that if a particular English verb the user wants a Tamil equivalent for is not one of the main entries, inputting the search item should take the user to the English synonym file, which will give the user the Tamil verb. For example, we do not have a main entry for 'pounce' but this item does appear as a synonym for 'jump, leap', and some other verbs, so searching for 'pounce' will get the user to a Tamil verb. The search engine provided for the web version on the DVD also allows the user to search for Tamil verbs with advanced search methods such as ends in, begins with, part of, contains etc., and this is where some interesting new insights about the structure of the Tamil lexicon can be found, and which I want to concentrate on in this paper.

Syntactic Complexity of the Verb Phrase Because the Tamil verb is morphologically complex, and the verb phrase therefore syntactically *very* complex, we decided to focus only on the Tamil verb. Tamil nouns are, in contrast, morphologically fairly simple and the noun phrase is remarkably uncomplicated--Tamil nouns have no gender distinctions (except where there is biological gender), no agreement, and no marking of adjectives as to number or gender. The Tamilnadu government has spent much time and energy creating lexica and glossaries for various modern usages for Tamil, but from what we can gather, these have mainly generated new nominal terminology, not verbs. This is partly because Literary Tamil cannot borrow verbs easily, i.e. it cannot take a 'foreign' word and add Tamil morphological material to it, such as tense marking and person-number-gender marking, which all Tamil finite verbs must have.

So what does Tamil do if it needs a new verb? In the past, Tamil did borrow verbs from Sanskrit, but that is now frowned upon, and it no longer does so. It also, according to Fabricius (1972), has a few borrowings from Telugu, but that also seems to have ceased.

Past participle plus main verb. One simple way to make a new verb in Tamil is to take the past participle of another verb and preface it to a main verb. Such examples as *collikkoṭu* 'teach' are constructed by taking the past participle of *collu* 'say' and prefacing it before the main verb *koṭu* 'give.' Other examples, such as *taḷḷi vai* 'postpone', *taḷḷippooḍu* 'put off', and many others like this abound in Tamil and can be found by perusing our entries.

Noun plus verb. Another way to make a new verb is to take a noun and follow it by a main verb. In the past, only Tamil nouns were used, but increasingly, borrowed nouns (Tamil can borrow nouns, even if it can't borrow verbs) are used. An example of the older type might be *kuṟṟam collu* 'say blame', which of course gives us a verb 'blame.' That this phrase is closely bound together is shown by the fact that even though *collu* is transitive, the noun *kuṟṟam* is not marked for accusative case. More 'modern'

examples of this type of noun-verb compounding would be *ṭaunlooṭ paṇṇu* which of course combines a noun borrowed from English (download) plus a common verb meaning 'make' or 'do': *paṇṇu*. Searching our database for examples of this type, using either *paṇṇu* or *cey*, both of which mean 'make' or 'do', which reveal dozens if not hundreds of examples.

Making an intransitive verb Transitive

Most grammars of Tamil have discussed the transitivity status of Tamil verbs as being a case of either transitive or intransitive, i.e., as if this distinction were exactly parallel to that of English or some other western language. Actually any cursory examination of the Tamil verb will reveal that the semantic distinction so clearly marked in the morphology, i.e., the distinction between pairs like *ooṭu* and *oo□□u* which is usually glossed as 'run' vs. 'cause to run' or 'run of one's own volition' vs. 'run something' is not as simple when all the verbs of the language have been taken into account. Some researchers on Tamil, such as Paramasivam 1979, have rejected the dichotomy between transitivity and intransitivity as inadequate for Tamil, and have opted for a distinction known as 'affective' vs. 'effective', which is felt to more adequately capture the distinction. We have opted to stick with the transitivity/ intransitivity distinction, however, because it is our experience that American students, at least, if they have any familiarity with this distinction, know it in this way, rather than as 'affective/ effective.'

In fact Hopper and Thompson (1982) show that verbs must be scaled for their *degree* of transitivity, since 'blaming' or 'seeing' is in some sense less transitive than 'breaking' or 'killing', actions which have a definite effect on an object, whereas to be blamed or seen does not affect the 'target' of the action in the same way. Thus to refer to *uṭai* as an intransitive kind of *breaking* since the process or person who caused the breaking is not known is also not as neat a distinction as one would like, even though the morphology of Tamil gives us two *uṭai*'s--one 'intransitive', i.e. without known agent, as in *kaṇṇaaṭi uṭaintatu* (spoken *kaṇṇaaṭi oṭenjadu*) 'the glass broke', the other 'transitive', as in avan *kaṇṇaaṭiyai uṭaittaan* (spoken avan *kaṇṇaaṭiye oṭeccaan)* 'He broke the glass.' These 'intransitives' are also usually possible only with a third-person, often neuter, 'subject,' i.e. 'glass.' Yet to think of glass as the 'subject' of 'intransitive' breaking but as the object or target of transitive breaking (when the agent of the action is known), is illogical.

Our solution to this problem is to issue caveats but not to attempt a wholesale reclassification or scaling of transitivity for the Tamil verbs. We continue to use the (probably archaic) bipolar scale of transitivity, with the two *uṭai*'s above given the traditional intransitive/transitive' labels, often with information about restrictions on person and number of 'subject.' Were it not for the fact that Tamil usually marks the distinction between intransitive and transitive morphological differences in the *tense*-marking of the two types, and that there are tense markers for *all* tenses in Tamil (unlike English, where only the past is morphologically marked) it would not be obvious to most non-Tamils that distinctions must be kept separate. English, for example, has only a small set of verbs that are paired in this way, one being transitive and the other intransitive. Even these (sit/set, lie/lay, fall/fell, rise/raise) are not kept separate by many speakers. In Tamil either the stem itself is different (such as the (c)vc/(c)vcc- type exemplified by ooḍu/ooṭṭu 'run' vs. 'drive' or there is an alternation (c)vNC-/(c)vCC- (as with tirumpu/tiruppu 'return'), or the differences are marked in the tense markers, usually with weak types for intransitive and strong types for transitive. Similarly, there are verbs with ngu/kku contrast as in *aṭangku/aṭakku* 'control', *toṭangku/toṭakku 'begin'*. There are also some occasional cases of verbs with vu/ppu contrasts like *paravu* 'fan out' vs. *parappu* 'spread'. More work

needs to be done on the ways that Tamil marks the distinction between transitive/ effective and intransitive/affective verbs; since the database for this dictionary can be easily searched, we hope future researchers will use it to look at various lexical patterns that have yet to be analyzed or described for Tamil. A search I did many years ago in the Fabricius database to see how many pairs of the *tirumbu/tiruppu* type existed came up with hundreds of pairs[11].

This feature of making a new verb with past participle of main verb, like *collikkoṭu* 'teach' and *taḷḷi vai* 'postpone', *taḷḷi pooḍu* 'put off' as noted above, is common in the verbal system of most Indic languages and is often referred to as creating a 'compound' verb. By this is meant the use of two verbs adjoined in such a way that only the last one has tense and person-number-gender marking, while the previous one(s) occurs in a form known in Tamil as an 'adverbial participle' (which is commonly referred to by the abbreviation AVP.) Thus where English or other languages might conjoin two sentences such as 'I went to the store' and 'I saw him' to get 'I went to the store and saw him' Tamil (and other Indic languages) typically has a sentence like 'Having gone to the store, I saw him', i.e. *naan uurukkup pooy, avaraip paartteen*. To complicate matters, aspectual verbs are also adjoined in this way, with the aspectual verb marked for tense and PNG, but not the lexical verb, which occurs in the AVP form. Beyond this, we also find that verbs are compounded in this way to in effect create new lexical verbs; since Tamil does not borrow ***verbs*** easily from other languages, it creates new ones by combining existing verbs, e.g. the verb 'teach' can be rendered as *collikkoḍu* 'say and give; having said, give.' Sometimes, such forms make homonymous pairs between lexical compounds and their corresponding verbal inflections as in *koḍuttu viḍu* 'send' vs. 'give away'; *eḍuttu viḍu* 'untuck' vs. 'take away' etc. Former type of meanings are a case of compound formation where as the later are verbal inflections with aspectual auxiliary *'viḍu'*. Interestingly, spoken version of these forms have a way of distinguishing this meaning distinction by lengthening the final vowel for the cases of compound forms but not for inflections. Thus, *koḍuttuuḍu* is for 'send' and *koḍuttuḍu* is for 'give away'; *eḍuttuuḍu* is for 'untuck' and *eḍuttuḍu* is for 'take away'.

The process of 'derivation'. One of the ways languages have to innovate new vocabulary is by the grammatical process known as ***derivation***. The term 'derivation' is also used to refer to deriving something from something else historically, but by morphological derivation I mean the process of creating a new form, e.g. by making a verb out of a noun, or a noun out of a verb. English is very good at this type of thing, e.g. the verb 'to fedex' which of course is derived from the noun Fedex, which is an abbreviation of 'Federal Express.' Tamil has a number of derivational processes that are semi-productive, such as ways to make nouns out of verbs by the addition of a suffix: *veṉu* 'hate' + *ppu* → *veṉuppu* 'hatred.'[12] What has not been studied so well in Tamil is the process of derivation of new verbs from nouns or from combinations of nouns, verbs, and various derivational ***suffixes***.

A study of the verbs in this dictionary will show that a large number of them have been 'created' this way, either with aspectual verbs, or with other lexical verbs, or both. Certain lexical verbs tend to recur often in these combinations, especially when the result is a transitive verb:

- *aakku* 'make s.t. become';
- *uuṭṭu* 'feed, nourish'
- *celuttu* 'make s.t. go'.

[11] This is summarized in "Causativity and the Tamil Verbal Base" (Schiffman 1976)
[12] See Schiffman 2005 for more on deverbal nominal derivation in Tamil.

The last example here is instructive, because it itself is an example of an intransitive verb made transitive by the addition of ***–uttu***, which is a common way to create transitive verbs.

But it is even more interesting because *cel* alone does not occur in Spoken Tamil; but as a derived transitive, *seluttu* is acceptable in spoken when combined with other ***verbs***, though not with nouns as the object. Again, this phenomenon needs to be studied; attention to it will reveal other interesting patterns, such as the fact that when *viḻu* 'fall' is made transitive by adding *-uttu* what we get is a form with a long vowel, but with only *ttu* suffixed to it: *viiḻttu* 'bring down, make s.t. fall, defeat,'. Another common verbalizer is *paṭuttu* 'cause to be made' even though *paṭu* as a marker of passive is not used in ST. This verb then becomes a general 'causativizer'[13] in Tamil, which, combined with other verbs in their AVP form, is found widely throughout the entries here. Another very common example of this is the verb *naṭa* 'run, walk' which can be made causative by adding *-ttu*, i.e. *naṭattu* 'run something, make s.t. go, operate'. What was not obvious to me beforehand was the existence of many other verbs like this, such as the following:

- *taaḻ* 'be ruined, decline' → *taaḻttu* 'ruin, destroy'. By the addition of aspect markers, such as *koḷ* 'self-benefactive' we can get *taaḻttikkoḷ* 'make lower, discredit, degrade, debase, devalue; cheapen, abase, humble, humiliate, disgrace, dishonor; behave unworthily; humble o.s.'
- *aaḻ* 'be deep, profound' can be transitivized by adding *ttu* to get *aaḻttu* 'put to shame; further, with various aspect mark*ers*, other f*orms ca*n be derived, such as *aa□ttikkoḷ* 'involve o.s. deeply in; throw o.s. into s.t., immerse o.*s. in.*'
- *nikaḻ* means 'resemble, be similar to'; by the addition of *ttu* we get *nikaḻttu* meaning 'create, form, work (a miracle), deliver (a speech)'
- *kaviḻ* means 'turn upside down, invert (o.s.)(intr.)' and by the addition of *ttu* we can get *kaviḻttu* 'derail; overturn or upset s.t., as a boat; turn over, turn upside down, upend, flip/tip/keel over'

Notice incidentally that in the last few examples, the last sound in the basic stem is the 'retroflex frictionless continuant' *ḻ*, symbolized in the Tamil orthography as ழ. Why this sound should be so commonly found in these kinds of verbs seems strange, but needs perhaps to be investigated. Another transitivizer already mentioned is the verb *uuṭṭu* 'feed, nourish, imbue, instill, infuse, provide, nourish, inject or introduce new life or interest into s.t.' which, in combination with certain verbs (or nouns) expressing emotions, makes new verbs that mean something like 'propagate, contribute to, create or intensify an emotional state'. In the examples below, we either get a lexical noun such as *uyir* 'life, life-breath' compounded with *uuṭṭu* or we get nouns that have been derived from verbs, plus *uuṭṭu.*

- *veṟuppu* 'hatred' → *veṟuppuuṭṭu* 'fan the fires of hatred'. Note that *veṟu* is itself a verb; *veṟuppu* is a nominalization formed on the base of *veṟu* which is a transitive verb (6 tr) meaning 'hate.'
- *uyiruuṭṭu* 'animate, breathe life into, enliven, spark, perk up, liven up, freshen (up)'
- *ninaivuuṭṭu* 'call forth/up; bring back to (the) mind; remind; recollect'
- *makiḻcciy*uuṭṭu 'cheer up; inspire or encourage with cheer; make happy; gladden; inject some life into s.o. or s.t., stimulate; (inf.) tickle s.o. pink'

[13] By this is meant that it can be used to make an intransitive verb transitive, or a non-causative verb causative.

- *aruvaruppuuṭṭu* 'cause loathing, aversion or nausea; cause s.o. to feel hate; offend the senses or sensibilities; make dizzy'
- *mayakkamuuṭṭu* 'drive mad, crazy'
- *kacappuuṭṭu* 'embitter, make bitter; cause to feel disappointed, hostile or bitter'
- *caktiyuuṭṭu* 'energize, give energy to; make energetic'
- *calippuuṭṭu* 'irk, weary or annoy; bother; irritate, gall, pique, nettle, exasperate, try s.o.'s patience; anger, infuriate, madden, incense, get on s.o.'s nerves; antagonize, provoke.'[14]

Other Verbalizers: the case of *aṭi*

Another interesting verbalizer involves the use of the main verb ***aṭi*,** which of course means 'beat, tap'. When combined with nouns or other verbs, however, we get some interesting examples. An older use of ***aṭi*** that retains the notion of 'beating' or 'tapping' is *tandi aṭi,* which means 'send a telegram' (literally 'beat-wire'). But other uses of *aṭi* are more interesting. Another common usage is *veyil aṭi* as in *veyil aṭikkudu* '(sun) beats down'

Compare the following:

boor paṇṇu 'bore (a hole)' vs. *boor aṭi* 'be boring'
kaappi paṇṇu 'make a copy' vs. *kaappi aṭi* 'cheat; copy illegally'
taṇṇi kuṭi 'drink water' *vs. taṇṇi aṭi* 'drink (alcohol) in excess'

Other uses, such as *romba ḍall aṭikkriinga* '(you) seem gloomy, downcast' show that certain usages of *aṭi* are definitely negative, or at least pejorative, and that we should not be surprised to find other examples like this.

The verb *eṭu* used 'inchoatively.' Another interesting usage is that of the verb eṭu, which has the basic meaning 'take.' But in combination with certain nouns, it means 'begin to experience X', e.g.:

daaham eṭu 'begin to feel thirsty'
paci eṭu 'begin to feel hungry' and
vali eṭu 'begin to feel pain.'

Enrichment of lexical stock

Some verbs tend to expand their shades of meanings using one of the verbalizers as noted above especially to empower the use of language in various genres such as in poems, novels, speeches etc.

taaḻ 'come down' - *taaḻvu aṭai - taaḻvu koḻ - taaḻnduviṭu etc.*
vaaḻ 'live' - *vaaḻvu peṟu - vaaḻkkai peṟu etc.*
aḻu 'cry' - *kaṇṇiir viṭu* 'shed tears' - *kaṇṇiir malku* 'fill with tears' etc.

A Tamil Thesaurus?

Another idea for future research that emerged during the preparation of this dictionary was that while Tamil lacks a thesaurus, i.e. a dictionary similar to Roget's *Thesaurus of English* (Kipfer and Chapman 2001), which groups words by their similarity of meaning, into 'fields of knowledge', the database of this dictionary could be used to construct a first-step towards a Tamil Thesaurus.

[14] More examples of combinations involving *uu□□u* are found in Appendix 1.

This could be done by sorting words according to the general synonyms they are provided with. One of the main features of this dictionary is that most main entries are provided with one or more synonyms--verbs similar in meaning to the main entry. Originally many of these synonyms were separate entries but because of considerations of space and volume needed for sound files, we consolidated many examples into synonym files. But some general features were planned in advance. When I was in the early stages of planning this dictionary, and when studying the verbs in Fabricius' *Tamil-English Dictionary (*1972*)*, I noticed that for every verb that had some kind of meaning associated with sound, he provided the synonym *cattam poo*ḍ*u*, i.e. 'make a sound.' We have continued this tradition, so every verb that involves making a sound is provided by the same synonym *cattam poo□u*. If the database were searched for this synonym, a large number of verbs having to do with 'making a sound' would emerge, and could be brought together under one rubric for the purpose of the thesaurus. Similar studies could be done for other verbs, by first calculating the frequency of certain synonyms, and then sorting by synonym, rather than main entry. Thus a rudimentary Thesaurus for Tamil would emerge, which could be enlarged by consulting other electronic resources for Tamil.

References

1. Fabricius, Johann Philip and Johann Christian Breithaupt. *A Malabar and English Dictionary,wherein the words and phrases of the Tamulian language, commonly called by Europeans the Malabar language, are explained in English*. Vepery (Madras): 1779. iv, 185. (Revised 1809, 1911, 1933, 1972.)
2. Hopper, Paul and Sandra A. Thompson, (eds.) 1982. *Studies in Transitivity*. New York: Academic Press.
3. Kipfer, Ann and Robert L. Chapman (eds.) 2001. *Roget's international thesaurus*. New York: Harper Resource Press.
4. Paramasivam, K. 1979. ''Effectivity and Causativity in Tamil." *International Journal of Dravidian Linguistics* 8(1):71-151.
5. Schiffman, Harold F. 1976. 'Causativity and the Tamil Verbal Base.' *International Journal of Dravidian Linguistics* Vol. V, No. 2, pgs. 238-248.
6. Schiffman, Harold F. 1999. *A Reference Grammar of Spoken Tamil*. Cambridge: Cambridge University Press.
7. Schiffman, Harold F. 2005. "Deverbal Nominal Derivation in Tamil" *International Journal of Dravidian Linguistics*, Vol. 34, No. 2, June 2005, pp. 159-166.

Appendix 1: Examples involving derivation using *uuṯṯu*

1. 'alert' *eccarippuuṯṯu*
2. 'comic, *be' sirippuuṯṯu*
3. 'irk' *salippuuṯṯu*
4. 'dishearten' *soorvuuṯṯu*
5. 'dreadful, be' *tikiluuṯṯu, ericcaluuṯṯu*
6. 'flavor' *vaasaneyuuṯṯu*
7. 'animate' *uyiruuṯṯu*
8. 'mammal' *paaluuṯṯum piraaṇi*
9. 'gun, accelerate' *veekamuuṯṯu*
10. 'happy, be' *uṟcaakamuuṯṯu*
11. 'hopeful, be' *nambikkai uuṯṯu*
12. 'imbue (with)' *peṟuppuṇarcciyai uuṯṯu*
13. 'incendiary, be' *koopamuuṯṯum (peeccu)*
14. 'infuse' *uṟcaakattai uuṯṯu*
15. 'irritate' *ericcaluuṯṯu*
16. 'localize' *iyaluuṯṯu*
17. 'magnetize' *kaantavicai uuṯṯu*
18. 'embellish' *aḻakuuṯṯu*
19. 'shock' *atirciyuuṯṯu*
20. 'sour, disgust, disenchante' *veṟuppuuṯṯu*
21. 'mislead, throw s.o. off' *kuḻappiyuuṯṯu*

Appendix 2: derivation involving *celuttu*

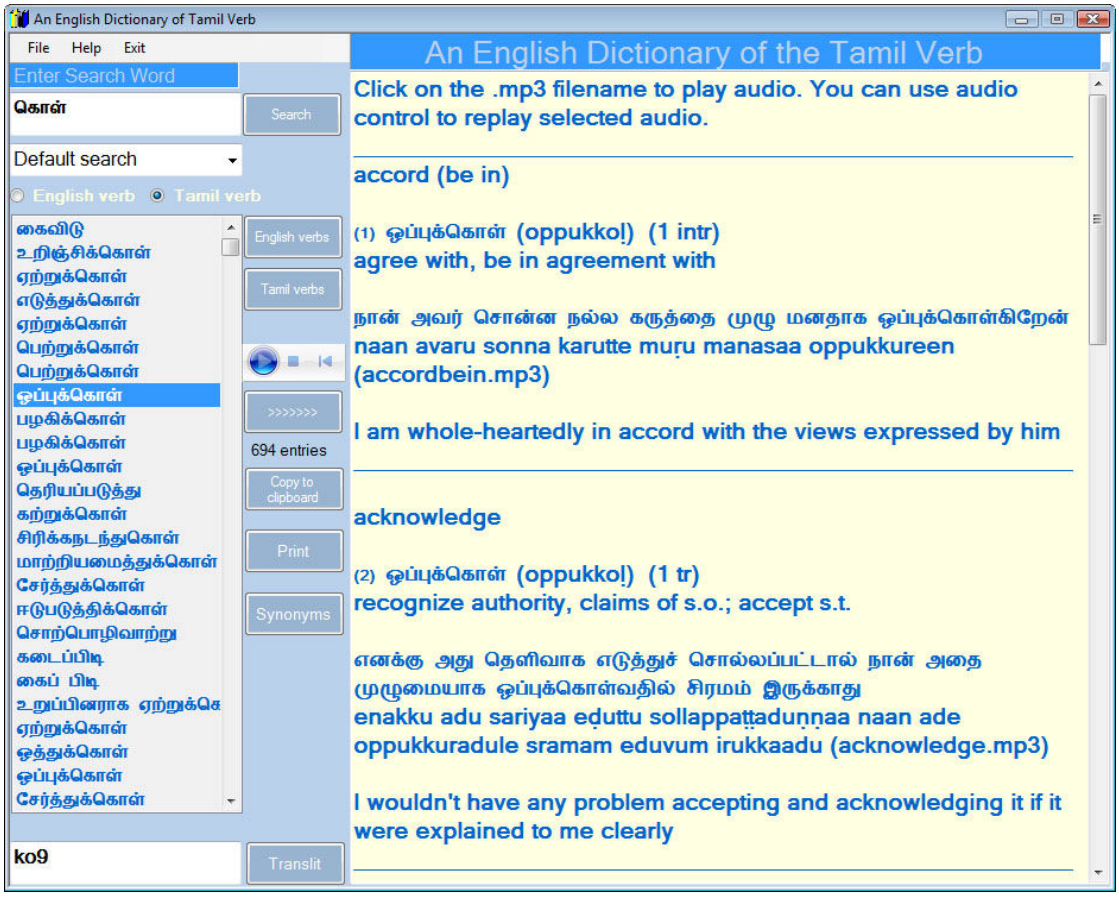
1. 'active, be' *gavanam seluttu*
2. 'skim' *kaṇṇoottam seluttu*
3. 'thrust, insert' *uuciyai seluttu*
4. 'venerate' *mariyaatai seluttu*
5. 'command' *atikaaram celuttu*
6. 'dominate' *aatikkam seluttu*
7. 'make payment' *paṇam celuttu*
8. 'reverse, back up' *(vaṇṭiyaip) pinnaale celuttu*

Appendix 3: derivation involving aḍi

1. 'apply whitewash' *sunnambu aḍi*
2. 'speak assertively' *aḍiccu peecu*
3. 'blow (hard)' *(kaattu) veehamaa aḍiccadu; viici aḍi*
4. 'belt (s.o.)' *oongi aḍi*
5. 'bilk (s.o.)' *koḻḻe aḍi*
6. 'blast' *sedayaḍi*
7. 'rate (storm)' *viici aḍi*

8. ‘bore (s.o.)’ *boor aḍi*
9. ‘bustle about (cooking)’ *padariyaḍiccu kiṯṯu (same)*
10. ‘butter up; apple-polish’ *aayil aḍi*
11. ‘catcall, heckle’ *ciiṉkkai aḍi*
12. ‘caterwaul’ *puunai poola aḍi*
13. ‘defeat’ *tooṟkaḍi*

Appendix 4: Software to search/browse Dictionary



Conference book brought to you by

இலக்கியம், கலை, அறிவியல், அரசியல், சமூகம், வரலாறு, மானுடவியல், வாழ்க்கை, கேளிக்கை என பல்வேறு துறைகள் சார்ந்த ஆழமும் அக்கறையும் மிக்க படைப்புகளை தமிழில் தொடர்ந்து வெளியிட்டு வருகிறது தமிழின் நம்பர் ஒன் புத்தக வெளியீட்டு நிறுவனமான கிழக்கு பதிப்பகம்.

An imprint of
NEW HORIZON MEDIA PRIVATE LIMITED
33/15, Eldams Road, Alwarpet,
Chennai - 18. Tamil Nadu, INDIA
Phone : 91 + 44 + 42009601, 03, 04
www.nhm.in, Email : support@nhm.in

www.infitt.org